The impact of the Nitaqat programme on the Saudi labour market

Formal assessment report for the economics PhD program

By

Tahani Bagazi

BA (Honours) in Economics (King Saud University, Riyadh, Saudi Arabia)

MA in Economics (King Saud University, Riyadh, Saudi Arabia)

Supervisors:

Dr Michael Nolan

Dr Gabriele Amorosi

September 2020

○ Dedications

To Wajd and Wafi.

To those who gave me hope and motivation.

○ Acknowledgement

Foremost, thanks to my parents, who taught me persistence and patience. I also acknowledge the souls of my father and my husband (may God have mercy on them) for their encouragement to complete my education.

Certainly, I would like to thank the government of the Kingdom of Saudi Arabia for providing me with this valuable educational opportunity and being supportive of the empowerment of Saudi women. My sincere thanks and appreciation to the Ministry of labour and social developments (MLSD) for providing this unique data. Thanks, are also extended to King Saud University (KSU), especially the economics department representatives who supported me overcome all obstacles I experienced.

Doubtless, I am glad to express my sincere thanks and gratefulness to my supervisor Dr Michael Nolan for the endless academic and personal support during my PhD research journey. I did not regret choosing him as a supervisor when I was searching for a PhD offer; it was an excellent choice for his knowledge, guidance, wisdom, and humanity. Dr Gabriele Amorosi has also been an effective member of this journey. Also, I thank the University of Hull for the friendly studying environment and student support services.

○ Abstract

This research evaluates Saudi Arabia's Nitaqat quota policy within a labour market context, where the target group (Saudis) is earning double the pay of the untargeted (non-Saudis) group[1]. We aim to examine the wage gap between the two groups to address its sources and discover how the quota can change this gap. This allows an understanding of how the quota affects Saudis' welfare. Moreover, understanding the gap in firm status allows an evaluation of the layoff risks, firms' employment behaviours and the earning structure changes. The Oaxaca–Blinder decomposition, a standard econometrics tool, is used empirically. This methodology is based on estimates from the ordinary least squares method. Data for two years, 2013 and 2017, was available, and a pooled sample for the two years was used. We find that the gap could be explained by around 34% in 2013 and 71% in 2017, while the remaining gap was due to the differences in earnings structures between the groups, especially the starting wages. The remainder of the gap was unexplained by a high percentage due to the different wage structures, especially the starting wage (via intercept), which implies that unknown variables could explain the gap. However, the explained percentage jumped to 80% and 85%, respectively, when we included consumption as a new exploratory variable. Additionally, we found heterogeneity in the wage gap among workers' origins. This implies that the gap could be closed if we considered some other unobserved variable, such as the wages in the sending countries. Additionally, the wage gap between workers in localised and non-localised firms was in favour of workers in localised firms in both years. Furthermore, workers responded to the Nitaqat quota in the opposite direction of that predicted by the hedonic wage model when layoff risks existed, but they were consistent with this wage model when unemployment benefits

[1] **Quota** is an official application for **affirmative** action. It is a compulsory number or percentage that institutions need to follow to meet the government requirement; otherwise, they will be punished. It is applied in various economic areas, such as schooling, trade, and employment. Employment quotas are targeted in this research; a particular number of employees are taken into consideration for firms' hiring decisions. This quota is normally imposed on two levels: first, immigration visas to limit immigration numbers; and second, employee percentages of targeted groups. The latter is known as a **hiring quota**. **Affirmative action** is a government policy applied to help disadvantaged groups, such as minorities, women, and blacks, in schools and workplaces. The latter is the research target existing on different levels: pre-jobs and in jobs. This means that its applications vary, such, beyond ending discrimination practices that are taken to promote equal opportunity and ensure that discrimination will not recur. **Hiring quota** is a formal policy that affects the demand side of the labour market. It is designed to employ a specific compulsory proportion from a target group that has received hiring discrimination from employers, regardless of the employers' orientations towards those employees.

were imposed. Accordingly, non-Saudis in non-localised firms could be described as a double negative group. Moreover, the gap narrowed between the groups because Saudis' wages decreased twice as much as non-Saudis' wages. This resulted from a massive redistribution of Saudi workers towards the quota minimum wage, which coincided with a reduction in education returns. In other words, linking the policy with a specific wage allowed firms to replace qualified Saudis with unqualified Saudis rather than Saudis with non-Saudis. Thus, firms balanced their costs by offering low-quality jobs for Saudis, aiming to achieve reasonable profits. We conclude that the Nitaqat quota policy negatively impacted Saudis' wages; however, it achieved some of its aims, such as decreasing the gap and controlling the unemployment rate. This resulted from the responses of firms and workers towards the policy; the policy lost its power because of the redistribution of firms' sizes and the redistribution of workers among occupations. An effective gap decrease requires the policy to make distinctions between occupations to ensure the job quality is provided. Moreover, adding some criteria or fees to Nitaqat could shift the replacements among non-Saudis, taking the substituted relationship and the labour intensiveness into consideration.

○ Table of Contents

○ Index of Figures

○ Index of Tables

○ Acronyms

GaStat: General Authority for Statistics

GOSI: General Organisation for Social Insurance

MLSD: Ministry of Labour and Social Developments

OS: Observed sample

OB*:* Oaxaca–Blinder decomposition

MCAR: Missing completely at random.

MAR: Missing at random

***MNAR/NMAR*:** Missing not completely at random.

CC: Complete case

IC: Indicator category

FR: Frequency replacement

MI: Multiple imputations

RE: Reweighted equation method

DV: Drop variable

UHR: Unit hourly rate

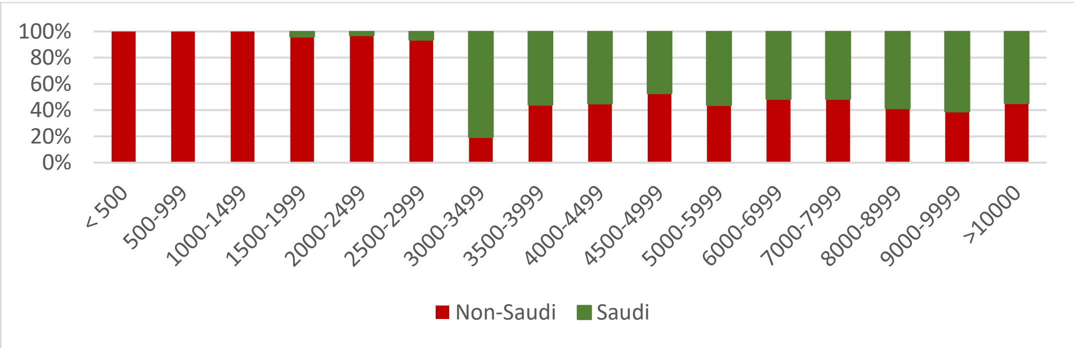
Chapter 1 General Introduction and background

1.1 Introduction

Saudi Arabia's economy is a special case among countries with the highest number of immigrants, including some of the Gulf Cooperation Council countries (GCC); which consists of Saudi Arabia, Bahrain, Qatar, Kuwait, Oman, and the United Arab Emirates. According to the World Bank statistics from 2015, the Saudi labour market is the fourth most popular destination for migrants, after the US, Germany, and Russia. Compared to those countries, one finds that immigrants comprise a high percentage of the total population and the labour force in Saudi Arabia. Indeed, immigrants in Saudi Arabia cover approximately 32% of the total population, compared to roughly 8% in Russia and 14% in the US and Germany. Moreover, those immigrants account for approximately 78% of the total labour force in Saudi Arabia, compared to 15% in Russia and 28% and 29% in Germany and the US, respectively.

Although the Saudi economy has been growing due to oil making up a major source of Saudi revenue, the labour market has failed to provide sufficient jobs for Saudi workers due to structural problems where non-Saudi workers were distributed intensively among lower wage categories. Figure 1-1 illustrates that non-Saudis mainly work under 3000 SR while Saudis mainly work between 3000-3500 SR. The worker's distribution of both groups seems to be convergent after 3500 SR. There were almost 400,000 workers working above 10,000 SR. This could imply two things: **first**, segregation. **Second**, labour market dependency on the low paid job.

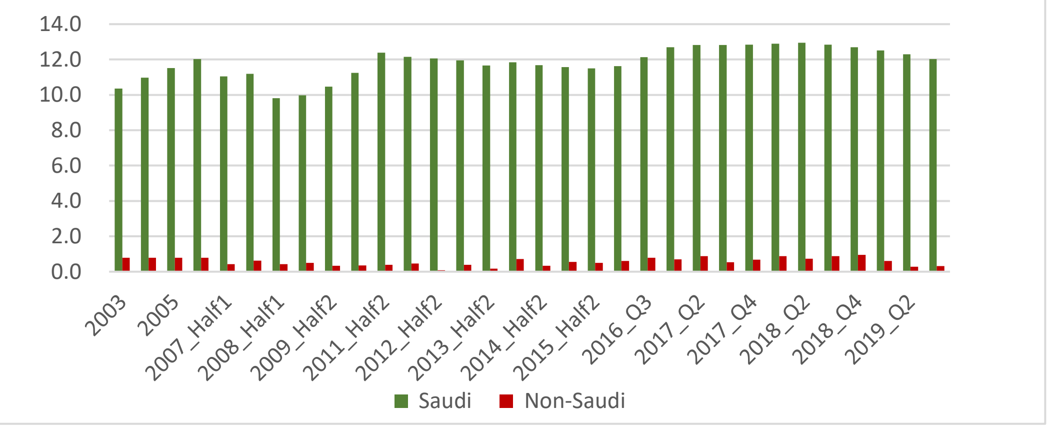
Figure 1-1: workers distribution among wage categories.



Source: General Institution for Social Insurance, 2015, p. 87

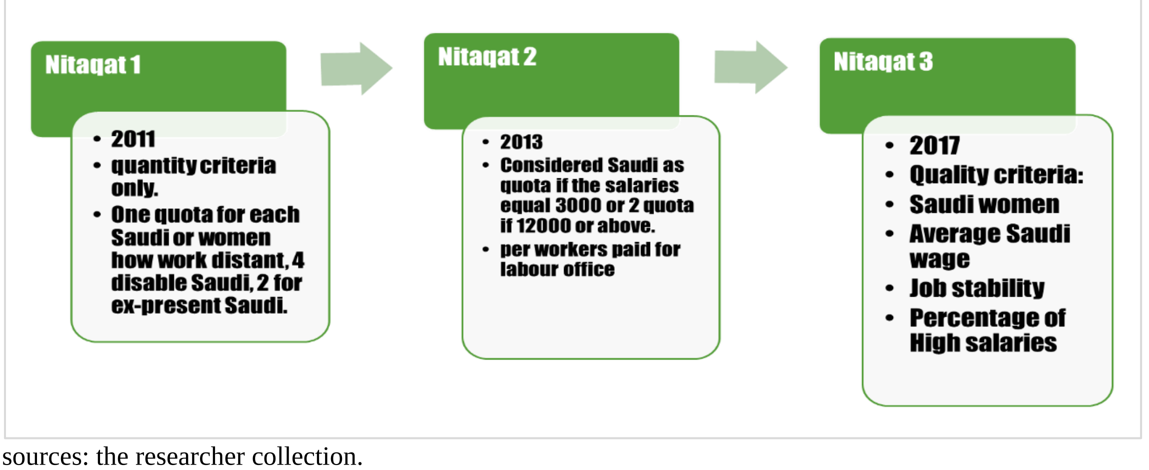
Additionally, Saudi Arabia's labour market has suffered from another problem, a high unemployment rate for Saudi workers. The unemployment between Saudis reached 12% in several years, while this ratio did not exceed 2% for non-Saudis (see Figure 1-2). Notice that the unemployment rate among Saudi women is higher than Saudi men. For example, in the third quarter of 2019, the unemployment rate was 12% among Saudis, 30.8% for Saudi women, and 5.8% among Saudi men. In terms of non-Saudis, this ratio was 0.3% in general. 0.2 among non-Saudi men and 1% among non-Saudi women.

Figure 1-2: Saudi and non-Saudi unemployment rate.



Source: The researcher's collection from several files published by GaStat. Non-Saudi follows the secondary axis.

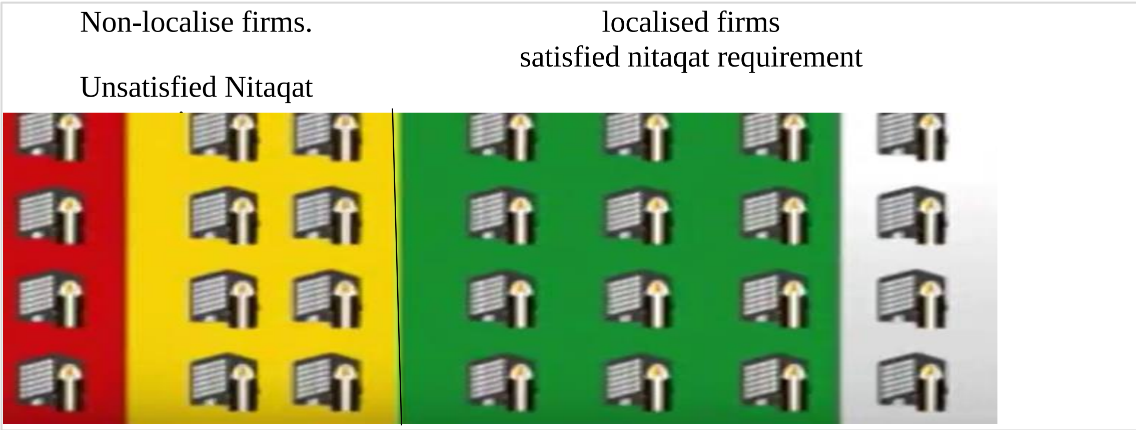
Figure 1-3: brief demonstration of Nitaqat editions.



sources: the researcher collection.

Briefly, the percentage of Saudi employees (quota) is calculated according to a certain mechanism respecting firms' size and activities. On this basis, firms have been classified into four zones – nitaqat: red and yellow represent non-localised firms (with a proportion of Saudi workers that is seen as insufficient), while firms with an acceptable proportion of Saudi workers called localised firms are represented by green and platinum (see Figure 1-4). This classification used to reward localised firms and penalise non-localised firms. For example, localised firms would be able to issue new visas, were given extensions to pay government payments, while non-localised firms would be banned from issuing new visas or renewing visas for ex-pat workers, and they would not have access to any government services such as renewing employee occupations licences.

Figure 1-4: firms type according to their achievements of Nitaqat.



Sources: arranged by the researcher.

Since that policy was initiated, there has been a remarkable change in the labour market structure, such as firms collapsing or downsizing. Foreign workers have been deported, and new workers have arrived, Saudi workers were employed and then laid off again, in a continuing sequence of cycles.[2]

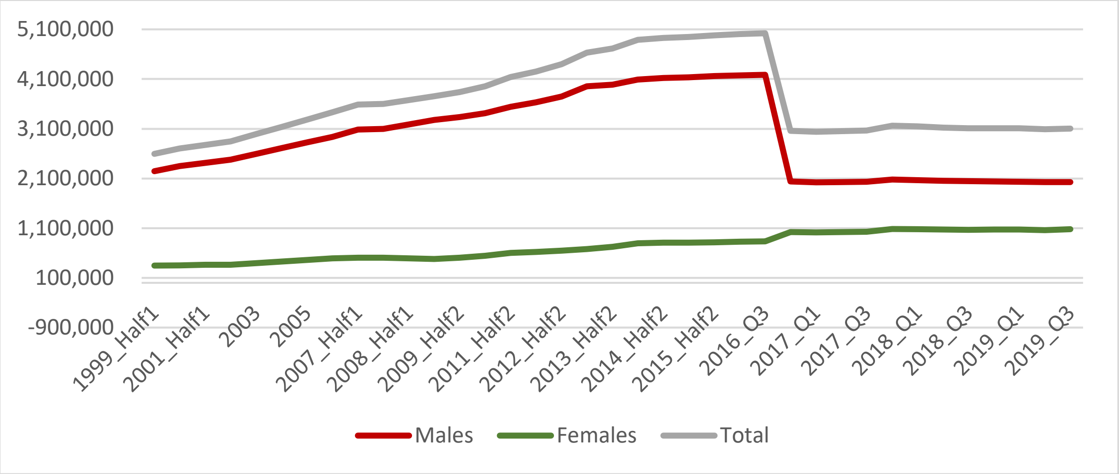
1.1.1 The research problem

The research problem stems from the contradictory of Nitaqat with quota principle, where the disadvantaged groups have lower wages and employment, Saudi workers with higher earning averages have become the policy target group. Thus, the quota could have a perverse outcome rather than enhancing Saudi workers in both wages and unemployment. By that, I mean, Nitaqat could not guarantee increasing wages and employment rates for Saudis. In this context, Nitaqat provided a slight insignificant

[2] See the background for more information.

increase in Saudi employment (Peck, 2014). However, this slight increase in the employment was affected by the announcement of non-Saudi fees and significantly decreased male employment by 2,143,739. This contrasted with the number of women employed, which increased by 183,557 (see Figure 1-5). Notice, this decrease in Saudi men's employment did not reflect on the unemployed numbers or labour force. Frankly, this could imply that those, who were removed from the employed statistics and did not register in the unemployed statistics, were transferred out of the labour force. In other words, the increase in the number of Saudi men since Nitaqat was not reflecting the success of the policy where employers depend on dummy Saudisation. For illustration, to meet the quota requirement, an employer will register Saudis (for example, students, retired, or disabled) and pay them (for example, 1500 SR) despite them not having a real job. This escape plan from Nitaqat was of benefit to both the employee and employer before the fees were announced. However, after these fees, it was to some extent costly for the employers depending on the dummy Saudisation. Thus, these fees that applied for the non-Saudi group indirectly played a role in curing dummy Saudisation for the Saudi group.

Figure 1-5: Employed Saudis between 1999 and 2019



Source: Collection from several published labour market reports

Because it is very unusual to introduce quotas when the majority group are lower-paid, its effect on wages could be contradictory to the expected effect (increase wages) for two reasons. **First**, the position of most non-Saudi workers appears to contradict the compensating-differentials theory[3] since non-Saudis are more commonly found in low-paid jobs that are also associated with negative job qualities, such as risky environments, long hours, and distant locations. This implies that the layoff risk which stem from Nitaqat could have a similar negative relationship. **Second**, the distribution of the workers tends to be concentrated in the bottom of wage scale, which increases the possibility of the replacement accord on the bottom of the scale (see Figure 1-1 above). Accordingly, the wage gap between the two groups could decrease or increase depending on the workers' responses to the layoff risk stemming from the Nitaqat and workers redistribution by firms. Therefore, considering the wage gap reduction as an aim of Nitaqat could be changing. The reaction would be in three scenarios: **First**: the average wage of Saudis would increase if Saudis were employed in high-paying jobs, which requires an increase in non-Saudi wages by a higher percentage to narrow the gap, which would be costly for the firms. **Second**, Saudi average wages would decrease if Saudis were employed in low-paying jobs (or at least around the quota's minimum wage). This could be associated with an increase in non-Saud wages. This could imply a spill-over effect where non-Saudis benefit from the policy compared to non-Saudis. In the **Third scenario**, both groups' average wage would decrease, and Saudi average wage reduction should be higher to narrow the wage gap. Thus, the success of the Nitaqat program's goal of reducing the wage gap would be potentially indicated if it were associated with a reasonable replacement toward the quota minimum wage for Saudis or increase both groups average wage, although it would be costly if associated with similar quality of non-Saudi workers.

1.1.2 The research hypothesis and questions

In this study, we examine the Nitaqat programme by exploring the wage gap between Saudis and non-Saudis, considering a firm's status (localised or non-localised) according to Nitaqat. In addition to estimating the role of Nitaqat affecting this gap in the Saudi labour market, the research hypothesises that Nitaqat will create unemployment risk which will lead workers to respond to the changes in the layoff risk

[3] The compensating-differences theory states that the more negative the jobs, the higher the wages. This reasoning could be explained by the heterogeneity between the two groups. However, under this theory, even if heterogeneity was observed, a wage differential would still exist (Tachibanaki, 1996).

and then the average wage gap will be affected. Thus, the efforts have been dedicated to answering the following questions.

I. What factors explain the wage gap between Saudis and other workers?

II. Does this gap result from worker characteristics or discrimination?

III. How does the Nitaqat programme contribute to change in the existing wage gap? Is it widened or narrowed? Are there any changes in Saudi wages as a target group? Do wages go up or down?

IV. Is there any difference between firms' behaviours towards the wage gap according to their colour bands? If yes, do localised firms have a lower wage gap than non-localised firms?

V. Does the layoff risk resulting from Nitaqat increase or decrease wages? Is this change equal in both firm statuses?

1.1.3 Research objectives

This research aims to analyse and evaluate the economic impact of the Nitaqat programme on the Saudi labour market regarding workers' wages. Its specific objectives are as follows.

I. Explain the main elements that construct the wage gap between Saudi workers and others and understand the structural differences and determine if there is any distinguishing variable that could explain this gap, such as layoff risk and consumption.

II. Shed light on programme outcomes regarding a change in wage gaps because narrowing the gap between those two groups is one of the programme's aims. If the gap is narrowed, determine if this is associated with an increase in Saudis' average wages. This would provide an implication regarding worker welfare, measured by wages.

III. Compare the wage gap among firms using the principle of localisation status to clarify if the wage gap is greater or smaller in localised firms than other firms. This might provide an implication towards a firm's capability to employ Saudis.

1.1.4 Data sources and limitations

The study benefits from two separate data cross-sections for the years 2013 and 2017. This was provided by the MLSD and restricted by the privacy policies. The data was rich in terms of observations; however, it was limited in terms of the number of variables. Regardless of this limitation, we determined the main variables to estimate earnings functions which require estimating our models. The Oaxaca–Blinder (OB) decomposition is the key empirical method used.

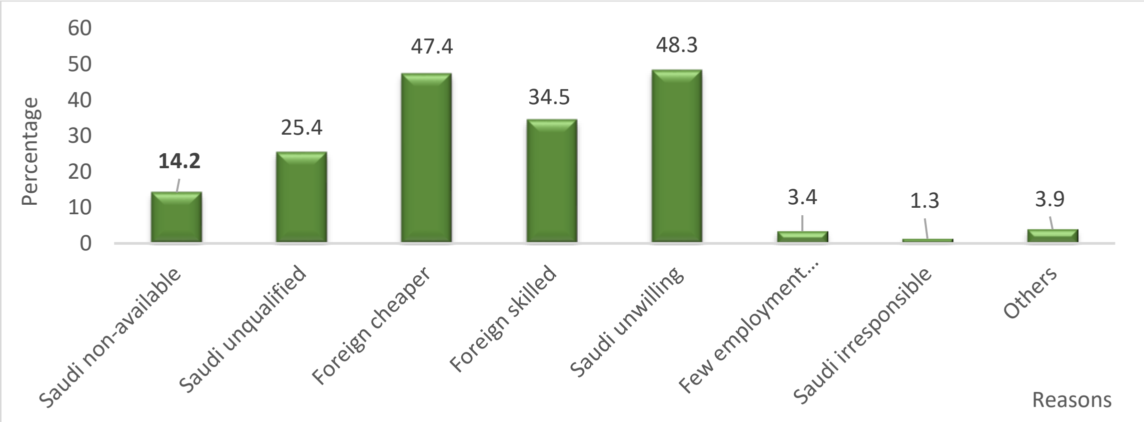
1.1.5 The importance of the study

This study sheds light on using affirmative action, i.e. quotas and policies as a remedy for high unemployment rates among Saudis compared to non-Saudis. Although Saudis suffered from a high unemployment rate, they accepted double wages and a full day of work, whereas foreign workers accepted any wage and any number of work hours.[4] This is because non-Saudis maximise their financial goals when they decide to work in Saudi Arabia. For example, foreign workers benefit mainly from the different exchange rates between Saudi Arabian currency and their home countries' currencies, even if they received lower wages than Saudi. Our viewpoint is that GCC countries should be cautioned about applying the quota policy investigating both firms' benefits and workers' drawbacks as the labour market depends heavily on cheap workers who work for less than 1,200 Saudi riyals (SR) per month, which is worth around $320. In this case, the more Nitaqat led to successfully employing Saudis, the higher the risk of employing Saudis in low-paying jobs. Therefore, wages are the main determinant for employers; Figure 1-6 shows two reasons that explain this. **First,** Saudis are unwilling to work, which could be due to low wages or harsh work atmospheres. **Second**, foreign labour is cheaper compared to Saudi labour. For these reasons, this research claims that affirmative action theory is not suitable to be applied equally to the whole wage level distribution without distinguishing between occupations and labour relationship. This could pull Saudi average wage down by replacing them in a lower paid job. This highlights the importance of this study. Ruppert (1999) claimed that the success of the Malaysian quota policy could

[4] It could be described as unpaid overtime jobs.

apply in the Gulf Cooperation Council (GCC) countries. Malaysian case (Ruppert, 1999) might be better developed to draw out unique features of the Saudi experience.

Figure 1-6: reasons for employers preferring foreign workers in Saudi Arabia.



Source: Ministry of Planning, private establishment survey, Table 3, 16, 1999. Adapted from Cordesman and Obaid (2005).

1.2 Background

The labour market regulations are heterogeneous between Saudis and non-Saudis. For example, labour mobility regulations where the movement of non-Saudis is restricted, unlike Saudis. Moreover, non-Saudis have a sponsor's agreement to switch to another job[5]. Thus, non-Saudis are the preferred type of workers for firms in general. However, Saudi wages could increase due to voluntary mobility, which could contribute to the wage gap, which is consistent with the finding of Brenzel and Reichelt (2017). Another unique regulation is recruiting policies being open for a long time, corresponding with the high demand for non-Saudi workers due to the lack of supply of Saudi workers. Regardless of the myth that Saudi workers are reluctant to do some types of jobs, firms tend to recruit cheap labour from less developed countries to achieve the minimum costs.[6] Additionally, there were no specific wage scales or wage premiums, unlike in the public sector. Moreover, the hourly wage rate is not known for either skilled or unskilled workers as the government has never regulated the wages of private-sector

[5] From March 2021when a non-Saudi's contract ends, they can move to another job without the employer's permission.
[6] Those Saudi day workers are found in some jobs, such as waitressing, that generate a reasonable salary, unlike the myth of Saudis being reluctant to work in this job.

occupations. Similarly, regarding the lack of minimum wage regulations for both Saudis and non-Saudis, although Nitaqat uses minimum wage terminology for the quota benchmark, there is no real use for minimum wages when workers could work with any agreed wage.

Furthermore, there is a lack of wage determination regulations, such as union bargaining power. For example, non-Saudi workers usually have no bargaining power. They can only choose between several jobs offered by several host countries. The power can be found in the recruiting offices, which can increase or decrease wages. A clear example is that of a Filipino housekeeper whose government prevents the recruiting office from sending any person wages less than 1,500SR per month and offering food, shelter, and a clear holiday system.[7] This is unlike other nationalities, such as Ethiopians, where there are no restrictions in their salary. As a result of this lower bargaining power, Ethiopians earn lower wages compared to Filipinos. This would apply to other jobs as well. In contrast, Saudi bargaining power could be better as there is no movement restriction. When firms offer fixed wages, Saudis can accept the job or wait for a better offer. Moreover, there is heterogeneity in the pension regulations and health insurance options for Saudis and non-Saudis. From one angle, non-Saudis are ineligible for a retirement pension, unlike Saudis. Thus, Saudis seek to increase their wages to the limit that guarantees achieving a minimum pension. Accordingly, employers are required to pay 12% of the registered salary as social insurance to the General Organisation for Social Insurance (GOSI), and 10% must be paid by the Saudi workers. Therefore, Saudis' disposable wages are 10% lower than what they receive, which increases their reservation (gross) wage. However, employers are required to pay 2% of the registered salary for risk insurance for non-Saudis in some occupations. Thus, from the employer point of view, Saudis seem costly compared to non-Saudis if considering identical worker characteristics and wages. In terms of health insurance, firms must provide health insurance for Saudis, unlike non-Saudis.[8]

[7] The housekeeper job does not count in Nitaqat. Thus, all drivers and houseworkers are excluded from the sample. However, we are trying to explain the source of bargaining power of the recruited non-Saudis.
[8] In 2015, insurance for non-Saudis was made compulsory and joint with Iqama renewal; the workers are responsible for this payment.

1.3 Saudisation

Due to the extensive impact of immigration on the Saudi labour market, *Saudisation* was implemented aiming to create jobs for Saudis in the private sector. Saudisation is a term that refers to restricting workers to Saudis and gradually substituting existing workers for Saudis (Manpower Council, 1422 H). The first Saudi plan (1970) pointed out to develop the local human capital to engage Saudis in the labour market. The first declaration on the substitution between terms was in the sixth edition of the fifth plan, 1995–1999.[9] This terminology was used earlier in the public sectors, where it achieved its goal directly. Saudis represented approximately 96% of the total public sector workforce in the first quarter of 2019, according to the Saudi Arabian Monetary Authority (SAMA). However, it was not sufficient to absorb the increase in the Saudi labour force; the public sector employed around 12.745% of total employment (AlShik, 2008). Surprisingly, this percentage has not significantly changed over time, as the percentages were 12.40% in the first quarter of 2019, according to SAMA.

Unfortunately, Saudisation did not succeed in the private sector, where the Saudi unemployment rate has persisted. The obstacles to achieving Saudisation were the main concerns of the researchers and the government. Several obstacles were pointed out in the seventh edition, such as the wage gap between Saudis and non-Saudis and the continuous, unlimited flow of immigrants. A meta-analysis identified 48 obstacles, five of them being classified as the most relevant: a low Saudi supply to some low-skilled jobs, Saudi education-jobs mismatch, less secured jobs, the al-tasatur[10] and Saudis' preferred the public sector (Riyadh Economic Forum, December 2013).

1.4 Nitaqat

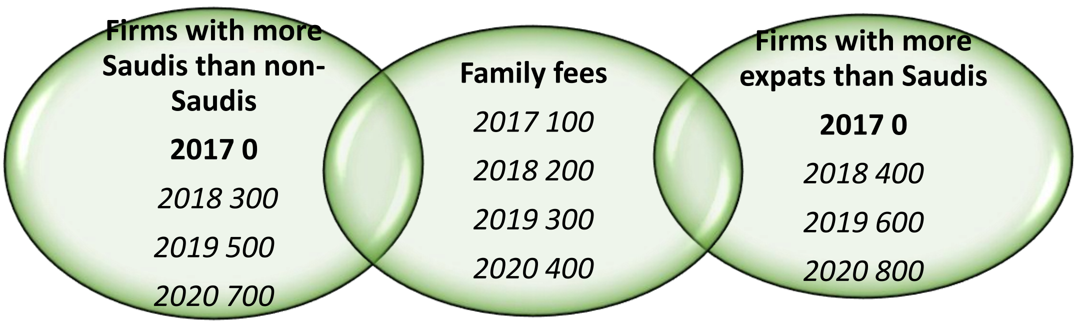
According to the unsuccessful results of the Saudisation previous years', the MLSD proposed a new employment strategy (Nitaqat) in 2009, by agreement of the Council of Ministers' Number 260.[11] Nitaqat is a hiring quota programme that encourages firms to

9 The Ministry of Economics and Planning issued those fifth-year plans; each plan covered five years.
10 al-tasatur refers to firms outwardly owned by Saudis while the real owners and operators are non-Saudis; a financial amount is paid to the Saudi owner. This is illegally practised, and non-Saudis could register as investors if they fulfilled the terms and conditions.
11 This is 05/08/1430 H.

employ Saudi workers for the entire private sector. The key objective of Nitaqat is to engage Saudi workers (whether male or female) in the private sector and achieve three general outcomes: full employment for the labour force, an increase in the contribution of Saudi human capital, and an increase in Saudi productivity.

Nitaqat is a 25-year plan divided into three phases: short-run (two years), medium-run (three years) and long-run (20 years). Each phase has a specific aim. The first phase aims to control unemployment through policies. The second phase aims to reduce the unemployment rate by encouraging employment rate growth, participation, and productivities. Finally, the third phase aims to achieve a competitive advantage for Saudi human capital. However, Nitaqat was frequently updated according to the needs and responses of the labour market. In the earlier version, it focused on Saudi quantities and then some quality criteria were updated. The **first** version of this programme was established in June 2011, while **Nitaqat2** was established in 2013, the key feature of which was setting the minimum wage for Saudis at 3,000SR to be considered as a quota because of the critique that Nitaqat pulled the Saudi wages down to less than 3,000SR. In November 2016, **Nitaqat3** was initiated to improve job quality for Saudis. Qualitative criteria were considered when the quota was calculated, such as work stability, female participation, and wage quality. However, the implementation of this version was delayed and was made available for firms to calculate their percentages on the MLSD portal in December 2017. Additionally, a new fee system was announced for non-Saudis; they gradually increased (see Figure 1-7). This fee is in addition to the existing fees as foreign workers pay 200 SR per month (2,400 SR each year) for the labour office and compulsory medical insurance.

Figure 1-7: Timeline for non-Saudi labour fees



Source: Researcher’s information collection

Several government bodies such as the Saudi Human Resources Development Fund (HDF),[12] the Ministry of Education, the Saudi Credit and Savings Bank and the GOSI were cooperating to support Nitaqat supplement programs. For example, **Taqat** programme provides a comprehensive electronic platform for Saudi employees and employers, including training. **Leqa'at** provides meeting rooms in each city to allow Saudi job seekers to meet a represented company for examinations and interviews. **Jahs** is a programme aimed at supporting students who are in higher education by improving their skills and providing jobs in an early stage before graduation, while **Enjaz** is a programme that aims to prepare high school students for the labour market. **Masarat** is a programme meant to financially support small family projects, thus encouraging self-employment. **Hafz** programme aims to provide monthly unemployment insurance of 2,000SR and engage Saudi job seekers with training programmes. Their responsibility ends by providing a suitable job. **Supporting Saudisation Growth** is a programme that reduces the cost of employing Saudis by subsidising Saudis' wages with firms. Also, in September 2014, the GOSI cooperated to introduce the **SANED** programme, which cares for Saudis who lose their jobs. The compensation for this one-year programme is around 60% of the average wage for the first three months after losing a job and 50% for any extra month without working.[13] Moreover, Saudis that qualify to register for this programme benefit from training and job search services.

Moreover, as women's participation is very low, specific programmes have been put in place to improve women's participation rates, such as **Maher**, which provides training support and finds a range of jobs for women, and **Wsol**, which covers around 80% of transport costs for Saudi women, but not more than 800SR monthly. This specific support is limited to up to 36 months after first registering, and the woman's monthly wage must be less than 8,000SR. The **Qurah programme** subsidises the cost of childcare for two children for up to four years when a woman works in the private sector. The maximum amount is 800SR in the first year and decreases to 600SR in the second year, 500SR in the third year and 400 in the third and final years.[14] Moreover, it

[12] This was launched in 2000 as a financially and administratively independent agent to increase Saudi worker quality and support job Saudisation.
[13] @SaudiGOSI 07/04/2014 and
[14] https://www.hrdf.org.sa/Program/466/%D9%82%D8%B1%D8%A9__

supports women working remotely from home and has introduced part-time job rules. Furthermore, under the “al-twtteen al-mowazi” programme, in Nitaqat3, firms can achieve the required Saudisation percentage by paying a monthly fee for each Saudi unit required; this could be considered virtual Saudisation[15]. This programme was proposed to avoid the system drawbacks on firms in the two old versions. However, the fees are paid for the HDF to improve Saudi workers through training and education.

Additionally, there are two other programmes interested in collecting data: **National Labour Force Observatory,** which provides a vital index of the labour market for who is interested, and the **wage protection** programme, which is responsible for collecting wage data for both Saudis and non-Saudis to ensure that the wage giving matches the wage contract and wage payment on the due date, aiming to reduce the conflict between employers and employees. This programme requires all information related to the basic wage, such as housing allowances and other allowances. This programme has 17 phases that follow a timeline. The first compulsory implementation was September 2013 for large firms that employed 3,000 workers or more, according to the MLSD website.[16] The final phase will be applied in October 2020 for small firms with up to four employees. Firms with 2,000 workers were required to join this programme in December 2013. The final deadline for small firms was in November 2018. To ensure this programme was activated, the MLSD imposed a penalty: if a firm is late to provide the requirements, the MLSD services will be stopped unless work licences are issued and renewed. This programme has a security angle, though, evaluating suspicious banking cases and reducing the risk of theft resulting from cash exchange for wages.

1.5 Worker’s criteria

Nitaqat started with quantity criteria, and then it was developed to account for quality criteria. Therefore, the calculation method of Saudisation percentages – the quota – is different in each version. In the initial version, Nitaqat1, the method was quantity-

---

[15] Dummy Saudisation refers to a registered Saudi person who is not working to achieve the programme percentage. Virtual Saudisation refers to a scheme proposed by the MLSD allowing employers pay a fee to achieve the required percentage of Nitaqat.

[16] https://mlsd.gov.sa/ar/initiatives/%D8%A8%D8%B1%D9%86%D8%A7%D9%85%D8%AC-%D8%AD%D9%85%D8%A7%D9%8A%D8%A9-%D8%A7%D9%84%D8%A3%D8%AC%D9%88%D8%B1

based. It considered each Saudi male/female, husband/wife, daughter/son of Saudi women or Saudi female distance workers as one quota; each Saudi ex-offender was considered two quotas; each disabled Saudi with working ability and willingness to work was considered four quotas; Saudi students who worked part-time were considered half of a quota, and any foreign workers who had received an offender's statement or were already in prison did not count.[17] After this quota was collected, it was entered in the moving average equation, as follows.[18]

$$Saudisation\ percentage = \frac{number\ of\ saudi\ quota\ at\ 13\ weeks}{Total\ employees\ number\ at\ 13\ weeks} * 100 \qquad 1\text{-}1$$

Although this Saudisation percentage equation was valid in Nitaqat2, the quota system accounted for the wage impact based on the wages of Saudi workers. This means the quota count was according to Saudi wages; 3,000 SR was the minimum monthly wage of the quota or the cut-off wage of one quota. For example, a Saudi who worked at 3,000 SR or above would be considered a full quota, while a Saudi who worked at 1,500SR would be considered a half quota. Saudis who worked with wages less than 1,500SR would be considered zero quotas, while Saudis who worked with wages above 1,500SR and less than 3,000SR would count as a percentage of the quota, based on the equation below.

$$quota\ percentage = 0.5 + \frac{saudi\ wages - 1500}{3000} \qquad 1\text{-}2$$

The sum of these quotas would be used as the numerator to calculate the Saudisation percentage equation.

However, Nitaqat3 uses the Saudisation index, which is built according to five qualitative criteria: Saudi percentage, average Saudi wage, percentage of Saudi women, job stability, and the percentage of high Saudi wages versus high non-Saudi wages. Each criterion was given specific points; the summation of these points specifies the percentage of Saudisation index and a firm's Nitaq accordingly. For each 1% of Saudis,

[17] Non-Saudi workers do not count in the Saudisation percentage until they enter the country.
[18] Quota refers to units for each person to count in the Saudisation percentage, using a moving average until 2015. Then, the MLSD changed the moving average policy, so Saudis count as a quota from the first day of the registration with GOSI, and the average across 26 weeks was considered.

a firm receives ten points; the firm receives six points for each 100SR in Saudi average (monthly) wages, two points for each percentage point of Saudi women from total employees, four points for Saudi stability status, and 0.01 points for the high wage percentage. The summation of those points is used to classify a firm's nitaq, rather than the Saudisation percentage that was used in Nitaqat1 and Nitaqat2. The MLSD provided an automatic calculator on its website before running the new edition formally.[19] Moreover, in Nitaqat3, firms can achieve the required Saudisation percentage by paying a monthly financial fee for each Saudi unit required, which is known as the "al-twtteen al-mowazi" programme, which is the improved version of "Nitaqat al-Mosaned". These fees are paid for the HDF to improve Saudi workers through training and education. Therefore, this could be considered virtual Saudisation. This system depends on the impact ratio (**IR**) for each Saudi unit, which is calculated by dividing the sequence of a Saudi unit by the total number of Saudis. For example, if a firm employs ten Saudis and needs three extra Saudis to move to another Nitaq, the **IR** for the first unit is 1/10, while it is 2/10 for the second unit and 3/10 for the third unit. The final fee would be equal to the summation of those three units: (0.1 + 0.2 + 0.3) * 100 = 60%. Thus, according to the **IR** cost, the monthly fee would be 6,600SR (see Table 1-1).

Table 1-1: The cost of virtual Saudisation.

IR	1–10%	11–20%	21–30%	31–40%	41–50%	51–60%	61–70%	71–80%	81–90%	+ 91%
Cost (SR)	3,600	4,200	4,800	5,400	6,000	6,600	7,200	7,800	8,400	9,000

Source: The MLSD website, government document

This version was proposed to avoid the system drawbacks, unlike the two old versions. However, this will increase the number of women where the new criterion depends on number of women. Moreover, virtual Saudisation would not end temporary or dummy Saudisation[20]. From firms' angle, if the firms used temporary or Saudisation for a Saudi who received a wage higher than 3,600SR, which is the lowest **IR** cost in Table 1-1,

[19] https://www.slideshare.net/Taqno/ss-70271269

[20] Dummy Saudisation refers to a registered Saudi person who is not working to achieve the programme percentage. Virtual Saudisation refer to what is legally proposed by the MLSD allowing employers to pay a fee to achieve the required percentage of Nitaqat.

they would prefer the virtual Saudisation, while if a Saudi earned a lower wage, they would continue in the temporary and dummy Saudisation practice. From the Saudi workers' angle, both methods require workers to register with the GOSI, while the new method – virtual Saudisation – does not need any registration. This means that when the dummy or temporary Saudisation is for a relevant person, it would be more valuable because they could increase years of registration with the GOSI and then benefit from the pension in the future while in temporary or dummy Saudisation in addition to the monthly finance they will receive. This implies that incentives to attempt to evade Nitaqat still exist.

1.6 *Firms' criteria*

Nitaqat has two criteria for firms: sizes and activities. Therefore, firms that are similar in terms of activities and size are supposed to achieve similar Saudisation percentages. Nitaqat1 has several mechanisms.

Mechanism 1: Firms with ten or more workers are classified into one of four segments (Nitaqat): red, yellow, green, or platinum.[21] The green firms are divided into three levels according to the best percentage achieved. This classification is based on the percentage of Saudi workers achieved concerning firms' sizes (see Table 1-2); for example, there are 50 firm activities and four firm sizes, which creates 200 groups, and the number of activities increased to 73 in the next version of Nitaqat.

Table 1-2: Segment based on Saudisation percentage in oil and gas extraction activity.

	Non-localise		Localise			
	Red	Yellow	Green 1	Green 2	Green 3	Platinum
Small 10–49	4%	5–9%	10–24%	25–39%	40–54%	55%
Medium 50–499	9%	10–29%	30–46%	47–63%	64–79%	80%
Large 500–2,999	14%	15–34%	35–51%	52–68%	69–84%	85%
Giant 3,000+	14%	15–34%	35–51%	52–68%	69–84%	85%

Source: Nitaqat guide, issue 1, 2014, p. 68.

Mechanism 2: If firms are owned by one person, despite undertaking more than one activity, and if each activity has less than ten workers, the number of workers is

[21] Localised firms are in green and platinum, while non-localised firms are red and yellow.

considered, regardless of the firms' activities, the only four colour segment policy is applied for all activities.

<u>*Mechanism 3:*</u> Firms that employ less than ten workers are treated differently. They were not subject to Nitaqat regulations in Nitaqat1 and were given a white colour at that time. However, in Nitaqat2, they were regarded as micro firms classified into two colour segments only, red or green. However, they are required to employ at least one Saudi, including the owners. Nitaqat2 updated firm size classification, medium firms considered as three categories: medium A firms employ 50–99 workers, medium B firms have 100–199 workers, and medium C firms have more than 200 employees, up to 499. Thus, MLSD's classified firms into localised or non-localised firms considered to apply rewards or punishments (see Table 1-3).

Table 1-3: benefit and drawback on firms considering Nitaqat classifications.

Localized firms	Platinum	Green
Rewards	Issuing new non-Saudi visa unrestricted by number or occupations.	Unlimited new visa subject for some occupations unless low green level they are banned.
	They are given an extension for a year if social insurance or zakat certificate were expiring.[22]	They are given an extension for six months if social insurance or zakat certificate were expiring.
	Flexibility in renewing existing visa and changing their occupation titles.	
	They benefit from HDF programme of subsidising wages with a limit of 4,000 SR per Saudi worker. This subsidy might be continued for four years, subject to the MLSD's regulations (Al-Ahsa Chamber, 2014).	
	Localised firms can have electronic Saudisation certificates to apply for government tenders;	
Non-localise	Red	Yellow
Punishments	Disable the MLSD's main services; they cannot open new branch.	Ability to renew foreign workers' licence and visa subject to workers time spending in Saudi Arabia if six years or less and 2 years.
	Banned from issuing new visas, renewing existing visas, renewing licences, and changing works occupations titles.	Banned from issuing new visas and changing work occupations titles.
	Worker free to move to another local employer.	
	Non-localised firm's ineligible for tender opportunity.	

[22] It is like a tax declaration.

1.7 Thesis structure outline

The thesis is structured in seven chapters to understand the effect of Nitaqat on the labour market outcome, wages. The **first chapter** “general introductory” provided a general introduction to the Saudi labour market and the Nitaqat program background. The general introduction exhibits the high volume of non-Saudi workers inflow. This inflow was associated with a high unemployment rate among Saudis, which encouraged the authorities to introduce Nitaqat. However, we expected inconsistent outcomes compared to other quota systems because non-Saudis were heavily distributed under the minimum wage of the quota, and they are associated with negative workplace characteristics usually. This was the core of the study problem where the quota principle was violated. Thus, the research objectives are focusing on the wage gap between Saudis and non-Saudis. The importance of this study is shedding light on how this policy could pull Saudi average wages down associated with weak improvements in the unemployment rate. The data was provided by the MLSD for two years, 2013 and 2017, which were the end of Nitaqat1 and Nitaqat2 application period. This can give an implication to improve the program. Moreover, Oaxaca–Blinder decomposition (OB) is the methodology used to achieve the research objective. In the background section, we display some historical regulations and policies concerning the labour market. Additionally, Nitaqat program criteria, rewards and punishments, and associated programs were provided in a simple and clear order. The literature review was the **second chapter**, which helps us to find the research gap theoretically and empirically in terms of the wage differential between Saudis and non-Saudis. We shed light on neglected theories that could explain the wage gap between natives and immigrants, such as the modern research theory, which considered consumption as an appropriate unit to analyse immigrant studies. Moreover, we exploited the hedonic wage theory to refer to workers layoff risk stemming from applying Nitaqat. Empirically, there were almost no studies conducted looking at the wage gap using such data. We fill a gap by providing new explanatory variables: consumption and labour policy. Moreover, we are addressing the wage gap among firms of a given localised status, which was ignored in the quota policy research. **The third chapter** provided details on the data sources. Moreover, the data, which will be used on the estimations was described thoroughly. The descriptions were in terms of the distribution of the data compared to the GOSI dataset to justify the missing data mechanism. Moreover, the normality of the data distribution was also discussed which give us a clear indication on the wage gap and different starting wage (reservation wage) for both groups in both years. Moreover, the outlier observation was diagnosed to ensure if there were real outliers, or a rare observation was observed. Indeed, this description could provide a useful overview for other researchers who might wish to have access to the data. The **Fourth chapter** displayed the method and methodology. We provided a simple, authentic framework to understand how the quota policy (Nitaqat) could affect the native immigrant wage gap. This framework was constructed by exploiting modern research theory and hedonic wage theory through a workers' choice framework, while the existing literature usually uses firm's choice to evaluate the effect of the quota policies on the labour market outcome. Then, we discussed OB methodology thoroughly to provide an empirical base of measuring the effect of Nitaqat policy on the wage gap where the policy caused direct and indirect layoff risk for both groups according to their firm's localised status. In addition, we provided a proposed solution for the identification issue in Oaxaca-Blinder decomposition caused by categorical explanatory variable. This proposed method depends on distributing the constant and added the value to the average coefficients of one categorical variable. This method ensures the stability of the summation of a categorical variable on the unexplained part using any base categories. We conducted the empirical analysis in **the fifth chapter.** This chapter answers the research question based on earnings function estimation, and also a decomposition where we analyse the wage gap from several angles. We found that the gap was significantly explained by education and occupation. Adding the consumption capture the some of the unobserved variable via reducing the intercept value, which indicated that the heterogeneity between the groups in behaviour would explain the existing gap. This requires further information, which we are limited on. However, we cannot refer to the wage gap as discrimination as much as a structural difference where the unexplained part elements (excluding the intercept) indicated discrimination toward Saudi (nepotism). However, Nitaqat played a small direct role to explain the wage gap while it plays a great indirect role through workers redistribution among occupation. This contributed to narrowing the wage gap between the two groups by reducing Saudi average wages by double the reduction in non-Saudi wages. Although the gap was reduced, the policy aim affects Saudi welfare which will increase Saudi dependency on the benefit system in the future. Moreover, the wage gap was higher in localised firms because they employed a higher proportion of Saudis who receive higher wage compared to the non-Saudi average. **In the Sixth chapter,** we take the possibility that using complete cases (CC) provides biased results into consideration, although the data was following a missing at random (MAR) mechanism. We follow the re-estimation procedure using Inverse probability weighting (IPW). Indeed, we expected unequal coefficients, but we expected that the coefficients from using CC and IPW are consistent in following a similar direction and significant level. We found that both methods were consistent when the total sample was considered, while there was a small inconsistency between the two methods when a finite sample used on the unexplained part. However, IPW showed sensitivity among the decomposition approaches and the weight used where it was completely consistent in some approaches or weights and inconsistent in others (one coefficient sometimes). Thus, we advise to be cautious when the IPW is applied regardless of whether CC could be biased. More research in this respect is required. **The final (seventh) chapter** provided a general conclusion for the research, which summarises the overall research on a chapter-by-chapter basis. Then we give some recommendations according to our findings. Additionally, we gave some future research suggestions we expected will be helpful. Finally, we discussed the research contribution to the literature.

Chapter 2 Literature Review

2.1 Introduction

Wage differentials are commonly used to test wage determination theories (Hellerstein et al., 1999). The well-known OB exploits the Mincer equation to understand the differentials between two groups according to their characteristics, such as education, qualification, or age. The human capital theory is tested in this approach. For example, when workers have similar education levels and receive different wages, this could be discrimination. However, the non-discriminatory assumption was developed later in the OB context, taking a group's weight into account. Similarly, the productivity theory was tested by Hellerstein et al. (1999), who used the production function to analyse both the wage and productivity differentials, assuming non-discrimination. When there is no justification for these differentials based on characteristics or productivity, this is considered discriminatory behaviour. Accordingly, affirmative action policies have been imposed to treat this discrimination. These policies were discussed thoroughly in the literature. However, Oaxaca methodology is rarely used to evaluate whether these policies explain wage gaps. Thus, this research combined three types of economic literature: earning determination, affirmative action, and OB methodology. Those three literature types are reviewed in this chapter.

2.2 Earnings function literature

Workers vary in their earnings due to differences in their characteristics, their employers' characteristics, and their job environments. This motivates understanding wage determination throughout the economics literature, both theoretically and empirically.

2.2.1 The theoretical background of wage determination

Several theories have been used to explain how wages are determined, and such theories can be classified historically at three periods: classical, productive and contemporary (Dunlop & Segrave, 2016). The second two periods focused on the micro-level analysis, while the classical period focused on the macro-level analysis.

Macro-level theories are mainly social theories, including classical theory, subsistence theory, wage-fund theory, and residual claimants' theory. The classical theory claimed that wages could be determined through Smith's "invisible hand theory", where supply and demand equate which consisted of the Marshallian modern theory perspective, where wages are determined when the demand and supply of labour are equal (Books, 2015). However, the classical theory was not successful in defining wage determination; it implied that the differentials between workers' skills were the heart of the human capital theory.

Moreover, Smith assumed that compensation would be enough for the workers' needs. This was the point of the subsistence theory, which was developed by David Ricardo and Thomas Malthus (Wood, 1991; Blaug, 1997). Those econometrics assumed that the subsistence price would be around the natural wage assumed by Smith. The increase in wages would cause an increase in the population, which would lead to more employees, then fewer, decreasing wages to the subsistence level in the long run. For the short term, however, John Stuart Mill expressed the wage fund theory, where the demand and supply of labour were equal, taking capital into account. However, this theory considered fixed labour compensation with respect to previous years. Thus, wages could not increase unless the amount of labour decreased. This theory is one of labour demand more than one of wages (O'Brien, 2004). Walker argued that this labour compensation should come from the residual amount after paying for other factors, which was known as the residual claimant theory. However, Karl Marx proposed that labour was the most valuable factor in the production process, and thus the surplus between labour costs and operators' revenues should be given to the labour (Bhattacharya, 2009). Unlike all the classical approaches mentioned above, the Keynesian theories suggested government intervention to keep wages at a full-employment level. Similarly, the new Keynesian approach believed that wage levels could not be determined by market power alone.

There are some important theories at the micro-level, such as marginal productivity theory and human capital theory. The marginal productivity theory refers to the second historical period; it was an extension of the marginal utility of the factor of production (Dunlop & Segrave, 2016). Clark developed this theory, assuming that the wage would be equal to labour's marginal productivity. The additional labour would increase the output; thus, the wage would increase for all labour doing similar jobs (Books, 2015). According to this theory, the marginal productivity of labour and capital are separated, which is not consistent with Taussig's theory. Taussig assumed that capital is created by labour and that those two factors are not separated. Thus, wages can be received in advance from an employer's capital because workers need to cover their consumption. This theory is called the discount marginal utility of labour (Rodgers, 2009; Lokanathan, 2018). The third period contains the contemporary theory, which assumed that the wage was determined by workers and employers in negotiations; this is called the bargaining theory (Dunlop & Segrave, 2016). This theory is consistent with the human capital suggestion, where the returns to education would be higher through the bargaining process when the skills were required for jobs. This implies that human capital theory is reliable and robust (Gottschalk, 1978; Purnagunawan, 2007).

2.2.2 Empirical applications

Those theoretical points of view were modelled in the economic literature. At the microeconomic level, empirically, the earnings function is usually estimated through the human capital theory or the productivity theory. The two theories are connected as the human capital theory assumed that investing in education would yield returns on the wages given by the labour productivity in competitive labour markets.

Admittedly, there were several attempts to determine wages through the productivity theory; the starting point of this research was the production function. The assumption was that wages are equal to the marginal productivities of workers. Earlier, it was found that the differences between wages and marginal productivities were sizable, which supports the human capital theories (Gottschalk, 1978). Purnagunawan (2007) obtained a supporting result; he found that there was no significant change in schooling and experience when the earnings motivation was added. He stated that this was a strong indication of the robustness of the human capital theory. The latter theory was mentioned earlier in Smith's classical theory and productivity theories. The human capital theory highlights that cognitive skills, education, knowledge or any abilities can increase production (Fleischhauer, 2007). For example, wage differentials can be found as a result of different effort requirements (Cahuc et al., 2014). This can explain the gap between different groups in similar jobs, for example, different wages between male and female teachers. If effort can be measured between (groups of) teachers using student performances, then wage differences can be logically understood.

However, Becker and Mincer both explained the return to human capital in their early works. Becker illustrated these dimensions and extended his analysis to human capital investment; some other factors can affect wages, such as turnovers and layoffs (Chiswick, 2003). According to Becker, human capital can be measured as expenditures on education, job-training and labour mobility (Fleischhauer, 2007). Becker also stated that an investor expects a higher return on the labour market and higher productivity to cover the investment cost of human capital if the firm is the investor. Similarly, Mincer indicated that several years of experience could equal specific schooling years. In other words, returning to job training could equal returning to school for several years (Chiswick, 2003). Mincer assumed that when the cost of education is only the alternative cost of earning and when the earnings increase proportionally on a constant base over one's lifecycle, the logarithmic form of earnings depends linearly on the schooling. This is known as the Mincerian wage equation, which is very widely used. His first application of this function was in 1974. He used the ordinary least squares (OLS) on a cross-sectional data of the United States from the 1960 census. He found that 11.5% was the additional wage return for an extra schooling year (Fleischhauer, 2007).

However, the original Mincer equation has been adapted to measure other human capital explanations and other demographical characteristics. For example, education level is significantly affected by education returns, which depend on the majors required in the workplace. Moreover, women and men have different majors that increase their returns (Arcidiacono, 2004). This result reflects a different gender distribution pattern among occupations. Furthermore, educational returns can vary by gender as well. For example, the proportions of return on educated women can be higher than the proportions for men. (Vignoles, 1999) revealed that male wages increased between 12% and 18%, while female wages rose between 34% and 38% in the UK. This result could have come from a higher number of educated men compared to women at that time, which could have decreased the proportion of the return on education. This idea was mentioned during the 1970s when education returns generally fell due to the higher supply of educated employees (Acemoglu, 2002). Educated workers maximise their returns on education by working in their specialities; otherwise, they earn less than the expected return on education. For example, if an accountant works as a receptionist, they will earn a similar amount as an employee who holds a secondary school certificate or lower because the latter gained experience in the secretary job while accountant was studying. Therefore, education mismatch is not profitable (Acemoglu, 2002; Rubb, 2003). The literature investigated the possibility of other factors that could affect the return on education, such as innate ability, parent education level, number of siblings, the ratio of teachers per students, school characteristics and family incomes (Fleischhauer, 2007). Additionally, several control variables could be included in the original equation. For example, firm characteristics, such as age, size and ownership, are also considered significant wage determinants, as a wide range of literature has pointed out for the context of the employee-employer sorting and matching process (John et al., 1999).

The earnings function, or Mincer equation, was estimated with an OLS approach. However, there are greater numbers of researchers who assumed endogeneity. Thus, they used the instrumental variable (IV) or the two-stage least squares (2SLS) approach to be consistent with these assumptions. Using those approaches would increase the return on education compared to the OLS. However, endogeneity is a controversial issue in the literature. Accordingly, some researchers displayed that the limitation of endogeneity was the result of a convergence between the OLS and 2SLS (Angrist & Keueger, 1991). Others found that the OLS had smaller coefficients compared to the 2SLS, which caused the endogeneity issues (Bazen, 2011). Usually, the expected argument is that the OLS estimates might be biased and inconsistent in the presence of endogeneity, but IV estimates can be expected to be inefficient (and they might also exhibit bias in finite samples). Admittedly, there is no rule for choosing an approach; it depends on the nature of the variable itself. When a variable was exogenous and treated as endogenous, it could overestimate the coefficients. Therefore, the argument about the OLS producing less-representative coefficients would not be rejected in this case, and vice versa.

2.3 Affirmative action literature

Smith suggested that any competitive economy would be in equilibrium as a result of supply and demand forces, without any governmental interventions. This is well-known as the invisible hand theory. Under this theory, the equilibrium of an open economy could settle into a non-preferable situation: multi-supply. Firms prefer the cheapest and most productive workers to maximise their profits, which could result in multi-equilibrium and cause a wage differential. However, firms may also discriminate against workers for a variety of reasons, such as race, gender, religion, and disability. This requires a political decision on whether or not to intervene. If governments intervene to tackle this unfair treatment by employers, they have a variety of policies, for example, taxation, wage subsidies, incentives and applying affirmative action on a wide scale. The last policy is well-known, and the first recognition of this policy was in 1961 in the United States (US) by President John F. Kennedy. He introduced what is known as Executive Order 10925, which suggested employing people in the public sector regardless of their background, race, gender or ethnicity (MacLaury, 2010). This led to the introduction of more and more regulations to help people in the US. The Equal Employment Opportunity (EEO) was also set up, although it was not identified by this exact term until 1965. The committee was the beginning of the emergence of the affirmative action policy. Following that, in 1965, the government expanded the remit to include women, and it became known as Executive Order 11246, which was established by President Lyndon B. Johnson. It was the first clear indicator of affirmative action. Over the years, many practices have been introduced in this scope, with different names and approaches. Quotas, as a type of affirmative action, were first introduced in business job opportunities in Norway in December 2003.[23] Companies were required to give 40% of the seats on the board to women, and they had to achieve this percentage by 2008 (Pande & Ford, 2011). It became a well-known policy that was dependent on selecting a percentage or number as a goal to be achieved by targeted groups. These policies were adopted in the Saudi economy between Saudis and non-Saudis; Saudi workers, who earned more, were considered the minority group. With this

[23] In the political world, it began earlier, in 1990, when a law was imposed for women to fill 30% of the seats by 1995.

fact, applying this policy could show dissimilar results compared to other applications in terms of employment, wages, distributions and the wage gap. This chapter exposes the literature with regard to using this policy in several contexts and a variety of dimensions to provide an overview of the expectations of this policy when it is applied as intended.

2.3.1 The impact on target groups' employment

The main goal of affirmative action programmes is to provide jobs regardless of employee characteristics and increase the employment rate for the targeted groups. This policy was attractive to researchers earlier when the programme was set out; noticeably, it has been examined in several pieces of research to see whether it has increased the employment ratio of the targeted groups. Admittedly, there is significant evidence that affirmative action has succeeded in providing jobs for targeted groups. Ashenfelter and Heckman (1974) found that the policy helped increase the ratio of Black males compared to White males in the short term, in 1966–1970, by approximately 3.3% in firms contracted by the US federal government compared with uncontracted firms. However, in the long run, they expected this ratio to be 12.9%. The firms that responded positively to the affirmative action and employed the targeted groups were satisfying the policy, and this led to the opportunity to win government contracts. Therefore, the policy motivation is government contracts. Similarly, Goldstein and Smith (1975) researched firms from 1970 to 1972 to evaluate an affirmative action programme and see whether it was beneficial for anti-discrimination behaviour in terms of race and gender. The result supported the argument that affirmative action helps increase the employment rates for those groups. Moreover, the results showed that race is more responsive than gender to affirmative action: specifically, the employment rate for Black males was increased. Unfortunately, this increase was small compared with the previous period of 1966–1970, as (Ashenfelter & Heckman, 1974) explained. Smith and Welch (1984) reported results that were consistent with other studies, where the descriptive analysis showed that the employment of Black men and women increased throughout the decade from 1970 to 1980. Leonard (1984b) reported similar findings, confirming that affirmative action boosted the employment rate faster for minorities and females in firms under affirmative action regulation in 680,000 plants. His approach empirically compared more than 16 million employees across several demographic groups – Blacks, Hispanic males and White females – that changed in firms throughout 1974–1980. Kurtulus (2012) supported a result similar to Leonard's. He found that the employment average increased among all minorities and women in the contracting firms when estimating the impact of EEO on 100,000 contracting firms in the US between 1973 and 2003. Specifically, the employment share increased by around 4% on average in professional jobs for Black women compared to 1973. Similarly, the share of professional White women increased by 7.3% during the application period. However, both Black men and White women increased their shares in technical occupations. In Kurtulus (2016), he found that this policy increased the employment share on average by 0.871% for Black women, 3.883% for Native American women, 0.603% for Black men and 4.011% for Native American men. He concluded that this policy increased workplace diversity. Beaurain and Masclet (2016) conducted an experimental study aimed at understanding the impact of quotas on women's employment. They applied three scenarios: first, they imposed no quotas, selecting two employees without regard to their gender, education, or age. Second, they imposed a low penalty; of the two employees, one compulsory woman had to be employed as a quota, with a penalty applied if they did not comply. Third, they imposed the same scenario but with a higher penalty. The result emphasised that women were ranked as less preferred without any quota, and this ranking improved after the quota was applied. The policy increased women's employment with no significant difference between high and low penalties. A similar result was obtained from applying a levy-grant scheme in Japan, where the number of disabled workers increased (Mori & Sakamoto, 2018). However, most of the studies supported the belief that the quota programme succeeds in increasing the employment rate of the target groups when the quota places a mandatory percentage for firms to achieve with a motivational reward or levy.

2.3.2 The impact on employment of non-target groups

Even though there is an agreement about affirmative action's capability as a job provider, whether for high- or low-paying jobs, the debate is raised whether affirmative action's programmes affect original employee groups, such as White men, in firms' employment decisions. Affirmative action programmes force firms to employ a special percentage of minorities as compulsory policy. It is expected that the targeted groups' employment rates will increase at the expense of the original groups. This might come from jobs being reallocated between targeted and original groups or jobs being lost for original groups. The latter is the more likely choice with respect to fixed job opportunities in the short term. It is argued that discriminatory treatment in such a policy will result in reverse discrimination (Newton, 1973). This argument is consistent with the majority of white men's opinions views. They worry that discrimination towards them will make them struggle to find jobs. They think that unqualified Blacks will benefit from the policy. In general, Americans assume there will be reverse discriminations (Steeh & Krysan, 1996). Moreover, there are cases registered in newspapers presented as reverse discrimination; for example; 100 Black scientists and engineers received settlements of $3.75 million (Sokoloff, 2014). Also, as a result of EEO, firms employ minorities and women in high-skill occupations, which leads to a decrease in the men's share of these jobs (Kurtulus, 2012). Holzer and Neumark (2000) found a similar result. Their study confirmed that jobs for the targeted groups – White women and Black men – increased by approximately 15%, whereas jobs for White men – the original group – declined by roughly 20%. This reallocation was to benefit women and other minorities. On the contrary, Ashenfelter and Heckman (1974) found a 1% increase for Black workers under contractor firms in the short term, but they expected that this percentage, in the long run, would decrease by 2%. This could imply that the quota system is an effective policy for redirecting firms to allocate jobs concerning government orientation, which makes governments more interested in imposing quotas to combat discriminatory behaviour in the private sector and guarantee vacancies for minority groups. The percentage of non-White employment increased by nearly 50% from 1966 to 1977, and they reported no supportive evidence of reverse discrimination. However, Sokoloff (2014) stated that the minority groups' employment increases being smaller than the White's employment loss is not usually the case when applying such a policy. Corresponding to that, Leonard (1984a) supported the argument that affirmative action does not usually result in reverse discrimination. Beaurain and Masclet (2016) considered that the substitution relationship was a limitation of their experimental study. When the relationship between two groups is substitution, reverse discrimination can occur. Thus, knowing the relationship between the workers in this respect before introducing a quota policy could be beneficial.

2.3.3 The impact on job quality

Firms vary in their responses in terms of giving targeted groups extra preference compared to the original groups when affirmative action programmes are applied. This raises the argument of whether this preference also varies with respect to job quality considerations, that is, whether it provides high- or low-skilled jobs. Kurtulus (2012) found that EEO successfully moved women and minorities into both white-collar and blue-collar jobs in 100,000 contracting firms in the US between 1973 and 2003. These are promising results if an employer requires any level of skill. However, some argue that increasing employment for minorities is true in blue-collar jobs, where employed Black and White women and Black men are more likely to be in the contractors' establishments (Heckman & Wolpin, 1976). This result is accepted logically if the targeted groups are less educated. Holzer and Neumark (1999) found evidence under affirmative action that the hired members of the target group had fewer qualifications from the target groups. Moreover, Indian tribal groups were more likely to choose low-skilled jobs than high-skilled jobs. This supports the idea that minorities might be less educated. Therefore, their reallocation is related to their educational backgrounds, which are reflected as less qualified jobs when quotas are respected (Howard & Prakash, 2012). However, it could be invisible discrimination towards the target groups. Employers may deliberately hire less-qualified workers from minority groups. Therefore, it could be claimed that employing someone from a minority group in a high-position job requires a sharply focused version of affirmative action, such as quotas, to increase the likelihood of being hired for these types of jobs. Supporting this view, a 2003 Norwegian law required 40% of women to be employed on a board (Miller, 2014). As a result of this binding quota, women held high-position jobs. Women being directors also plays a role in improving women's chances of holding high positions on a board (Wang & Kelan, 2013). Similarly, in 2005 in Malaysia, high-position jobs comprised 52.5%, 22.4% and 30.8% of Bumiputeras, Chinese and Indians, respectively, as a result of applying affirmative action in Malaysia. This is consistent with affirmative action's goal to allocate jobs on a needs-based level rather than an ethics-based level to promote a middle layer in Asia (Kabe et al., 2016). Howard and Prakash (2012) found that, in India, scheduled caste groups (an official designation) were more likely to be in high occupational jobs under the quota system. This might relate to quota design, which respects women's positions (Pande & Ford, 2011). From the dissection above, the economy's response to the policy varies depending on the policy design. It helps to engage the target groups but does not guarantee quality according to the groups' characteristics or invisible discrimination.

2.3.4 The change in the wage gap

Even though high-quality jobs can be provided through affirmative action policies, the gap between the target group and other groups can persist. (Ransom & Megdal, 1993) found that although women' wages improved from 1970 until the middle of the 1970s, the relative wage gap between women and men still existed and remained unexplained even after 1977. As their research was on academic workers, it could be considered that the different results stemmed from the differences in publications, which were then reflected in their wages. If this statistic were available, it could be a substantial contributor to explaining the wage gap between the two groups. Ransom and Megdal study was partly consistent with the recommendations of (Patrinos & Sakellariou, 1992). They found that North America Indians were less educated and experienced than others in Canada and recommended educating them in an affirmative action effort to reduce the gap. However, the question still arose of whether the gap would narrow if the government followed Patrinos and Sakellariou's recommendations. One could expect that the gap would narrow to some extent because this gap is explained by only 0.060 through education, according to the study, which means the wages could increase by only 6%. Therefore, education can contribute to enhancing the demographics of North American Indians but cannot solve discriminatory issues. Thus, one can argue that, if North American Indians were educated, they might become less employed in Canada's labour markets, which is not an aim, while their wages could increase slightly. Inconsistently, women who hold higher degrees face less discrimination than those who have not completed their degrees (Montgomery & Powell, 2003). In the main model, however, they found that women's wages improved by 22%, while men's wages improved by only 12%. There was a negative unexplained part on the OB, which indicated that women's educational returns are higher than men. Affirmative action did not reverse the discrimination in earnings when both groups' earnings increased. One can deduce that education is more likely to narrow the gap between two groups, and a supportive policy is needed for the equal employment process. Burger et al. (2016) supported this deduction; education increased Black returns, which narrowed the wage gap. However, using a fixed female proportion under the employment equity policy can help reduce the wage gap by 20% (Reilly & Wirjanto, 1999). This study included a fixed proportion in the regression, which was 26%. Their recommendation passed on the analysis of Arrow's model for heterogeneity and applied a new method of decomposition at the establishment level. They explained that unskilled workers were paid their marginal productivities regardless of their gender, whereas the gender wage gap was observed in the skilled workers. Despite the high percentage assumed in Reilly and Wirjanto's study, Burger and Jafta (2006) reported a similarity. They pointed out that there was a slight improvement in the wages as the upper wage distributions resulted in narrowing the gap. However, Burger and Jafta (2010) found that the reduction in both employment and wage gaps was marginal as a result of affirmative action when compared to open education access for Africans. Correspondingly, affirmative action policies could help reduce the gaps; however, the gaps remain large due to segregation (Groshen, 1991). Groshen discussed how women are distributed in low-paying jobs, where the wage gap is found; however, the gap is around 1% in higher-paying jobs, where the education access policy is found. Therefore, one can state that affirmative action's ability to reduce the wage gap between groups is not guaranteed.

2.3.5 The impact of affirmative action on firms

The discussion above supports the view that affirmative action increases the employment rates of minority groups. The debate is about the impact of this policy on job quality, the wage gap and reverse discrimination. This debate carries many possibilities of affecting firms, especially when targeted groups are being employed at a high percentage despite having fewer qualifications than the original groups. This is not a concern if the two groups have similar job skills, which is the main factor affecting firms' performances under affirmative action. Thus, firms will be influenced by the policy, negatively or positively, depending on the circumstances. It is difficult to generalise (Kaletski & Prakash, 2016)

One critique of affirmative action relates to the influence on reducing productivity and firms' performances or efficiency as a result of increasing the employment of the targeted groups. This is part of the truth if a firm is forced to employ a group regardless of what skills are needed there. Moreover, this practice could reduce productivity and might confirm employers' beliefs that those individuals are unwilling to be employed (Coate & Loury, 1993). Despite the criticism, employing groups based on their skills and wages would help keep firms' productivity stable, meaning that increasing the employment of targeted members would not harm productivity, according to Leonard (1984a). Productivity seems to be the best theoretical evidence to measure targeted groups directly (Leonard, 1989). However, if employment increases, the costs to firms would increase. Firms' responses are consistent with this view; for example, firms downsize employee numbers to avoid being under the policy regulations (Bøhren & Staubo, 2014; Peck, 2014). Supportive evidence was found in Norwegian firms that changed their legal forms to avoid the quota law (Bøhren & Staubo, 2014). This supports how the quota system is costly, and firms seek to maximise value. This contradiction between the firm's aims and government policy leads firms to have inefficient board members or inefficient forms, as (Bøhren & Staubo) suggested. This is about the cost function when it is subjected to additional factors as well as outcomes. Griffin (1992) estimated a contractor firm's costs, and he found that their costs were higher than non-contractor firms by 6.5%. This increase might have resulted from a wage differential. Frankly, this results in increasing the firms' total expenditures when the wages of targeted groups are high, which would influence the firms' performances, as stated by Ahern and Dittmar (2012). In their work, the stock price dropped significantly; similarly, Tobin's Q in Norway after a quota for women was imposed. This quota restriction contributed to employing younger and less-experienced women, which was reflected in firms' performances. Furthermore, there is some evidence confirming that applying affirmative action programmes increases the wages of the targeted groups; for example, women's wages quickly increased when applying the policy in women's job zones (Leonard, 1989). Males in the targeted groups also obtained higher salaries when sectors applied affirmative action (Leonard, 1984).

However, Beaurain and Masclet (2016) found no evidence that affirmative action affected firms' performances. Additionally, one could expect the policy could affect the prices of the product through the increasing wages, as mentioned above. Corns and Schotter (1999) carried out an experiment and they concluded that if the policy were set up without optimal price-preference consideration, purchasers' prices would increase. In contrast, Denes (1997) found that service costs did not increase in small businesses if the number of bidders was not reduced.

Although Becker discussed why firms discriminate under a competitive market, no study includes competitiveness as a control variable. Thus, firms' efficiency studies under affirmative action have different conclusions depending on the market circumstance (competitive or not). According to Becker, firms in a competitive market maximise their profit through discrimination when the prices are given. He suggested that equal treatment could reduce firms' profits if there was discrimination among workers or customers. He also implied that applying affirmative action might harm the efficiency of firms if they worked in a competitive market. To illustrate firms' responses under the competitive market, sharp affirmative action programmes (such as a quota) can be assumed; these are successful in increasing the number of target groups, which shift labour demand for both the target and original groups. Considering the substitution relationship between the groups, in this case, the wages of the target groups would increase; Leonard's result supported this point of view. The supply might be increased for the same groups as well, which would pull the wage down. Regardless of how the original groups responded, the wage could increase as a result of increasing the demand for this group, with an ambiguous effect on the wage from increasing demand and supply. The impact on firms' costs and efficiency could then be uncertain. This could be noticed when the quota is imposed at a specific wage, in other words, when the quota asks for equal wages.[24] When the quota does not impose any wage restriction, firms can reallocate their workers to satisfy the policy and stay efficient. Firms are free to choose the replacements for the lower wages and benefit from their lower wages if they are labour-intensive or, for the upper limit, hire the most educated and qualified workers if

[24] According to the literature, the gap is still large between the groups even though it was narrowed when the affirmative action was applied. See above literature regarding the wage gap.

they require human capital.[25] However, the economic theory has suggested that a government's intervention would restrict firms' choices, which would increase the total cost of the bill (Mazaheri & Mazumdar, 2005). Admittedly, the replacement process is costly. However, firms would cope in the long run. Theoretically, LeChatelier's principle suggested that firms will return to the equilibrium by adjusting their mix of factor inputs. Therefore, if the programme is successfully applied, efficiency will not be harmed in the long run, even if it is in the short run. To summarise, the programme's success depends on balancing the firms' environments and the relationships between worker groups.

2.4 Oaxaca-Blinder (OB) decomposition literature

Since the OB was introduced by Blinder (1973); Oaxaca (1973), it has come to be used as a standard tool in many labour market studies. Mainly, this methodology is used to investigate the discrimination between two groups using a unique approach built initially on using the OLS regression results. This approach separates the mean of the interest variable for two demographic groups into two components: the explained and unexplained parts. The explained part reflects the differences between those groups according to the variations in their characteristics, such as qualifications, education, experience, ages and other variables in the model. The remaining value is the unexplained part, which is the difference due to the coefficient of the model structure. This latter part was known as the discrimination towards the lower group's mean in the labour market (Oaxaca, 1999). However, the discrimination aspect is only part of the unexplained component. For example, it must be acknowledged that there may be differences in the unobservable, too. This methodology has been extended and applied in several contexts.

[25] The literature recommended quotas in education. This education quota would provide minorities with educational opportunities that would help them engage at the upper wage limit. This would not affect firms' performances as they would qualify regardless of their ethnicity, gender or other considerations, with respect to only having the quota restriction.

2.4.1 Methodology developments

As the first indication of the OB depends on the OLS approach, other approaches were developed using this methodology. Researchers have admitted that this methodology is an important tool in the discrimination literature through their great effort to develop beyond the linear regression of earnings functions. Twofold or threefold decomposition provides similar explained parts; however, the unexplained part in the twofold decomposition is equal to the sum of the second and third parts of the threefold decomposition (Blinder (1973). This part is known as the discrimination part. However, the main difference between the twofold and threefold decompositions is that the twofold decomposition involves a reference coefficient's vector.

The first development mentioned above assumes a non-discriminatory vector to capture unobserved differentials alongside the discrimination. This idea was mentioned by Oaxaca (1973) as the index problem. Reimers (1983) pointed to the non-discriminatory possibility. In his three-part decomposition accounting for selection between Hispanic and Black men in America, he used D as an index between the discriminatory and non-discriminatory assumptions, where D equalled one for the latter and zero for the former. Moreover, when D equalled 0.5, it was the average weight of both groups. Likewise, weight was the coefficient for the average group size. This coefficient was further improved by Neumark (1988) and Oaxaca and Ransom (1994); both studies derived different theoretical frameworks proposed for estimating this coefficient vector from a pooled model. The results from the pure-discrimination coefficient or non-discriminatory weighted vector generated similar results under some restrictions; otherwise, it would have generated dissimilar conclusions (Oaxaca & Ransom, 1994). However, this study acknowledged the issue of the reference groups being left out. Adding the reference group dummy to the pooled regression as a correction was proposed by Fortin (2006) and Jann (2008b). The theoretical derivation was different in each approach, yielding different results for the explained and unexplained parts. Fortin used a mixed-data quantitative and qualitative method, while Jann used a quantitative method. However, Jann argued that the pooled estimation of the OB transferred a great amount of the unexplained part to the explained part, while Fortin found that the gender pay gap was roughly closed. Fortin stated that this result came from the soft index in the pooled regression since both approaches had similar additions in the pooled regression yet reached contradictory conclusions. Jann's derivation was weak, capturing the weighted power on the pooled regression of the OB, unlike Fortin's derivation. Jann's method captured the heterogeneity, which explained the part equal to the index coefficient in the pooled model. Thus, the difference between the two groups was captured by the pooled regression, not by separate group regressions. The results from Oaxaca's approach and Fortain's approach were convergent, although the pooled model included the group's indicator in Fortain's approach (Abdullah et al., 2020). Tyson et al. (2013) supported this point of view; they found that the individual pooled regression was sensitive to the lower-paying jobs, which would reflect a lower explained part in the Oaxaca decomposition proposed by Jann. The pooled model proposed by Oaxaca and Ransom generated bias as a result of leaving the indicator out; however, we disagree with the point that it transferred much of the unexplained part to the explained part for two reasons. First, empirically, leaving any variable out of the regression would generate bias; adding this variable would slightly reduce the constant and increase the other variables' coefficients. Thus, the sum of the explained and unexplained parts would increase or decrease depending on the addition of the new variable. Second, the effect on the explained and unexplained parts would be zero in terms of the groups' indicators as a regressor. However, any other variable would impact the explained or unexplained parts because the coefficient of the indicator in the separated estimation would be zero. The result from Elder et al. (2010) denoted that the unexplained part was out of the discriminatory boundary in some cases, while non-discriminatory analysis required the result to be located between the discrimination boundaries. Moreover, in the worst scenario of the pooled male-female wage gap for 1985, which they mentioned, the unexplained part comprised 74% under the pooled approach and 97% in the proposed OLS approach. In this case, the OLS approach missed the group's weight. Moreover, the White and Black wages for a similar year were 50% and 51.6%, respectively, while Group 1's share was 0.927. In this case, the OLS method lost its weighting power, and the unexplained gap exceeded the unexplained part when considering Group 1, which indicated that the amount transferred to the unexplained part was overestimated. Briefly, Jann, and Elder et al. approaches seem to change the assumption of the non-discriminatory structure to solve the clear bias that existed in the original non-discriminatory approach. This bias decomposition allowed for an interpretation of the explained part regardless of the heterogeneity of the groups. This amount is usually left unexplained in most decomposition methods and is interpreted as unobserved discrimination or treatment. The unexplained part is linked with the programme evaluation literature or so-called average treatment on the treated (ATT). Barsky et al. (2002) proposed a nonparametric model as an alternative to the OB to study the wealth gap between Black and White workers. They stated that the unexplained part in the nonparametric decomposition model was identical to the parametric model, ATT. However, they clarified that their explained part had a different interpretation. Similarly, Black et al. (2006) demonstrated that their nonparametric model had the spirit of the Oaxaca decomposition; the first term of treatment was considered the unexplained part, while the second term was the explained part. Black et al. distinguished between two decompositions: the treatment on the treated and the treatment on the untreated. They stated that the first type is more appealing when studying market discrimination against the minority.

Despite the development of this methodology, the linear model faced two challenges: reporting a standard error and the sensitivity of the base. First, the standard error and interval confidence statistics were not reported for a long time when the OB was performed. The first development in this context was achieved by Oaxaca and Ransom (1998); they used the delta method to estimate the variance for the decomposition. The formulas for variance and the standard error were generalised using another decomposition method. This method is still valid in big samples, although it has a weakness; it assumes fixed regressions, which leads to a significantly understated variance (Hill et al., 2018). However, Jann (2005; 2008b) calculated the variance for linear regression concerning the stochastic regression. Kline (2014) revealed that variability ignorance leads to wrong inferences when estimating a variant's asymptotic distribution for the linear model. For non-linear regressions, Fortin et al. (2011) used bootstrapping to estimate the variance and suggested that this was the only approach that could be applied to non-linear regression. Hasebe (2016) calculated the variance of the asymptotic distribution for those non-linear regressions and stated that his approach yielded similar results to those that used a bootstrap. Second, choosing a base category affects the decomposition result. Jones (1983) was the first to address this issue in the literature. He found evidence contradictory to Blinder's suggestion that the result would not change when changing the base categories. According to Jones, the main effect – when the categorical base changed – was on the unexplained part, which was constructed by the coefficient and interaction effects. This did not affect the conclusion of the discrimination part, which was the summation of the two effects, especially when the twofold decomposition was considered. However, Jones and Kelley (1984) ignored that and focused on their in-detail decomposition in their research (see Column 3, Table 2). Thus, the identification issue is inevitable on the detailed decomposition, and it can lead to changing an amount to explain a variable effect (Oaxaca & Ransom, 1999). In this context, Nielsen (2000) extended the methodology to overcome this issue. In the original Oaxaca decomposition, the used parameter faced each value, and Nielsen's approach calculated the constant for both the men and women in the study. This study recommended sticking with the original decomposition if the indicator's set was summed. Moreover, Horrace and Oaxaca (2001) compared Fields and Wolff's approach to the industrial gender pay gap with an alternative measure they proposed. They highlighted the main drawback of Fields and Wolff 's approach: the results varied when the omitted group's category changed, unlike the result from their proposed method, which was invariant. This proposed model was recommended for use on several wage gap concepts, such as native–immigrant and race by regional, occupational or industrial groups. New efforts were made by Gardeazabal and Ugidos (2004); Yun (2005) to find the average effect of an indicator; they ignored the base category approach as it was the cause of the identification issue. The 2004 approach was dependent on the affected coding or deviation contrast coding methods (Jann, 2008b). However, the 2004 approach was the simple average of the other indicator results (Table 1, Column 6); there was no need to estimate several results and average them. The 2005 approach had a similar advantage; it was a great effort, unlike the standard method, which depended on omitting the constant. If this approach could be generalised on the initial OLS regression, as they share similar identification issues, it would be an acceptable method for considering where the decomposition mainly stems from in a regression. Yun's motivation was that this issue stemmed from the lack of agreement on choosing a reference category. He found that average values were a solution for avoiding misinterpretation when the omitted category's variable changed. This was an outward solution, not a substantial alteration of Oaxaca decompositions. Because the OLS is an estimation method while the OB is a calculation method, the differences in their natures generate vital differences in the results. In other words, depending on one base category to estimate the average values of each group of a categorical variable would result in said variable affecting the explained part differently, depending on the omitted base. Once other bases are considered to average the categorial variable, the variable's effect will change accordingly.[26]

2.4.2 Methodology extension

Although this methodology was proposed for linear regression, it is extendable. Several extensions were made to apply this methodology when the OLS approach could generate bias. Despite the dependent variable types – binary, censored or limited – this methodology could be applied through non-linear methods. For example, logit and probit models were suggested to be used for a binary dependent variable. Therefore, these two approaches were an extension of the OB by Fairlie (1999); (2006). Fairlie found a substantial difference between the results of using the OLS and using logit and probit techniques. Moreover, Bauer and Sinning (2008) theoretically extended the OB for several non-linear models, such as logit, probit, tobit, truncated and count data models.[27] Probit and logit techniques were applied in several pieces of research. Belman and Heywood (1990) applied the OB on the two probit equations model to understand the effect of the union on the fringe benefit. They found a small effect of the union on this benefit, unlike other studies that used a single equation. They argued that the results would be the same if linear regression was used. Burke et al. (2009) used the logit model to estimate the differences in self-employment as a dependent variable between the southern and northern UK. The study found several results in this context and emphasised the importance of a regional study, which requires policy variations across the region. The proportion of the people who are the targets of policies vary across

[26] The sum of all categorial variables is fixed in the explained part, while it is changeable in the unexplained part.

[27] It includes Poisson (P), negative binomial (Negbin) (NB) models, zero-inflated (P) models, zero-inflated (NB) models, hurdle models (P) and hurdle (NB) models. The Negbin model is an alternative to the Poisson model when the assumptions of mean and variance are equality violated. Unlike the Poisson model, the Negbin model assumes a quadratic relationship between the mean and variance (Bauer & Sinning, 2008).

regions. However, we partially disagree with (Yang, 2017) recommendation. He used a logit regression to investigate the differences in employment levels between Korean women and immigrants' wives. He found that approximately 30% of the gap was explained and focused on asserting that the language and education would not close the employment gap between Korean women and immigrants' wives. Thus, he recommended a strong affirmative action policy. The employment gap comes from the immigrant restriction rules; dependent family members cannot legally work.[28] Initially, they are out of the labour force. Thus, removing the working restrictions placed on them would change this gap significantly, rather than applying a strong anti-discrimination policy.

However, when the dependent variable is found to be censored, the OLS can generate bias results as well. Thus, the tobit model is recommended to be used. However, Bauer and Sinning (2010) argued that their proposed tobit model was more reliable than conventional OB. They stated there was a slight improvement compared to the OLS results for the original variable when this proposed technique was used. They used an artificial censored dependent variable to examine the wage gender gap in Germany. Moreover, Bauer and Sinning (2011) applied the tobit model on the saving differences between Germans and immigrants. They used the tobit model to show that some people do not save. They stated that adding the remittance to the regression made a vast difference. This corresponds with our theoretical model, which states that future consumption matters. However, using a full-information maximum likelihood tobit model provided satisfactory results, supporting the assertion that the gender wage gap can be smaller when women have high GMAT scores, unlike the OLS result, which did not support the researchers' hypothesis (Montgomery & Powell, 2003). The difference in the result might be because they used the selection tobit model compared to the OLS, but this expects that the selection model was considered in the tobit, unlike the OLS.

In the context of the limited dependent variable, Aristei (2013) extended and applied the double hurdle model for a non-Gaussian specification. His results showed that the remittance gap between permanent and temporary immigration was explained by the

[28] See https://www.ibs.re.kr/eng/sub05_02_03.do

migrants' preferences, not by their characteristics. The expected result was that temporary migrants remit more than permanent migrants. This result is inconsistent with the result obtained by Bauer and Sinning (2011), who found that 70% of the savings gap between permanent and temporary migrants could be explained by socioeconomic characteristics. Moreover, Bauer et al. (2007) extended and applied the count models on smoking differences between males and females. They used the logit and the truncated Negbin[29] to study the determination of the number of cigarettes per day. However, they used several count models to decompose the cigarette consumption: Poisson, hurdle Poisson, zero-inflated Poisson, Negbin, hurdle Negbin and zero-inflated Negbin. They found the gender differences in smoking mainly came from heterogeneity in behaviour. Therefore, the anti-smoking policies would be affected by taking this behaviour heterogeneity into account. They recommended that when policymakers are designing policies, they should take this heterogeneity into account. They recommended using these models on health conditions and workplace hazards as well. However, the study did not provide heterogeneity measurements. They used the unexplained part of the OB to generate their conclusion.

Yun (2007) made an additional extension to solve for some econometric issues when zero was not the sample average. He used Heckman's textbook selection model, where the residual effect was a presented component in the decomposition. This residual solved for the selection issue, and the researcher (Yun) generalised this result when the model suffered from endogeneity or simultaneity. Reimers (1983) extended the two-step Heckman estimation. Yun found that 22.1% of the gap between Black and White females was explained by the residual. Frankly, ignoring this issue would have led to a different conclusion. For example, Albrecht et al. (2004), after extending the Machado–Mata decompositions, found that the gender gap widened in the high wage bracket when accounting for full-time as a selective variable for women, which indicated glass ceiling issues. However, the gap between them was explained by their characteristics when the selection was ignored. Hence, by using the quintile Machado and Mata (2005) decomposition, using the counterfactual distribution of wages rather than the mean

[29] See footnote number 27.

effect used in the original OB, they found that the higher the education, the higher the gap, which contradicts intuitive beliefs. Noticeably, this is an implicit reference to the glass ceiling issue in Portugal and the Netherlands. In Portugal, this result was found without using the selection method, while in the Netherlands, the results were obtained when the selection model used assumed that women tend to work full time when paid more.

From another angle, OB can be extended to a group's indicator, unlike the original version, which required a binary indicator. Allanson et al. (2000) proposed a multilateral decomposition using aggregate regression for the whole workforce as a reference group. This innovative methodology allowed an examination of more than two racial groups compared to the mean of the workforce. This methodology was efficiently used to evaluate innovative policies in South Africa (Allanson & Atkins, 2005). On the continuum variable, Ñopo (2008) extended and applied a continuum indicator for the OB. His indicator of interest was racial groups. He used the racial indicator as an independent variable in the regression. However, Ulrick (2012) argued that his proposed approach was more effective than Ñopo's method. Upon comparing the unexplained part of the two proposed approaches, it is seen that the latter approach allowed variation by income in the unexplained part, while the earlier approach did not. He argued that his proposed model was advanced because his study used a cubic term for the parental income to capture the non-linearity to be estimated easily with the OLS, while Ñopo used an interaction term. This was the strength of Rios-Avila (2019), who proposed an extension of the OB in terms of the continuous indicator. He used a varying coefficients model in his semiparametric proposed approach; thus, his approach was more flexible than the last two approaches. His paper mainly extended the OB using body Mass Index (BMI) as a continuous group variable to address the relationship between wages and workers' weights. His varying coefficient model with a semiparametric approach was more fixed than Ñopo's extension. Rios-Avila tentatively concluded that there are benefits of using this methodology with the endogenous treatment effect when the treatment of interest varies. This methodology could be used alongside the methodologies proposed by both (Samuele & Jeffrey, 2017) and (Delgado et al., 2020). Both studies discussed the endogeneity on the semiparametric model but not in the OB context.

2.5 The impact of immigrants on the labour market

Labour mobility is recognised around the world as having an impact on labour market outcome. It is recognised that immigrants from multiple origins compete differently with natives. Accordingly, if immigrants wage was less than the reservation wage of the natives, the gap in employment and wages will exist. Labour demand and supply interact to determine the equilibrium wage and employment level that affect the wage and employment levels of natives. These features will be discussed below.

2.5.1 Native-immigrant relationship

The assumption of the multiple supply emerges because immigrants come from many economic backgrounds. Recruitment offices are responsible for making wage offers for workers from the sending countries and employers in the host countries. Those offer wages are different among the sending countries for similar skills and occupations. This heterogeneity of origin creates a disaggregated supply based on the origin of immigrant groups. Theories of labour supply are silent regarding how origin affects labour supply (Stier & Tienda, 1992). They commented on the importance of the workers' origin to the labour supply, and they divide their sample accordingly. They aim to study the supply of Hispanic females. They use two- steps estimation, as wage is an endogenous variable in their supply function. They find that immigrant females have less education than their peers born in the US. Moreover, their results show that the labour behaviour of Hispanic immigrant spouses is exceedingly responsive to their earning potential, and, unlike that of U.S.-born white spouses, it is less restricted by their familial roles as moms. Similarly, Bratsberg et al. (2014) point out this issue of multiple sources in Norway when they investigate the wage effect by origin in developed countries, developing countries and neighbouring Nordic countries. They estimate the log wage daily, weekly, and annually on the immigrant share of unemployment and their length of time in the labour force. Additionally, they estimate the elasticity of labour between natives and immigrants. They find that immigrants from the Nordic area have a substantial negative effect on natives. This result is consistent with the factor demand theory of elasticity, where immigrants who are considered close substitutes for natives have a negative effect. Disaggregated immigrant supply would affect labour markets negatively or positively depending on whether the native–immigrant relationship is substitute or complementary (Viseth, 2020). The negative influence would be on the labour market outcomes of employment levels and wages associated with the substitution relationship. Although the disaggregated supply could affect one type of worker such as skilled workers or those from a specific nationality or region, the aggregate labour supply increases and is shaped by this supply.

In this context, Borjas (2003) gives a theoretical explanation known as the national market approach to address the impact of immigrants on the USA labour market. He estimates the elasticity of substitution for workers based on their education, experience, and skills. He defines imperfect substitution as workers having similar education levels but unequal experience. Borjas found that factor elasticities were between -.3 and -.4, which means that immigrants affected the native labour markets negatively because they were substituted for native workers. In Australia, however, Bond and Gaston (2011) find that immigrants have a positive effect on natives, and immigrants who do not speak English have a larger positive impact on natives compared to English-speaking immigrants. This result illustrates the meaning of the multiple supply we assume. However, their results draw from estimating the logarithmic form of weekly wages and working hours using OLS and IV to capture the endogeneity. They use the logarithmic form of citizen labour force as the independent variable.

These substitute and complementary relationships have been examined through labour demand as well as through production or cost function, using each labour group as independent inputs. These relationships received attention in the immigrant–native literature. This effect is calculated through the translog production function by Akbari and DeVoretz (1992) who find that immigrants in Canada are not complementary to apital. Immigrants were not complementary with, or substitutes for, natives both before and after 1971. This result implies a segregation between the immigrant and native labour markets. Bean et al. (1988) used a different methodology. They used a generalised Leontief production model in Mexico. They found a substantially different relationship in the status of natives and immigrants. Undocumented immigrants were complementary to native workers, while legal immigrants were substituted for native Mexicans. Moreover, the aggregate male labour force (unless they are illegal immigrants) are considered a substitute for female workers. Thus, they suggest that illegal immigrants have a small positive effect on the earnings of other groups, while legal immigrants have a small negative effect on the earnings of white (non-Hispanic) workers. Borjas (1983) conducted several studies in the USA using the Leontief approach, and he found a complementary relationship between black and Hispanic groups while a similar relationship between white and Hispanic groups. He suggests that the heterogeneity of Hispanic groups should be taken into consideration. This implies that the demand for immigrants can be disaggregated by origin as well. In 1986, Borjas attempted to understand the labour demand on blacks from changes in the labour demographic. He found that females have a substitution relationship with black men and other men's groups including immigrants, while black men have greater substitution with other men. He found a complementary relationship between blacks and other immigrants, Hispanic or not Hispanic. This indicates that employers' demand for labour is affected by the supply of immigrants, and this will be reflected in wages and the level of employment. Moreover, Greenwood et al. (1997) found that increasing the supply of unskilled immigrants reduces the demand on low- and medium-skilled natives in the US, causing a small decrease in those natives' wages. They used symmetric normalised quadratic semi-flexible function to estimate both production and costs, aiming for the demand elasticity of unskilled workers.

2.5.2 Effect on wages and employment

The final effect on wages and employment from of increases in the aggregate labour supply depends on changes in the aggregate labour demand as well. There are three possible scenarios. **First**, a fixed labour demand associated with an increase in supply because of immigrant entry would shift the supply down. This will affect native wages and employment negatively according to the theory of supply and demand. **Second**, there can be an increase in the demand for labour along with an increase in supply. Because immigrant consumption increases the demand for products and services, there will be an increase in labour (input) demand, according to the theory of factor demand. If that is the case, immigrants will not affect wages and employment negatively. They might have a positive effect or no effect, depending on how this increase compares with the supply shift. **Third**, there can be a decline in the demand for a specific group of workers. This is assumed to have a negative effect on that group. Thus, empirical studies have tried to address this effect, since no specific conclusion has been reached in the literature (Friedberg & Hunt, 1995).

Nevertheless, the final effect of the immigrant is ambiguous. Labour demand will be different for each skill group, wage level and the proportion of participation and employment of native workers (Bratsberg et al., 2014). This implies that the response to the immigrant inflow will vary across heterogeneous groups. For example, lower-skilled native employment increased when the immigrant supply increased, while there was a decline in wages of around 1.2% for less-skilled natives (Altonji & Card, 1991). That study uses first differences IV approaches on cross-sectional data for 1970 and 1980 in the USA. A similar result was found in Germany where native blue-collar wages decreased. However, a small positive effect was found for white collar workers. The researchers assume that the net effect on workers is negative if both effects are weighted (John & Zimmermann, 1994). They estimate wage function through a fixed panel data effect. They use experience and its square, marital status and industry. However, in New Zealand, there was evidence that immigrants had limited impact on native wages and working hours if the researchers control for more educated workers (bachelor's degree). Tse and Maani (2017) extend the national approach of Borjas, adding the area dimensions. They estimate the elasticity and supply shock on wages and working hours, distinguishing between immigrants in terms of the entry age: adult or child in some estimations. They find that current workers' wages increase by less than 1% when immigrant supply increases 10%.

However, increasing the immigrant supply could change native internal mobility – whether there is inflow or outflow from the labour force (Del Carpio et al., 2015). They use survey data in an exclusive analysis of the labour market in Malaysia. Then they use an IV approach to consider whether the immigrants' location is an endogenous variable that can correlate with unobserved variables. By comparison, there is no evidence that immigrants change internal movement or the unemployment rate in Thailand. However, there was a negative association between immigrant inflow and wages. When immigrants increase the total supply by 1%, it causes a decline in Thai wages of .5% (Bryant & Rukumnuaykit, 2013). That study uses a unique variable: the proportion of the poor when estimating the earning function. In Turkey, Tumen (2016) finds that the informal supply of immigrants is an advantage for labour cost in the informal sector, and that is reflected in the prices of products in that sector but not in the formal sector. Immigrant inflow in the formal sector increased by around 0.46%; however, informal immigrants reduced the chance of natives to get an informal job by 2.26% and the unemployment rate increase by 0.77%. That study used a difference-in-difference approach between control and treatment regions before and after the movements of Syrian immigrants (refugees) in 2012.

In addition, there is some empirical evidence that immigrants have no impact on the labour market. Easton (2001) estimates nominal wages, seeking evidence of a negative effect on natives' wages from immigrants. He finds no evidence of an effect from immigrant supply shock on wages. Moreover, when he controls for inflation, the positive correlation between wages and the number of immigrants disappears. He uses the two-stage method from Dickens and Katz (1987). The first stage estimates the wages for individuals and the second stage estimates the relative wage. Cortes (2008), on the other hand, states that the focus of the literature is that wages for low-skilled workers decreases, when he found there was a gain for them. He finds that a 10% increase of low-skilled immigrants decreases service costs by 2%. Moreover, it increases the purchasing power of high-skilled labour. This means that less-skilled immigrants have a net benefit by decreasing the cost of living in the USA. Moreover, Addison and Worswick (2002) conclude that immigrants to Australia did not affect native real wages negatively. To address the native respond to the immigrant's inflow, they use several approaches. They use panel data from 1982 to 1996 to estimate IV and consider demand and supply simultaneously. Also, they use first differences IV with cross-sectional data for 1980 and 1990. Similarly, in the USA, some studies find no evidence of a negative impact from immigrant, contradicting the findings of previous research. For example, Card (2005) reviewed the effects of immigrants on less-skilled native workers, and he found a slight effect on their wages. Moreover, the relative wages of natives who are educated to those less educated remained stable for long time. The less-educated immigrant reaches the average wage of natives, while the second generation seems to be as educated as the natives. After all, Card concludes that the effect of immigration did not seem negative. He states that immigration policy is a world-wide concern. Accordingly, the effects of immigration on the labour market, which vary based on period, country, and methodology, require continuous evaluation to help policymakers to develop the correct policies toward immigrants, where balancing the gain and pain from immigration is needed.

2.5.3 Reservation wage

The reservation wage is an important factor for the employment of immigrants (according to their origin). This is especially true for who come through recruiting offices (the first generation) where the actual wage is set through the bargaining power of the recruitment office. That is, the recruiting office bargains to satisfy the reservation wage of workers who want to work abroad. This wage considers several factors in the sending counties: the average wage they expect to earn in the mother country, the unemployment rate they experience, the standard of living they have, and the safety level in their countries. All these factors could lead immigrant to accept a lower wage than that of natives in the receiving countries. This reflects on the employment (demand) on those immigrants and appears as a wage gap. The farther the reservation wage is from the actual wage in the receiving country, the more attractive the worker is compared to those from other origins. Indeed, this is a type of self-selection by the immigrant.

Constant and Zimmermann (2005) found evidence to support this point of view. They found that Germans receive higher wages, and they have a higher reservation wage compared to other nationalities: Poles, Lebanese, and Turks. Furthermore, Poles earned more than Turks while Lebanese earned less for full-time work in relation to workers who earn less. In addition, if an immigrant holds a degree in the home country, the reservation wage and participation will increase. However, in Denmark the human capital was not important for the expected reservation wage, and they found that full-time Polish workers have a lower reservation wage compared to Turks. Constant and Zimmermann (2005) point out that the most important factor to choose between working status is the difference between the expected reservation wage and the expected earnings from a full-time job. If this difference is large, the pensility of working is high. To reach this conclusion, they use data from two receiving countries Germany and

Denmark. They use probit and structure probit to account for earnings and reservation wage selectivity. They use several variables: age, language, nationality group, religion and pre-emigrant experience and education. A similar conclusion was reached in Australia (Beggs & Chapman, 1990). Among the less educated, immigrants have a lower unemployment rate because of the lower reservation wage because they have a high cost for the job search and a low rate of having a job. Although English-speaking migrants have a higher reservation wage than non-English speakers, they have a lower unemployment rate because their cost for the job search is higher. The researchers use the probit model to reach their conclusion, and the result was in line with expectations of job search theory.[30] This can explain why immigrants are absorbed in a distant labour market where Card (2005) said that all unskilled workers are absorbed in industry in cities with high numbers of immigrants.

By contrast, the reservation wage for second-generation immigrants is higher compared to first-generation immigrants. The reason is that second generation values their earnings according to the country they live in. They try to maximise their utility without looking at the standard of living in their home country or the average earnings there, unless they planned to move back (which is rare) or if immigration policies force them to leave the country. Constant et al. (2017) support this viewpoint. They find that the reservation wage increases over time between immigrant generations in Germany. They estimate net of hourly reservation wage using an OLS approach, controlling for variables such as schooling abroad, ethnic background, age at immigration, time spent in Germany, the duration of unemployment and the last employed wage. They were agreed with the suggestion of Algan et al. 2010 that the native–immigrant gap would not disappear over generations. Unlike the explanation by Card (2005) they state that the second generation could be educated above the level of the native children, implying that education would decrease inequality. Moreover, inequality could be reduced if the immigrant is hired in a skilled occupation because of the reduction of wages for the top jobs in the UK. However, if demand increases, this reduction does not apply

[30] The job search theory expects that when job search cost is low, the worker will increase his reservation wage and stay longer unemployed (Beggs & Chapman, 1990).

(Wadsworth, 2011). In the next section we discuss the wage and employment gaps between natives and immigrants.

2.5.4 Wage and employment gaps

Empirical evidence indicates a general agreement on the existence gaps in wages and employment between natives and immigrants. Some studies used Oaxaca decomposition and others did not. In term of wage gap, the general view is that being an immigrant usually means being disadvantaged. For example, Americans and Canadians earn more than immigrants to their countries. In America, this gap disappears when human capital is controlled for, but it persists in Canada (Smith & Fernandez, 2017). That study uses the multinomial logistic estimation for occupations to understand if the immigrant works in low occupations in both countries. The results showed that immigrants were found in large numbers in the two lower occupations. This study uses the monthly wage gap as the independent variable in the logistic regression. Moreover, between 1995 and 2000, in Germany, immigrants from the rest of Europe were disadvantaged compared to German natives, although the wage gap was heterogeneous based on the country of origin in Europe (Lehmer & Ludsteck, 2011). This study uses Oaxaca decomposition with log daily wages as the outcome variable. The study uses a set of categorical variables such as qualifications in four categories: not being skilled; the size of the firm, depending on the cumulative distribution; the industry level, based on the WZ 03 classification and the classifications for occupations, following Blossfeld (1985). Also, they use region. They also use continuous variables such as tenure and age with squared forms.

However, migrants are not usually disadvantaged when the migrant’s origin is considered. In Spain’s labour market, Simón et al. (2008) find that those who come from developing countries experience disadvantages and work under segregation. This is different for immigrants from developed countries. They earn more than native Spanish workers because of their characteristic endowments. The logarithmic hourly wage is used, calculated as the annual earnings divided by the number of yearly hourly wage. The age is limited to between 16 and 65, and independent variables include 11 education categories, full-time work, and experience (and its square). Firm size is divided into five categories, and the Gini index and the Theil index are used for descriptive analysis.

However, in terms of how background affects the employment gap, there is a variety of results. For example, Kee (1995) finds discrimination in the Netherlands by around 11.85% in favour of Moroccan immigrants. Moroccans can earn more than natives when their education and experience was obtained in Morocco. However, when their education and experience occur in the Netherlands, they earn less than natives. Immigrants from Suriname have the lowest discrimination coefficient in the wage gap – 1.29%. Antillean Turkish immigrants are treated similarly; they face higher discrimination in general. This could be related to the higher participation of Moroccans who are recruited compared to those who are second generation, unlike other immigrant groups origin. Kee assumes that education and experience accumulated in the source country helps to explain the gap. Kee uses OB between natives and each immigrant group separately; he estimates four decomposition regressions to address the origin differential. Moreover, he uses the weighted approach proposed by Oaxaca and Ransom (1994) with sample selection of monthly earnings of a wife or other family members. The study uses log hourly wage as the dependent variable. For independent variables, Kee separates them. For example, he considers if schooling and experience (its square) occur in the home country or the Netherlands. Also, he uses dummy variables such as marital status, speaking Dutch with difficulty (subdivided into some difficulty and much difficulty), and living in Amsterdam. Like Kee (1995), Longhi et al. (2012) find that both first- and second-generation Hindu Indian workers earn more than White Christian British workers. They assume that the wage gap is not based on discrimination against religion but that it is related to other unexplained characteristics. Moreover, they find that the second generation earns more than the first generation in each minority group. The study uses log hourly wage as the outcome variable and age is limited to between 34 and 64 years in the categorical set. Dummy variables are full time, public sector and qualifications (set at several levels including ‘no qualifications’ and ‘other qualifications’). Firm size and region size have three categories based on the number of employees. For occupations, the standard occupation classifications (SOC) are used. The researchers use some third-digit division in areas of immigrant intensively, in addition to the first-digit division. Nielsen et al. (2004) find support from Denmark for the idea that the second generation earns more than the first generation, as Longhi et al. (2012) found in the United Kingdom. Nielsen et al. (2004) conclude that the immigrant–native wage gap narrows as the number of years spent in the host country increases. They use the consumer price index 1995 to deflate hourly wages. For explanatory variables, they use occupation, age (limited to between 20 and 59) and age squared, years of experience in Denmark and number of children. Moreover, they use missing occupations as separate categories. The findings by Frank et al. (2013) are very different from the literature; they find that recent immigrants are severely disadvantaged. However, the immigrants are disadvantaged based on demographic and ethnic characteristics, while the wage gap advantage is expressed in term of human capital. Log yearly hourly wage is the independent variable. For dummy variables they use gender, marital status and whether the person voted in past elections. They use age as a continuous variable, and there are several categorical variables: minority status or origin, language spoken, area in Canada where they live and education level.

Following the general belief that immigrants are disadvantaged groups, some researchers explore if women face a double negative. Hayforn (2002) uses Oaxaca decomposition to discuss this possibility in Norway. He decomposes the annual log earnings on age and years of schooling. Dummy variables are marital status, whether they work in manufacturing, industry, region, and country of origin. The interaction variable is between gender and ethnicity. The results show that females experience the double disadvantage of being women and immigrants. However, the gender pay gap has much effect compared to the ethnic wage gap. The gender gap contributed between 79% and 93% while the ethic gap contributed between 12% and 21%. The result of the double negative mirrored the findings by Boyd (1984), who uses multiple classification analysis. Boyd aims to address the wage gap in terms of occupation status for immigrant women compared to Canadian women. He finds that females work in the lower occupations and immigrant females are lower when place of birth is considered. In several Europeans countries, Tverdostup and Paas (2019) found that a gap still exists even though natives and immigrants have similar occupational skills, cognitive skills and demographic characteristic. They assume that the wage gap between immigrants and natives can be explained because they use different skills in the workplace, with immigrant using the lowest skills. To clarify, this study does not use the decomposition methodology, it used multivariate regression analysis. However, the gap is substantially explained when they control for job mobility (Brenzel & Reichelt, 2017). This study uses four mobility dummy variables in the person-specific fixed effect method.[31] The wage gap is measured by the logarithmic hourly wage. Length of employment, unemployment and experience and its square are used as continuous variables. Firm size is a categorical variable.

An employment gap between natives and immigrants also has been recognised. In Sweden, Luik et al. (2018) find substantial heterogeneity when the admission background is taken into account to evaluate the employment gap between males from Sweden and immigrant males. The lowest gap is in immigrants' labour compared to those who were family or humanitarian immigrants. They use unconditional decomposition alongside separate decomposition. Moreover, they use a fixed effect for each country in their Oaxaca probit decomposition. That study uses employment status as the outcome variable to measure the employment gap. Although it is not relative to the wage gap, it uses similar independent variables such as categorical variables for education levels and education types. ("Unknown type" is used as separate type in the education category). Continuous age is used, and marital status is a categorical variable, unlike most of the previous studies that use it as a dummy indicator. Another categorical variable is country of origin. In this context, native Swedes are less likely to be self-employed compared to non-Western immigrants (Joona, 2010). Joona finds that the exit rate is higher among those immigrants compared to the Swedes, and this gap remains unexplained. That study also uses a binary variable for the outcome, but it still differs from wage as an independent variable. Place of birth is used to denote the origin of an employee and wither he was born in Sweden. He uses education levels, not the actual education achieved. One categorical variable is main industry, and a category called 'unknown' also is used, in addition to the continuous variables of age and number of years in Sweden. Notice the outcome variable includes being self-employed. From another angle, Kil et al. (2018) find that whether it is a full-time job or a part-time job

[31] They assumed three main types of mobility or job changes: voluntary, involuntary-layoff and internal. They used 'other' to capture other reasons for changing jobs.

makes no difference with respect to the migrant's generation and origin compared to the native Belgian. This indicates that the differences between natives and immigrants is the choice to continue in the labour market after parenthood. Accordingly, from another angle, the possibility for being employed for a native mother after maternity is higher than for immigrant mothers according to their origin and generation. This difference is greater for first-generation mothers than for second-generation mothers compared to natives. The researchers expect that first-generation women had weaker networks to help them resume work. Moreover, Turks and Moroccans have strong family formation unlike immigrants from southern or western Europe. This could explain the low participation by Turk and Moroccan mothers. The study uses Belgian panel data from 1999 to 2010 for females between 15 and 50 years old. It uses the binary outcome of employment versus unemployment regressed in hierarchical regression models. The independent variables include the dummy variables of marital and motherhood status, region, industry and pre-birth employment.

2.6 Model specifications

In this section, we review the theoretical and empirical studies that relate to our research from three angles: affirmative action, earnings, and Oaxaca decomposition. Then we discuss some related studies to approach the model specification.

Its terms of the dependent variable, it should be noted that wages can be measured under two theories. **Production theories** assume wages are equal to the marginal productivity of workers, so wages could measure for example by sales per worker (John et al., 1999). However, Gottschalk (1978) considers Mincer's earning function as an appropriate way to measure earnings compare to the productivity function under **human capital theory**. There are studies that follow Gottschalk's consideration for using Mincer's approach to estimate the earning function (Purnagunawan, 2007; Bazen, 2011). This function is used in wage gap literature presented by Blinder (1973) and Oaxaca (1973). Since then, wage gap depending on human capital theory has been measured as a dependent variable (wage in the logarithmic form). Wages take several forms in the literature, with some studies using log hourly wage as the outcome variable. (Longhi et al., 2012; Brenzel & Reichelt, 2017; Abdullah et al., 2020). By contrast, Nielsen et al. (2004) use the consumer price index 1995 to deflate hourly wages. Simón et al. (2008) follow another approach. They calculate logarithmic hourly wage from the annual earnings divided by the number of yearly hourly wage. Frank et al. (2013) use log yearly hourly wage as the outcome variable, while Reimers (1983) uses annual earnings data. Daily data in logarithmic form also has been used (Lehmer & Ludsteck, 2011) as has the logarithmic form of monthly data (Mahdi, 2005; Smith & Fernandez, 2017). This indicates an agreement in the literature for using the logarithmic form of earnings, although they use several levels of wage: hourly, daily, monthly, or annually according to the data available.

Wage determined in Mincer's equation by the independent variables such as experience and qualification. However, over time other dimensions have been added to these human capital variables. We discuss how other studies use the independent variables to help us to specify our models. **First**: worker's characteristics, qualification and education are used in the literature to measure the return from education on wage. Agrawal (2014) uses education categories such as illiterate, primary, secondary until graduated, and Razzolini, et al. (2017) classify qualifications as apprentice, white collar and blue collar. Some studies use qualifications, others use education, and still others use both. For example, Luik et al. (2018) use two categorical variables including these education levels: pre-secondary, post-secondary, and scientific. Then they use nine education categories, such as health, humanities and technical. They use 'unknown' as a category to capture those whose education type is missing. Notice that they use an indicator category (IC) to treat those workers who are unregistered or whose qualifications or education are unknown. Lehmer and Ludsteck (2011) use a method to capture those with missing qualifications. They combine them into three qualification levels: low-skilled, skilled, and highly skilled. Notice thyhe depend on the secondary education level to select the skilled and highly skilled categories. Longhi et al. (2012) use their own classification and describe qualifications as levels. They include 'no qualifications' and 'other qualifications' as categories. However, unlike qualification, education can be measured either as a continues variable or as a categorical variable using years of schooling. Kee (1995) uses number of schooling years spent in Netherlands and the number of schooling years in the home country, while Fernandez (2017) uses three variables to measure the effect of education on wages: years of schooling, numeracy test and literacy test. Moreover, education has sometimes been used as binary data, as well, where it equals one if the worker has had higher education and zero if not (Joona, 2010).

Age is continuous, but it is sometimes used as a categorical variable instead (Longhi et al., 2012; Luik & Steinhardt, 2016). In rare occasions, its logarithmic form is used. For example, John and James (1999) use this for measuring the production function where the outcome variable is log sale per worker. The continuous form has been use in several studies as a demographic variable (Joona, 2010; Lehmer & Ludsteck, 2011; Frank et al., 2013; Luik et al., 2018). However, some researchers use it as a proxy for experience, but this has been found to generate biased coefficients. Therefore, some tend to calculate years of experience from age information to reduce bias. Reimers (1983) follows this approach when he subtracts highest school grade from actual age. Age could link with wages in a linear relation such as the study by Luik et al. (2018). They limited their sample to those between 25 and 59 years old, while Longhi et al. (2012) limited their sample to those between 34 and 64 years old. This specification helps the researcher to know the wage trend when workers get older. The signalling theory could be a valid logical reason to include this variable in the regression, where a linear term gives a direct link between age and wage. A positive relationship means workers signal to the firm that they have developed their expertise through experience accumulated over the years. If it is negatively related, instead, the workers signal that they were more productive when they are younger, particularly for jobs that require physical effort. However, age is usually used as a non-linear variable where the linear term is combined with age-squared, cubed or to the fourth power. Most research does not go beyond the age-squared term. This helps to address the concavity and convexity of the age curve. A concave curve indicates that a person's wages increase until a specific age and then they decline. In Pakistan, Siddiqui et al. (1998) find a significant concave age relation for both male and female earning estimations, while Javied and Hyder (2009) find a significant convex increasing trend when they estimate the earning function. Unlike in post-Soviet Russia, Borisov and Pissarides (2019) find a convex relationship between age and wage in the selection earning estimation.[32] Age could be cubed, as well. Gottschalk (1978) finds that significant results are generated for all polynomial age variable address the age-productivity profile. Indeed, choosing a linear or non-linear term depends on understanding the data and using a statistical test to ensure the correct version of the age.

Gender and country of origin are treated similarly in the literature. Some studies use them as independent variables in earning regressions, and some studies prefer to estimate the earning function separately for each gender or country of origin. Hayforn (2002) uses three categories to refer to workers' origin: Nordic, developed and less developed in separate gender regressions, while Frank et al. (2013) include both gender and origin in their earning function estimations. However, to achieve homogeneity females are excluded from the estimation that condition to countries of immigrants (Kee, 1995; Lehmer & Ludsteck, 2011; Longhi et al., 2012). The focus of those studies addresses the immigrant wage gap among males only. Other studies add gender only. Rand and Torm (2013) use a dummy variable if a male hold the job to estimate the wage function. Similarly, Smith and Fernandez (2017) and Abdullah et al. (2020) use gender as a regressor in the estimation, and they do not distinguish on the basis of an immigrant's origin. Simón et el. (2008) estimate using both strategies, gender is included as an independent variable in a separate immigration origin (developed countries and developing countries), and a separate gender estimation is conducted concerning the country of origin.

Second, workplace characteristics. These might be expected to explain a lot about the earnings function, and they play a great role in inequality studies as well. For example, firm size is included in some research as a categorical variable. Longhi et al. (2012) put firm size into three categories according to employee number: 0–25, 26–250 and 250+. This differs from the UK definition even though his study is about the UK. Moreover, Hofer et al. (2017) use firm size as an explanatory variable to study the immigrant gap in Australia. They use five categories: 0–9, 10–19, 20–49, 50–499 and 500+. Lehmer &

[32] Borisov and Pissarides (2019) use dummies in the estimated equation, and they use continuous age and age-squared in the Heckman selection model. This approach helps them use similar variables in the estimation.

Ludsteck (2011) defined firm size according to the cumulative distribution. Although firm size is defined differently in each study, there is agreement in the literature that wages are linked positively with firm size. In Germany, Schmidt and Zimmermann, (1991) find that high wages are associated with large firms. This could explain why large firm survival under high-cost labour. Bílková (2019) comes to a similar conclusion even when he excludes the salaries of management. Furthermore, firm age is an important explanatory variable to indicate the owner's experience (Rand & Torm, 2012). The belief is common in the literature that there is a positive relationship between wages and firm age. Brown and Medoff (2003) found that firm age was associated positively with wages when they control for the employer's characteristics, but it is linked negatively when they control for workers' characteristics. Another interesting result is that wages decline as firm age increases, and this relation reverses direction for older firms. The study uses age and logarithmic age in the model, and this might explain the non-monotonic relationship between firm age and wages. Firm age could be used as a dimension for policies rather than firm size because many firms struggle in the first seven years (Coad, 2018). Moreover, we rarely find firm age use in decomposition research. For Rand and Torm (2012), firm age does not play a significant role in explaining the wage differential between formal and informal households in Vietnam. This study uses the logarithmic form of the firm age.

A firm's economic activity and occupation are considered vital job characteristics. Studies control for these variables as categorical variables. Lehmer and Ludsteck (2011) use the industry level, based on the German classification WZ03, and they follow Blossfeld's (1985) occupation classification. Both variables are used as control variables in their decomposition. Hofer et al. (2017) include both variables in their regression, and their classification seems close to international classifications for occupation and industry. Hofer describes those variables as labour market characteristics, and he finds that immigrants are in low-occupation categories. Similarly, Longhi et al. (2012) use the standard one-digit classification for occupation, also they use a three-digit occupation code when immigrants are involved. They find that occupations could explain the high and significant advantage for the wages of both first- and second-generation Indian Hindus. Longhi et al. (2012) use occupation only as a control variable, while Joona (2010) uses only industry as a control. Joona's classification contains 'unknown industry', and she finds that industry provides the best explanation of the employment gap by around 30%. Unlike these studies, Rand and Torm (2012) use a dummy variable of one if the company belongs among high-tech industries, which are defined as the sector between 30 and 35 in the international industrial classification, otherwise, it gets a zero.

Firm location and geographic location also have been used in the literature. Burke et al. (2009) focus on the variation in England between north and south in terms of job creation by the self-employed. In the north, they find that less-educated people did more job creation by being self-employed. This is compared to well-educated people in south. This implies that firm location is an important variable, particularly for policymakers. However, immigrants are usually found in large cities. Hofer et al. (2017) use Austrians' federal state and city size as control variable. The main finding of their study is that discrimination occurred in the highest wage category. Similarly, Lehmer and Ludsteck (2011) aggregated the German Federal Office classification for Building and Regional in 5 categories. Moreover, firms' area plays a role in explaining the wage gap (Rand & Torm, 2012). That study uses a dummy variable if the firm is urban or rural and if it is in the north or the south.

Additionally, there is the quota policy indicator. This is used in the literature as a dependent variable (Kurtulus, 2016). That study uses several regressions where the dependent variable is percentage of males or females according to their ethnicity regressed on the federal contract status. The study shows that black and native American men and women benefit from federal contract under an affirmative action policy. However, the quota variable has been used as an independent variable as well. Wang and Kelan (2013) perform several regressions to address the effects of female quotas in Norway. They use the proportion of females among the total employees and a dummy variable such as female chair. Similarly, female proportion is included in the regression to decompose gender gap and find the effect of the quota on female wages. It found that the female quota helped to increase women's wages at a high percentage – 26% (Reilly & Wirjanto, 1999). Similarly, Groshen (1991) uses the percentage of women in both the workforce and the occupation to examine the effect of the quota on the gender gap. However, the proportion of female was used to investigate the gender wage gap but not for the quota context. It is used as a control variable instead of being a gender dummy variable (Rand & Torm, 2012). Similarly, the share of citizens in a region is used when the researchers look for the reasons for the wage gap between Germans and immigrants (Lehmer & Ludsteck, 2011).

Finally, some variables are used to serve the research aim or because of data availability. Examples are the public sector (Longhi et al., 2012), mobility variables (Brenzel & Reichelt, 2017), living in Amsterdam and difficulty speaking Dutch (Kee, 1995), voting in past elections (Frank et al., 2013), number of children (Nielsen et al. 2004) and the Gini index and Theil index (Simón et el., 2008). Studies use the interaction of education and time or between cohorts (Burger et al., 2016), and the interaction idea is used for gender and ethnicity to explore whether immigrant women face a double negative effect (Hayforn, 2002). However, marital status is used frequently as a dummy variable (Kee, 1995; Frank et al., 2013), but Luik et al. (2018) use marital status as a categorical variable.

2.7 Theories explaining the wage gap.

The wage gap between native and foreign workers was thoroughly discussed theoretically and empirically in the economic literature. Empirically, significant evidence for a native–foreign wage gap exists in different countries (Mahdi, 2005; Simón et al., 2008; Lehmer & Ludsteck, 2011; Parodi et al., 2012; Himmler & Jaeckle, 2017; Razzolini et al., 2017). When individuals' characteristics are similar and earnings different, according to their home origin, this can be interpreted as discrimination (Lehmer & Ludsteck, 2011). However, the gap could be, contradictory to the common belief, in favour of immigrants (Kee, 1995; Simón et al., 2008; Longhi et al., 2012).

Several pieces of research tend to explain the wage gap through discrimination literature when the unexplained part is substantial (Arrow, 1973; Agrawal, 2014). However, there were some research explained the gap the classical theory under the contexts of human capital (Becker, 2010; Collard, 1972). The human capital theory (specifically, educational level or qualification) can explain substantial amount of the wage gap between native and immigrants. Kee 1995 showed that over 60% of the gap was explained by the education and experience for each group. Moreover, Cahuc et al. (2014) in his book, suggested that wage differentials can be found because of different effort requirements. This can explain the gap between different groups in similar jobs, for example, different wages between male and female teachers. If effort can be measured between (groups of) teachers using student performances, then wage differences can be logically understood. Similarly, in the US, cognitive skills explain a substantial amount of the stockholding gap in the financial market, which can explain the superior wealth of those of US origin compared to immigrants (Luik & Steinhardt, 2016). Accordingly, human capital theory suggested, this wage gap can be understood and closed when education is increased for immigrants. However, this suggestion is not true in all countries. In Canada, for example, the gap was present even though immigrants were highly educated (Smith & Fernandez, 2017). Although Becker's suggestions can partially explain the wage gap through the differences in education, age and experience, this suggestion is not usually sufficient to explain the differences between native and foreign workers. This belief stems from the fact that natives earn more than others, even when they have similar characteristics. This applies to the productivity theory, as well.

This implies that the wage gap cannot be explained only through microeconomic theories, such as the productivities theory and human capital theory. Therefore, the human capital and productivity theory can be combined with other theories to explain native–immigrant gap. Accordingly, some research associated this gap with the wage structure, where immigrants engaged in the labour market in the low job categories; this might be due to the recruitment policies (Hanson, 2009)[33]. This could explain through immigrant self-selection (Smith & Fernandez, 2017), or entry quota (Ruppert, 1999). Undoubtedly, this last point indirectly suggests that the possibility of changing of wage structure when hiring quota considered. Moreover, duality theory could contribute to a further explanation for this gap. In the dual labour market, where there are usually

[33] In this thesis, **recruitment policies** mean the policy and the process of bringing non-Saudis to the country, not the general employment policies.

primary and secondary jobs, foreign workers work in secondary jobs, where the nature of the pay is low. Meaning, foreign workers might negatively select themselves jobs or industries, even in the secondary sectors. This could also be a result of the recruitment policies, as mentioned above. It is not surprising that immigrant workers are found in the secondary sector if we take into consideration the expected wage gap between the first wage in the home country and the wage in the distant country, as the neoclassical theory suggested. In the extreme case, when the expectation of this gap is very high, foreign workers might accept jobs under their qualifications. For example, one could work as a salesman when his qualification was accounting. This mismatch mentioned might be an individual decision for the foreign worker. However, the mismatch theory could partially explain why the wage gap exists between native and foreign workers as unexplained components. Royalty et al. (1993) modelled matching and turnover behaviour between men and women and interpreted their results in the context of the unexplained gap. They found a key result that matching behaviour can be explained by the wage gap. The case of negative selection and mismatch can also end with a similar conclusion of the migration theory; the native–immigrant gap stems from the rule of the international movement. Therefore, the international wage gap increases the migration return, regardless of how much they receive compared to natives.

Recently, in this respect, there has been an attempt to find other theories to explain the native–immigrant wage gap. For example, signalling theory have been used to explain part of this gap through a differential in movement rates. Brenzel and Reichelt (2017) found strong evidence that the differences in transition behaviours between natives and immigrants explain a substantial amount of the gap. They used Pearson's fixed-effect method, which was an approach for measuring wage differentials proposed by Schmelzer (2012). This study did not explain why those immigrants accepted being highly involuntary mobile; they could self-select on those types of jobs. Brenzel & Reichelt stated that involuntary mobility signals that an employee has less ability. This does signal decreased costs for the employer as there is less turnover. We partially disagree with this theoretical explanation. Accordingly, this theory was more acceptable for explaining the employment gap. Moreover, the job shopping theory implies that those who are voluntarily mobile can achieve higher future wages. This implies that natives can be voluntary mobile, which means they can seek jobs with higher wages. This mobility is not compulsory, which discourages mobility if workers find satisfactory job environments compared with a slight wage variation (Sousa-Poza & Henneberger, 2004; Wheeler et al., 2007; Hinz & Lechmann, 2018). The expectations of their responses would vary if the movement restriction were removed; if they were in higher-wage jobs, the probability of finding an employer who paid a higher wage than they already received would be low (Sousa-Poza & Henneberger, 2004). According to Brenzel and Reichelt (2017) the uncertainty wage reward because of job mobility could explain the wage through the job-shopping theory, while the signalling theory could imply that workers who had less mobility were more preferable, but this did not explain the gap.

In contrast to (Brenzel & Reichelt) study explanation, which assumed four mobility levels and concluded that different rates of movement caused the wage differential, the difference in voluntary movement indicates differentials on group policies, which relatively implies heterogeneity. However, this heterogeneity is considered the core of hedonic wage theory, which considers that the differentials in wages stem from the differentials in jobs characteristics. This theory assumed that, risky jobs have higher wages than other jobs, which counters the wage gap. Bloch (1979) find that layoff risk associated positively with wage rate. He used OLS regression where number of layoff risk for firm dependent variable and average hourly wage independent variable with several control variables. Similarly, Hutchens (1983), found positive relation between layoff and wages using logits regression where the layoff a dummy variable. He theoretically developed a framework to examine this relationship using the expected utility function while other study used the Cobb-Douglas function (Lin et al., 2019). However, there was some theoretical evidence that hedonic wage theory relation could be reversed where workers could move to lower layoff job although they would cut their actual wage (Pinheiro & Visschers, 2015). This theoretical viewpoint was supported empirically, where it was found that workers with less possibility of layoff earn higher compared to those high risk in both method OLS and IV (Theodossiou & Vasileiou, 2007). However, Scicchitano et al. (2019) found that Italy workers in aa secure job can earn more than workers in insecure job. Furthermore, they found that insecure job is more likely to be found on the bottom wage scale. They used Oaxaca decomposition where wage was regressed on several independent variables such as age, gender, occupation mobility, stability, and some interaction variable. This relation was found in Germany (Hübler & Hübler, 2006).

According to the nature of the native/immigrant studies, it would be appropriate to explain the differentials through migration literature, for example, neoclassical theory. This theory suggests explaining the wage gap through individual choices, where immigrants are motivated by the earning and employment gaps between the two countries. It considers the earning gap between countries as a fundamental component for international movement, unlike the modern theory of immigration, which argues that the wage gap between countries is not considered a necessary condition for movement to occur. The general belief of this theory is that labour movements would continue even if the international earning gap did not exist because of individual risk diversity (Massey et al., 1993). We agree with this point of view when there is a high unemployment rate in the sending countries compared to the receiving countries. However, this theory considered consumption, for example, an adequate unit for investigating the migration research instead of individual choices. These two theories explained the reason behind the international movement's initiation phase at the macro level. Correspondingly, at the micro level, when the net return for the individual was positive, the movement would exist regardless of the native–immigrant gap.

In general, foreign workers do not earn as much as natives do; this attracts employers and signals that they are similar products at lower costs. Indeed, for both high and low wages and whether immigrant policies are present or not, any wage higher than the wage the immigrant expected to earn back home is their reservation wage. If the international movement were free, they would move for several reasons, not only wages, which is not the case for most of the world. However, the neoclassical theory discussed how wages allowed international movement. We could name that as the reservation wage for international movement. This wage would create the wage gap as it was expected to be less than the native reservation wage, especially if those foreign workers came from low background countries where their reservation wages would be less than the host country's offering wage through the recruitment offices. The different reservation wage means immigrants could accept any wage higher than the compensation in their own countries, as the neoclassical theory suggested. This could explain why the first generation of immigrants has a bigger wage gap than the second generation (Nielsen, 2000; Longhi et al., 2012). The international restriction of immigrant movements encourages the neoclassical theories as immigrants must leave after they complete their contract period, or they will be ruled by legal penalties; migrants tend to benefit from the gap between the international prices (wages). Even if they tended to vary their risks, as the modern theory suggested, they would be legally forced to return to their own countries. This restriction makes the neoclassical theory more convincing for explaining international movement. The reservation wage should be fulfilled by their current and future consumptions, which explains why the gap is in favour of immigrants if they come from high background countries, as (Kee, 1995; Simón et al., 2008) found.

2.8 Saudi labour market literature

The Saudi labour market is considered a unique economy where migrants have formed the majority, unlike other economies, where migrants have formed the minority. Thus, Becker did not distinguish between the different origins of immigrants and considered them minority groups. An open economy like Saudi Arabia faces an infinitely elastic supply side, which makes employers' choices more flexible. Firms demand migrants as they are often more skilled than Saudis or, at least, have similar qualifications with lower reservation wages when they are recruited from less-well-paying countries. Therefore, the problem does not come from discrimination based on gender, race and minority status.[34] It is a matter of optimal choice for productions where wages are an important criterion for maximising profits. Thus, it is expected to hire more foreign workers than Saudis as the economic theory suggested; when factor prices decline, the labour demand should increase (Ehrenberg et al., 2016). The density of Asian workers could be explained through this theory as well since those workers earn less than other non-Saudi groups, according to Al-Farhan and Al-Busaidi (2019). This study

[34] This could be religion or disability, for example.

decomposed the wages of three types of immigrants: Western, Asian and Arab from the GCC.

Accordingly, there is an expected gap between the two groups in both employment and wages. This gap stems from the existence of two supply functions,[35] one for Saudis and another for foreign workers. Thus, the equilibrium wage is affected by international wages, which makes Saudis remain in their reservation wages. Unlike Nuri (2012); labour (2014); Kabli (2015) discussed the argument that the high Saudi reservation wage generated from the gap between wages in the public and private sectors, causing high unemployment rates among Saudis. They claimed that the public sector paid more than the private sector, with the public sector paying 3,000SR as a minimum wage. However, this is a weak argument to rely on because the public-private wage gap is in favour of the private sector, unlike their assumptions, especially when education returns are taken into consideration. Alfarhan (2015) supported this point of view empirically, using the Oaxaca decomposition. He stated that his result was contradictory to the common opinion. Therefore, it is inaccurate to consider the wage gap between the private and public sectors as determining the high unemployment rate. Admittedly, the minimum wage of 3,000 SR per month, for example, is the wage level to help Saudis have an acceptable living quality, whereas non-Saudis accept less than this limit and have a harsher living quality; however, they benefit from the remittance, which is the basis for future consumption. Due to this fact, the reservation wage is unequal when analysing low-paying jobs for those from a low background. This discussion would be weak if it were moved to the upper part of the wage distribution or high background countries. Therefore, the higher reservation wage for Saudis comes from the higher marginal substitution rates between leisure and consumption, not, as (Chugh, 2015) claimed, from any wage distribution level for low background countries. Workers from low background countries, even if they are educated, have a lower reservation wage than Saudis. They would prefer to work rather than spend time in leisure as they earn more than they would in their original countries. They understand that they will leave in

[35] There are many literature points on the segregation market. However, this might not be the case as foreign workers are found in both the upper and lower wage limits. Imported segregation is likely coming from the supply function differential.

a certain amount of time; thus, they maximise their earnings in that time for a similar reason; they would prefer future consumption to actual consumption.

Note, the future consumption is worth $3.75 per SR at any point as the value of the Saudi currency is connected to the American dollar. This provides non-Saudis with an accurate and clear plan for the future. Moreover, they have accepted living in a lower-quality life compared to Saudis, maybe even below the subsistence level of Saudis, which could explain the lower reservation wage for low background countries. Beng (2017) suggested that the reservation wage for poor people was somewhat less than the subsistence wage. This suggestion supports our view that those coming from low background countries seek better income, even though it is not considered enough for Saudis to live. For non-Saudis, a lower reservation wage and a willingness to work many hours lead to a lower unit hourly rate (UHR).[36] This is one of the key factors that leads the labour market to absorb non-Saudi workers and causes a higher unemployment rate among Saudis. In other words, the difference in the UHRs between the two groups in the same country is assumed to be the criterion behind firms' intensive demands for non-Saudi workers. The rare data on the UHR for each group makes using it an explanatory variable that is much harder to apply among the literature, especially in Saudi Arabia.

Moreover, this implied that the two employee groups were imperfect substitute factors, while Nitaqat was set for all economic activities, expecting all non-Saudis to be perfect substitutes, which would contribute to the increase in Saudi employment. This point would minimise the chance of employing Saudis in observed percentages. In these contexts, Bin-Obaid (2003) found that the relationship between foreign workers and native workers was negative in the GCC private sector, especially in Saudi Arabia and Oman. This reflects strategy weaknesses, such as tax and visa fees when the relationship between the workers is substituted; increasing the tax by 10% decreases the number of foreign workers by 2.1% in Saudi Arabia and Oman. In addition, using taxation

[36] Sapsford and Tzannatos (1993) assumed, for simplicity, that the hourly wage was the only cost for employment. However, they adopted other variables to measure employment. This simplified assumption could be applied for this study’s analysis as well.

increases the private wage cost by 8% in these countries. However, Nitaqat combined fees for all non-Saudi workers to increase the cost of replacing them with Saudis.

(Mahdi, 2005) conducted a survey to describe and analyse the Saudi labour market. He claimed that the segmentation in the labour market created by policymakers caused the wage gap between foreign and Saudi workers. He used mobility status as one of the heterogeneity policy indicators and concluded that mobility increases the logarithm of earnings for Saudi workers by roughly 2%. This does not conflict with the view that the wage differentials between Saudis and others come from heterogeneity in the sources of foreign workers as a natural consequence of an open economy that has fewer recruitment restrictions associated with the sudden economic boom because of the discovery of oil resources. However, narrowing the wage gap between Saudis and other workers depends on equalising their utility functions without guaranteeing the Saudi employment level. Applying fees for non-Saudis as it was done could be beneficial for narrowing the wage gap. However, these fees might contribute to increased deportation of non-Saudi workers more than narrowing the wage gap. Therefore, increasing the cost of non-Saudis through the regulation of semi-skilled jobs would alter firms' demands as non-Saudi are costly with the reservation wage discussed above. Furthermore, it would offer acceptable quality jobs for Saudis, and they would appear profitable for firms under Nitaqat. Moreover, the government seeks to improve Saudi workers' skills through education and training, which are associated with the Nitaqat programme. This has been done with several programmes, such as the King Abdullah scholarship programme and Human Resource Funding programmes. They have offered an opportunity for Saudis to have jobs with regulated helpers.

In this context, this might affect firms' performances, especially if the firms are more labour intensive. Thus, a change in wages because of the fee, when many immigrants work under the minimum wage (as imperfect substitutes), would increase firms' costs in the short run. This is consistent with (Peck, 2014) finding; it used the firms' downsizings or exits as the cost measurements. Although (Peck) result was expected, it did not account for if the downsizing was associated with downsizing in their non-Saudi recruitment. Admittedly, the lack of provision of the required data could have been a barrier for more investigation in her research. However, she discussed the disparities in the downsizing effect among sectors. This might be the fact that the MLSD rewarded localised firms with fewer recruitment restrictions when they aimed to extend their jobs, which might imply that the Nitaqat programme is costly for some firms and profitable for those that enlarge their jobs within the sectoral level. According to Peck, this downsizing results in a decrease of non-Saudi employment in non-localised firms. However, Peck's study mentioned that Saudi employment increased overall by 2.73%, and non-Saudi employment increased by approximately 9%.[37] Moreover, Peck pointed out that the Nitaqat did not significantly increase in some firms due to temporary Saudisation. This term refers to temporarily employing Saudi workers to achieve the quota, then laying them off. Therefore, the employment of Saudis is still expected, and non-Saudis still observe vacations. This is closely consistent with the findings of Purdie et al. (2006) regarding indigenous Australians; there was a small change in employment when the government implemented affirmative action in their favour. The downswing in Peck's research needs a complementary study to know if the firms' downsizings or exits affected the firms' efficiency. It could be an efficiency signal, where the inefficient firms previously existing in the market were not able to compete with other firms; a significant number of new establishments have started in the market. Moreover, as it was a one-year study, which is short, capital is expected to be fixed, and a decrease in employment might be beneficial if the firms work under diminishing returns. It could increase the marginal productivity of the workers where firms downsizing means a decrease in the number of employees to meet the Nitaqat percentage rather than adding new Saudi workers. Those firms were likely labour intensive, whereby adding new Saudis would be costly, and they would gain from laying off non-Saudis to meet the quota. However, this aspect was neglected in the literature.

To summarise, the Saudi labour market is as complex as it is open. The economy benefits from non-Saudis whether they are skilled or not – either from their very cheap hourly rates or their qualifications. The effect of the policy on the micro-level is affected by the macroeconomic level. Unfortunately, the lack of microeconomic data

[37] The study denoted that the increase was 60.02%, and the decrease was just over 50%.

makes research on the labour market very rare. Admittedly, studies need to be compared with each other for a clear picture of the Saudi labour market to emerge.

2.9 Conclusion

The discussion above can lead to some critical issues and the possible contribution of this research. In terms of critique, there has been little research conducted within this area from an economic perspective since Nitaqat was set up. However, (Peck, 2014) provided clear evidence that Nitaqat has managed to slightly increase the employment of Saudi workers, which is consistent with Holzer and Neumark (2004). Accordingly, the expectation is a reduction in the unemployment rate of the targeted group because of affirmative action. However, Peck's study was a limited investigation on the unemployment side, and programme outcomes require extra evidence to ensure that this increase was not temporary. Moreover, the policy needs further support evidence in terms of wages as an additional outcome. The substantial value of this element reflects the total welfare of Saudi workers, especially for those known to have no other income source. This study intends to investigate the wage differentials as a proxy for the success of Nitaqat; reducing the differentials between Saudi and non-Saudi workers is a general aim of this programme. Moreover, understanding the source of the differential allows consideration of whether there is any discrimination between the two groups. Justifying if the Nitaqat policy decreases the gap between the two groups and if it increases the discrimination between them. The study aims to address this gap among different background origins. This would be considered a proxy of future consumption, where the SR equals $3.75 at any point. This would create the multi-supply in the Saudi labour market and generate the gap, which could contribute to reducing the chance of the programme's success.

Although hiring quotas have been applied in other economies, Nitaqat is considered a unique programme. Nitaqat tends to be an incentive policy for firms rather than being an exhaustive policy. There are several rewards for firms who employ Saudis without any quality or salary restrictions. In contrast to other international programmes, this will have a deep impact on Saudi workers' welfare since Nitaqat is designed to help them engage the private sector while surprisingly ignoring wage structure. Statistically, more than a million foreign workers are operating below subsistence wages.[38] As Smith suggested, working at this level of wages would reflect on social welfare and worker well-being (Stabile, 1996). However, there are other support programmes intended to develop Saudi human capital. Together with Nitaqat, this demonstrates an understanding of the nature of the Saudi economy and attempts to engage Saudis in the market effectively. Nitaqat could change the wage distribution for both Saudis and non-Saudis because of the employment system.

2.10 Summary of the key feature on the literature

Author	Data sources	Model applied	Key finding	Relevance
Gottschalk 1978	From 1969 to 1983. Over 300,000 employee form 1000 establishments in the USA.	This study used the first order condition for the estimation.	Mincer approach is more appropriate than the productivity function in wage estimation.	Including age on earning function. Cubic age could be used in wage function.
Purnagunawan 2007	Australian Income and laboured dynamic, household. 2 cross sectional data 2001, and 2004 cover 13,696 and 12,408 full time workers, respectively.	This study using OLS estimations. The instrumental variable IV method used as well when the ability variable added.	Additional schooling increases the wage to 6% in 2001 and 7% in 2004.	This study confirms the important of using a human capital variable on earning function such as schooling and experience.
Fleischhauer 2007	review study of the human capital theory.	Descriptive analysis.	Human capital theory could be adopted in wage gap research.	Education/ earning in logarithmic form can be reflected on Mincer equation.
Arcidiacono 2004	Higher Education Directory 1973–1974, Tripariate Application Data 1973–1974, HEGIS Finance Survey 1972–1973, and the ACE institutional characteristics File the 1972.	full information maximum likelihood to regress log earning on several variable such as grad on subject, gender major specification, and the SAT reflect ability with respect to major.	He two keys finding of the earning function are: educational major strongly affects earning. Moreover, major can explain the gender differences in earning.	However, it confirms the important of adding independent variable could measure the workers major on the earning function. This can capture the earning differences between two heterogenous groups as well.
Blinder (1973)	The data collected from Michigan survey research centre 1967. The research conducted in the United States US	He used two of second procedure. First wage function reduced form, and second structure estimates of wage. He used OLS in both procedures.	He found that third of the differential of white gender gap explained by their attribute and 40% of the white/ black gap refer to discrimination.	This study inspires us to understand the basic of the decomposition approach. This initial paper used the three-fold decomposition approach.

[38] A subsistence wage is a wage sufficient for maintaining basic needs, such as food and shelter. In the Saudi labour market, it is common feature for foreign workers' jobs (not human capital) to have shelter and food for free, regardless of the quality. Thus, workers' future values in this case are quite high as they come from a poor country with a relatively low cost of living, and small wages by the Saudi currency equal a large amount in their original currencies.

Oaxaca 1999	The study benefit from mailing a survey among university and college facilities 1989 in the US. The participants were 2624 male and 892 females.	He applied to specification once using the PhD holder as a reference group and once non advance degree as a reference group.	Main finding that unexplained part changed with the reference category choice. when PhD is the reference, the constant 0.037 and .219 otherwise.	This study gives us an inspiration to propose a solution of the identification problem.
Yun 2005	Theoretical article	He followed suits (1984) for OLS. Total categories divided by number of categories including the reference (b1+b2+b3/4)	He found that applying normalised regression approach give a fixed value for each category and solve the identification issue.	It outwardly fixed, require fix base categories at the split regression. Moreover, the summation for one variable is variant, accordingly.
Griffin 1992	Firm level data from New York stock Exchange (NYSE), US, 1980. This data collected for EEO-1 report, affirmative action policy.	Theoretical and empirical study. The model estimates the cost function and inputs elasticities.	The elasticity of labour demand is low and lower substitutable relation between inputs factors. Firms cost increased 6.5% on average.	This study inspires us in terms of the multi-supply effect where firms dealt with the input factors (black/white) under this policy as separate inputs.
Kee 1995	The study used quality of live survey (QLS), Netherlands. native counts 1275, Antilleans formed 109, Surinamese counts 216, Turks counts 627, and 283 were Moroccans.	Oaxaca decomposition methodology was used with selection biased approach to capture other family income sources.	He found that Moroccans explained part indicate that there were advantage groups, and their wage could exceed native if there were similar characteristic.	Give us overview on applied Oaxaca decomposition. Inspire us to search behind workers origins.
Lehmer et al. 2011	Data collected from German federal employment Agency registration (BEH), 1995-2000. This covered 80% from Germany labour market.	Oaxaca decomposition was used for several European countries.	All immigrant: from European countries were receive lower than Germany. The gap shows heterogeneity among European group	We agree that immigrant would have lower wage than native depending on their background. This implies there is unobserved variable make all unexplained part high in some immigrants.
Longhi et al. 2012	British labour force survey (LFS). The data start from the 2ed quarter 2002 to 3rd of 2009.	Generalised Oaxaca decomposition was used. The study distinguishes between first and second generation for Pakistanis and Indian with respect religions. This yield 6th minorities groups. The reference groups is British white Christian.	That result was varying among the minorities groups. However, second generation received higher wage compared to first generation although the explained gap through characteristic was not improved in some minority's groups.	This study inspire us to start our theoretical analyse through the utility function as consumption would be essential factor could explained the native-immigrant gap. The heterogeneity on the utility function caused the heterogeneity on received wage.
Frank et al. 2013	The data sources is the report of the annual earning 2002, Canada.	Oaxaca decomposition.	Immigrant were disadvantage in ethic and demographic characteristic. They	Using a groups of variable as one attribute was a good

			were advantage in terms of human capital.	strategy to follow. We follow this strategy.
Smith & Fernandez 2017	The US data collect from International Assessment of Adult Competencies, 2016	Multinomial logistic regression for occupation	Immigrant located on the two bottom categories of wage scale.	This result agreed to search behind the substituted relation when use quota policy.
Massey et al. 1993	Theoretical review stud.	Review and appraisal theories of international migration	That consumption considers an appropriate unite for immigrant research.	Including consumption could explain substantial amount of the wage gap when immigrant considered.
Jann 2008	Stata command file	Develop Oaxaca decomposition commands	Useful tool to perform Oaxaca.	Help use to perform the empirical part.
Scicchitano et al., 2019	Survey of quality of workers, 2015, Italy. 15000 observation	Oaxaca decomposition OLS and quantile	There is a negative relation between job security and wage.	We agree that layoff risk could decrease wages.
Bratsberg et al 2014	Administrative registration data, Norway. From 1993 to 2006.	Basic estimation was used OLS for earning function and the labour elasticity.	They found that labour origin is important dimension in the labour market.	Disaggregate supply by origin. Agree with our multi-supply idea

Chapter 3 Data Sources and Description

3.1 Introduction

In the history of Saudi economic studies, the wage profile has been discussed thoroughly as a key issue in the labour market literature. The data has been the biggest obstacle to studying the Saudi labour market outside the macroeconomic level. Thus, some of these studies were qualitative, while others were quantitative; however, all were at the macroeconomic level. Admittedly, at the microeconomic level, there is a lack of studies that discuss the wage equation. Moreover, studies at the microeconomic level usually depend on surveys to achieve the research aim. This has distinct advantages, such as being able to customise variables according to the research aim. The drawback is the limited number of observations, usually because of the time limitations on collecting data or the lack of responses. However, this drawback was avoided in this study as the research benefitted from the secondary data at the individual level. This data was provided electronically to the MLSD by firms, giving the MLSD access to several key characteristics, such as age, nationality, and gender, for almost all workers. These data were linked to firms' characteristic such as activities, classification in Nitaqat, and size. Accessing all firms' and workers' characteristics at this data level was not feasible due to the privacy policy. However, obtaining the main characteristics helped with the research aim. The data contained two cross-sections for 2013 and 2017. The employee data contain two numerical variables: first, basic monthly wages in Saudi Rial SR which did not include allowance, or bonuses. Second: workers age. Additionally, there were two dummy variables: gender and Saudi. Gender refers if the employee is male or female. Saudi refers if the employee Saudi or non-Saudi. In 2013 dataset, the nationalities of non-Saudi were provided, unlike 2017, thus we had a categorical variable of around 37 nationalities.

In terms of employers' data, we got one continuous variable: **firms age** which was provided only in 2017 dataset. However, other variables were discrete (categorial) variables. For example, **educational qualifications** giving a particular qualification number 1 otherwise 0 for each qualification. Notice that the two data set was not identical in those classifications. There were 25 categories in 2013 and 24 categories in

2017. **Education** is presented in 12 ascending categories. The data benefit from the **administrative area** classification which contain 13 areas: Riyadh, Makkah, Easter province, AL-Baha, Northern Borders, Najran, AL-Jouf, Qassim, Madinah, Tabuk, Jazan, Hail, and Asir. Although, we got firms' **colour:** a categorical variable of Nitaqat. The variable provided the classification of firms under the quota system reflecting the percentage of Saudi in a firm. It is ascendingly classified: red, yellow, green, platinum. This classification is linked to firms' size, for example red small and red, green small and green which divided in 3 levels A, B, and C. Therefore, this variable end with 9 categories in both datasets. **Firm's size** was classified ascendingly according to number of the employee. The MLSD definitions is micro between 1 to 9 workers, small from 10- to 49 workers, medium from 50 to 499, big from 500 to 2999 workers, and giant had 3000 workers or more. Because size was criteria in Nitaqat, it is updated frequently, thus the two datasets was not identical. Medium firm's classification changed which was divided into 3 categories A, B, and C. The classification update was available for firm online to check which criteria they fill in. Similarly, the **activity** variable, it was used as a Nitaqat criteria thus, it changed several times it was reach 73 categories in the 2017 dataset. However, we link them to the national classification for the economic activities (ISIC4) which based on the international classification. This link yield identical categorial variable for both datasets contained 11 categories. **Occupations** was giving similarly over 2000 detail jobs titles then we link them to the standard Saudi occupation classification, accordingly occupation being 9 categories for both data set. It was classified descending starting with higher occupation manager and end with the basic engineering.

Unlike with other secondary data, dealing with this data was not straightforward; we experienced several issues with using them.[39] Furthermore, there was a high number of missing wages in the 2017 data set, which required an understanding of the nature of the missing data. As far as data is concerned for quantitative research, **first**, this chapter discusses the data in-depth the missing data issue is discussed by comparing the data to The GOSI to approximate the similarities of both missing and observed data

[39] For more details see (appendix A, section 9.2)

distributions in the MLSD dataset compared to the GOSI data. Moreover, the missingness in qualification and education is discussed, which helps identify the missingness mechanism and how to manage it. The **second** section imposes some limitations and discusses the outlier issue. **Finally**, the third section covers testing the samples graphically and statistically to understand the dependent variable distribution.

3.2 Missing data in the datasets

Although the datasets were sufficiently large, they contained missing observations for key variables, such as the continuous dependent variable – wage – and independent categorical variables – qualifications and education. The problem with the missing data was that the accuracy of the conclusions derived from the analysis was affected if the missing data was not treated following the mechanism. Thus, the distribution of missing and observed data was crucial (Bouza-Herrera, 2013). The missing distributions are thoroughly discussed to ensure which mechanism they followed.

There are three types of mechanisms. The **first** mechanism is known as missing completely at random (MCAR), which means the missingness does not depend on other variables in the dataset for both missing and observed values. However, this mechanism type is rarely found. The **second** mechanism is called missing at random (MAR), which means the missingness is conditional upon another variable unless the variable contains the missingness. The **third** mechanism is known as missing not at random (MNAR/NMAR), which means the data depends on the missing data itself (Scheffer, 2002). The statistical test is known as Little's MCAR, and it states that the data follows the MCAR mechanism if the test result is insignificant (Li, 2013).[40] If the result was significant, the graphical distribution would be involved to distinguish between the remaining mechanisms. Once the mechanism is known, one of the common methods to handle missing data can be used. For example, complete case (CC), indicator category (IC), frequency replacement (FR), multiple imputations (MI) and reweighted equation (RE) methods are considered novel (Henry et al., 2013). The CC method involves regressing only the CCs or making listwise deletions. The IC method recodes the

[40] For the test formula and detailed result (see Appendix A, Heading 9.3).

missing data with a fixed value. FR requires replacing the missing data with the frequency value or the means counted from the CC. MI impute the missing data from a model depending on the observed variable. RE estimates the logistic model and uses the inverse probability as sample weights.

3.2.1 Missing wages

For the 2013 dataset, this was not a recognisable issue, since only 155 observations (around 0.16%) from the sample were missing. It was considered a fairly low percentage as we allowed for 5% in the larger sample (Vieira, 2017). At this percentage, the data was considered MCAR. The Little's MCAR test supported this argument. The chi-square was 0.4015, and the p-value was 0.5263, which means we could not reject the null hypothesis that stated that the data was MCAR. Thus, the 2013 missing wages were considered ignorable (VanGeest et al., 2017). Accordingly, using CCs (only) or so-called listwise deletion would not generate a biased estimation.

Contrastingly, wage's missingness comprised over 50% of the 2017 dataset from non-Saudi only. This because non-Saudi were not required to register their wages at GOSI for pension unlike Saudi. thus, Saudi wages has not missing. However, both statistical and graphical methods were performed to investigate this. **First**, statistically, we completed Little's MCAR test, and the result was significant, with a p-value less than 0.05. Accordingly, the missingness could follow the MNAR or MAR mechanism, and the MCAR was eliminated. This made the challenge more difficult since most of the imputation or replacement methods might be considered for this type of missing data mechanism. For example, listwise deleting or replacing missing values with the mean value or the frequency categories could generate unbiased estimations if the data was MCAR or MAR (Enders, 2010; VanGeest et al., 2017). Even though some methods of imputing require MAR, such as maximum likelihood and MI, this method could give an unbiased result with NMAR data, as well. However, there are unique methods for NMAR data, such as the Heckman selection model and pattern mixer method. It should be noted that all missingness in wages was for non-Saudi observations, which means the missingness depended on another variable. Upon combining this information with the statistical implication, initially, the data would be MAR. However, the distribution of the missing and observed data should be considered before making the final decision on whether the data is MAR or NMAR to choose a satisfactory replacement method. To do so, missing and observed wages are discussed in terms of the conditional and marginal distribution against the GOSI aggregate level. Conditional distribution is the percentage of the cell from the total row or conditional to another value. In the calculation, the row must be 100%. Marginal distribution is the percentage measure of each cell from the total column. The summation of this percentage must be equal to 100% (see *Table 3-1* below).

3.2.1.1 2017 missingness distribution

This section starts with the above note on the data; all missing wages were found on non-Saudi observations (Table 3-1). The implication of this case is that the data could not be NMAR unless the observed wages depended on the missing wages. Otherwise, the data followed the MAR mechanism. If this was the case, ignoring the missing cases where wages exhibited a value was considered the appropriate method for a continuous dependent variable.

Table 3-1: Missing wages according to Saudi origin status

	Saudi	Distribution Conditional to Saudi	Non-Saudi	Distribution Conditional to non-Saudi	Total	Marginal distribution
Missing	0	0.00%	5,234,584	66.54%	5,234,584	54%
Observed	1,910,918	100.00%	2,631,826	33.46%	4,542,744	46%
Total	1,910,918	100.00%	7,866,410	100.00%	9,777,328	100%
Marginal distribution	20%	-	80%	-	100%	-
Distribution Conditional to missing	0%	-	100%	-	100%	-
Distribution Conditional to observed	42%	-	58%	-	100%	-

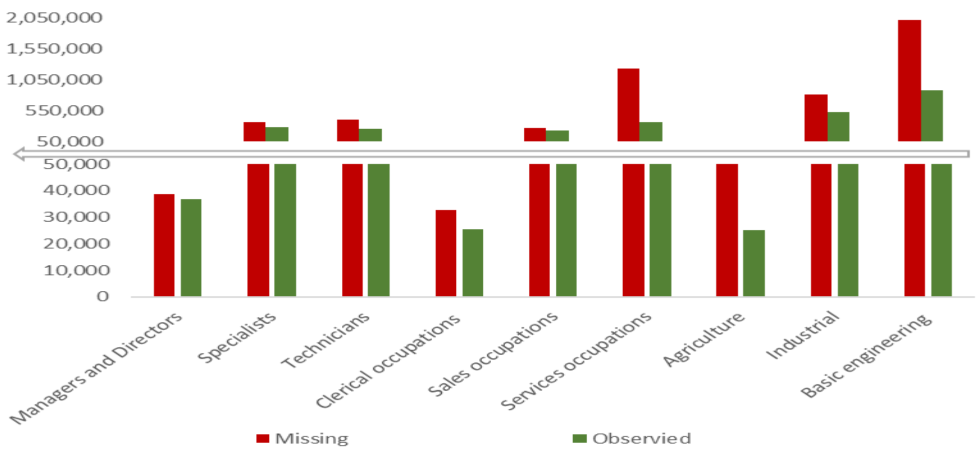
Missing cross-tabulations were provided to ensure that missing wage observations did not depend on the wages themselves. To do so, we used the occupation as a proxy to identify the missingness in wages among the occupation categories and to understand the mechanism of the missing data between those two variables. Table 3-2 shows the conditional distribution of missing and observed values. The missingness was not a presence on only one occupation category; the percentage of missing data varied across the occupation categories. The least missing data was found in the clerical occupation, where Saudis who had complete data were more likely to be found. There were 540,120 Saudi workers, while non-Saudi workers numbered only 58,013, and around 32,517 of them were missing data. This means the clerical jobs displayed only 5% missing observations in total observed cases, even though they formed 56% of the total non-Saudis. Contrastingly, the agriculture occupations had the highest percentage of wage missingness as a result of the lowest Saudi participation. The reason behind that could be that most workers in this category work in areas well away from cities, on farms or pastures, which creates an unpleasant working environment, contributing to Saudi workers abstaining from such jobs. Following this category, service occupations and basic engineering had high percentages of missing data, at 63.27% and 64.34%, respectively; those categories have a similarly harsh environment, decreasing Saudi workers (who have complete data).

Table 3-2: Conditional distribution of observed and missing wages by occupations.

	Missing	Percentage	Observed	Percentage	Total
Managers and Directors	38,625	16.37%	197,279	83.63%	235,904
Specialists	371,998	47.15%	417,007	52.85%	789,005
Technicians	402,475	48.11%	434,160	51.89%	836,635
Clerical occupations	32,517	5.44%	565,616	94.56%	598,133
Sales occupations	270,244	34.88%	504,560	65.12%	774,804
Services occupations	1,224,913	63.27%	711,030	36.73%	1,935,943
Agriculture	64,458	69.52%	28,266	30.48%	92,724
Industrial	812,347	58.90%	566,929	41.10%	1,379,276
Basic engineering	2,017,007	64.34%	1,117,897	35.66%	3,134,904
Total	5,234,584	53.54%	4,542,744	46.46%	9,777,328

By looking specifically at the distribution of non-Saudi observed and missing data, the missing data was spread across all categories. Furthermore, the cases of missing data exceeded the observed data in all categories, regardless of the variation of the amounts in each category (Figure *3-1*). This feature gave the impression that the observed data was sampled randomly by thirds from the non-Saudi group. Evidence was found by exploring the wage missingness distribution conditional upon the occupation categories, for example, technical occupations (38.58%), industrial occupations (39.12%) and basic engineering (30.35%). However, this percentage was a total of approximately 33.46%, roughly one-third.[41] The conclusion for this category was that there was no specific occupation categorised as responsible for the missingness; however, the percentage varied between occupation categories. The variations tended to be larger on the lower categories.

Figure 3-1: non-Saudi worker distributions by occupation categories



The above occupations were spread across economic activities; thus, it was necessary to understand the distributions of missing data throughout these activities. The amount of missing data exceeded the observed data in most activity categories, except education, mining, other activities and wholesale-retail (see Figure 3-2). In construction, around 88% were non-Saudis, forming approximately 60% of the missing data, with only 28% observed; the other 12% comprised the Saudi observed data. The wholesale sector behaved in a similar manner; non-Saudis formed roughly 77% of the total, of which, around 45% was missing data; Saudis comprised the remaining 23%. Therefore, the observed values were 40% and 55% for each category, respectively (see Table 3-3). From that table, we can also note that the category with the most missing data was professional, with 70% missing. The mining and quarrying sector had the least missing data, with 38% missing. We are aware that the representation percentage was affected due to the existence of the missingness; the percentage of observed cases of Saudis exceeded non-Saudis in some categories, such as other activities. Decreasing the Saudi

[41] The percentage was calculated as *(observed non-Saudi / total non-Saudi) * 100*.

observations to reach the representation percentage in the observed sample (OS) would not have developed the estimation result.[42] Thus, we preferred to accept this variation in the representation percentage when the missing cases were ignored.

Figure 3-2: Distribution of observed and missing wages throughout the economic activities

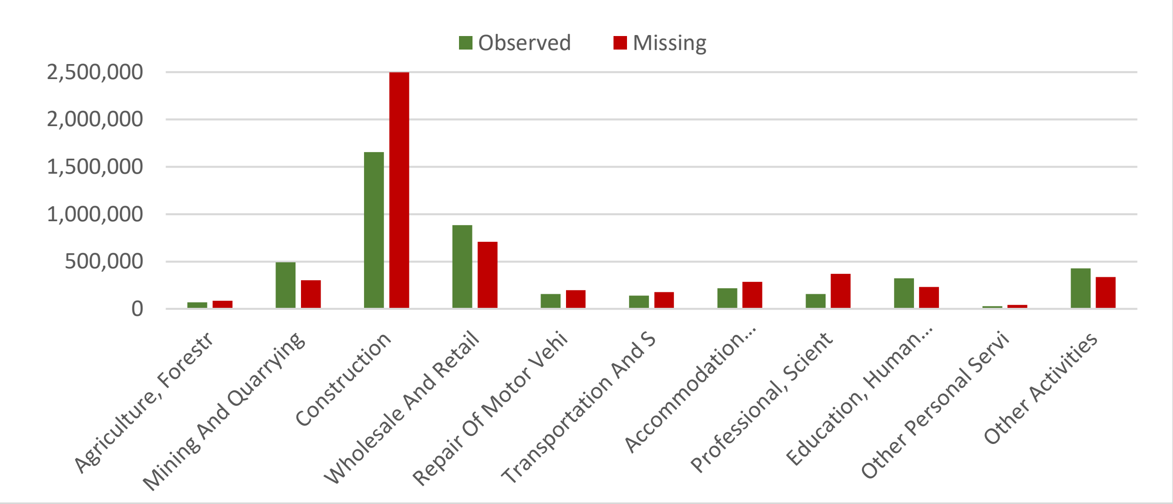


Table 3-3: Conditional distribution of observed and missing occupation data for both Saudis and non-Saudis

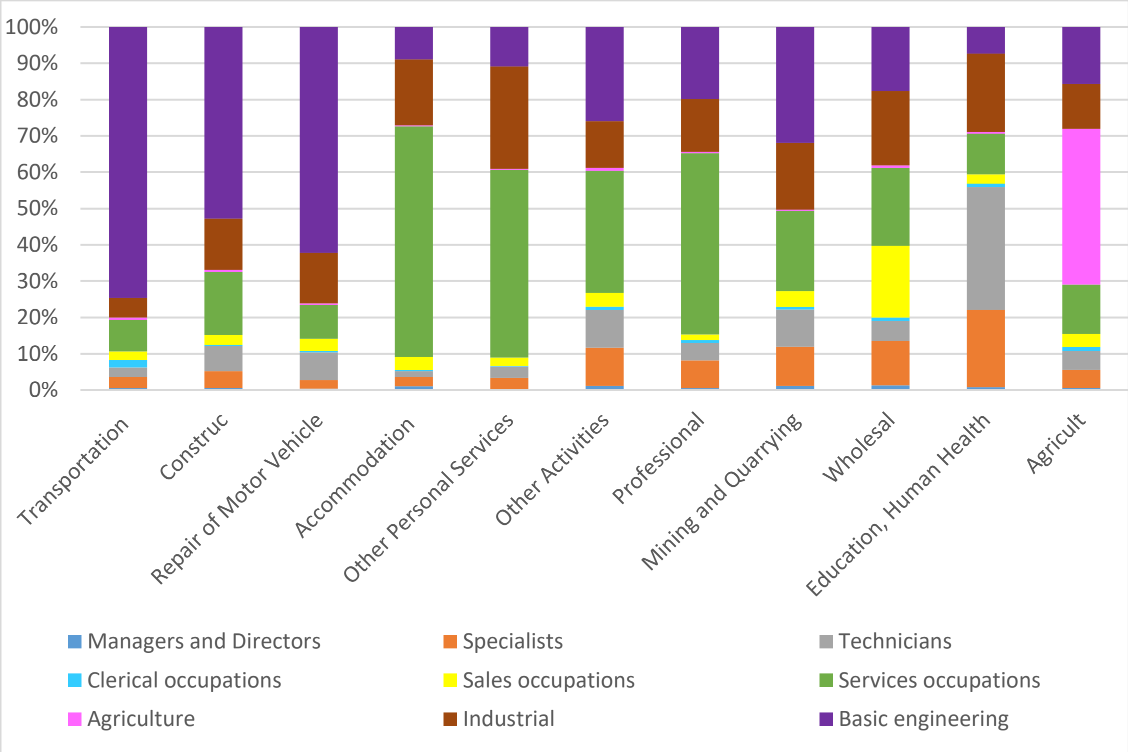
	Non-Saudi		Saudi
	Observed	Missing	Observed
Agriculture, Forestry	30%	55%	15%
Mining And Quarrying	24%	38%	38%
Construction	28%	60%	12%
Wholesale And Retail	32%	44%	23%
Repair Of Motor Vehicle	31%	56%	13%
Transportation	24%	56%	20%
Accommodation	24%	57%	19%
Professional, Scientist	16%	70%	13%
Education, Human Health	23%	42%	36%
Other Personal Services	24%	60%	16%
Other Activities	26%	44%	30%

Figure 3-3 shows that clerical occupations and managers and directors had unnoticeable missingness in all activities, with just less than 2%. In comparison, one of three

[42] A double sample technique can be used to develop the estimation in some conditions (Avery & Burkhart, 2015). Note, **observed sample** refers to the data subset of the MLSD dataset after removing missing wages.

occupations – basic engineering, industrial and services – formed the highest missingness in all activities except for agriculture and education, which followed a unique distribution. This implies that the missingness was high in low occupations. Notably, the missing cases in the industrial and specialist occupations were roughly similar in most activities, even though the earning expectations of both groups were different. Low paid workers were reported on the dataset and formed around half of the non-Saudi observations. Thus, we could not generalise that missingness was associated with low wages.

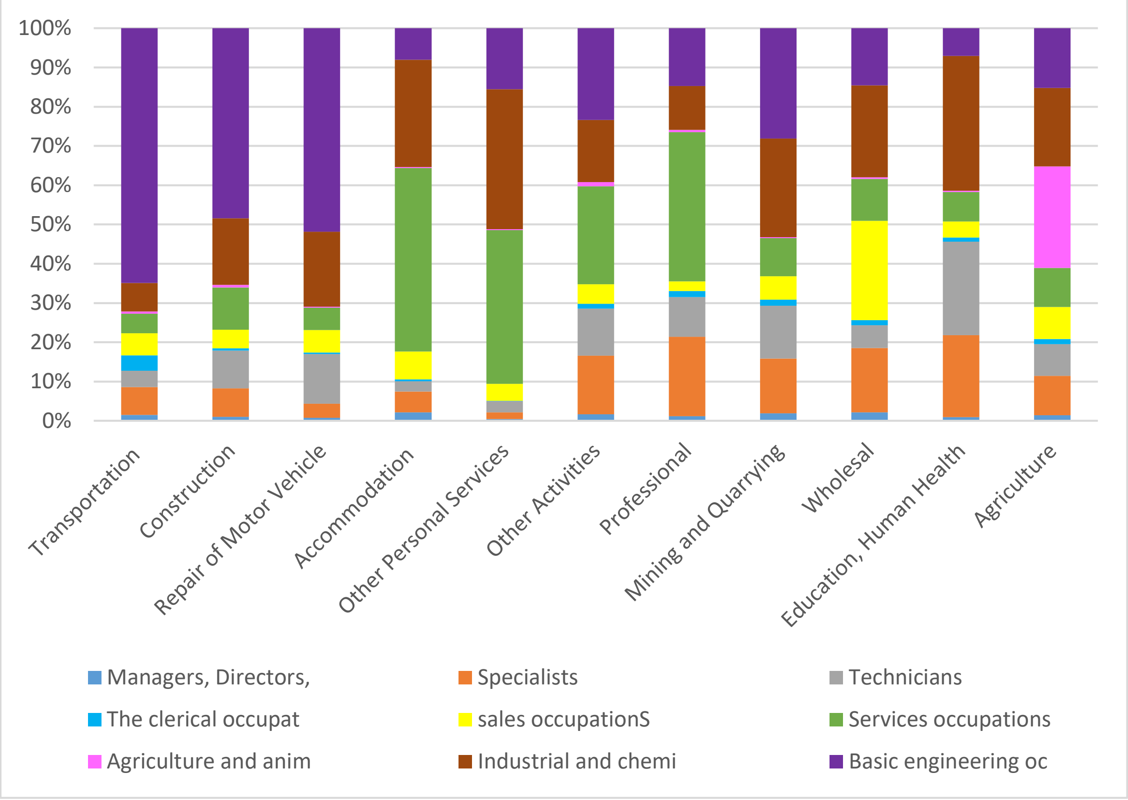
Figure 3-3: Wage missingness distribution among occupations conditional to activities



It seems that missingness was directed by the distribution of occupations among the activities. It increased when an activity heavily depended on this occupation category. For example, the missing cases for basic engineering in transportation numbered around 132,312, forming approximately 75%. In contrast, missing cases in industrial occupations numbered only 9,641, even though both occupation categories were classified as low-paying occupations. Similarly, the agriculture occupation's missingness was approximately 43% in the agriculture activity, while it was less than

1% in all other activities. To get a clear picture of that missingness, we compared them to observed cases for non-Saudis as the missingness only existed in this group (see Figure 3-4). Regardless of the variation in the percentages, observed cases had roughly the same trend as the missing cases.

Figure 3-4: Observed distribution of non-Saudis among occupations conditional to activities.

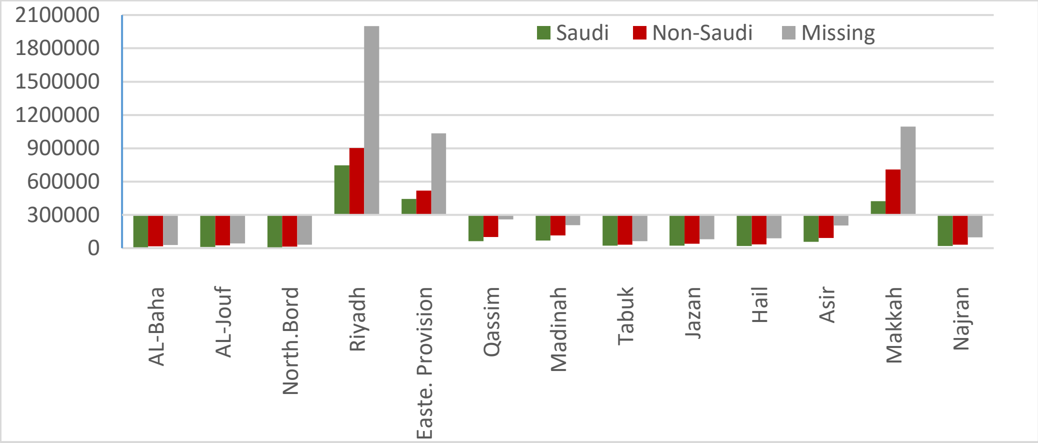


The geographical distribution of missing and observed data followed a similar trend. The highest amount of observed data was found in the area with the highest missingness (see Table 3-4). For example, Riyadh had over 3.5 million workers, representing approximately 37%; around 2 million were missing (38%), and around 1.5 million were observed. Although Riyadh and Makkah had the most missing data, the differences between Saudi and non-Saudi workers were still noticeable (see Figure 3-5). The missing data was spread among all the geographical areas. The distribution of the missing data gave the impression that it had been sampled randomly for only non-Saudis, even though the data for all Saudis was observed.

Table 3-4: Marginal and conditional distribution of wage missingness among the geographical area

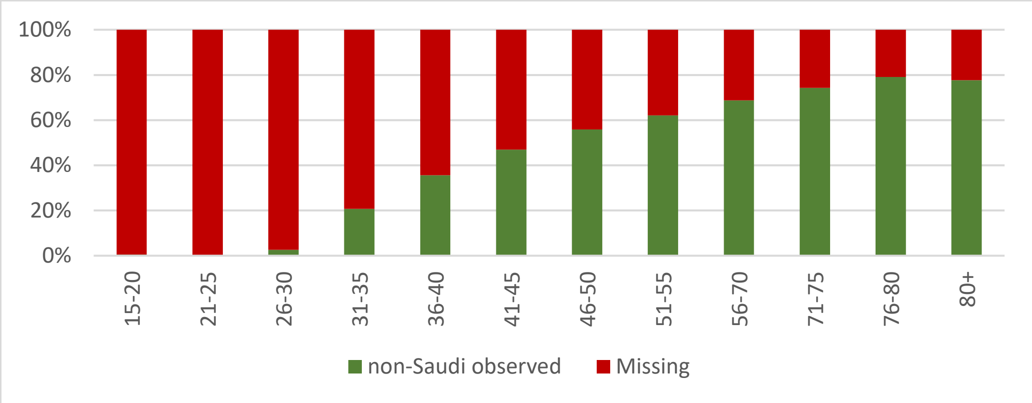
Distribution type	Conditional		Marginal
	Observed	Missing	Total
Al-Baha	0.53%	0.54%	0.54%
Al-Jouf	0.76%	0.80%	0.79%
Northern Borders	0.50%	0.57%	0.54%
Riyadh	36.33%	38.22%	37.34%
Eastern Province	21.19%	19.75%	20.42%
Qassim	3.55%	4.94%	4.29%
Madinah	4.07%	3.97%	4.02%
Tabuk	1.15%	1.19%	1.17%
Jazan	1.39%	1.56%	1.48%
Hail	1.20%	1.71%	1.47%
Asir	3.30%	3.92%	3.63%
Makkah	24.94%	20.95%	22.80%
Najran	1.09%	1.89%	1.51%

Figure 3-5: Distribution of observed and missing data for Saudis and non-Saudis among the geographical area



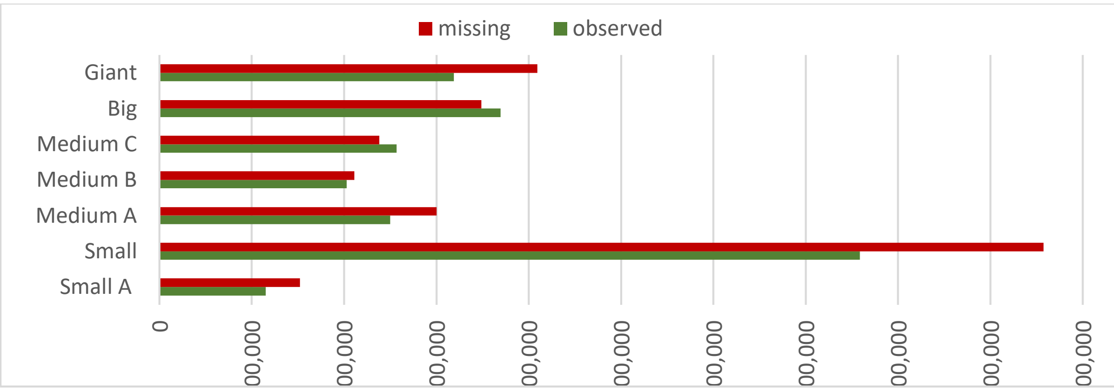
Another dimension that could be investigated is age groups. Figure 3-6 illustrates that missingness was spread across all age categories. However, this missingness was concentrated in the first three columns, with approximately 97% or higher. This percentage decreased substantially when age increased. It seems that wages were systematically missing with age groups, meaning the intensive missingness of wages tended to be higher when the ages were younger, not when the wages were lower. Thus, we could conclude that the missingness of wages followed a MAR mechanism, which means that we could recognise a clear relationship between missing wages and ages, which was inconclusive evidence that this missingness was related to a specific wage group.

Figure 3-6: Distribution of missing and observed data along with the age group categories.



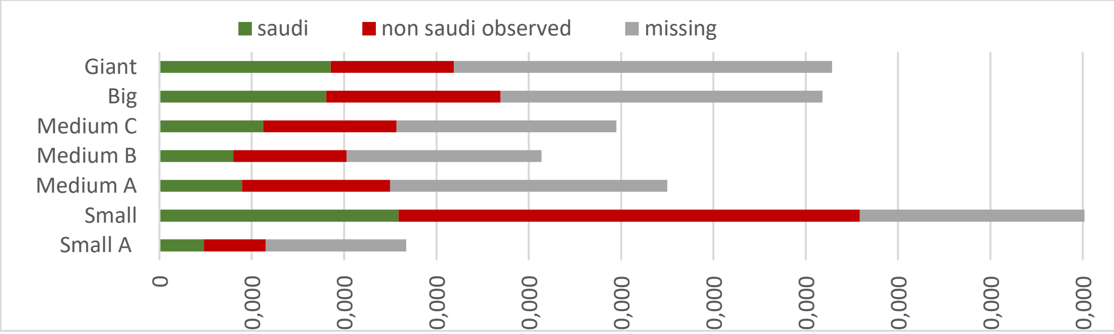
From a firm size angle, Figure 3-7 shows that when the number of observations was higher, the probability of missingness was higher. For example, small firms had the highest amount of both observed and missing data, while the small A or micro firms had the lowest amount of both observed and missing data. At this point, we cannot deny that there was no systematic missingness in terms of firm size. In other words, missingness in wages was randomly found, which means that the missingness was MCAR according to the firm size angle. However, missing data cases were more numerous than observed data cases in all categories except big and medium C firms, where the observed data slightly outnumbered the missing data.

Figure 3-7: 2017 observed and missing data according to firm size



The observed data for non-Saudis was larger than for Saudis in all firm sizes (see Table 3-8). Admittedly, this sheds some light on the underrepresented data concerning the Saudisation dimension. This means firms in the green zone, which are supposed to employ more Saudis, could show a high number of non-Saudis. Thus, we explored the missingness data for the firms' colour zones in the Nitaqat classification.

Figure 3-8: 2017 distribution of missing and observed data for both Saudis and non-Saudis.



To explore missingness under the Nitaqat programme as mentioned above, we divided the firms into two groups: localised and non-localised. The first group included the green firms or above, and the second group included the red and yellow firms. Table 3-5 indicates that the missing data was very similar for both firm types, although there were vast differences in the total number of employees. Accordingly, the missingness was independent of firm status, which implies the possibility of the missingness following the MCAR mechanism.

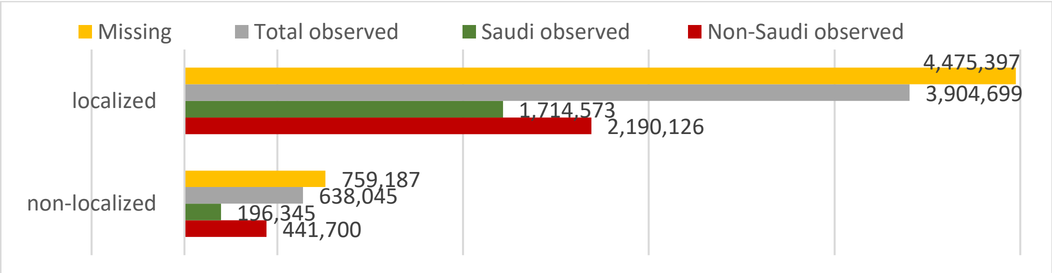
Table 3-5: Conditional distribution of observed and missing data among firms' status.

	Observed	%	Missing	%	Total
Non-localised	638,045	45.66	759,187	54.34	1,397,232
Localised	3,904,699	46.59	4,475,397	53.41	8,380,096
Total	4,542,744	46.46	5,234,584	53.54	9,777,328

Even though the missing data formed roughly half of the total non-Saudi employee population, the non-Saudi observations were still higher than the Saudi observations (see Figure 3-9). Indeed, ignoring the missing data would have influenced the proportion in this dimension. At localised firms, Saudis comprised around 43.91%, while non-Saudis comprised around 17.54% when considering missing data.

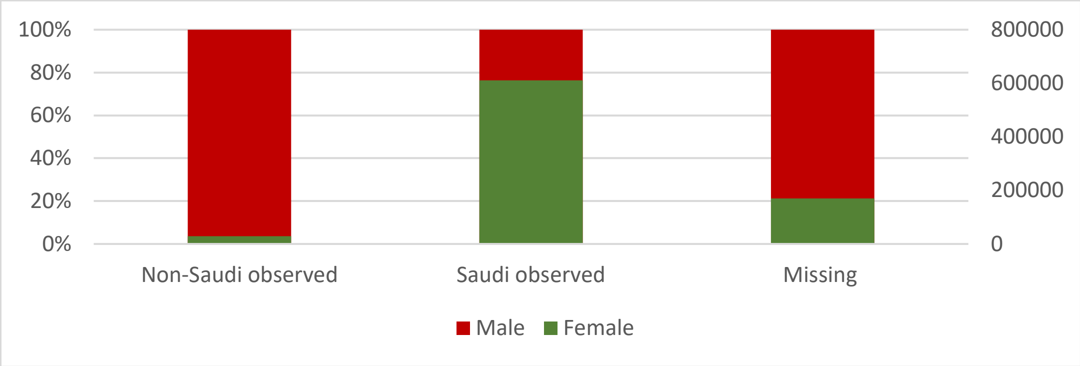
Correspondingly, only 196,345 Saudis were in non-localised firms, forming around 2.01%, whereas this percentage increased to 30.77% concerning the observed data. However, the percentage of non-Saudi workers decreased substantially when the missing data was ignored. For the total, however, Saudis were considered overrepresented; they encompassed roughly 19% of total data set and increased to 42.07%, which decreased the non-Saudi percentage to 75.93%, although they comprised 80.46% of the entire data set. Indeed, this led to a link between the total given sample and the labour force survey or social insurance annual statistics to ensure that the data was still representative. This helped us decide if the data needed to be randomly sampled to meet this percentage or if it was fair enough to keep the data provided.

Figure 3-9: Saudi and non-Saudi missingness according to firm status



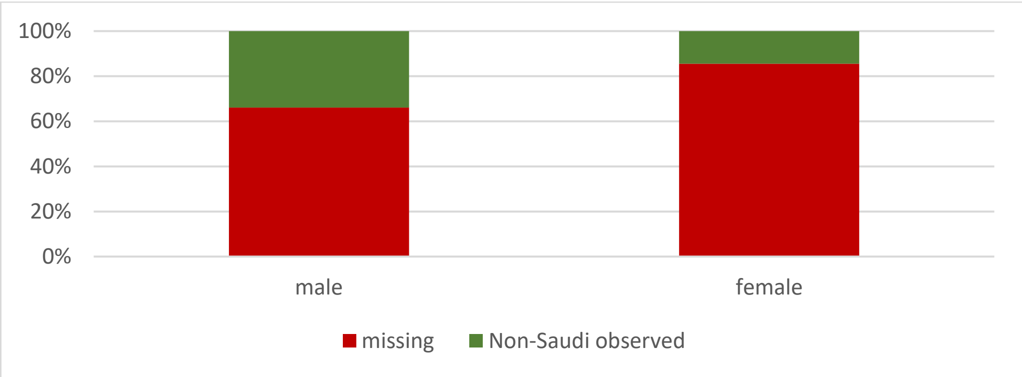
The next dimension was gender. Female participation comprised around 8.3% of their observation of less than one million, whereas men formed approximately 91.7% of their observations of nearly nine million. Figure 3-10 shows that the trends of males and females were opposite when considering origin. Female participation was low in general, and it seemed lower if they were non-Saudi, while for men, participation was high, and it was higher if they were non-Saudi. To illustrate, this figure shows that Saudi women's participation was 75% of the total female participation, while non-Saudi women formed only 25%, and 21% of them were missing data. However, Saudi men formed only 15% of the total male participation, whereas non-Saudis formed the majority, with 85%; around 56% were missing data. In the same graph, Column 3 shows that both males and females had missing data; women lost less than 200,000, forming 20.99%, compared to men, who comprised over five million missing wages, forming 43.52% of the total dataset.

Figure 3-10: Wage distribution according to gender, conditional to Saudi status



Although women had fewer missing data than men, they had a higher percentage missing, around 85.49%, so the remainder was a quite small observation (see Figure 3-11). However, non-Saudi men were missing around 66.05%. The high missing number of men did not change the fact that both groups lost a substantial amount of data. Admittedly, we cannot deny that there was no systematic missingness in terms of gender. Wage missingness seemed to be distributed randomly according to gender, which relatively implies that wage missingness followed the MCAR mechanism in this dimension.

Figure 3-11: Comparison between male and female observations and missing data



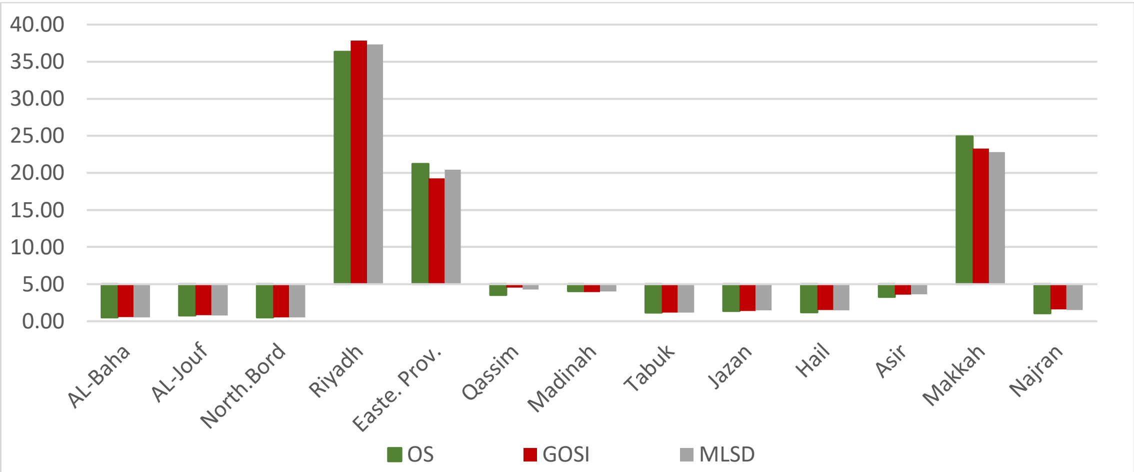
3.2.1.2 2017 observed wages.

Admittedly, we cannot deny that there was systematic missingness only in terms of age. This indicated that the data followed the MAR mechanism. However, the sample representativeness was broken because of this missingness in the Saudi versus non-

Saudi dimension. Non-Saudis still formed a higher proportion than Saudis in general. Thus, exploring the observed data was beneficial, especially if linking the three data sources: the GOSI data, the MLSD dataset and the OS. The reasoning behind this stage was to ensure that the sample distribution followed a pattern like the published data, which provided a suggestion about the mean wage value after this deletion. This covered several dimensions, according to their availabilities.

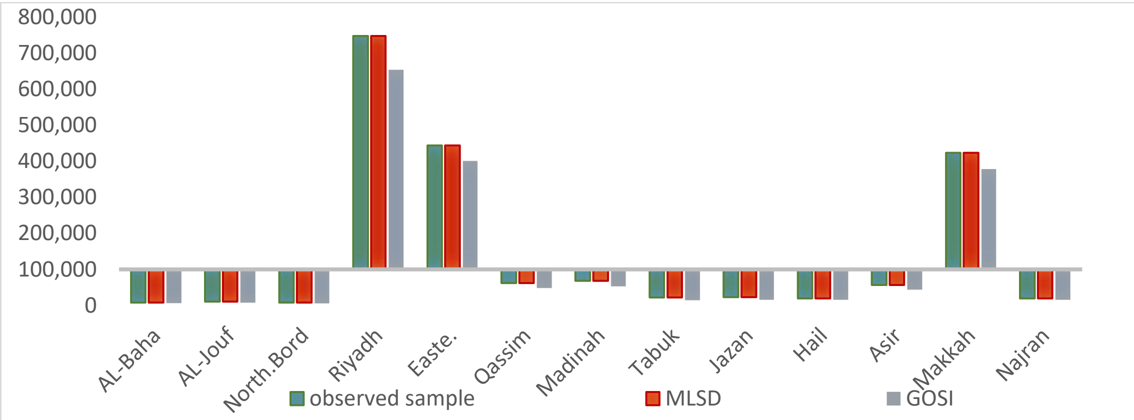
The GOSI report for 2017 focused on the geographical area as the main dimension, as well as ages and wages. In geographical areas, the proportions of workers were unevenly spread across the three datasets. It was insensitive in Riyadh, then Makkah, then the Eastern Province (see Figure 3-12). This means the workers were centred on the three biggest areas, and other geographical areas had less than 2% of the workers each, except Madinah, Qassim and Asir, but their representation rates were still less than 5%. Moreover, the proportions were similar in the three datasets, with some minor variations, although the OS raw count was around half of the two other datasets. For example, in Riyadh, there were around 1,647,459 workers, forming approximately 36.35% of the OS; this percentage was 37.86% for GOSI and 37.34% for MLSD. Interestingly, in the OS, Tabuk represented a percentage equal to the percentage provided by the GOSI at 1.14%, but the count of the workers was 51,807 in the OS and 112,944 in the GOSI. To conclude, the OS represented a satisfying percentage in each geographical area compared to the GOSI and MLSD datasets.

Figure 3-12: The distribution percentages of registered workers across areas



For Saudis, the observed data was identical to the MLSD data (by definition), whereas there was a small variation between those two data sources and the GOSI data, although it could be considered equal if the numbers were rounded (see Figure 3-13). In total, the MLSD dataset provided a higher number of Saudi workers compared to the GOSI dataset. The biggest differences in those two datasets were found in Riyadh, Makkah, the Eastern Province and Madinah. For example, in Riyadh, around 652,994 Saudi workers were registered in the GOSI data, and around 747,227 Saudis were registered in the MLSD data, a difference of 94,233. Despite the variation, those workers represented 39.33% and 39.10%, respectively, for the two datasets. All of the variations between the two data sources comprised less than 1% in each geographical area. For example, the Eastern Province had the greatest difference in the datasets, with 0.90%, and Makkah was underrepresented by 0.61%. Fairly, the three datasets had a convergent representation percentage for Saudis.

Figure 3-13: The distribution of Saudis among the geographical areas



However, considering gender to know who was responsible for this small difference proved interesting. Looking for Saudi data concerning gender allowed us to discuss the distribution of those workers. This did not appear to be a great issue to discuss (see Table 3-6); both groups followed a similar distribution of the total. However, this led to knowing the total representation of each group. There was no major difference between data sources concerning both dimensions: Saudis and gender. The variation between male and female participation seemed to be the most vital issue, as the data represented almost the total number of Saudi workers.

Table 3-6: The conditional distribution of registered Saudi workers by gender

	Female			Male		
Zone	OS	GOSI	MLSD	OS	GOSI	MLSD
Al-Baha	0.27	0.22	0.27	0.51	0.47	0.51
Al-Jouf	0.33	0.34	0.33	0.64	0.56	0.64
Northern Borders	0.34	0.30	0.34	0.42	0.38	0.42
Riyadh	43.94	43.67	43.94	36.83	37.47	36.83
Eastern Province	15.30	15.58	15.30	26.91	27.76	26.91
Qassim	3.22	2.88	3.22	3.24	2.97	3.24
Madinah	3.52	3.34	3.52	3.63	3.16	3.63
Tabuk	1.21	1.09	1.21	1.11	0.83	1.11
Jazan	1.47	1.21	1.47	1.08	0.85	1.08
Hail	1.46	1.41	1.47	0.83	0.77	0.83
Asir	2.30	1.98	2.30	3.30	2.92	3.30
Makkah	25.46	26.86	25.47	20.57	20.98	20.57
Najran	1.19	1.10	1.19	0.93	0.88	0.92
Total	100.00	100.00	100.00	100.00	100.00	100.00

There was a convergent distribution between men and women among the geographical areas (see Figure 3-14 and Figure 3-15). The two charts shed light on the participation trend similarities between the two gender groups concerning the areas. For example, according to the MLSD dataset, the lowest participation rates for both genders were in Hail. In this region, Saudi men had a rate of approximately 0.56%, while women had 0.47%. However, in Riyadh, both men and women had the highest participation rates, with roughly 25% and 14%, respectively. In total, the participation of Saudi men was around 68.05% in the MLSD data and 69.9% in the GOSI data; the women formed the rest. The relative gender contribution of Saudi women's participation was 47% of the sample and 43.04% in the GOSI data, which was much higher than the percentage provided for the total economy, which was around 30%, according to the World Bank.[43] Regarding the expected reasons for the variation of female participation between regions, it could be family financial status, availability of attractive jobs for women in some areas, childcare facilities, or other reasons. Keep in mind that supportive policies for women were equally imposed in all areas.

[43] It is calculated *as (31.95 / 68.05) * 100* and *(30.1 / 69.9) * 100*.
The World Bank link access as of 18/6/2019:
https://data.worldbank.org/indicator/SL.TLF.CACT.FM.ZS?end=2018&locations=SA&start=1996

Figure 3-14: The distribution of Saudi men across the geographical areas

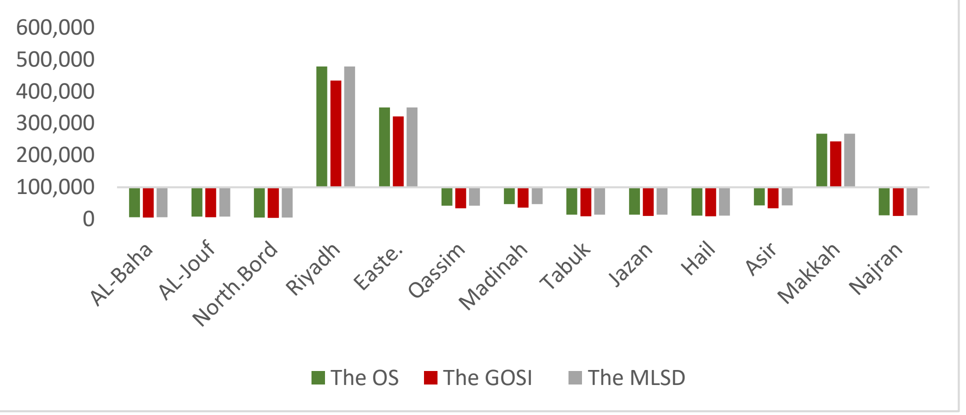
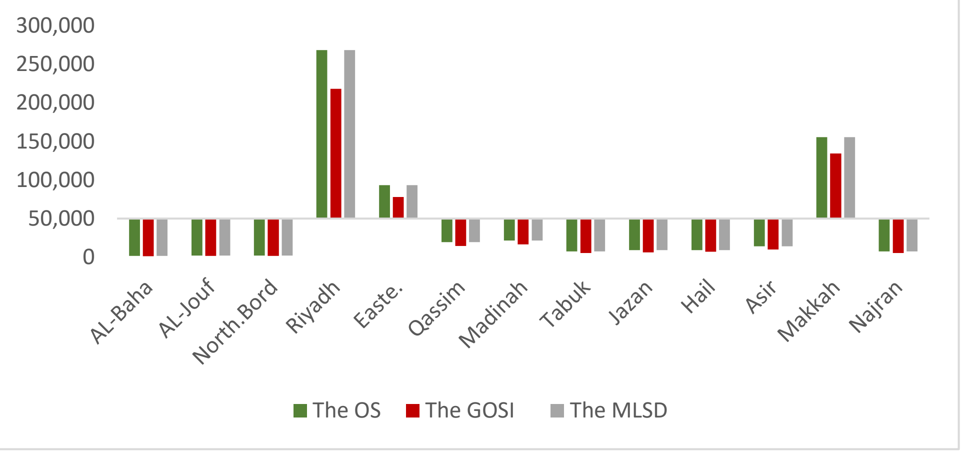
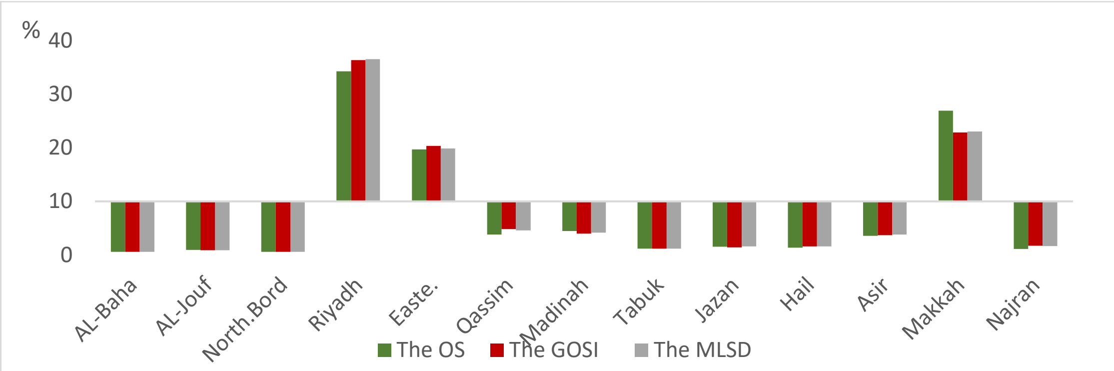


Figure 3-15: The distribution of Saudi women across the geographical areas



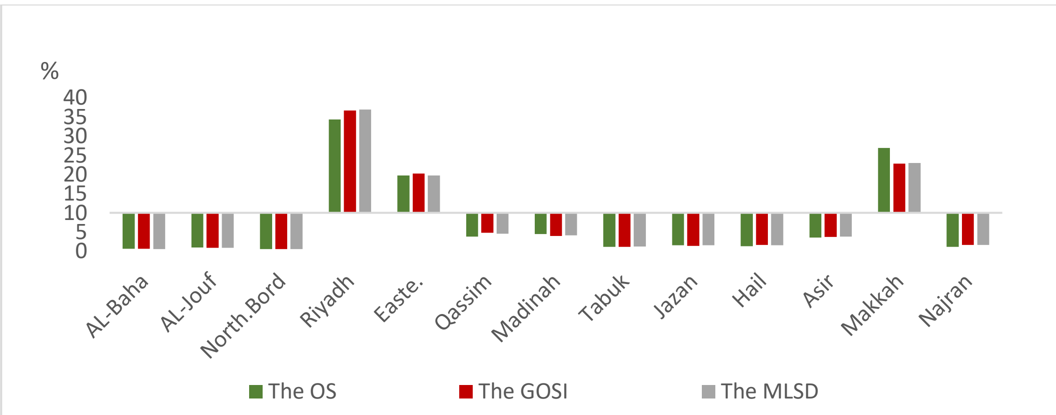
For non-Saudis, the three datasets had similar representation percentages in some regions, such as Al-Baha, Al-Jouf, the Northern Borders and Tabuk, with 1%, while the Eastern Province had 20%, and both Madinah and Asir had 4% (see Figure 3-16). In other geographical areas, such as Hail, Najran and Qassim, the OS revealed a 1% lower representation percentage compared to the GOSI and the MLSD, except for Jazan. In this area, the GOSI representation percentage was lower than the MLSD and OS by 1%. However, Makkah was the area affected the most by missingness. There were 1,926,050 non-Saudi workers in the GOSI dataset and 1,806,295 in the MLSD dataset, whereas the OS provided only 706,334. Although it was a small number of workers compared to the other two datasets, it represented 27%, compared to 23% for the two other datasets. Makkah was overrepresented in the OS by around 4% (relatively 17.4%). In comparison, Riyadh showed an underrepresentation percentage of roughly 3%. There were 900,410 workers, representing 34%, in the OS, compared to 3,088,803 workers, forming 37%, in the GOSI dataset and 2,903,838 workers, representing 37%, in the MLSD dataset.

Figure 3-16: non-Saudi distribution across the geographical areas



In terms of gender, a similar feature was found for Saudis; men were more intensive in their participation than women. Moreover, both gender groups were intensive in Riyadh, Makkah and the Eastern Province (see Figure 3-17 and Table 3-7). For men, the GOSI dataset exceeded the data from the MLSD, apart from the case of Jazan. However, the representation percentage was not affected. It appeared exactly like the above descriptions, with one exception in the GOSI dataset for Riyadh (see Figure 3-17). The percentage was 37% for total non-Saudis and 36% for men; otherwise, the approximate percentages were equal. Regardless of deleted missingness, the non-Saudi male sample was roughly representative across the geographical areas.

Figure 3-17: The distribution of non-Saudi males across the geographical areas



For females, the number of observations in the MLSD data exceeded the number in the GOSI data in five geographical areas: Al-Baha, Riyadh, the Eastern Province, Hail and Jazan. It was the opposite of the other areas. However, the representation percentage was affected in those areas as much as others, apart from the Northern Borders and Al-Jouf, where the representation rate remained the same (see Table 3-7). There were some areas in the OS that appeared overrepresented. In the Eastern Province, for example, women represented approximately 21%, while the representation percentage was approximately 15% in the other two datasets. A similar percentage variation trend was found in Makkah, as well. However, Al-Baha, Qassim, Madinah and Tabuk were overrepresented by approximately 1% or less. In the other direction, the areas underrepresented by around 1% or less were Asir, Jazan and Hail. Riyadh displayed the highest underrepresentation percentage in the OS, compared to the GOSI and the MLSD, with approximately 10% and 11%, respectively. The female data displayed small variations between the GOSI and MLSD data in approximate percentages in Riyadh and Makkah, regardless of the opposite direction of the representation. However, we cannot deny that the representation percentage for the OS was significantly affected in Riyadh, Makkah, and the Eastern Province.

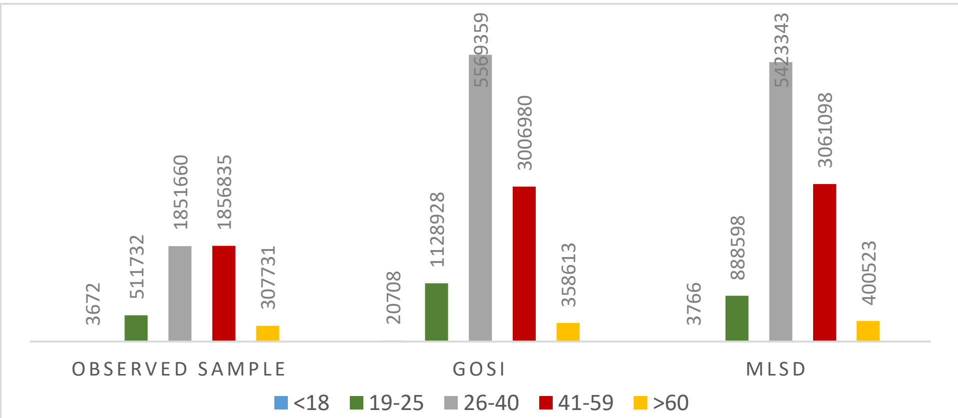
Table 3-7: The conditional distributions of non-Saudi females across the geographical areas

Zone	OS	GOSI	MLSD
Al-Baha	0.7	0.3	0.4
Al-Jouf	0.4	0.5	0.4
Northern Borders	0.3	0.3	0.3
Riyadh	38.7	48.4	50.1
Eastern Province	21.2	14.8	15.4
Qassim	4.1	3.2	3.0
Madinah	3.6	3.2	3.0
Tabuk	0.7	0.6	0.6
Jazan	0.7	0.9	0.9
Hail	1.0	1.1	1.2
Asir	3.3	4.1	4.4
Makkah	25.0	21.8	19.4
Najran	0.5	0.8	0.7

Age groups were another dimension provided in the GOSI dataset concerning geographical areas. The overall age range was divided into five groups: under 18, 19–25, 26–40, 41–59 and over 60. The major deleted data was in the third age group, 26–

40, where the labour force was concentrated (see Figure 3-18). There were roughly 5.5 million workers, which became less than 2 million. Accordingly, this group formed around 40.86% of the OS, which is an underrepresentation of the percentage of roughly 55% in the other datasets. Similarly, the group under 18 exhibited an underrepresentation rate in the OS, although all three datasets formed less than 1% of the total sample. Contrastingly, in the fourth age group, the OS showed an overrepresentation percentage of approximately 41%, compared to the GOSI and MLSD datasets, with 30% and 31%, respectively. Similarly, the over-60 group also displayed an overrepresentation rate. Although the second group showed that around half of the data was missing, the representation percentage was not affected as much as the fourth group, which represented roughly 11% in the OS; this was like the GOSI dataset and only 2% different than the MLSD dataset. Although the highest missingness was found at the early age groups, deleting them affected the middle age groups severely. It seems this group was weighted heavier in the dataset compared to the early age groups.

Figure 3-18: Total worker distribution among age groups



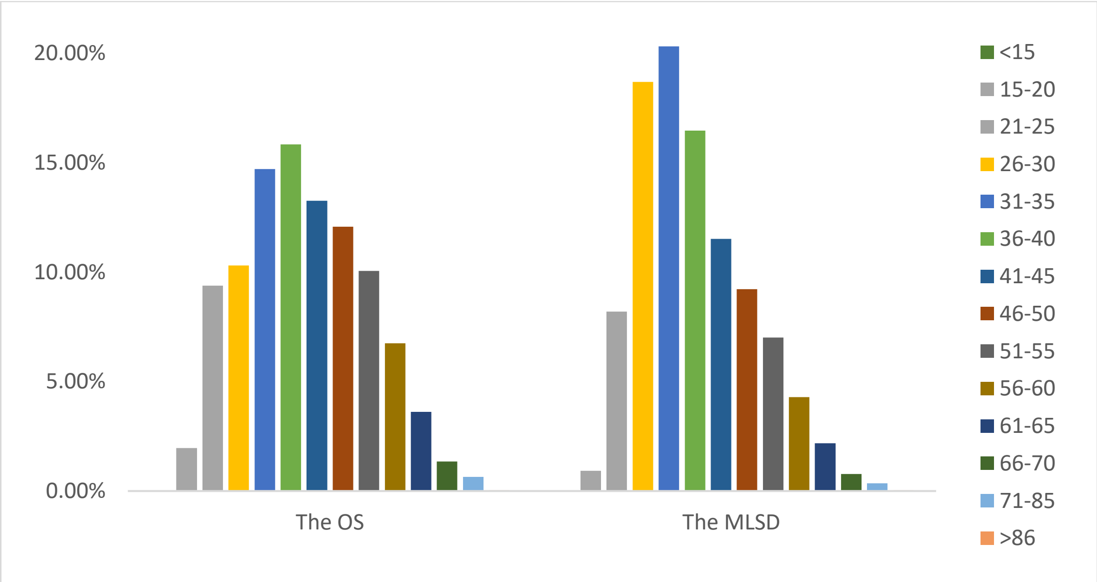
From another angle, the fourth age group was not as significantly affected as the third age group. Therefore, whereas the GOSI and MLSD datasets displayed a normal distribution among the age groups, the observed data reflected more of a bimodal distribution (see Table 3-8). This is because the data cut mainly affected the third age group, which contained the highest number of workers.

Table 3-8: The total representation percentage according to each age group

	< 18	19–25	26–40	41–59	> 60
MLSD	0.04	9.1	55.5	31.3	4.1
GOSI	0.2	11.2	55.2	29.8	3.6
OS	0.08	11.29	40.86	40.97	6.7

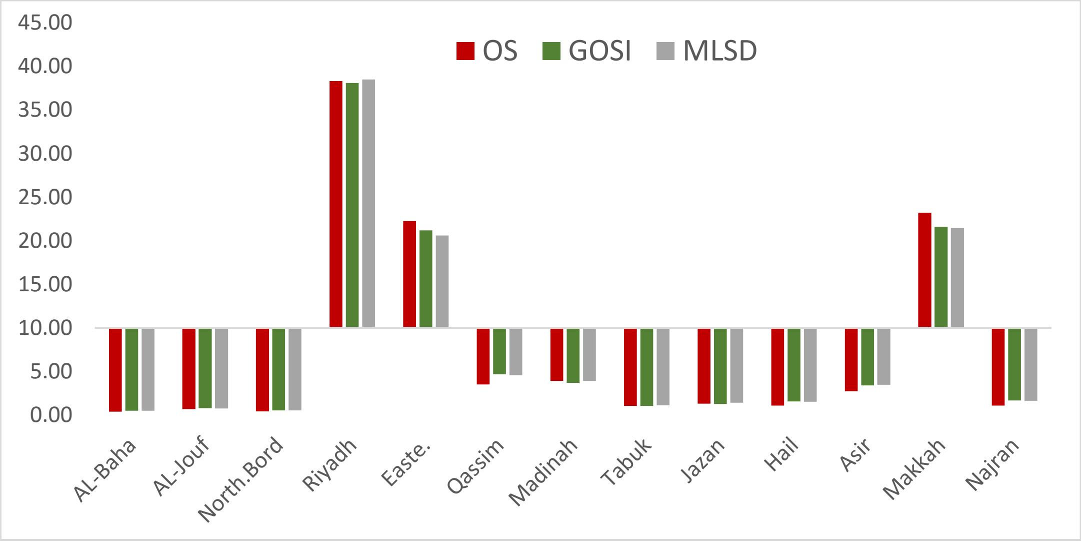
The table shows that age groups in both the OS and the MLSD followed a similar trend when classified principally in five-year age bands, as follows: 15–20, 21–25 and so on, until 81–85. However, three age groups seemed not to maintain the same percentage as before the data reduction. Surprisingly, the OS followed a normal distribution when considering the detailed classification (see Figure 3-19). This lognormal distribution did not reflect the same distribution in the MLSD dataset; however, it was still acceptable. It stemmed from combining Saudis and non-Saudis in the graph. Saudis were concentrated in the early age categories, which made the representation higher than expected in these groups. Excluding Saudis who had no missing values provided an accurate picture of the remaining observations, especially in the early age categories. Figure 3-20 shows the similarity in the representation rates after the age of 40. Despite the intensity of the missingness in earlier age groups, there were still a reasonable number of observations. Thus, we assumed that the observation numbers were satisfactory unless a clear reason was shown.

Figure *3-19: The OS and the MLSD distributions with the researcher's classifications*



Despite the variation display in Table 3-8 above, there was similarity shown in the totals of some geographical. This implies that the observed data in total concerning the geographical area could provide similarity. Thus, the third and fourth age groups concerning a geographical area require a discussion. We chose those two groups because they were considered the most affected groups due to missingness in wages (see Table 3-8). Surprisingly, in the 26–40 range, the variation was not high and followed the total age groups' behaviours when the conditional distribution was considered (see Figure 3-20). From the graph, the OS was overrepresented, with around 1% and 2% in the Eastern Province and Makkah, respectively, compared to the other datasets.

Figure 3-20: The conditional distribution of the age group bands in the 26–40 range



However, the OS was underrepresented, with 1% in the Northern Borders, Qassim, Hail and Najran; the other areas were roughly similar, including Riyadh. In the four bands in the 41–60 range, the OS was underrepresented, with approximately 1% in Riyadh and Qassim, while it seemed overrepresented in Makkah, with a similar approximate percentage (see **Error! Not a valid bookmark self-reference.**). However, other areas were equal with the rounding consideration. Frankly, it seems that wage missingness conditional to age was distributed fairly among the geographical areas. This was an expected and satisfactory finding.

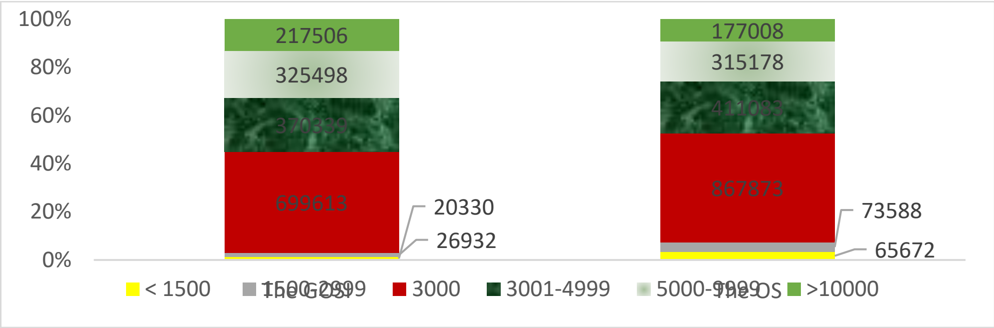
Table 3-9: The conditional distribution in percentages for the four age groups 41–60

	OS	GOSI	MLSD
Al-Baha	1%	1%	1%
Al-Jouf	1%	1%	1%
Northern Borders	1%	1%	1%
Riyadh	34%	35%	35%
Eastern Province	21%	21%	21%
Qassim	3%	4%	4%
Madinah	4%	4%	4%
Tabuk	1%	1%	1%
Jazan	1%	1%	2%
Hail	1%	1%	1%
Asir	4%	4%	4%
Makkah	26%	25%	25%
Najran	1%	1%	1%

A key angle was the wage distribution according to wage category and the geographical area compared to the GOSI dataset.[44] Following the GOSI classification, at all datasets, there was a convergent conditional distribution on the wage categories of Saudi workers who had accurate data because they did not have missing data (see Figure 3-21). Although the distribution showed similarity in the fourth wage group, with around 22% in all datasets, the observations of the GOSI were slightly higher in the highest category groups, while they were slightly lower in the lowest categories compared to the OS. For example, the third group formed the highest percentage, with 42% in the GOSI dataset and approximately 45% in the OS. Saudis who received more than 5,000SR represented around 26% in the OS and 33% in the GOSI data. There was a gap of more or less than 4% between the two datasets because the OS had around 260,258 extra observations for those who earned 3,000SR or less, while the GOSI had 10,074 extra observations of those who earned more than 3,000SR. In general, the Saudi worker distribution did not show a radical difference between the two datasets, whether measured in frequency or percentage. However, the heterogeneity of the wage distribution for Saudis and non-Saudis could be noticed (see Figure 3-21 and Figure 3-22).

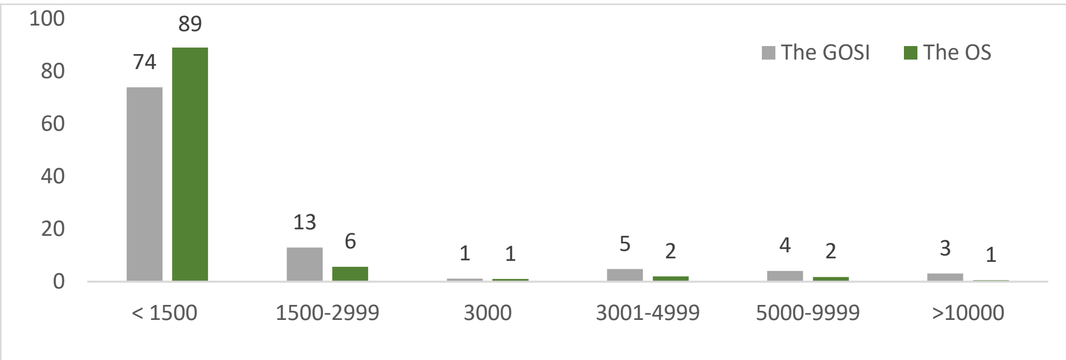
[44] Data variation was expected as it was collected from two different organisations. However, it provides the reader an idea of how much the OS was similar to the GOSI data. This could be an index of the strength of the data.

Figure 3-21: Saudi frequency distribution and percentage among wage categories



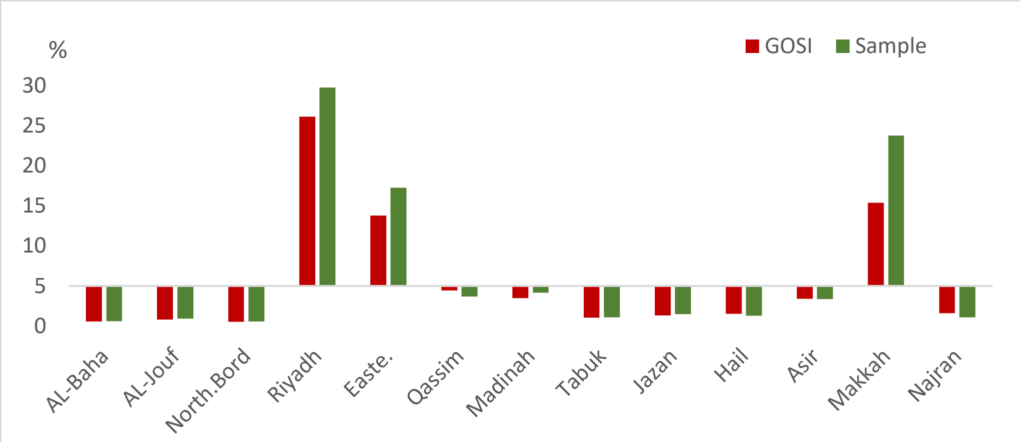
For non-Saudis, the two datasets followed a similar trend (see Figure 3-22). The data skewed to the left mainly for the first category. The first wage categories formed around 75% of non-Saudis, with approximately 74% in the GOSI dataset and 89% in the OS. However, in the next wage group, the sample was underrepresented around 6%, compared to 13% for the GOSI dataset. This trend was found in the highest three categories, as well. Non-Saudis formed approximately 1% of the two datasets where the Saudi minimum wage was located. However, there were five million cases observed in the minimum wage groups and under, using the rounding approach, and 900,000 in the higher wage categories. The missingness did not belong to one category, whether low or high. Thus, the highest missing percentage was found at the highest observed frequency, mainly the first wage category (see Figure 3-22).

Figure 3-22: Non-Saudi distribution across the wage categories



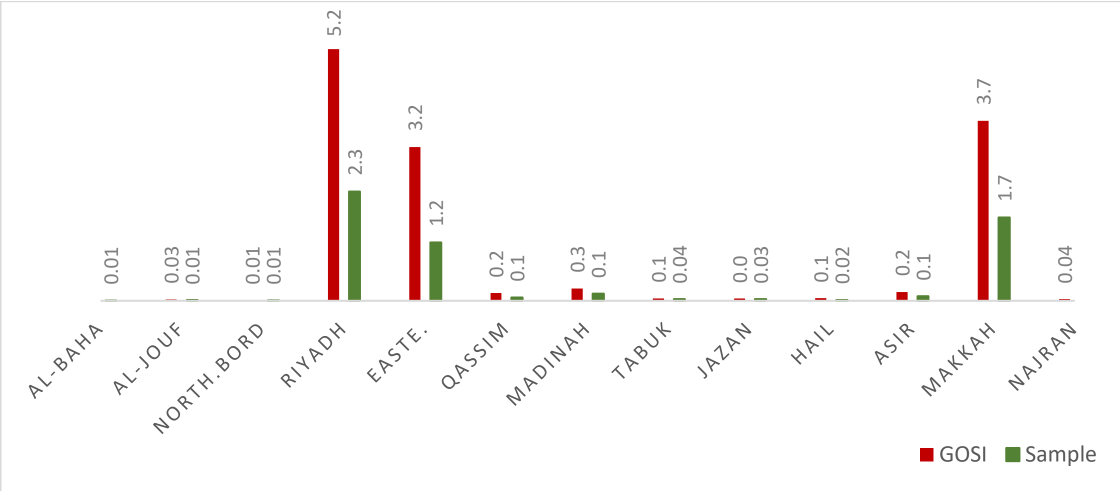
The highest densities of non-Saudis were found in Riyadh, Makkah, and the Eastern Province. For the first wage group, for example, the OS was overrepresented in Makkah by about 9%, Riyadh by 4%, Madinah by 1% and the Eastern Province by 3%, compared to the GOSI. However, Najran and Hail were underrepresented by 1%. However, a similar distribution was shown in the remaining regions (see Figure 3-23).

Figure 3-23: The distribution of non-Saudis in the geographical areas (percentage of the total, solely for those with wages below 1,500SR per month)



Unlike the first groups, which were overrepresented in total, other wage categories were underrepresented. The highest group among them was the second group, containing those who earned between 1,501 and 2,999SR. Similarly, Riyadh, the Eastern Province and Makkah revealed the highest variations. The OS was underrepresented by 2.8% in Riyadh and approximately 2% in the Eastern Province and Makkah (see Figure 3-24). Indeed, this was a small percentage, and it was centred in Riyadh, Makkah, and the Eastern Province, which had the highest populations. However, the rest of the wage groups formed small percentages of both datasets. This small representation did not mean they were not important; it reflected the non-Saudi wage distribution in the Saudi labour market. It seems that the non-Saudi missingness increased when the population increased.

Figure 3-24: Geographical distribution of non-Saudis in the 1,500–2,999SR wage group



Concerning the average wages in the geographical areas, Table 3-10 displays how the average wages in the OS and the GOSI dataset were roughly convergent in those areas, apart from Riyadh, Tabuk, Madinah and Makkah. This variation could result from the additional Saudi observations provided by the MLSD. Figure 3-13 shows extra observations on the MLSD dataset compared to the GOSI. Remember that no missing Saudi wages were found. This variation in the datasets should be taken into consideration when comparing non-Saudi workers, who had all the missing data.

Table 3-10: Saudi average wage (SR per month) for the OS and the GOSI datasets

	Male		Female	
Zone	OS	GOSI	OS	GOSI
Al-Baha	3,213	3,331	3,113	3,240
Al-Jouf	3,250	3,333	3,354	3,586
Northern Borders	3,603	3,991	3,045	3,182
Riyadh	6,154	4,684	3,764	3,451
Eastern Province	6,912	6,666	3,854	3,745
Qassim	3,376	3,569	3,092	3,243
Madinah	4,854	6,333	3,230	3,406
Tabuk	4,572	3,588	3,535	3,432
Jazan	3,496	3,832	3,094	3,217
Hail	3,502	3,783	3,041	3,200
Asir	3,883	3,990	3,151	3,228
Makkah	5,111	4,414	3,570	3,574
Najran	3,254	3,447	2,967	3,079

When considering non-Saudi males, a similar conclusion was found; there was not much variation between the two datasets. However, the non-Saudi female average wage was influenced in most of the geographical areas, except Al-Baha and Al-Jouf. Although the females had fewer missing observations, they affected the average wage (see Table 3-11). However, in the OS, non-Saudi females earned 1,869SR on average, while the men earned 1,171SR on average.[45] Similarly, in the GOSI dataset, females earned extra. They earned approximately 2,321SR, whereas males earned around 860 SR. The average difference between the two datasets was 453SR for males and 311SR for females, which was a reasonable variation compared to the fully observed Saudi cases. The GOSI datasets presented Saudi male's wages as less by 1,564SR and Saudi female's wages as less by 557SR. the dataset of OS might be overrepresented, or the GOSI datasets might be underrepresented as they were two different sources. However, we found that the OS was in the middle of the GOSI dataset and the GaStat.[46] This dataset followed a different survey method because that we did not compare the OS with GaStat data source. The main variation was that this dataset considered a representative sample to search, while the MLSD and GOSI datasets considered all the workers in the market.

Table 3-11: Average wages for non-Saudis (SR per month)

	Male		Female	
Zone	GOSI	OS	GOSI	OS
Al-Baha	748	1,153	2,751	1,063
Al-Jouf	633	637	1,307	945
Northern Borders	749	747	1,925	959
Riyadh	1,090	1,261	2,034	1,519
Eastern Province	1,685	1,289	3,470	2,633
Qassim	708	704	2,227	1,160
Madinah	1,149	1,001	3,076	1,753
Tabuk	874	984	2,778	1,778
Jazan	787	745	2,739	1,363
Hail	663	814	2,541	1,252
Asir	766	941	2,252	1,770
Makkah	1,330	1,193	3,079	2,009
Najran	629	698	1,924	1,175

[45] The average wage was not provided by the GOSI dataset. It was calculated by the author, and this could carry a variation as it was calculated from the areas' average wages. Thus, the average wage was approximated while the OS had its mean calculated by Stata.
[46] The GaStat provided the average wages (see Appendix A, Table 9-1).

All the above deep discussions imply that the missing data followed the MAR mechanism more than the MNAR mechanism, according to the statistical results and based on the analysis above. The wage missingness for non-Saudis had a recognisable relationship with age, whereby the missingness decreased when the upper groups were considered. Moreover, there was no evidence that missingness could be found in specific earning groups. Therefore, it was satisfactory to use CCs only when excluding missing data under the MCAR and MAR mechanisms to generate unbiased estimations, which is known as ignorable missing. Using some replacement method for the missing data would risk generating less rigorous estimations, according to this mechanism (VanGeest et al., 2017).

3.2.2 Missing qualification and education

These two variables were categorical, unlike wages. This section discusses the possibility of using those two variables as providing a separate category for missing data. The advantage of keeping this category is that it captured workers who had qualifications less than secondary school. A separate neutral category could be used in an opinion survey (Houser, 2009; O'Sullivan et al., 2016). Removing this category from qualification could generate bias. Henry et al. (2013) supported this point of view; the CC usually generates biased coefficients, even in the MCAR mechanism, when the proportion of data with missing cases is large. However, Vittinghoff et al. (2012) found that using missing as an indicator could generate biased results, even with MCAR, and needed to be imputed unless the adjusted and unadjusted estimation results were similar when the number of missing cases was small. Unbiased results could be generated with imputation, even with large amounts of missing data (up to 90%), with a satisfactory result when the data was MAR (Madley-Dowd et al., 2019). However, they argued that the proportion of missing cases could be used as a guide for choosing an auxiliary variable. Increasing these variables would not increase the efficiency of the regression generated by MI. Henry et al. (2013), by contrast, concluded that MI could generate bias and recommended that REs be used in both MCAR or MAR data mechanisms. However, they recommended using the IC method with great caution because it could generate bias. Groenwold et al. (2012) supported this recommendation and limited the use of IC with randomised trials. Moreover, Pampaka et al. (2016) found that imputing for missing cases did not change the results and conclusions. They pointed out that the benefits of using imputation in their research were obtained as significant compared to the original models. Fairly, they assumed that increasing the sample size through reduplication of cases and the imputation method would generate a significant result. Enders (2010) displayed that imputation might not be necessary and could generate biased coefficients as well. This viewpoint quite agreed with Vittinghoff et al. (2012), who suggested a comparison of the estimations' results between methods. Thus, we compared the common imputation method against the IC. Our data was randomly collected; it was collected by the MLSD to cover almost the whole labour market. However, the 2013 data was chosen randomly by the MLSD software.

3.2.2.1 Qualification 2013

The mechanism of the missing qualification data could be MAR or MNAR when Little's MCAR test was significant. To find out which mechanism the missingness followed, the marginal and conditional distribution of a proxy – occupations – was investigated to have an indicator of whether the missingness was related to specific qualification groups. If yes, the missingness followed the MNAR mechanism; therefore, ignoring the missingness would generate bias. Otherwise, it followed the MAR mechanism. The expectation was that the less qualified workers would comprise more of the unregistered qualification categories because there was no category that was considered a good method to partially capture them.

By looking at Figure 3-25, it seems that the lowest amount of missing data was in three occupation categories: managers, directors and senior officials; the clerical occupations; and agriculture and animal husbandry, with less than 1% missingness. Those require different qualifications; for example, managers are expected to be qualified, unlike those who work in the clerical and agriculture occupations. Around 44% missingness was found in the service occupation, and 27% was found in the basic engineering occupations. The density of those two categories could be related to the occupation structures; meaning, qualification missingness followed the distribution of the occupation; when an occupational category had high observation numbers, it was more likely to have the highest number of missing qualifications (see Table 3-12). For example, basic engineering comprised approximately 25.72% of total occupation categories and was associated with roughly 23.95% of the missingness in this occupation. Moreover, the marginal distribution for service occupations was around 42.06%, while approximately 39.90% was more likely to be missing. This trend was found in almost all categories. Surprisingly, the highest observed occupation categories after specialists were found in service occupations and basic engineering. These joint distributions indicated that missing qualifications were not related to the qualifications themselves. The missing qualifications were spread across all occupation categories, approximately following the occupation structure. The data was not systematically missing in occupations; however, there was a relationship between the missingness in qualification and occupation.

Figure 3-25: Conditional distribution of qualification missingness among occupation categories

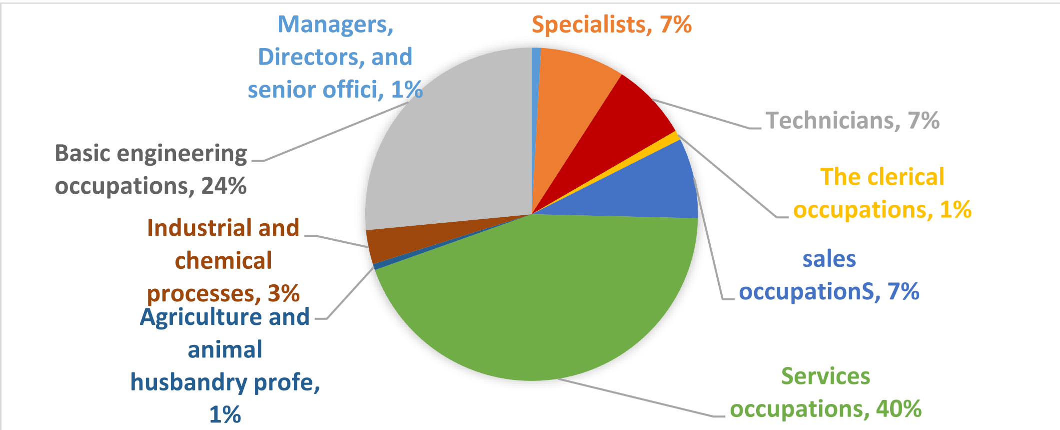


Table 3-12: Joint and marginal distributions concerning missing qualifications

	Missing	Observed	Total
Managers, directors, and senior officials	0.82%	0.53%	1.35%
Specialists	7.42%	2.49%	9.91%
Technicians	6.80%	1.37%	8.17%
The clerical occupations	0.84%	0.29%	1.13%
Sales occupations	7.03%	0.89%	7.93%
Services occupations	39.90%	2.16%	42.06%
Agriculture and animal husbandry	0.53%	0.02%	0.55%
Industrial and chemical processes	2.99%	0.19%	3.18%
Basic engineering occupations	23.95%	1.77%	25.72%
Total	90.28%	9.72%	100.00%

We cannot deny the fact that the conditional distribution could end with a different summary. For example, the manager categories captured approximately 39% of the qualifications; the remaining were missing (Figure 3-26). Specialists and clerical occupations captured roughly 25% and 26% of the worker qualifications, respectively. Around 17% and 11% were observed in the technique and service occupations, while other occupation categories registered less than 10%. However, in total, less than 10% of the occupation cases were captured. The first implication we could suggest is that when the occupation required a highly skilled worker, it was more likely to capture their qualifications. However, we cannot state that the missingness in qualifications depended on the observed qualifications for two reasons: **first**, less-skilled workers might have qualifications not listed on the qualification categories, such as elementary primary education. Therefore, it was expected to be unregistered if they did not have a real qualification. **Second**, qualified workers must register their qualifications by law, and they still had just over 60% missingness, meaning that both qualified and less qualified workers had missingness, which relaxes raising the issue of the MNAR.

Figure 3-26: Conditional distribution of missing qualifications among occupations

By looking in-depth at job titles, we found that the missingness in qualifications covered two worker types in general: **first**, workers in jobs with low qualification requirements, such as drivers, cleaners, tailors, and marketers, and **second**, a mismatch of job and qualification. For example, observation 7,040 was a dentist working in the health services sector, which indicated real missing data for education. However, observation number 5,482 was missing a qualification, and his job title was a doctor in blood diseases working in construction. This observation indicated the mismatch issue.

Although missing qualifications followed the MAR mechanism, the use of this category, called a neutral category approach, within data collection surveys has generated great debate in the literature (Weisberg, 2009). Therefore, we compared the common imputation method coefficients, following Pampaka et al. (2016). Table 3-13 displays several regression results concerning the common missing data handling method. Frankly, the dropping the variable (DV) method has not been used in the literature. It usually results in omitted variable bias. we used this scenario only for comparison purposes. Generally, all approaches generated an approximately similar result. To compare those methods, we looked to the well-known F-test for all regressions except MI. The results indicated that dropping qualifications would be a worse method and would generate biased results compared to the FR method and the indicator category method (IC). The DV method generated better regression results than using the CCs approach. This outcome was expected as the missing data formed a large proportion of the overall sample, and the mechanism was not MCAR (Henry et al., 2013). When the DV approach was better than the CC method, the CC method was eliminated. As shown in the table below, DV and MI had approximately similar coefficients. Indeed, this was clear evidence that the MI were biased; the qualification variable was nearly equal to zero. Even though the MI were biased, they still performed better than CC; it was identical to the DV regression. This result was similar to Pepinsky (2018) conclusion. Accordingly, MI were eliminated. It is noticeable that regression with CC was superior to FR, where the adjusted R^2 was higher in the first regression. This result was supported by the findings of (Henry et al., 2013). Moreover, the FR coefficient was like the DV coefficients, which caused a serious omitted variable bias issue. It could be seen that regressions that used FR generated worse results than the two other regressions using the CC or DV approach. Accordingly, we preferred the regression using the IC method, which was superior among other regressions. This method was expected to be the best when the data was MAR, as it was in the analysis above. In addition, the sample was randomly chosen by the MLSD, which met

Groenwold et al. (2012) recommendation. Indeed, choosing the IC method captured all of the unregistered workers compared to the qualified registered workers.

Table 3-13: Comparing the common methods of imputation for the 2013 dataset

Coefficient	Complete cases (CC)	Mean replacement (FR)	Indicator categories (IC)	Drop variable (DV)	Multiple imputations (MI)
Age	−0.1506 (−16.67)	−0.1113 (−39.17)	−0.0999 (−36.47)	−0.1132 (−39.09)	−0.1132 (−39.09)
Age-squared	0.0015 (16.07)	0.0013 (43.78)	0.0011 (41.35)	0.0013 (43.66)	0.0013 (43.65)
Zone	−0.0009 (−0.40)	0.0163 (29.43)	0.0175 (32.76)	0.0160 (28.38)	0.0161 (28.40)
Colour	0.0850 (12.48)	0.0629 (38.22)	0.05266 (33.08)	0.0652 (38.90)	0.0652 (38.91)
Size	0.0662 (7.56)	0.0632 (30.23)	0.0536 (26.58)	0.0621 (29.13)	0.0622 (29.21)
Qualifications	−0.0335 (−23.40)	−0.0608 (−59.30)	−0.0512 (−106.05)	-	−0.0034 (−8.76)
Nationality	−0.0074 (9.84)	0.0017 (10.23)	0.0016 (9.63)	0.0017 (9.93)	0.0017 (9.94)
Occupations	−0.1279 (−40.53)	−0.0985 (−114.62)	−0.0881 (−105.53)	−0.1073 (−124.49)	−0.1069 (−123.88)
Activities	−0.0028 (−0.95)	−0.0015 (−2.00)	−0.0021 (−2.81)	−0.00002 (−0.02)	−0.0001 (−0.12)
Constant	11.9013 (53.61)	10.1213 (140.29)	10.1694 (148.13)	9.2985 (128.95)	9.3418 (129.28)
Adj R^2	0.3079	0.2469	0.3021	0.2189	
Observation #	9,163	94,312	94,312	94,312	94,312

3.2.2.2 2017 dataset

This dataset had missing information for qualifications and education. Both variables could not be MCAR because the null hypothesis in Little's MCAR test was rejected. Therefore, the missing mechanism could be MAR or MNAR. To identify which mechanism those variables followed, this section discusses the distribution of those variables, supporting the analysis with some tables and figures.

3.2.2.2.1 Qualification

Although the percentages of the observed values conditional upon occupations were around 25.16% for clerical, 8.57% for managers and 11% for basic engineering, we could not generalise that the lowest qualification was the highest observed (see Table 3-14). The first four occupation categories were much more heavily represented among those with existing qualification information, whereas the last three occupation categories were much more heavily represented among those with missing qualification information. The joint distribution revealed that the percentage of this missingness was expected to be high in the high-density occupation categories and vice versa. For example, agriculture, which had the lowest occupation density, had the lowest missingness percentage. Similarly, basic engineering had the highest occupation density and the highest missingness, with 21% (see Columns 3–5, Table 3-14). This revealed that the observed value was significantly spread among all categories. Although we could not deny that missing qualifications were smaller in the highest occupations, we could not generalise that the higher occupations also formed the largest observed percentage (see Table 3-14). Thus, missing qualifications were not dependent on the observed qualifications, although there was a relationship between total occupation and missingness. Therefore, it could be said that the data followed the MAR mechanism.

Table 3-14: Marginal and conditional distribution of missing and observed qualifications across occupations.

	Conditional Distribution		**Joint and Marginal Distribution**		
Occupation categories	Missing	Observed	Missing	Observed	Total
Managers, directors and senior officers	2.61%	8.57%	1.84%	2.53%	4.37%
Specialists	7.96%	11.85%	5.60%	3.50%	9.11%
Technicians	8.14%	13.28%	5.73%	3.93%	9.66%
Clerical occupations	7.69%	25.16%	5.42%	7.44%	12.86%
Sales occupations	10.12%	13.89%	7.13%	4.11%	11.24%
Service occupations	16.73%	12.14%	11.78%	3.59%	15.37%
Agriculture and animal husbandry	0.79%	0.17%	0.55%	0.05%	0.60%
Industrial and chemical processes	15.78%	3.22%	11.11%	0.95%	12.07%
Basic engineering occupations	30.18%	11.72%	21.26%	3.47%	24.72%
Total	100.00 %	100.00%	70.43%	29.57%	100.00%

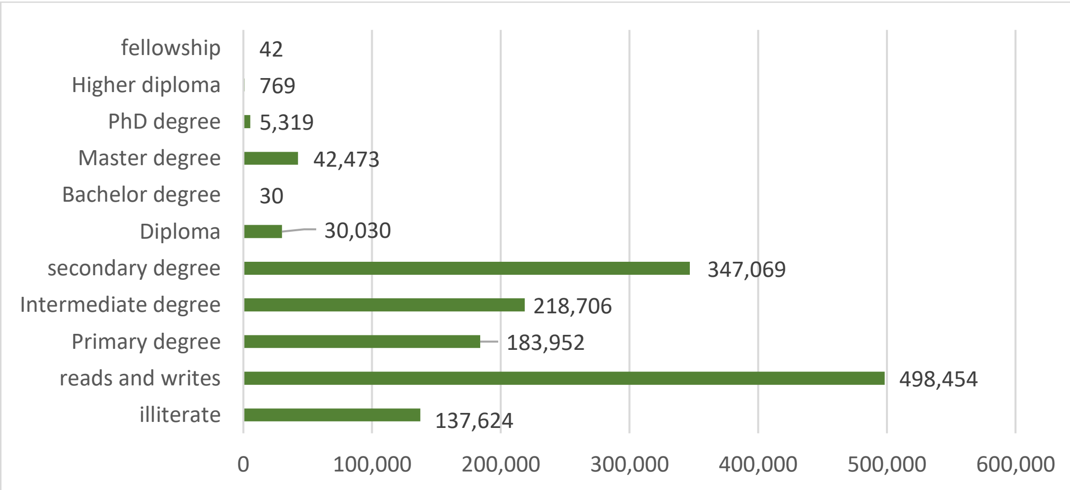
Table 3-15 displays the missingness distribution of the qualification variables, conditional upon each occupation, in turn. Observed qualification information dominated the first and fourth occupation categories, whereas there was a preponderance of missing qualification information for the last three occupation categories (and overall). This conditional distribution provided a clear picture that the missing qualification information was spread across all occupation categories.

Table 3-15: Missing and observed qualification distribution among occupations

	Missing	Observed
Managers, directors, and senior officers	42%	58%
Specialists	62%	38%
Technicians	59%	41%
Clerical occupations	42%	58%
Sales occupations	63%	37%
Services occupations	77%	23%
Agriculture and animal husbandry	92%	8%
Industrial and chemical processes	92%	8%
Basic engineering occupations	86%	14%
Total	70%	30%

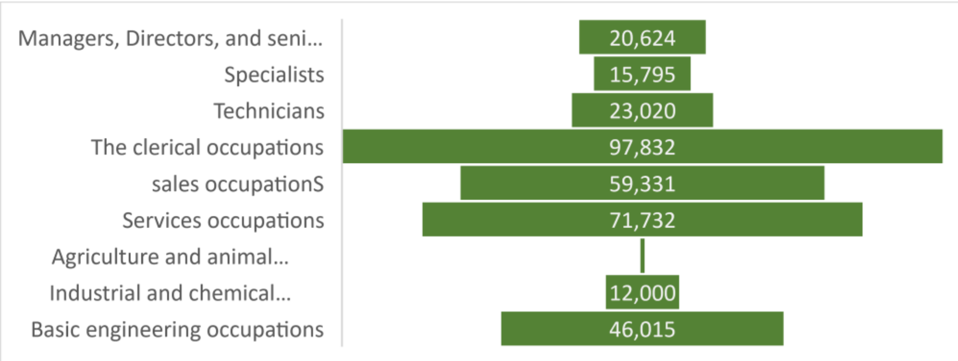
As mentioned above, this trend of missingness in the lowest occupation categories stemmed from two reasons. **First,** qualifications less than secondary school were not included. This dataset allowed us to partially see other qualifications as the education variable was provided. Regardless of the missingness in education, the focus was on the observed education. *Figure 3-27* illustrates that most workers with unregistered qualifications were less educated, especially those who could read and write but did not complete elementary school. This indicated the strength of our expectation above. **Second**, there were mismatches; the high unregistered qualification rates for those who graduated from secondary school implied a mismatch possibility, meaning that those considered semi-skilled did not expect to find themselves in skilled occupations; otherwise, they indicated a mismatched worker or outdated information.

Figure 3-27: The distribution of missing qualifications among education levels.



We found over 50,000 workers in skilled occupations with unregistered qualifications and secondary school education (see Figure 3-28). For example, observation number 2,661,641, a doctor, had a secondary degree; this was expected as a non-updated qualification more than a mismatch; it is impossible to be a doctor without proper qualifications. In another example, approximately 13,972 were managers; this could imply a mismatch where some managers were expected to be self-employed, for example, restaurant managers.

Figure 3-28: Secondary school workers who missed qualifications among occupation.



3.2.2.2.2 Education

This variable was the most challenging; there were 1,680,455 missing education cases, including 1,614,050 with missing qualifications. The other 66,405 observations had known qualifications. This led us to recognise a distinction between the more qualified and the unqualified. Although we expected that a high percentage of those unregistered in both qualification and education would be less educated, we could not prove that. Therefore, we followed a similar strategy to see if there was any dependency of missingness on observed education levels. The marginal distribution in percentage suggested that the lowest three categories comprised the highest missing education, with approximately 50%; following that were the semi-skilled occupations, with 27%, and the lowest missingness was found in the skilled occupations, with 22% (see Figure 3-29). This was unlike the observed percentage, which denoted that the highest observation was found in the semi-skilled job requirements, with 47%; following that were the unskilled job requirements, with approximately 29%. The lowest percentage was found in the skilled job requirements, with 23%. Since the skilled occupations had the least missingness and did not have the most observed data, we could not generalise that the missingness in education depended on observed education levels. However, the registration tended to be higher when the people were educated, either skilled or semiskilled, because of the MLSD orientation towards registering both qualification and education to avoid problems with forged certificates.

Indeed, the conditional distribution revealed that missingness in total formed about 38% on average. Most categories had a lower missingness percentage than the average point, apart from specialists and the three lower categories (see Table 3-16). However, there was no evidence of missing education being in one category; however, it was higher in some categories, which suggests that missing education followed the MAR mechanism.

Figure 3-29: Conditional distribution of education missingness upon the missing status

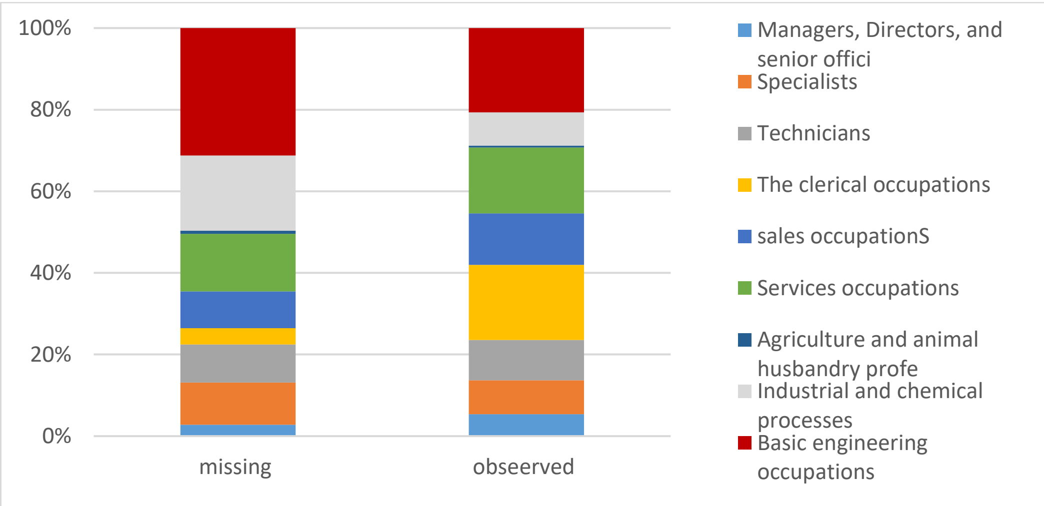


Table 3-16: Conditional distribution of education missingness across occupations

	Missing	Observed
Managers, directors, and senior officers	24%	76%
Specialists	44%	56%
Technicians	37%	63%
Clerical occupations	12%	88%
Sales occupations	31%	69%
Services occupations	35%	65%
Agriculture and animal husbandry	53%	47%
Industrial and chemical processes	59%	41%
Basic engineering occupations	49%	51%

For z-score result see (Appendix A, Section 9.4).

From the above analysis, the missingness mechanism was MAR. Given that, it did not directly lead to the preferred estimation method as it was a categorical value. Therefore, we compared the regression methods: CC, IC, FR, DV and MI.

Table 3-17: Comparing the common methods of imputation for the 2017 dataset.

Coefficient	Complete cases (CC)	Means replacement (FR)	Indicator categories (IC)	Drop variable (DV)	Multiple imputations (MI)
Age	0.02049***	0.0246***	0.02244***	0.0236***	0.0251***
Age-squared	−0.0002***	−0.0002***	−0.00014***	−0.00015***	−0.00017***
Education	0.10911***	0.0828***	0.02044***		0.05273***
Qualification	−0.00466***	−0.0099***	−0.01416***		−0.00286***
Female	−0.27922***	−0.3135***	−0.28595***	−0.2668***	−0.29569***
Colour	0.0354***	0.04104***	0.042595***	0.0453***	0.04287***
Size	0.05963***	0.0646***	0.06601***	0.0641***	0.06400***
Saudi	1.1300***	1.5638***	1.592999***	1.6094***	1.57632***
zone	−0.0031***	−0.00294***	−0.00282***	−0.0011***	−0.00197***
Firm age	0.00076***	0.0034***	0.00373***	0.0038***	0.0036***
Firm age2	−0.00001***	−0.00004***	−0.00004***	−0.0001***	−0.00004***
Activities	0.01001***	0.00418***	0.005343***	0.0082***	0.00603***
Occupations	−0.0556***	−0.05903***	−0.06362***	−0.0724***	−0.0648***
Constant	5.9794***	5.6491***	6.06428***	5.9087***	5.6851***
Adj. R^2	0.5153	0.7450	0.7388	0.7268	
Observation#	1,226,339	4,371,262	4,371,262	4,371,262	4,371,262

***significant at 1%.

Table 3-17 above shows the similarity between the methods in general. We used F-tests to determine which was the best model compared to the DV. The test indicated that the DV was the worst among the other three models, making identification of the best model challenging since the DV model was eliminated. The CCs method had the lowest adjusted R^2 – a reasonable reason to exclude this model. Those steps did not help us end up with one model. Thus, We completed the F-test between FR and IC as they were nested. However, this required the extended version of the regression; we found that the null was rejected and concluded that the IC method was better than FR, although it had a slightly lower adjusted R^2. The residual sums of the squares were 918,294.66 and 4,371,177.00 for FR and IC, respectively. The degrees of freedom for both regressions sequentially were 4,371,179 and 4,371,177. In comparing MI and IC, MI provided significant results for all coefficients, while IC showed insignificant results in one zone category, Makkah. The challenge was that this category was significant and positive in MI while it was negative and insignificant using IC. This was rather unlike Pampaka et al. (2016) conclusion, which assumed that MI would provide a significant result without changing the estimation conclusion. However, almost all other categories were identical with the rounded numbers, apart from qualification and education level. Those imputed variables had quite lower coefficients in some categories in the MI approach, compared to IC for those who were more educated. This could happen because the imputation excluded both variables from the imputation to force the imputation process to complete. Using MI to exclude the missing variables from the regression could somewhat lead to an avoidance of the use of MI. Thus, we selected the IC method to have the chance to capture those massive groups' behaviours in the dataset.

3.3 Detecting and dealing with outliers

Before diagnosing the outliers, the sample was limited to age and wage. **First**, the age boundary used a sample between the ages of 15 and 65. This limitation was based on an insertion error or impossible cases, but they were not rare cases. For example, they could be a typo mistake, dummy Saudisation or non-updated information. The minimum age of the labour force is 15. Moreover, the retirement age is 60, and it can be extended for a maximum of five years. This limitation was applied in several related pieces of research; for example, Manacorda et al. (2006) limited their sample to ages 26–60. **Second,** the wage cut-off was 400SR, which is the minimum monthly wage to register on the professional hazard insurance is 400SR for non-Saudi. Those under the cut-off might have illegal company owners or had a percentage share of the sales. Another expectation was relevant; the employers covered the labours' basic expenses, and the employees used their wages for remittance. This could be relevant for workers from low background countries. However, the cut-off for benefitting from social insurance services is 45,000SR; wages above this are not taken into account when a pension is due. Despite this fact, there are still employees who can earn more than this, according to their contract. Thus, we excluded the highest wages if outliers were influencing the regression. To do so, we searched behind the extreme values in the dataset to check the upper cut-off chosen for wages. Extreme values for wages were recognised in 12 observations that were located within the highest population areas: two in Riyadh, two in Makkah and eight in the Eastern Province. However, we kept the observations considered rare values when the occupation was in the manager and specialist category; other observations were displaying typo mistakes.[47]

3.3.1 Detecting outliers

After limiting the data, outlier observations needed to be checked (Aggarwal, 2016). There are two main techniques for detecting outliers: statistical and graphical. **First**, statistical techniques provide a numerical measurement for an outlier. Leverage value, for example, calculates the diagonal element on the predicted matrix. It is a value between zero and one; at least one unusual value can be found if it is near 1. As a rule, if the leverage value is greater than the cut-off value of 2L or 3L, the data is diagnosed with high leverage. L is the average value computed as the summation of the number of independent variables plus one (the intercept), divided by the number of observations (Hahs-Vaughn, 2016). Cook (1977) constructed another measurement to assess the influence of each case on the slope at least square (Cook, 1979). The cut-off point was $2 / \sqrt{n}$, where *n* was the number of observations (Fox, 2015). Furthermore, the studentised residual was a practical outlier measurement. This residual could indicate the presence of an outlier when it took a value greater than 3 (Hahs-Vaughn, 2016). However, Fox stated that the cut-off value was 2 (both cut-offs are for absolute values). The standardised residual is less sensitive than the studentised residual for outlier observations (Hahs-Vaughn, 2016) because the studentised residual considers the standard error of the regression with one case removed in its formula.[48] Thus, the smallest residual tends to be for high-leverage observations. Those observations pull the regression line towards them (Fox, 2015). **Second**, graphical techniques, such as the added variable plot and the residual plot, helped visualise those statistical tests. Aggarwal (2016) argued that the statistical methods outlined above experienced weaknesses like the simple assumptions regarding sample representativeness, algorithm poorness and difficulties with interpretability. Admittedly, the graphical method could not be achieved without statistical values. For example, a leverage plot for all observations could give a complete picture of this test, unlike the statistical value.

[47] Those could be typing mistakes; for example, observation number 94,337 was 379,900 rather than 3,799.00.
[48] For the formula (see the Appendix A, Equation 9-4).

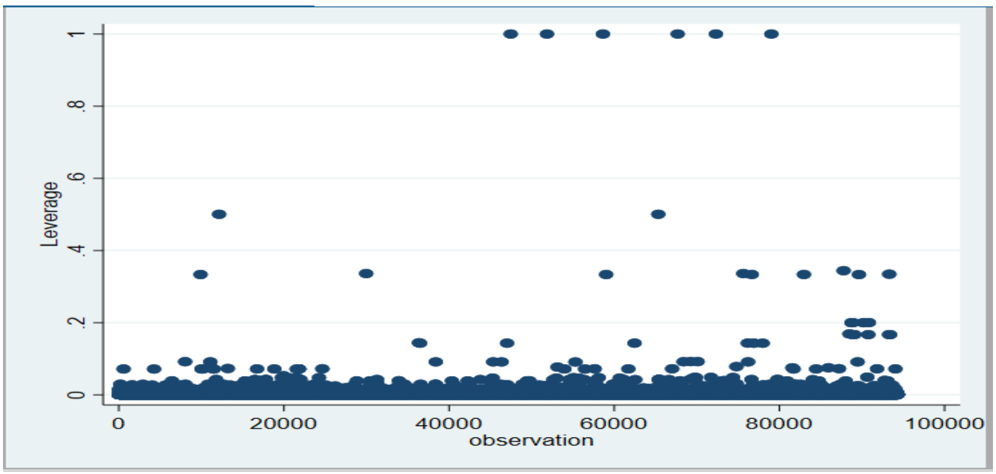
Indeed, this would reflect the weakness of the graphical method accordingly, if we accepted the argument on principle. Therefore, we do not fully agree with Aggarwal's argument. This can be generalised for any graphical plot provided as all graphs consider the statistical value, which is not simply a matter of wise judgement. Researchers should diagnose the data and make a wise decision so that no outlier will unduly influence the regression results.

3.3.2 Application on the datasets

3.3.2.1 2013 dataset

Considering the 2013 dataset, we first undertook an OLS regression, considering all variables. We predicted the statistical values needed. The leverage value maximum value was 1. Even though the mean was 0.00109212, there were some individual cases classified as unusual. There were roughly 3,956 observations greater than the leverage cut-off value.[49] Plotting these values helped evaluate whether the outliers had influencing points on the regression. We plotted the leverage values versus the observations. Figure 3-30 shows that there are several points away from other observations, certainly above 0.2, which was higher than the cut-off point. Dropping those values decreased the power of the model by 0.0187. Thus, we kept those values, which provided a better power of fit.

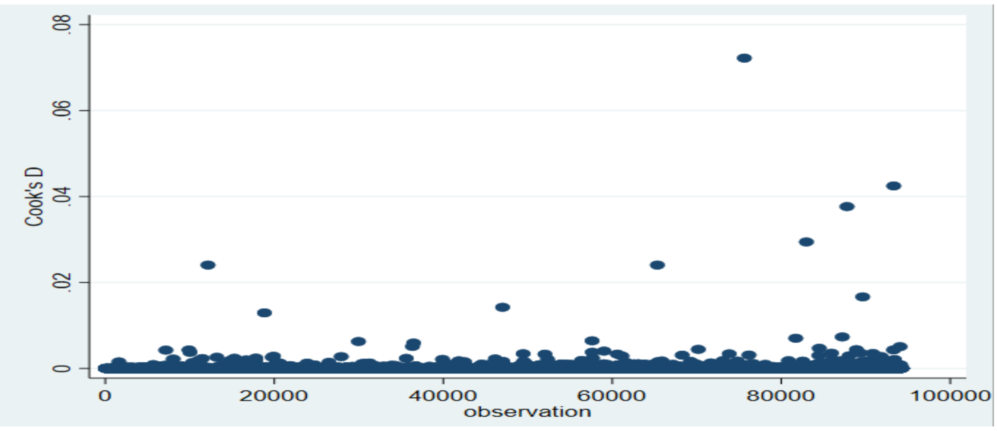
Figure 3-30: Observation and leverage value plot for 2013 data



[49] *3 * L = 0.0032763*, where *L = (102 + 1) / 94,312 = 0.00109212*, which should be equal to the mean of the leverage value.

The mean value of Cook's distance was 0.0000157, which indicated that most of the data was located around the cut-off value.[50] However, roughly 12 observations were quite distinct from the others; regardless, it was powered towards the model slope (see Figure 3-31). Noticeably, the Cook's D plot demonstrated a similar picture for unusual points with the statistical cut-off, unlike the leverage plot. However, dropping those values did not change the result; the power of the model was similar. Moreover, the coefficients were identical when rounding was taken into consideration. This might explain Aggarwal's caution about using the statistical terms; the indication of the unusual point cannot be taken seriously without diagnosing the model.

Figure 3-31: Observation and Cook's distance value plot for the 2013 dataset



Unlike the two tests above, studentised residuals have two cut-off values. Approximately 1,026 values exceeded the cut-off: 874 observations were over 3, and the rest were under −3. According to Hahs-Vaughn, observations considered as outliers could pull the regression towards them. Unlike the statistical terms, the plot shows that most of the data looked alike, even if they were outside the cut-off boundaries, unless there were very few observations (sees Figure 3-32). Those values seemed to influence the regression; dropping them improved the model power by 0.0263. Indeed, dropping them caused more problems than improvements. Thus, including them in the regression was necessary.

[50] C = 0.006512481735567 ≈ 0.01. For the formula details (see the Appendix A, Equation 9-3).

Regardless of the sensitivity, standardised residual results were approximately identical to those from the studentised residuals. This is evident from a visual inspection of Figure 3-32 and Figure 3-33. This conclusion was unlike that argued by Fox. The reason behind this result is that the removed cases formed only 1.13% of the total observations, and most of them had leverage above the cut-off point, far from 1.

Figure 3-32: Studentised residuals against observations plot for the 2013 dataset

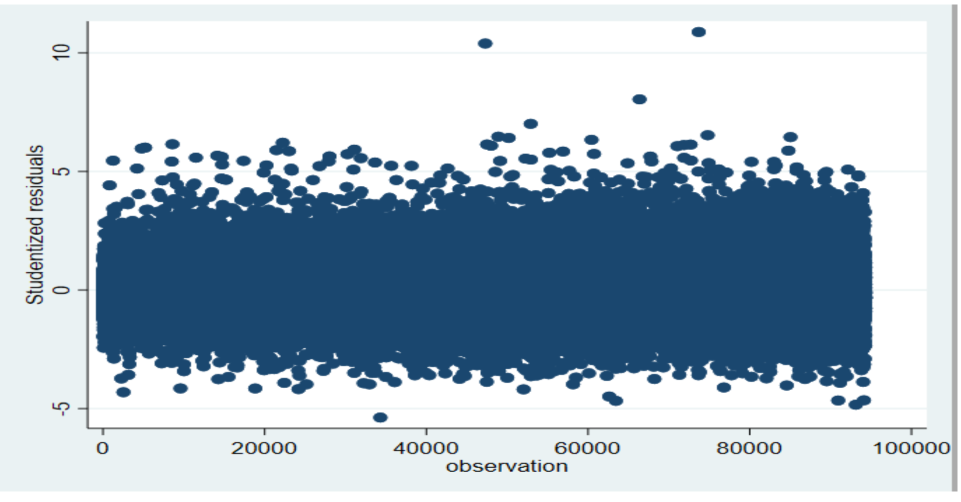
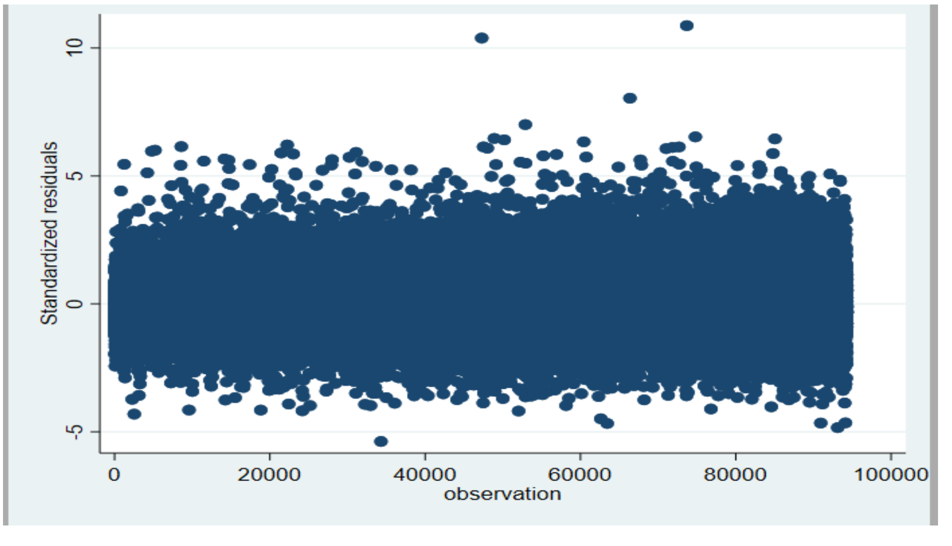
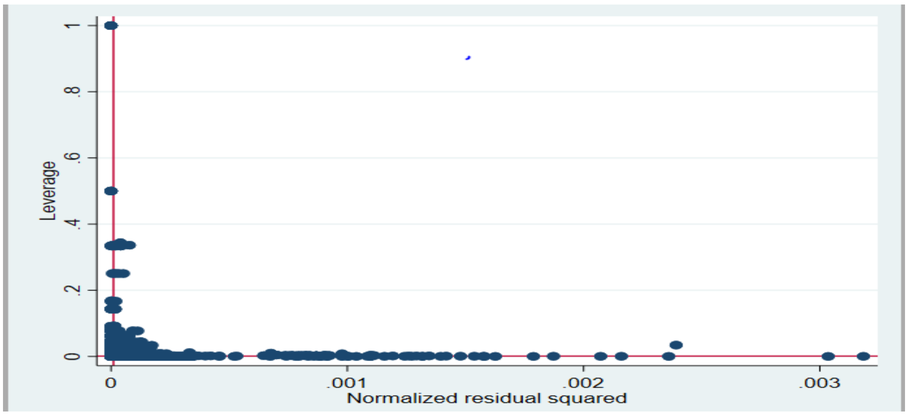


Figure 3-33: Standardised residuals against observations plot for the 2013 dataset



However, the studentised residuals were our preferred statistics, which could be used to recognise the outliers in the data, as Fox (Fox, 2015)stated. Moreover, he preferred to plot this value versus the predicted value. He considered plotting the fitted value against the raw residual unsatisfactory. According to him, there is a correlation between the observed value of the dependent variable and the residual term, given by $\sqrt{1 - R^2}$. Plotting leverage versus normalised residual is commonly used to diagnose a model issue, as well. Figure 3-34 illustrates that the data was spread along the leverage value as interrupted small groups of observations. The observations above 0.6, for example, were dropped, and the observations spread as interrupted clusters. This could be an indication of a heteroscedasticity issue more than an outlier problem.

Figure 3-34: Leverage versus normalised residual squared for the 2013 dataset



The observations that were diagnosed as unusual, through both statistical methods and graphical plot, were rare. Excluding them reduced the regression's power or deformed the regression's conclusions. Thus, it was wise to keep those less frequent observations for this dataset.

3.3.2.2 2017 dataset

A similar approach was applied to the 2017 dataset. There were 124,780 observations above the leverage cut-off, which was 0.0000582. The regression power was improved by 0.0006; however, several qualification categories disappeared as a consequence, for example, the College of Medicine and Pharmacy. Therefore, the observations that were rare needed to be included in the regression. Figure 3-35 shows that there was only one observation far from the other observations. Indeed, dropping this single observation produced an estimation identical to the original regression, meaning the regression outcome would not be affected whether we kept this outlier or removed it. However, only one observation was considered an outlier via the graphical method (see Figure

3-36).[51] Surprisingly, individual number 2,469,426 was considered an outlier in both leverage plots and Cook's D. We kept this value, as mentioned above.

Figure 3-35: The leverage plot against the observations for the 2017 dataset

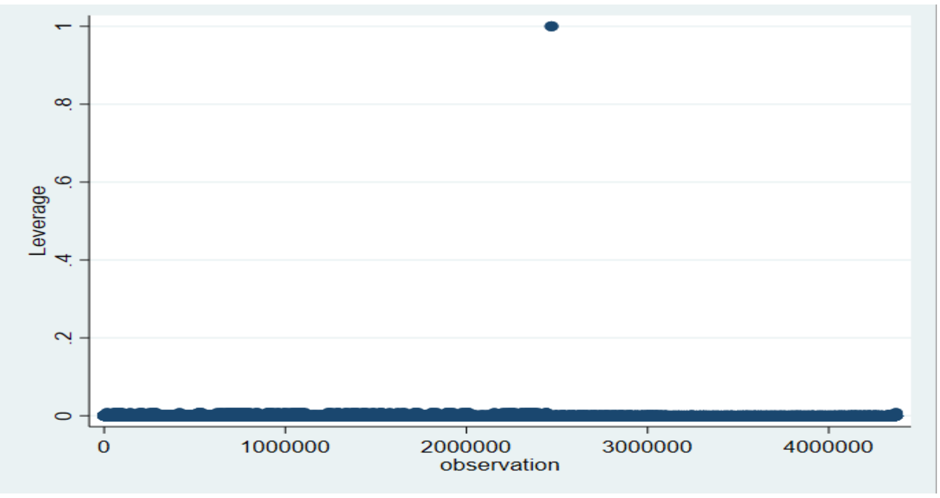
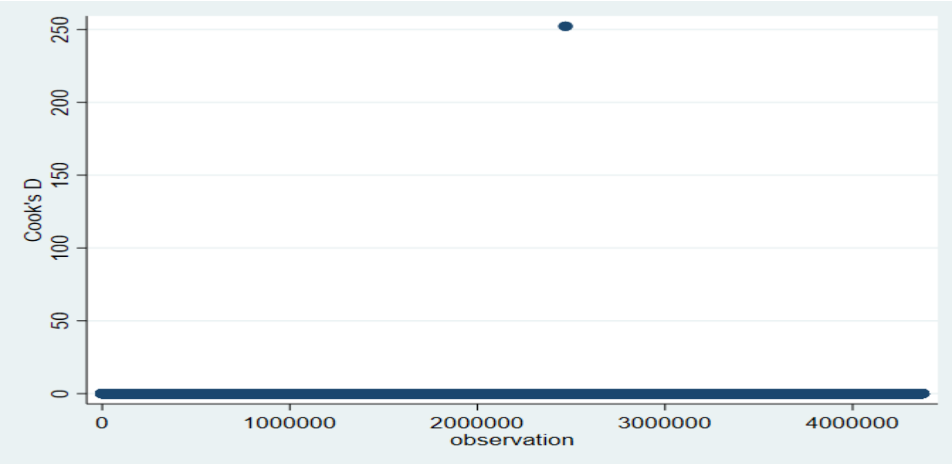


Figure 3-36: Cook's distance plot against the observations for the 2017 dataset

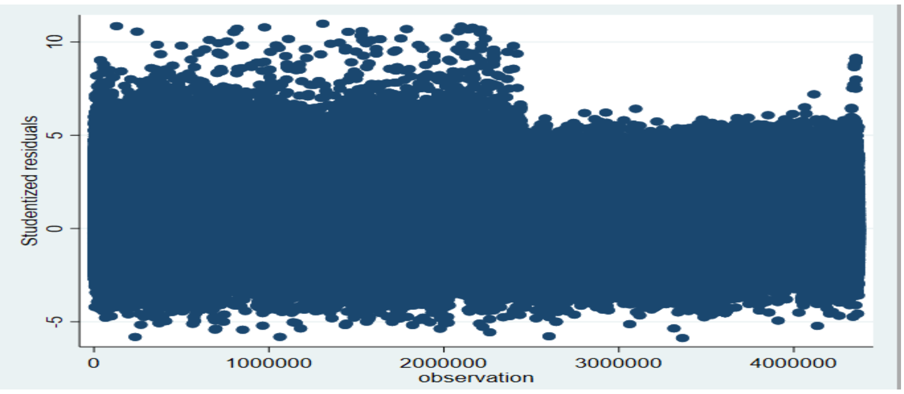


Furthermore, around 59,274 observations were nominated as outliers, according to the studentised residual. Dropping them increased the model power from 0.7809 to 0.8147, which was an improvement of approximately 0.0338. However, when those observations were dropped, the coefficient sign of the professionals was negative, which was an unrealistic conclusion. By visualising these values, it became clear that the

[51] The cut-off for Cook's distance was c = 0.00095659; accordingly, there were 12 unusual observations.

54,996 observations above the cut-off were not outliers, but they were above the value; the same can be said for the values under the lower cut-off (Figure 3-37).

Figure 3-37: Studentised residuals plot for the 2017 dataset



According to the studentised conclusion, depending on the statistical terms only would mislead researchers if they did not consider the graphical approach. In this context, the fitted value was plotted against the studentised residual (Figure 3-38). The graph indicated that there were no clear unusual values, with most of the values found to surround each other. Similarly, the leverage value was plotted against the square normalised residual; this plot indicated that there were no outliers detected apart from observation number 2,469,426, which was mentioned above (Figure 3-39).

Generally, those figures confirmed that the best solution to detect outliers combined the statistical and graphical approaches. This way was guaranteed to provide a fuller picture of the outliers' possibilities.

Figure 3-38: Studentised residuals against fitted values for the 2017 dataset

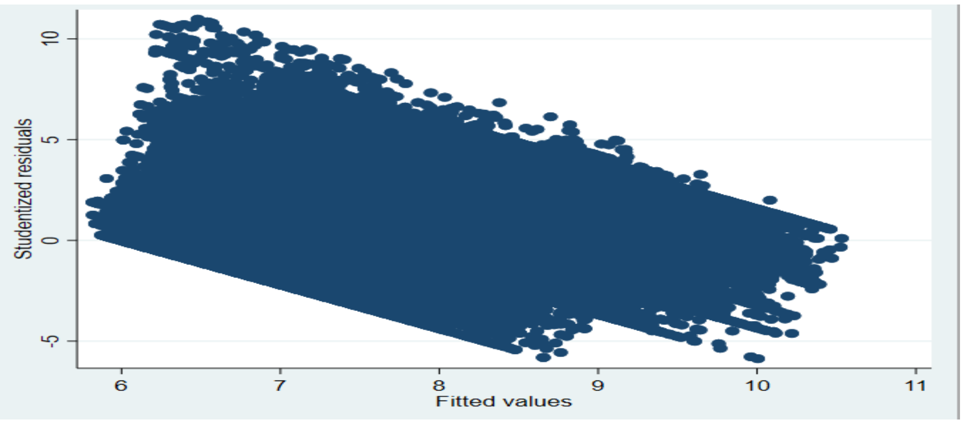
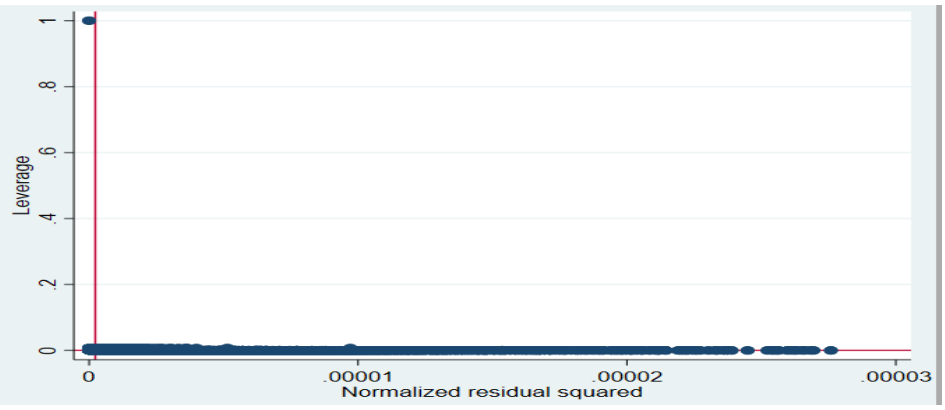


Figure 3-39: Leverage plot against the normalised residual squared for 2017 dataset



3.4 Sample test

Concerning the privacy policy applied in those years for the two separate cross-sections mentioned above, the two datasets were very dissimilar in observation numbers. The 2017 sample was more representative of the total labour market population compared to the 2013 dataset, which was provided as a random sample representing approximately 1% of the total workers, while the 2017 dataset was roughly the total worker population. Moreover, the 2013 sample had fewer Saudi observations, unlike the 2017 sample, where Saudis formed around 19% of the MLSD sample. This huge variation in the observation numbers for both groups reflected the density of non-Saudis in the Saudi market, which resulted from the huge inflow of foreign workers into the labour market when oil was discovered. Indeed, this inflow of labour affected the wages in the market. This could indicate that each sample had a distinct distribution. Therefore, the two samples were tested to understand each sample distribution separately.

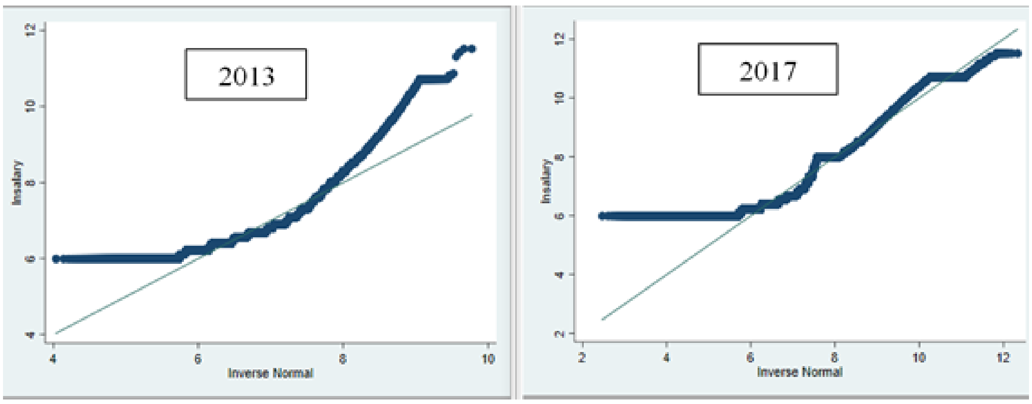
After filtering the above data, in 2013, there were 94,312 observations; 1,943 were Saudi, and 92,369 were non-Saudi, while in 2017, there was a sizable sample, with around 4,371,262 observations. There were 2,472,014 non-Saudi observations, forming roughly 56.55% of the sample, and 1,899,248 Saudi observations (approximately 43.45%). However, neither dataset reflected the true proportions of the group participation compared to the total population. Moreover, the overall sample salary's mean in 2013 was approximately 1,407SR (per month). The Saudi salary mean was roughly 8,622SR, whereas the non-Saudi salary mean was around 1,255SR. In 2017, the wage mean in this group was around 1,084SR (per month), unlike Saudis, who had a mean of 5,016SR, while the overall average mean was 2,483SR. Although there were non-Saudi workers who earned more than some Saudi workers because they were qualified, lower-paid workers pulled the overall wage mean towards them because they formed the largest fraction of market participation. This implies that the market was labour-intensive and employing those low-wage workers was profitable. Noticeably, the wage means decreased for both sub-groups between these years, but the overall mean increased because of a move towards Saudi workers in the composition of the overall sample.

To diagnose the data distribution, several graphical and statistical methods could be applied. This could help understand how the distribution of the data compared to the expected normal distribution. Moreover, it could give a picture of how the data distribution was influenced by the non-Saudi distributions that reflected some of the labour market's features. As the logarithmic form of wage is used in the literature as the key variable in earning regressions, the normality test was examined for this variable.

3.4.1 Graphical methods

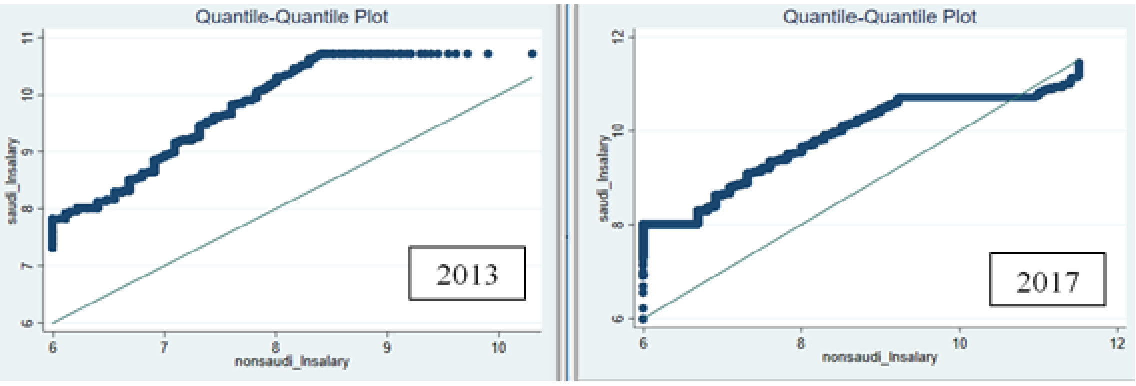
Several techniques could be performed, such as the Q–Q plot, P–P plot or cumulative frequency plot and density plot for a variable. Figure 3-40 shows that the logarithmic form of the wage from the 2017 dataset displayed more normality compared to the 2013 dataset, where it was more obviously positively skewed. Both plots show the limitation imposed earlier, where the lower wage cut-off was around six on the logarithmic form.

Figure 3-40: The Q–Q plots for logarithm wages for the 2013 and 2017 datasets



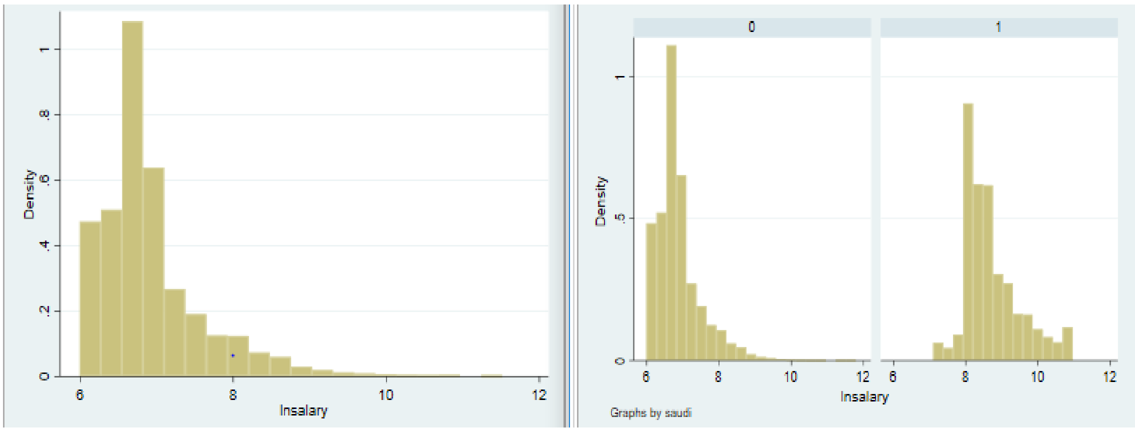
Regarding the subgroups of being Saudi or not, the Q–Q plot has another tool; it requires first dividing the sample into the two nationality groups. This version of the Q–Q plot used the horizontal variable (non-Saudi) as the reference distribution rather than the normal distribution, while it showed the variance distribution for the vertical variable (Saudi). Figure 3-41 shows a clear gap between the distributions of the two groups. For 2013, the Saudi variable was above the non-Saudi variable, roughly following the distribution. However, the 2017 plot shows a similar trend, but after the tenth quantile, the trend was reversed, meaning that non-Saudis could earn more than Saudis in the upper quantile wage categories, corresponding with the Nitaqat programme.

Figure 3-41: The Q–Q plots between Saudi and non-Saudi salaries in the logarithmic form



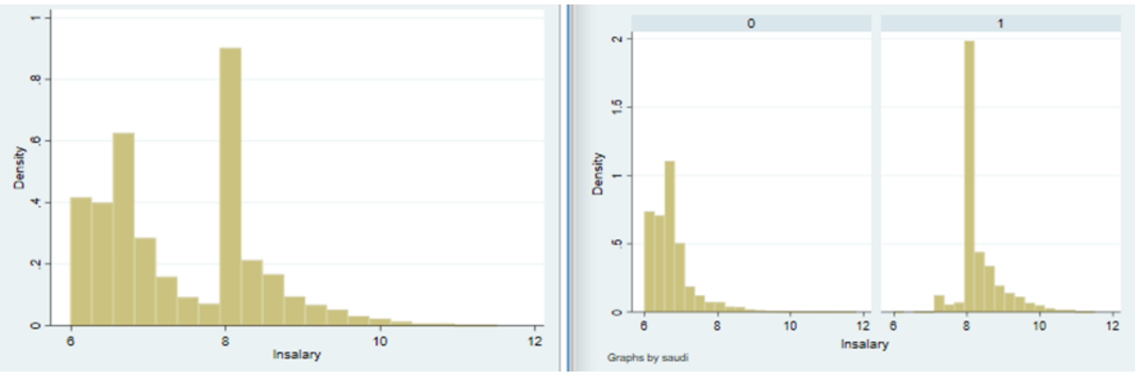
The density description shows that 2013 followed the lognormal distribution overall, following the distribution of the much larger non-Saudi subset, as mentioned previously. Saudis followed a similar distribution as well. However, there was a noticeable feature in the subgroup distribution, where the highest bar was different (Figure 3-42). For non-Saudis, it was at a log-salary of less than 7, which was where the Saudi bars began. This implied segregation in the Saudi labour market. Saudi workers were largely not employed below a wage of roughly 3,000SR (log-salary of 8) because this is considered a very low wage for full-time Saudi workers to live in a basic living style.

Figure 3-42: Total and subgroup density distribution for the 2013 dataset



For Saudis in 2017, it seems there was a change in the data registration policies; Saudis registered under and above the GOSI cut-off (Figure 3-43). Another change was captured; there was a cluster of Saudi workers around the salary cut-off indicated by the Nitaqat criteria. The density of workers at this cut-off point increased significantly; otherwise, it decreased. This could provide tentative evidence of the impact of the quota policy (Nitaqat) in the short run. This initial glance was consistent with our hypothesis, which assumes that introducing affirmative action associated with a low wage constraint would harm the target group's welfare. For non-Saudis, there was a small variation in the wage distributions between the two datasets (see Figure 3-42 and Figure 3-43). Neither datasets followed the normal distribution, according to all of the graphs discussed above. However, the density of the wage's logarithmic form seemed to show that wages followed the lognormal distribution in both subgroups. This was an expected result, given that the wage's logarithmic form occupied a renowned place in the literature. However, this was not the case for the overall wage distribution, which did not follow the specific distribution. Admittedly, the distribution perfectly captured the segregation in the labour market, even though it was not referring to a specific distribution type.

Figure 3-43: Total and subgroup density distribution for the 2017 dataset



3.4.2 Statistical methods

This approach could be taken via different tests for normality, such as the Shapiro–Wilk test, Shapiro–Francia test, Kolmogorov–Smirnov test, W/S test, D'Agostino's test, Anderson–Darling test and Jarque–Bera test.[52] Each method had strengths and weaknesses. The results were expected to support those from the graphical methods. With a sample as large this dataset, there is a tendency to reject the null hypothesis, which assumed the data was normally distributed. However, the Shapiro–Wilk test is considered an outstanding test for capturing non-normality compared to the Kolmogorov–Smirnov test and the Anderson–Darling test (Mellinger & Hanson, 2016).

Starting with the Shapiro–Wilk test, the most common normality test, the null assumption was that the data was normally distributed, and the alternative assumption was that the data was not normally distributed. This test is often denoted by W and requires the sample to be sorted in ascending order (Thode, 2002). The weakness in this test is that it requires a small sample size (*n*), recommended to be less than 50. However, Royston developed the test to cover sample sizes up to 4,000 observations. Both the original and extended versions assumed a small sample. Therefore, applying this test in our datasets was expected to provide a non-normality result, which would reject the null hypotheses according to the sample size. Supporting this point, the test results for the 2013 and 2017 datasets in total indicated that salaries were not normally distributed in either their raw or logarithmic forms (Table 3-18).

---

[52] For test results (see Appendix A, Section 9.6 and 9.7).

Table 3-18: The results of the Shapiro–Wilk test

	2013		2017	
	Statistic	p-value	Statistic	p-value
Salary	0.32562	0.000	0.52673	0.000
Ln (salary)	0.89038	0.000	0.94867	0.000

The Shapiro–Francia test was developed for a maximum sample size of 5,000 observations and was expected to have a similar conclusion to the Shapiro–Wilk test. This test was approximate to the W value, which used a similar approach using the least squares method. This test is considered simple compared to the Shapiro–Wilk test, which depends on the covariance and the mean. This means the formula is different in both tests, which makes the Shapiro–Wilk a powerful result, even though it requires fewer observations. Although the two tests used different methods to estimate the ordered sample on the sample variance, the results were roughly similar for our data (Table 3-19). The results again supported the non-normal assumption, where the null was rejected. This confirmed that, depending on the statistical test, the results might be misleading since the graphical method showed the lognormal distribution for the two datasets.

Table 3-19: Shapiro–Francia test results

	2013		2017	
	Statistic	p-value	statistic	p-value
Salary	0.32558	0.00001	0.52673	0.00001
Ln (salary)	0.89053	0.00001	0.94868	0.00001

Regardless of the different conclusions for the lognormal distribution between the graphical and statistical methods, they did agree that the data was not normally distributed. This conclusion did not mean that linear regression could not be performed, or the t-test used, as is widely believed. When the sample size is sufficiently large, linear regression and t-tests can provide reliable results. This sufficient size would be 100 observations or more (Lumley et al., 2002). Accordingly, the t-tests for our sample sizes provided reliable results, for example, using the t-test to examine the mean equality of two subgroups – Saudis and non-Saudis. The null hypothesis of this test assumed the means of the two groups were identical; the alternative was that the two groups' means were significantly different. The t value was calculated as follows.

$$t = \frac{(\bar{\mu}_s - \bar{\mu}_{ns})}{SE_{\bar{\mu}_s - \bar{\mu}_{ns}}} \qquad \textit{3-1}$$

Where $\bar{\mu}_s$ represented the Saudi logarithmic wage mean, $\bar{\mu}_{ns}$ illustrated the non-Saudi logarithmic wage mean, and SE represented the standard error of the two groups' means. When the sample is sufficiently large, the standard error is equal to the standard deviation. This led to the possibility of estimating the standard error depending on the pooled variance for both sample groups. Following Cohen (2013), the pooled variance, S_p, would be written as follows.

$$S_p^2 = \frac{(n_s - 1)S_s^2 + (n_{ns} - 1)S_{ns}^2}{n_s + n_{ns} - 2} \qquad \textit{3-2}$$

The standard error, $SE_{\bar{\mu}_s - \bar{\mu}_{ns}}$, for both groups' means would be written as follows.

$$SE_{\bar{\mu}_s - \bar{\mu}_{ns}} = \sqrt{\frac{S_p^2}{n_s} + \frac{S_p^2}{n_{ns}}} = \sqrt{S_p^2[\frac{1}{n_s} + \frac{1}{n_{ns}}]} \qquad \textit{3-3}$$

These results showed that for both 2013 and 2017, the wage means in the logarithmic form were significantly different between Saudis and non-Saudis, where the t values were −1.3 and −2.8, respectively, with a p-value equal to zero.[53] The wage gap between the two groups' means in the logarithmic form were 1.83 and 1.58 in the two datasets, respectively. This result supports the obtained result from the graphical method above (Figure 3-41).

Moreover, Hotelling's T^2 test, with a similar null hypothesis, confirmed this result, as well. The T^2 value of this test was considered a generalised form of the student t-test (Appasani & Visscher, 2016). The results indicated a wage gap between the two groups, which could be indicative of a discrimination issue. Indeed, the heterogeneity was a possible reason for this gap between the sub-samples. Several tests could be conducted, such as Levene's, Wilk's lambda and Pillai's trace. Levene's test is an alternative to Bartlett's test, using the F-distribution rather than the chi-squared distribution. This test assumes homogeneity between two groups, where the null hypothesis is that the groups'

[53] The null would be rejected if the t-calculated > t-critical, if the difference between the group mean was equal to zero or positive. The null would be rejected if the t-calculated < t-critical when the difference was negative.

variances are equal, while the alternative hypothesis assumes that the two groups have unequal variances. According to this test, both datasets rejected the null hypothesis for the mean, median and trimmed mean. This means that the variance was not equal for both groups, as well. In other words, the two groups were heterogeneous in their mean, median and 10% trimmed mean (Table 3-20).

Table 3-20: Levene's heterogeneity test for wage's logarithmic form

	2013		2017	
	F-test	p-value	F-test	p-value
Mean w_0	240.49	0.000	560.78	0.0000
Median w_{50}	170.21	0.000	5,253.99	0.0000
Trimmed mean w_{10}	191.94	0.000	262.2	0.0000

3.5 Conclusion

We experienced several issues with the datasets, starting from opening the data until the data was ready for estimations. The limitation on the number of variables was due to privacy policies. However, some of the important variables needed to carry out the research were included. Moreover, there was a high volume of missingness for both dependent and independent variables. For the dependent variable, the wage, we found that missingness did not severely harm the wage distribution. The wage distribution displayed similarities in trends among the GOSI dataset and the OS dataset, although the data originated from different organisations. Furthermore, the wage data in several aspects, such as area and colour zone, was roughly convergent to the GOSI dataset. According to the thorough discussion on missing and observed wages, the missingness in wage was MAR. As a continuous dependent variable, the CC method was chosen to deal with that missingness.

The data was trimmed in ages and wages. This was, first, to limit the analysis on the research to the workforce ages between 15 and 65; otherwise, there was an expectation of typographical errors, dummy Saudisation and non-updated information. This age range was used in GaStat statistics. Moreover, wages were trimmed as well, subject to the GOSI lower cut-off. The reason behind this is that we believed there was a different payment system for those who earned less than 400SR per month. We are aware that this wage was less than $100, which could not meet the subsistence limit, even if these workers also received job benefits. Thus, we assumed those workers followed another payment system, depending on their achievements, and this amount comprised the contract values. Consequently, we excluded them from the sample, where we focused on the monthly payment. The lower limit was set as the minimum wage that could be registered at the GOSI to cover the hazards stemming from that sort of job. This encouraged us to use this limit as a cut-off for low wages. By contrast, we did not impose an upper limit on wages to 45,000SR as some workers received higher wages than this. However, we removed values of more than 100,000SR as those were typos, as explained in the main text.[54] The outlier examination confirmed the importance of removing these values. The issue of outliers was addressed after the wages had been trimmed (for age and wage level). However, we had diagnosed the outliers before we set the limitation; those over 100,000SR were influential. Thus, the limitation generated more consistency in the dataset.

After trimming and dropping missingness, we investigated missingness in the key categorical independent variables: qualification and education. That missingness formed most of the OS, which reflected the weakness of the data collection on those two variables. However, the GOSI data did not follow a similar classification, and the GaStat did not follow a similar method in collecting data. It seems that those two variables suffered from some duplication in the recording process, whereby some were registered in one category but were missed in the other. Moreover, qualification did not include any categories for those with less than secondary school qualifications. Thus, they were unregistered. However, we used occupation as a proxy and found that the observed proportion was higher on the highest occupation categories. This was a reflection of the MLSD rules, which require specialists to announce and register their qualifications to ensure that the job and qualifications match. Thus, the registrations were higher in the highest categories. However, the missingness was not found in one occupation category. On top of that, the highest occupation groups showed some cases with less education. Therefore, the missingness was considered MAR. This mechanism considers data ignorable and provides a variety of choices to deal with missingness. Accordingly, we used the IC method to capture the large missingness population.

[54] (See Heading 4.3: Detecting and dealing with outliers).

The wage was found to follow the lognormal distribution, which suggested the use of the logarithmic forms of wages. Moreover, there was a clear separation between the Saudi and non-Saudi wage distributions, which emphasised the need to understand this differential.

Chapter 4 Method and the Methodology

4.1 Introduction

Although several empirical and theoretical studies have been carried out on native–immigrant wage determinations, affirmative action policies and wage differentials, there is still very little understanding of those issues for the Saudi Arabian labour market. Unlike other economies, Saudi Arabia and other GCC countries are distinctive because foreign workers form most of the population and the labour force. Thus, the policies should be derived from the labour market data. The foreign workers found extensively in the lowest wage categories are from lower background countries, which has created a large wage gap between the two groups, describing Saudis as the advantaged group. However, the employment gap was in favour of the foreign workers, which was the motivation of the Nitaqat programme. This was unlike other economies' motivations for applying affirmative action policies, where the disadvantaged groups usually suffered from low employment and wages. The wage gap between Saudi and non-Saudi workers is under Nitaqat policy addressed in a simple framework reflecting the research hypnotises.

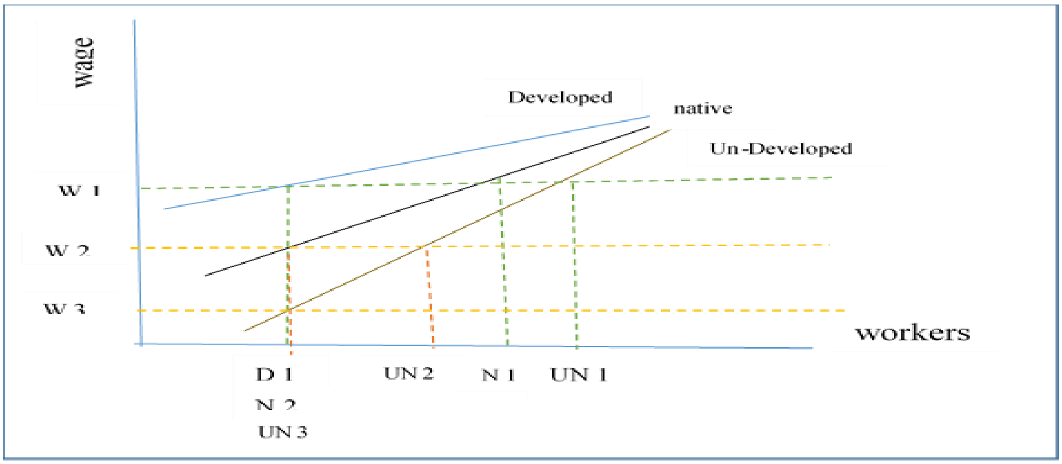
Unlike the previous chapter, which offered a fundamental descriptive analysis of the data, this chapter focuses on the empirical model and the theoretical analysis to provide a thorough understanding of this distinct market. Indeed, the wage gap between Saudis and non-Saudis cannot be denied.

4.2 Theoretical discussion

Although the general intuition relatively agreed that the wage gap could be reduced due to affirmative action policies (Burger et al., 2016), the gap will still be found and be at least partly unexplained (Groshen, 1991; Ransom & Megdal, 1993). Theatrically, however, researchers’ attention was motivated by the equilibrium models or firms’ choices. For example, job assessment model was used to assess this policy to evaluate the effect on labour productivities (Coate & Loury, 1993). Moreover, Leonard illustrated his point of view through supply and demand theory. He explained that the policy constraint will shift (increase) the demand of the target group which will increase wages and the gap will reduced (Leonard, 1984a). From other side in 1989, he looked to the policy as a tax constraint on firms which will reflected on the target group (women) progress, however, this will be negligible effect if the women's supply increases due to this policy (Leonard, 1989). However, Griffin (1992) focused on firms' choices optimality, showing that the constraint of this policy will affect firms' cost. He analysed the cost function of two types of firms: contractor firms which was restricted by the affirmative action policy, and non-contractor firms (unrestricted). His hypothesis was contractor firms will experience high production cost. Therefore, these firms will have lower demand elasticity for inputs and lower elasticity of substitution between inputs. Based on the cost function, he estimated the substitute relationship for inputs types: white, black, female, Hispanic, and capital. He found that the demand reduced and the elasticity of substitution between the inputs reduced as well.

Based on Griffin discussion, we deduced that firms distinguish between workers type as a separate type (origin) of inputs which inspired us to shed a light on workers' supply which could be driven from workers choices literature. We suggested that workers supply would be different according to their origin, thus firms distinguish between them. To illustrate this point of view, **first**, we assume no quota imposed yet. We exploited the modern theory of immigrants which suggested that consumption could be appropriate unit of immigrant research (Massey et al., 1993). Therefore, considering their substantial differences on their consumption behaviour would yield different utilities function for each group. Accordingly, disaggregated labour supply would be driven. This implies that employers, when demanding workers, would experience non-linear pricing schemes from the recruiting offices for each country of origin. This practice would reflect the multi-supply (or discrete supply function) available for employers, which means the employers would maximise/minimise their profits/cost through the combination of those alternative sources of labour and capital. Indeed, this situation will be reflected in the firms' demand for each type of worker according to the supply they experience. If we considered similar education level for each labour group's member, multi-supply, which stems initially from the differences in labour utility, could explain the wage and unemployment gap between immigrant and native labour. Figure 4-1 visualises this viewpoint, where workers demanded by employers would be influenced by the several levels of workers' supply by origin, which would yield the wage and employment gap among native and immigrant groups of origin.

Figure 4-1: multi-supply yields wage and employment gaps



Source: Researcher's original work.

For example, if firms had a horizontal demand function equal to W1, the demand would be filled by D1 of workers from developed countries, N1 of natives and UN1 of workers from undeveloped countries. In this demand, the gap in employment would be in favour of natives, then the workers from developed countries. However, if the demand function was horizontal on W2, the employer would not find a supply of any workers from developed countries at this wage. Thus, natives would supply at N2, and workers from developing countries would supply at UN2. However, if their horizontal demand was on W3, they would not find any supply at this low wage except those who came from low background countries. The three scenarios were applied for three different types of jobs, qualifications, and other required characteristics. The quantity of each groups' supply depends on the elasticity of the supply towards the wages. Accordingly, the employment gap was constructed.

Now, the vertical scenarios must be considered. If an employer needs a specific amount of work for a job, it is assumed that D1, N2 or UN3 would be paid different wages to match their supplies. This means that a similar number of employees doing identical jobs would cost the employer different amounts, and the cheapest choice would be the workers from developing countries. Accordingly, when the job is secondary and does not need any qualifications, then the cheapest costs will maximise the firm's profits, but when the job needs a professional, the employer will weigh the cost of the labour with the expected productivity. Therefore, immigrants do not select low wages for themselves. They supply each type of job and have a lower supply function, so they get chosen extensively when the other groups' supplies are costlier. The horizontal and vertical scenarios are rarely found in an economy; the demand for labour is a function of several variables, and the mean variable is the wages. Thus, the demand is a combination of those workers to maximise the firms' profits. Once the demand function is known, then the quantity and the wage will be determined. Frankly, when the demand and supply are activated, a wage gap will result, according to the immigrant supply levels.

Second, the quota policy was imposed. The quota will restrict the workers utilities as they will be under the layoff risk by the firms to satisfy the percentage require in the quota (workers redistribution). This risk would motivate workers to accept higher or lower wages than they do before the quota which will contrubited to explained part of the native immigrant gap. By that, I meant risk diversity between native and immigrant groups because of the quota policy could be considered as a source of explaining the existence of a wage gap according to the hedonic wage theory (Hutchens 1983). However, in terms of explaining the wage gap between native and immigrants, some researchers expected a greater rule of human capital theory explaining wage gap through education and experience (Becker, 2010; Collard, 1972). However, some expected that occupation structure could explain the gap between the native and immigrants' workers (Hanson, 2009). This difference in structure could result from self-selection of immigrants in low status jobs (Smith & Fernandez, 2017), or through entry quota restrictions policy (Ruppert,1999) where immigrants were not free to access a received country without permission. Moreover, modelled matching and turnover behaviour was exploited in this context where the gap between native and immigrants would exist because of the differences in international wages (Royalty et al., 1993). Workers movement types explained a substantial amount of the gap between native and immigrants which was explained through signalling theory and job search theory (Brenzel and Reichelt (2017)).

Unlike the previous research, we looked to the workers' choices to explaining the gap between native and immigrant. Afterward, we expected that quota policy will restrict workers' utility which will affect their choices and either reduce or increase the gap.

Accordingly, we are aiming to understand the effect of Nitaqat on the wage gap between the two groups considering Nitaqat as a source of layoff risk that could explain this wage gap. To do so, we built our simple theoretic framework exploiting two theories – the modern theory of immigrants and hedonic wage theory.

4.3 The suggested framework

This analysis can be captured by taking the Saudi labour market into consideration; the wage is given in the market for firms and workers as representative of demand and supply sides, respectively. Like other GCC countries, the Saudi labour market experienced a high demand for workers that exceeded the real supply in both quantity and quality of employees when oil was discovered. This discovery boosted the government's investments in several activities, which encouraged the private sector to grow rapidly. However, this sudden change in the economic growth required recruiting workers, whether they were skilled or non-skilled. Indeed, this created a type of selectivity among Saudi workers in the public sector as it was much more convenient and well-paying. Thus, the public sector labour demand would follow the equation below.

$$\mathrm{L_G} = \mathrm{L_f} + \mathrm{L_s} \qquad \textit{4-1}$$

L_G indicates the total demand in the public sector, which was formed by Saudis (s) and foreign workers (f), especially skilled workers. However, the high government expenditure on education increased the Saudi human capital. Gradually, the Saudi labour force became more educated, which helped the government substitute foreign workers in the public sector. However, this substitution process gave the government the ability to absorb the Saudi supply for a long time, until it reached the point that the government's demand for Saudi workers was not equal to the Saudi supply, following the equation below.

$$L_G < S_s \qquad \textit{4-2}$$

Unlike the public sector, the private sector depended on foreign workers as they recruited skilled and non-skilled workers as required, with reasonable global wages. Saudis owned most of the firms as the policy did not allow foreign firms to engage the market unless they fulfilled the government criteria.[55] Therefore, one can say Saudis held the job ownership, and the other employees were foreign workers. The private sector demand follows the equation below.

$$L_p = L_f + L_s \tag{4-3}$$

However, the total supply in the private sector was roughly formed by foreign workers.

$$TS_p \cong S_f \tag{4-4}$$

Several reasons encouraged the reluctance of working in the private sector. The lower wages compared to the public sector, especially in the unskilled and semi-skilled jobs, resulted from the vast inflow of foreign workers, which pulled the wages down compared to the public sector, which is consistent with demand theory suggestions. The increase in the supply (the total workers) caused a decrease in the prices (wages). On top of pulling wages down, a harsh workplace exists in the private sector; workers usually work more than eight hours daily, compared to seven hours in the public sector. One could justify that, foreign workers accepted working in harsh conditions because of their characteristics as single workers, even if they were married back home. They were more interested in the job than leisure, so they increased their compensation.[56] Moreover, employers could fire their employees, Saudi or not, without restrictions, which created high competition between workers to keep their jobs and caused less job security for Saudis. However, as mentioned above, the Saudi supply was involved later in the private sector.[57] Thus, the total supply in the private sector follows the equation below.

$$TS_p = S_s + S_F \tag{4-5}$$

As the private sector is our object, we ignore the public sector in the next analysis because of the lower possibility of a move from the private to the public sector. Accordingly, the labour market was already shaped, and the wages were given when Saudi workers became involved in the market employment process. The question is why

---

[55] Review of foreign investment policies and free loans for Saudi investors.
[56] Workers have the right to indemnity; this policy is for both Saudis and non-Saudis.
[57] That does not mean that Saudis did not work in the private sector, but the public sector observed the most Saudis.

Saudis asked for higher wages than foreign workers, which created large wage and employment gaps between Saudi and non-Saudi workers.

Following the modern migrant theory, we chose household consumption as a unit of target analysis from an economic perspective. To fulfil this point, workers accept a job that maximises their utility function. This maximum utility is achieved when their total consumer satisfaction is balanced with job quality. The well-known Cobb–Douglas utility function for i^{th} individual in j^{th} nationality groups follows the below equation.

$$Max\, U_{ij} = u\,(C_{ij}\,, Q_{ij}) \qquad 4\text{-}6$$

C_{ij} represents consumption, both current and future, through the marginal consumption slope. The future consumption is savings, which could be considered as the remittance for non-Saudis. Q_{ij} is job quality, measured in the literature by several job characteristics, such as a holidays system, layoff risk, location, job prospects, housing privileges, wages and worker's hours.

On the one hand, consumption would be discussed as a function of the income, starting with the Keynesian consumption function following the below equation.

$$C_{ij} = \varphi_{0j} + \varphi_{1j} Y_{ij}^{d} \qquad 4\text{-}7$$

Notice, C_{ij}, the consumption, is formed by autonomous (fixed) consumption, φ_{0j}. This consumption is inevitable even when a worker's income is equal to zero, which refers to basic needs consumption, which is assumed as wages, as determined in the subsistence theory of wages (Sharma, 2016).[58] The $\varphi_{1j} Y_{ij}^{d}$ is the induced consumption, which is dependent on disposable income, Y^d, which means the remaining income, which is the wages, W, is obtained by working after subtracting all of the fees, F, and adding the government benefits, B.[59] The disposable wage follows the equation below.

$$Y_{ij}^{d} = W_{ij} + B_{ij} - F_{ij} \qquad 4\text{-}8$$

According to the Saudi policies, Saudi workers could qualify for several benefits, such as unemployment, training, interest-free loans, wage subsidies and income benefits,

[58] The subsistence theory was known earlier in the classical economics.
[59] There is no income tax applied in Saudi Arabia.

while non-Saudis were ineligible.[60] Moreover, non-Saudi workers were required to pay fees to the labour office, the interior ministry and the MLSD. The φ_{1j} denotes the marginal propensity to consume (slope of the consumption function), and $1-\varphi$ is the marginal propensity to save (slope of the savings function). Therefore, φ captures the individual preference variations between consumption and savings, S_{ij}. Accordingly, the disposable income would be spent relatively between consumption, C_{ij}, and savings, where the accumulation of the savings would partially form the total wealth.

$$Y_{ij}^{d} = C_{ij} + S_{ij} \qquad 4\text{-}9$$

By substituting this equation on the consumption equation above and rearranging the formula, the modification function would follow the below equation.

$$S_{ij} = -\,\omega_{0j} - \omega_{1j}\, C_{ij} \qquad 4\text{-}10$$

Where S_{ij}, the savings, is a function of consumption. The ω_{0j} is a fixed value obtained by a negative value of the autonomous consumption divided by the marginal slope of consumptions on the modification process: $\frac{\varphi_{0j}}{\varphi_{1j}}$. This value assumes that when the marginal slope of consumption increases, savings will decrease by ω_{0j}. However, ω_{1j} calculates the percentage change in savings when the consumption slope changes. This is obtained on the modification process as the savings slope divided by the consumption slope. This reflects a negative relationship between consumption and savings, subject to the consumption slope. If the consumption increases under a fixed disposable income, the savings will decrease.

The aim of modifying the consumption function was to see how individuals would change their preferences when their disposable income did not change. An individual would select their standard of living to maximise their utility through a balance of savings and consumption. This differential would be reflected in the consumption function parameters. Indeed, this is a personal choice; individuals are responsible for

[60] The income benefit is called Hesab Al-Mowaten, or citizen account, where the benefit was applied to cover the increase in the basic consumption of oil and electricity as a result of the latest price increases, SANED program and Hafiz as well. For more program detail (see section 3.3.1 above).

selecting their standards of living, not employers (Figart & Marangos, 2013). This means that when an individual is keen to increase his standard of living, he will change his savings/consumption habits under a similar wage, which is determined in the labour market, and the individual can accept that wage or refuse it according to his savings/consumption criteria; this is known as the reservation wage. This implies that the heterogeneity in the consumption preference for an individual would create a gap in the supplied wage, which could be a core of the wage differential analysis. Admittedly, a substantial number of foreign workers in Saudi Arabia come from Asia, which could explain the difference if the consumer needs for Saudis and for the groups of foreign workers. On top of that, the savings amount is not an equal value, even if Saudis and Asians save an equal amount of money. Saving one SR for non-Saudis equals $3.75 at any point, which could be worth a great deal in the lower background countries. However, in countries such as Britain or Jordan, this would not be a large amount; their currency is worth more than this for one SR. From this point, we can understand the sources of different wage valuations according to nationality, which causes the multiple strata of the labour supply. Meaning, a large number of non-Saudis accept lower-quality jobs if it maximises their total consumption – the current subsistence consumption and the savings valued by their home currency. Therefore, they care about remittance to their mother country more than their current consumption, unlike those who come from higher background countries or Arabic origins. They maximise their total utility with the current standard of living at least equal to their home countries. Indeed, the remittance is valued differently according to their origins. However, Saudis require a current consumption at the standard of living, and the savings are set by the price for durable goods, such as cars and houses. This would make the savings undesirable if the prices were unaffordable for those goods. Therefore, it is reasonable to ask whether equal wages could be equal to those who come from high background countries – higher than other Arabic countries – which is like the analysis in Figure 4-1.

On the other hand, job quality can be captured by job characteristics, such as job insecurity or layoff risk as a limitation. The cost of the job quality is measured by the opportunity cost of layoffs, where the expected lifetime income would be affected and, thus, the total consumption. In some firms, according to the localised status, workers are less secure and feel under the risk of layoff to satisfy the programme criteria, unlike the period before Nitaqat was imposed. From this point, we assume that the job quality is equal to the layoff risk, holding other variables, such as hours, salaries and locations, constant. This led to $Q = R$ and the utility function, modified as follows.

$$Max\, U_{ij} = u\,(Cf(\,Y_{ij}^{d}), R_{ij}) \qquad \textit{4-11}$$

However, this layoff risk would result from the Nitaqat programme; employers could be downsized or, in the worst case, closed. Thus, the layoff risk possibility would be equal to one or zero or in between. We assumed two scenarios for the layoff risk. First, we assumed a direct effect for this restrictive policy on the job layoff risk. The policy would increase/decrease the chances of non-Saudi workers layoffs when they belonged to the non-localised firms. An employer could layoff one non-Saudi to satisfy the Nitaqat criteria and become a localised firm. This would be costlier than the layoff of a Saudi, from the employer's point of view.[61] Therefore, the employer could lay off a Saudi who works in a higher wage job and employ two Saudis in the first copy of Nitaqat. Accordingly, both groups would be under the layoff risk, depending on their firm's perspective about the localised firm status. However, under Nitaqat2, Saudis who earned 10,000SR were considered two quotas worth, which reduced the layoff risk as a substitution action between Saudi workers. This means Saudi workers were only secure in non-localised firms under Nitaqat2 if they were paid higher than this wage. Furthermore, the indirect effect of Nitaqat would slightly increase the layoff risk for both groups if they worked in localised firms. This would happen when the firms aimed to achieve a higher Nitaqat band, which would give the firms some encouraging advantages. This would create indirect and unexpected layoffs for Saudis more than non-Saudis because employers initially prefer to employ non-Saudis. Moreover, the layoff costs would vary among the workers themselves. For Saudis, the cost of a layoff would be less as they would receive unemployment benefits. The difference between the income when a person is employed and the income from the benefit would be the cost in the short-run until they started a new job. By contrast, non-Saudi workers would be deported in most cases, especially when their contract ended. This means they would lose all of the income they expected for their lifetime. The wide deporting action, which was shown in the

[61] There are no Saudi layoff costs compared to non-Saudis; the employers pay the recruitment cost.

descriptive analysis, would increase their layoff risk expectation, and the time spent in the host country (Saudi Arabia) would be a crucial factor. In other words, before Nitaqat, foreign workers spent as many years as their total consumption required, whereas, after Nitaqat, they would take work years into consideration, which indirectly increased the cost of layoff risks for foreign workers. Therefore, the workers would be under a direct threat of job loss if they worked in non-localised firms – deportation if they were non-Saudi and substitution if they were Saudi – while they would be under an indirect effect in localised firms.

We assume that Nitaqat affects consumption through the layoff risk shock, according to firm status and worker expectations. Therefore, the utility functions are subject to a hedonic wage constraint. Following Hutchens (1983), W is a function of the risk, and the constraint is as follows.

$$W_{ij} = \theta_{0j} + \theta_{1j} R_{ij} \quad \text{4-12}$$

Wage is determined under the natural risk at θ_{0j} when the Nitaqat risk is not present, while θ_{1j} denotes the change in the wages as a result of the change in the layoff risks resulting from Nitaqat. This value could be positive or negative: $\infty > \theta_{1j} > -\infty$. When the job loss risk is absent, then it is zero. The positive value means a positive relationship between wage and risk. Thus, when risk increases, wages increase (Bloch, 1979; Hutchens, 1983). That result consists of the original hedonic function, where a risky job should have a higher wage and then higher consumption. Bloch and Hutchens explained this result as follows: workers who are exposed to high layoff risks would be justified in asking for higher wages to maximise their lifetime earnings. However, this is not always the case; sometimes there is a negative relationship between wage and layoff risk (Hübler & Hübler, 2006; Pinheiro & Visschers, 2015; Scicchitano et al., 2019). The explanation for this result is that workers who are under a layoff risk could accept a lower wage to maintain job security. In terms of Nitaqat, people might change from risky firms to safe firms with lower wages if they appreciate staying safe. Alternatively, a worker could accept a similar wage and stay in his job if the employers are willing to employ new Saudis. However, Saudi workers do not have leverage over employers to request higher wages, even if a firm is in the red zone. An employer could depend on the temporary or dummy Saudisation to solve their Nitaqat zone issue. Accordingly, there is an ambiguous relationship between this type of risk and wages that depends on how each individual appreciates this risk.

Noticeably, increasing consumption increases the utility of a household, which is a function of disposable income, including wages and, similarly, job quality (Theodossiou & Vasileiou, 2007), which is represented by the job loss risk caused by Nitaqat. A secured job would increase the utility, unlike an insecure job. Therefore, workers choose jobs that maximise their utility, using consumption level and layoff risk as a bundle. The risk is usually theoretically modelled using uncertainty literature, especially through the von Neumann–Morgenstern utility function, which is based on the expectation theory. For example, Hutchens (1983) sought to maximise the supply expected function under the layoff risk. However, the Cobb–Douglas function was used, for example, in Lin et al. (2019) study. This study was on production, unlike Hutchens' study, which was on the labour supply. The Cobb–Douglas type was used as a first attempt in the layoff context. Notice that both methods are equivalent.[62] According to this, the Cobb–Douglas utility function of workers would be written as follows.

$$U_{ij} = Cf(Y_{ij}^{d})^{\alpha} * {R_{ij}}^{\beta} \qquad \textit{4-13}$$

By plugging the Keynesian consumption function into the utility function with the substitution of the disposable income, the Lagrange multipliers were formed with respect to the hedonic wage constraint. The Lagrange multipliers are the standard method for finding the optimal value under maximised or minimised problems and written as follows.

$$L = [\, \varphi_{0j} + \varphi_{1j}\big(W_{ij} + B_{ij} - F_{ij}\big)]^{\alpha} * R_{ij}^{\beta} + \lambda\, (Wij - \theta_{0j} - \theta_{1j}\, Rij) \qquad \textit{4-14}$$

To fulfil the maximisation, we took the first-order condition and derived the Lagrange multiplier on wages and risk to find the marginal rate of substitution between them as follows.

[62] Cobb–Douglas utility is a special format of the expectation utility function of two bundles equal to the expansional value of the expected function, where
$u(c_1,c_2) = \alpha \ln c1 + B\, lnc2; \; = e^{EU} \; = \; C1^{\alpha}{}_{,}C2^{\beta}$.

$$\frac{\partial L}{\partial w} = \frac{\partial U}{\partial w} = \alpha\, \varphi_{1j}[\, \varphi_{0j} + \varphi_{1j}\big(W_{ij} + B_{ij} - F_{ij}\big)]^{\alpha-1} *\; R_{ij}^{\beta} + \lambda = 0 \qquad \textit{4-15}$$

$$\frac{\partial L}{\partial R} = \frac{\partial U}{\partial R} = \beta\, [\, \varphi_{0j} + \varphi_{1j}\big(W_{ij} + B_{ij} - F_{ij}\big)]^{\alpha} * R_{ij}^{\beta-1} - \theta_{1j}\, \lambda = 0 \qquad \textit{4-16}$$

$$\frac{\partial L}{\partial \lambda} = W_{ij} - \theta_{0j} - \theta_{1j}\, Rij = 0 \qquad \textit{4-17}$$

To find the $MRS_{W,R}$, which is equal to $\frac{\frac{\partial u}{\partial S}}{\frac{\partial u}{\partial R}}$, the equation below is needed.

$$MRS_{W,R} = \frac{\alpha\, \varphi_{1j}[\, \varphi_{0j}+\varphi_{1j}\big(W_{ij}+B_{ij}-F_{ij}\big)]^{\alpha-1} * R_{ij}^{\beta}}{\beta\, [\, \varphi_{0j}+\varphi_{1j}\big(W_{ij}+B_{ij}-F_{ij}\big)]^{\alpha} * R_{ij}^{\beta-1}} = -\frac{\lambda}{\theta_{1j}\lambda} \qquad \textit{4-18}$$

For simplicity, $\alpha\ and\ \beta$ were symmetrically assumed. After simplifying the fraction above, we obtained the following.

$$\frac{\varphi_{1j}R_{ij}}{[\varphi_{0j}+ \varphi_{1j}W+ \varphi_{1j}\big(B_{ij}-F_{ij}\big)]} = -\frac{1}{\theta_{1j}} \qquad \textit{4-19}$$

However, to find the optimal wage under Nitaqat, θ_{1j} should be found. To do so, the $MRS_{W,R}$ would be written in a different format, considering the job loss risk function on wages, as follows.

$$R_{ij} = \frac{\varphi_{0j} + \varphi_{1j}\big(B_{ij} - F_{ij}\big) + \varphi_{1j}\big(w_{ij}\big)}{-\, \varphi_{1j}\theta_{1j}} \qquad \textit{4-20}$$

This relationship revealed a similar conclusion; the wage and risk negatively responded to each other. When the wage increased, the layoff risk decreased, which is similar to the findings of Scicchitano et al. (2019), who reported that the affected layoffs were found in the lower quantile. Moreover, from a worker's point of view, when benefits increase, the risk decreases, while increasing the fees affects the risk positively. Therefore, the risk expectation increases when a worker has no income source other than his wage, which motivates him to accept a lower wage rather than a layoff. However, for workers who have other income sources, the risk is lower. The previous risk equation was plugged into the hedonic wage constraint to find the optimal wage under risk. The optimal wage decreases when the risk increases.

$$Wij = \theta_{0j} + \theta_{1j}(-\frac{1}{\theta_{1j}}\, [\frac{\varphi_{0j}}{\varphi_{1j}} + W_{ij} + B_{ij} - F_{ij}]) \qquad \textit{4-21}$$

By simplifying the fraction and rearranging it, the optimal diminished wage follows the equation below.

$$W_{ij}^{*} = \frac{(\theta_{0j}\,\varphi_{1j} - \varphi_{0j} - \varphi_{1j}(B_{ij} - F_{ij})}{2\,\varphi_{1j}} \quad \textit{4-22}$$

The following equation is required to ensure that W_{ij}^{*} is a positive value.

$$\theta_{0j} \geq \frac{\varphi_{0j}}{\varphi_{1j}} + B_{ij} - F_{ij} \quad \textit{4-23}$$

However, the optimal risk probability is obtained by replacing W_{ij}^{*} in the R equation above, as follows.

$$R_{ij}^{*} = -\frac{\theta_{0j}\,\varphi_{1j} + \varphi_{0j} + \varphi_{1j}(B_{ij} - F_{ij})}{2\,\theta_{1j}\,\varphi_{1j}} \quad \textit{4-24}$$

To obtain the hedonic wage constraint function, both θ_{0j} and θ_{1j} are required. In the absence of Nitaqat risk, the parameter θ_{0j} would equal the wage given by $\theta_{0j} = W_{ij}$, while θ_{1j} would equal zero. However, when Nitaqat was imposed, R would equal one.

$$\theta_{1j} = -\frac{1}{2}[\frac{\varphi_{0j}}{\varphi_{1j}} + \theta_{0j} + (B_{ij} - F_{ij}) \quad \textit{4-25}$$

We needed to know the θ_{0j} when R equalled one to find out θ_{1j} . To do so, we replaced the value of θ_{0j}, which made W_{ij} equal to zero, as mentioned above in the R_{ij}^{*} function. The following relationship was obtained.

$$\theta_{0j} = -\theta_{1j} \quad \textit{4-26}$$

By plugging the θ_{0j} value when W_{ij}^{*} equals zero at θ_{1j} , it results in the following.

$$\theta_{1j} = -[\frac{\varphi_{0j}}{\varphi_{1j}} + (B_{ij} - F_{ij}) \quad \textit{4-27}$$

The hedonic wage parameters are constructed by the consumption parameters and other determined disposable income. The θ_{1j} could be negative or positive, depending on the benefit and fees subtraction. When this value was negative the wage was expected to be reduced under the risk, while this value would be positive if it were bigger than the first term, $\frac{\varphi_{0j}}{\varphi_{1j}}$. By substituting these parameters on the hedonic wage, we obtained the following.[63]

$$W_{ij} = [\frac{\varphi_{0j}}{\varphi_{1j}} + B_{ij} - F_{ij}] - [\frac{\varphi_{0j}}{\varphi_{1j}} + B_{ij} - F_{ij}]\ R_{ij} \qquad 4\text{-}28$$

The last equation supports the team who found negative relationships between job loss risk and wages (Hübler & Hübler, 2006; Pinheiro & Visschers, 2015; Scicchitano et al., 2019). This equation could provide the responses of workers under Nitaqat shock. If a worker was under a layoff risk according to his firm's localised status, where R = 1, this worker would be laid off and generate zero income. From another angle, workers in safe firms, when R = 0, would receive a wage equal to their (ω_{0j}), in addition to any benefits and excluding any fees. This relationship assumed that an individual's optimal wage would vary according to their consumption parameters; we assumed a convergent value of (ω_{0j}) for each national group. However, fees and benefits determined the differences in groups. Furthermore, if workers were feeling an indirect risk under Nitaqat, R could be a proportion between zero and one. This means that when a worker was under a low indirect risk, the wage would decrease by a small amount compared to high expectation of risk. This means imposing Nitaqat created indirect layoff risks alongside the direct risks, which could lower the wages in general.

From one side, the hedonic wage equation allowed us to capture the wage gap between Saudis and non-Saudis. First, we presented the gap between Saudis and non-Saudis when the risk was equal to zero. For Saudis, it followed the equation below.

$$W_{iS}^{*} = \frac{\varphi_{0S}}{\varphi_{1S}} + B_{iS} \qquad 4\text{-}29$$

For non-Saudis, it followed the equation below.

$$W_{iF}^{*} = \frac{\varphi_{0F}}{\varphi_{1F}} - F_{iF} \qquad 4\text{-}30$$

[63] The function assumes that the wage equals zero when the risk is equal to one and equal to w when the risk is zero. The uncertainty, whether high or low, would be assumed as a percentage, for example, 0.5 to show that the wage would be reduced if the indirect risk was imposed.

F indicated several non-Saudi origins, such as Arabic, African, Asian and European. According to the two equations above, we could state that the wage gap between Saudis and non-Saudis result from two reasons: first, the heterogeneity of the benefits policy, which could be generalised on all support programmes designed to help Saudis, and second, (ω_{0j}), which was determined by the consumption parameters. Indeed, this value varied according to each group that had a distinguished preference. This consumption variation could reflect the sources of the multi-supply issue found as a result of international recruitments. From this point, the gap existed between non-Saudis themselves, and the variations between Saudis and non-Saudis were assumed to be different concerning each group.

From another side, the hedonic wage equation could reflect the gap among firms' localised statuses for a single group of workers. For example, Saudis could receive less than their peers by around $-\theta_{1j} R$ if working in localised firms where the indirect risk could be assumed. This would be the reverse of the hedonic literature assumptions. Indeed, the benefits system could lessen the impact of the risk and increase the wages, which could raise the wages compared to their peers if the benefit targeted laid-off workers, such as unemployment benefits.[64] Similarly, in non-localised firms where the risk was high for non-Saudis, they expected to earn less than their peers in localised firms by $-\theta_{1j} R$. This reflects the reverse of the hedonic wage expectations, where the higher the risk, the lower the wage. However, under an intensive fee system, workers under a risk assumed they would receive less than they could previously as the F value increased. However, this expectation of wage reduction could significantly affect non-Saudis responses, assuming invariant consumption and prices through the substitution and income effects. Therefore, under the substituted effect, they would prefer the obtained wage compared to spending an extra hour in work, unlike the income effect, where workers could increase their work to maintain at least (ω_{0j}). The first effect would be found when the fees were small and the second when the fee was aggressive. Indeed, when the effect of income was applied, the wage could increase for those under

[64] Unemployment benefits have two programmes, Hafiz and SANED, which were explained earlier in the descriptive chapter. For more program detail (see section 3.3.1 above)

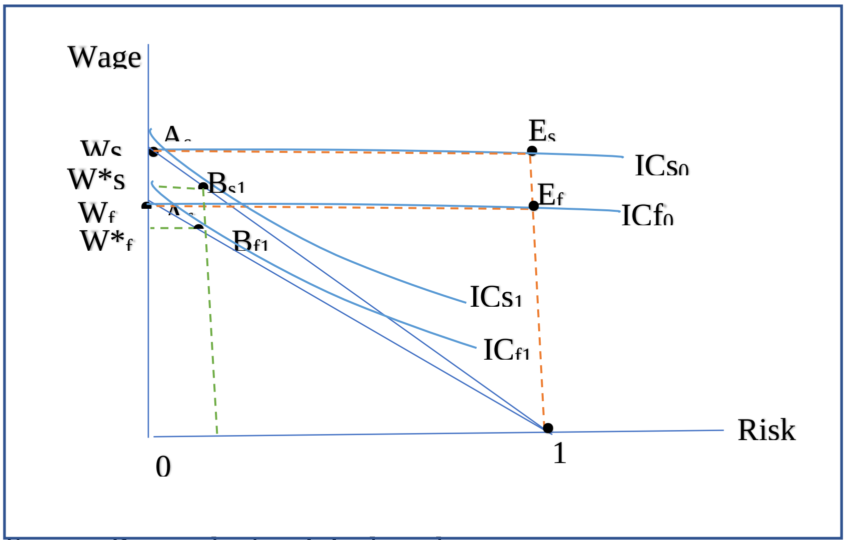
risk compared to their peers in safe firms which allowed the hedonic relationship to apply to non-Saudis.

To visualise that graphically, let us assume two workers, a Saudi and a foreigner, worked in a firm before Nitaqat was imposed, which means the risk equalled zero, and wage was equal to Ws and W_f at points As (0, Ws) and A_f (0, W_f). Due to the absence of the policy, they would enjoy the highest wage they could, according to the labour market equilibrium, where the risk of job loss was natural. The Cobb–Douglas utility function would have zero shares of the risk, and α would equal one for consumptions. In other words, the utility function is equal to the consumption. Thus, the indifference curve would be horizontal IC_{0s} and IC_{0f}.[65] Therefore, a corner choice in the hedonic wage constraint would represent this situation (see point A in the diagram below). Once Nitaqat was imposed, this employee could find himself in one of the two firm statuses: localised or non-localised. Thus, the direct risk would be binary, equal to one or zero, according to firm status and the worker's nationality, as discussed earlier. The worker could be classified into two risk circumstances. The first is no direct risk. For example, if the worker was in a safe firm, the direct layoff risk would be zero; accordingly, he would stay at the initial points. However, if he expected an indirect risk of losing a job, he would search for a new job. In this case, the worker would ensure his safety from layoffs at this new firm as he would be chosen with respect to the policy. Admittedly, the new job would be subject to the hedonic wage constraint on top of a lower utility function. As a result of the risk share on the utility function, the indifference curve would be IC_{1s} and IC_{1f}. In other words, the worker would accept a slightly lower utility function under uncertainty when the indirect risk was considered. This would result in wage being equal to W*s and W^*_f at points B_{s1} and B_{s2}, the second point being when the worker would be under direct risk. If a worker lost his job, he would be at point C, as mentioned above, whether he was Saudi or foreign. However, he could be under defined risk and still working. In this case, the worker had two options. First, he could ignore the risk because of his expectation that the employer would be keen to keep him and find a way to get rid of the quota restriction, then he would have a similar wage.

[65] Basically, this a utility function on consumption only.

Although this worker would have a high probability of job loss risk, he would have zero shares of the risk on his utility function, which would maximise his utility function under the risk restriction at points Es and E_f.[66] The indifference curve would be identical to the one before the quota was imposed at IC_{0s} and IC_{0f}, meaning that under any probability of risk, the wage before the quota would maximise the utility function. The second option would be to find a new job if he was quite uncertain regarding the employer's response to the policy. Under a high uncertainty of layoff risk, this worker would accept a new job subject to the hedonic wage w* and his lower utility, with indifference curves IC_{1s}, and IC_{1f}, as well. This means the expectation of layoff risk, whether direct or indirect, would decrease the wages, while a lower expectation of the layoff risk would keep the wages stable. This graph reflects how benefits and fees affect the acceptance; Saudis were more likely to accept a higher wage reduction considering the benefits compared to non-Saudis. This means non-Saudi wages would decrease but at a lower rate. The distance Ws to W*s formed the reduction in Saudi wages, while the line from W^*_f to W^*_f formed the reduction in non-Saudi wages. Accordingly, the gap would be reduced because of the higher reduction rate on Saudi wages.

Figure 4-2: Workers' wage choices under risk



Source: Researcher's original work .

[66] This would occur through dummy Saudisation or temporary employment, which usually happened when there was a relationship between the workers and the employees or illegal ownership.

Both the mathematical and graphical models illustrated that consumption was considered the source of the wage differential between the two groups, which created a multi-supply in the labour market, causing the initial wage gap. Moreover, the wage would be negatively affected by the quota policy through the layoff risk. The quota would affect the welfare of both groups, and the target group would be affected deeply.

4.4 The econometric model

After the descriptive analysis earlier in Chapter 2, econometric models were applied to give some inference results regarding the wage structure and the wage gap at two times. Empirically, there were two main aims to be achieved: the earnings function and the Oaxaca decomposition, which was derived from the earnings function. The econometric model and its specifications, the methods and the methodology are explained under this title.

4.4.1 Method

The OLS estimation method could estimate the parameters for both earnings functions, and the Oaxaca decompositions were capable of answering the research question. This method is one of the most common methods used to estimate the parameters of multiple variable linear models. Admittedly, it is the best estimation among other linear methods when its error term is independent, identically distributed and has zero expected value. This statement is known as the Gauss–Markov theorem. In other words, the OLS is considered the best linear unbiased estimator (BLUE) (Brown, 2019). Reliable and powerful results can be obtained if its assumptions are satisfied.

Several assumptions needed to be met, such as the **linearity** for the variable and the error term. Using polynomial variables helped specify a linear model for a curved relationship. Thus, good model specification follows the linear assumption. Another assumption was that a **zero mean for the error term** otherwise indicated an underrepresented model for observed values. Moreover, **homogeneity** was a vital assumption where the error term could not be predicted through the independent variables. In other words, the error term and the explanatory variable were independent. Furthermore, these error terms should be independent of each other, which means the error terms needed to be unpredictable through other error term observations. Thus, the

OLS assumed **no autocorrelation**; this issue is usually found in time-series datasets, unlike heteroscedasticity, which is usually found in cross-sectional datasets, especially earnings. However, **homoscedasticity** was assumed for the BLUE OLS estimation. This means the error terms' variance was constant across all fitted values. The heteroscedasticity occurred when the error terms were distributed differently across the fitted values. This issue could be treated by increasing the sample size or using a robust standard error. Moreover, **no multicollinearity** was assumed for a BLUE estimator. This means the explanatory variables needed to be independent of each other when a variable was partially or perfectly correlated with another variable; this would give biased coefficients for the correlated variables, while the other coefficients' variance inflation factor (VIF) was less than five, and some used ten as a threshold (Brooks, 2019).[67] Moreover, this issue was neglectable when the variable of the polynomial or an interaction variable was used. Finally, regarding **normality,** when the error term did not have a normal distribution, it did not harm the reliability of the OLS estimation. However, these features helped us obtain reliable confidence intervals and statistically test a hypothesis. The model residual usually acted as the error term; thus, when these assumptions were applicable to the residual, it could be generalised on the error term.[68]

The OLS parameters were found when the residual squared summation (RSS)[69] was minimised, where $\sum_1^n(\epsilon_i)^2$ = $\sum_1^n(w_i - w)^2$ = $\sum_1^n(w_i - \alpha - \beta X_i)^2$. Therefore, its small possible value depended on the choice parameters α and β. The estimated parameter $\hat{\beta}$ could be calculated as the product of the difference between the observed value and its meaning for both the dependent and independent variables, divided by the square value of the difference between the independent observed value and its mean, as follows.[70]

[67] VIF = $\frac{1}{1-R_i^2}$, where R_i^2 is the value of R^2 obtained from an auxiliary regression for each variable regressed on the constant and all explanatory variables of the main model.
[68] A model residual resulted from the difference between observed and fitted values.
[69] This is known as the unexplained sum square (Dougherty, 2016).
[70] There are number of $\hat{\beta}$, depending on the number of the independent variables. However, it is usually expressed as sum of these values.

$$\hat{\beta} = \frac{\sum_1^n (x_i - \overline{x})\,(w_i - \overline{w})}{(x_i - \overline{x})^2} \qquad \textit{4-31}$$

The α equalled the mean of the dependent value subtracted from the mean of the independent value, multiplied by $\hat{\beta}$, which could be written as follows.

$$\hat{\alpha} = \overline{w} - \hat{\beta}\overline{x} \qquad \textit{4-32}$$

After finding the coefficients, the model could be interpreted to be understood; it described the relationship between the wages and other independent variables (Dougherty, 2016). The fitness of the interest model could be measured by R^2, which could be obtained by subtracting the ratio of explained sum of squares (ESS) and total sum square (TSS), as in the following equation.

$$R^2 = 1 - \frac{RSS}{TSS} \qquad \textit{4-33}$$

The TSS was equal to the sum of the explained and unexplained sum squared (ESS + RSS), where the TSS could be calculated as $\sum_1^n (w_i - \overline{w})$, and the ESS was known as $\sum_1^n (\widehat{w}_i - \overline{w})$. Therefore, when an OLS parameter minimised the RSS, they maximised the R^2. The larger the R^2, the better, while the smallest R^2 implied a high variance value. The high value indicated the strength of the explanatory variables. Thus, an increase in the R^2 resulted from adding a new independent variable that could support this addition. These changes in the R^2 were important for measuring the joint F-statistic between two models as a result of including or excluding a variable[71] (Wooldridge, 2008). Moreover, the R^2 was reported with the adjusted-R^2. The latter value was obtained from the former value with respect to the number of observations, n, and the number of variables, k, as follows.

$$\overline{R^2} = \frac{1 - (1 - R^2)\,(n - 1)}{n - k - 1} \qquad \textit{4-34}$$

This value was useful for comparing non-nested models, such as different specifications or forms (Wooldridge, 2008). Moreover, the t-statistics and the standard errors, p-values and confidence intervals were important for measuring the coefficient goodness. A

[71] The F-statistic procedure was demonstrated in Chapter 2 to examine the nested models.

significant coefficient had a p-value less than 5%, or 0.05, which denoted a 95% confidence interval for the t-statistic. This tested the hypotheses of the possibility of a coefficient equal to zero. Thus, we evaluated the measurements in the total model and each variable.

4.4.2 Methodology

4.4.2.1 Oaxaca-Blinder decomposition

The decomposition methodology was chosen to fulfil the research aim and answer the research question. This methodology was used to disentangle sources of the pay gap between two groups and determine the inequality sources (Fortin et al., 2010). The decomposition has been widely used since Blinder (1973); (Oaxaca, 1973) introduced it into the labour market research. Since then, it has been considered a standard tool for examining economic issues (for gaps between two groups).

The main criticism of the OB is that it arbitrarily links the between-group wage differential to explain discrimination. Neumark (1988), however, provided a link between the theoretical background of employer discrimination and a derived Oaxaca decomposition. The critiques did not hinder Oaxaca from developing a wage structure decomposition for union and non-union workers. The procedure considered the competitive wage without unionisation, which yielded three terms to explain the wage differential (Oaxaca & Ransom, 1988; 1994). Additionally, critics have also argued that the traditional OB deals only with linear models, but researchers have advocated for covering other types of models, particularly those with a limited or discrete dependent variable. (Even & Macpherson, 1990) developed a decomposed probit model, and the reduced form was developed by (Gomulka & Stern, 1990). Similarly, a logit and probit decomposition model was developed (Fairlie, 1999; 2005). Fairlie reported that the estimation was not sensitive to the issue of whether the logit or probit model was used. Yun (2004) published a more comprehensive study for the linear and non-linear studies to decompose first moment differences. Another development was the generalisation of the decomposition methods in discrete and limited dependent variable cases (Bauer & Sinning, 2008). The effort to develop the Oaxaca methodology continued with the Bayesian models based on the Markov chain Monte Carlo method, which estimated a posterior distribution for characteristics and coefficients; this was considered the easy method in the Bayesian estimation (Radchenko & Yun, 2003). Also, Brown et al. (1980) developed an interesting work on the wage distribution concerning the occupational differential as a separate part of the equation. Brown et al. 's approach was applied in other papers and expanded the differential between and within different occupational types (Sung et al., 2001; Liu et al., 2004; Zhang & Wu, 2012). As is well-known, the OB was designed to compare binary groups. Recently, Ulrick (2012) generalised this decomposition using a continuous variable rather than a discrete variable.

Traditionally, the OB considered a Mincer equation and applied the OLS method on two separate earning estimations. This principle of the traditional tool is used in this chapter to estimate the wage gap equation for Saudis and non-Saudis.[72] Then, the equations are decomposed to understand the differential sources from several perspectives.

As the first step, let us assume two groups: Saudis (s) and foreigners (f). An outcome variable, W, points to log-wage, in addition to a set of explanatory variables denoting human capital characteristics and other worker characteristics, such as firm age, location, activity, occupation and job loss risk, given by x_{i}. The raw gap between those two groups in the mean log-wage could be counted from the expected value of the outcome, as follows.

$$W_{gap} = \bar{W}_{is} - \bar{W}_{if} \qquad \textit{4-35}$$

$$W_{gap} = E(W_s) - E(W_f) \qquad \textit{4-36}$$

Thus, the wage gap was constructed by subtracting the expected values for each group. This value could be obtained by the prediction of the group difference based on a linear model of wage predictions, following the Mincer equation as a standard, which was explained earlier under the model specifications. Separating the equation according to the groups yielded two different equations for each group, which allowed us to decompose the differential in the mean wages as follows.

[72] This function, connecting salaries and job characteristics, contained a risk variable.

$$\bar{W}_{is} = \alpha_s + \sum \beta_s \bar{x}_{is} + \epsilon_{is}, \quad if \qquad i \in s \qquad \text{4-37}$$

$$\bar{W}_{if} = \alpha_f + \sum \beta_f \bar{x}_{if} + \epsilon_{if}, if \qquad i \in f \qquad \text{4-38}$$

The two groups of interest were indicated by the subscripts s and f for our cross-section analysis. Thus, $\bar{W}_{is}$ and $\bar{W}_{if}$ denoted the log-wage mean for each individual in the specific group the worker belonged to. Both α_s and α_f indicated the intercept for the Saudi and foreigner equations, respectively. Similarly, $\sum\beta_s$ and $\sum\beta_f$ indicated the coefficients for each explanatory variable in the two groups, respectively. $\bar{x}_{is}$ and $\bar{x}_{if}$ represented the explanatory variables for individual characteristics, job characteristics and policies, especially the job loss risk in the two groups' equations. Finally, ϵ_{is} and ϵ_{if} referred to the error term in each equation. Therefore, the wage gap between s and f at a point in time from the programme's launch in 2013[73] or 2017 could be expressed by subtracting these two equations, as follows.

$$\bar{W}_s - \bar{W}_f = [(\alpha_s + \sum \beta_s \bar{x}_{is} + \epsilon_{is})] - [(\alpha_f + \sum \beta_f \bar{x}_{if} + \epsilon_{if})] \qquad \text{4-39}$$

The main method used in the OB was estimating the mean value using the OLS. Thus, for simplicity, $\epsilon_f - \epsilon_s$ was removed. One of the assumptions underlying the OLS method is a zero-conditional mean for the disturbances. This assumption should be satisfied when the expected value of ϵ equals zero: $E(\epsilon|x_i, \ldots., x_n) = 0$, where i = 1, 2, …. *n*. This assumption means that the error term was not correlated with any of the independent variables given by *x* (Wooldridge, 2015). Therefore, rewriting and rearranging the equation above and applying the decomposition methodology yielded the following equation.

$$\bar{W}_s - \bar{W}_f = (\alpha_s - \alpha_f) + (\sum \beta_s \bar{x}_s - \sum \beta_f \bar{x}_f) \qquad \text{4-40}$$

When adding and deducting $\sum \beta_s \bar{x}_f$ with a rearrangement for the right hand side (RHS) in the equation, and when renaming the left hand side LHS as $\text{salaries}_{\text{gap}}$, it could be rewritten as follows.

[73] The first year of Nitaqat2 was 2013. However, it was announced in early of 2012, and firms could know their colour bands before applying for Nitaqat2.

$$W_{gap} = (\alpha_s - \alpha_f) + (\sum \beta_s \bar{x}_s - \sum \beta_f \bar{x}_f + \sum \beta_s \bar{x}_f - \sum \beta_s \bar{x}_f) \quad \textit{4-41}$$

The rearrangement of the decomposed equation could be written as follows.

$$W_{gap} = (\alpha_s - \alpha_f) + \sum \beta_s (\bar{x}_s - \bar{x}_f) + \sum (\beta_s - \beta_f) \bar{x}_f \quad \textit{4-42}$$

For research purposes, we isolated the risk of job loss from other characteristics. Risk is usually characterised as an unsafe environment in the hedonic wage function. We considered the risk as a possibility for a job loss risk, depending on firm status and worker nationality. Therefore, the associated parameters with this variable reflected the Nitaqat contribution to the explained wage gap between Saudis and non-Saudis. Moreover, the unexplained value associated with the unexplained part parameter reflected the discrimination of this policy. The decomposition equation followed that arrangement.

$$W_{gap} = (\alpha_s - \alpha_f) + \underbrace{\sum \beta_s (\bar{x}_s - \bar{x}_f) + \sum (\beta_s - \beta_f) \bar{x}_f}_{} + \underbrace{\rho_s (\bar{R}_s - \bar{R}_f) + (\rho_s - \rho_f) \bar{R}_f}_{} \quad \textit{4-43}$$

The term ρ_s clarified the hedonic wage function theory. This could imply that risk could be an additional explanatory variable, which was neglected in another research. The risk hypothesis would be expressed in the explained part as follows.

$H_0: \rho_s = 0$: The risk of job loss could not explain the wage gap, or $\rho_s(\bar{\bar{R}}_s - \bar{\bar{R}}_f)$ was insignificant.

$H_1: \rho_s > 0$: The wage gap could decline because of the job loss risk differential encouraged by Nitaqat. The wage differential was narrowed between the two groups because the foreign wages increased or Saudi wages decreased, or both groups decreased/increased by different percentages (such that the wage gap narrowed overall).[74]

$H_1: \rho_s < 0$: The wage gap widened due to the job loss risk differential, which implies that foreign wages decreased, Saudi wages increased, or both groups followed a similar

[74] Assume $(\bar{\bar{R}}_s - \bar{\bar{R}}_f)$ was positive.

trend in different percentages. The risk differential, however, could be unexplained due to discrimination or nepotism. The equation could be rewritten in OB notation, as follows.

$$\text{salaries}_{\text{gap}} =$$

$$\text{Explained part: } \sum \beta_s(\bar{x}_s - \bar{x}_f) + \rho_s(\bar{R}_s - \bar{R}_f) +$$

$$\text{Unexplained part: } (\alpha_s - \alpha_f) + \sum(\beta_s - \beta_f)\,\bar{x}_f + (\rho_s - \rho_f)\bar{R}_f \qquad 4\text{-}44$$

Using the possibility of job dismissal as a proxy for job risk was equivalent to the case of hedonic wage theory; when the risk of dismissal increased, the wages increased, and vice versa. However, it has commonly been assumed that the total decomposition in risk level would not reverse the direction of this relationship; this means that the wage gap would increase with increasing dismissal risk. Recently, there has been some empirical evidence that this relationship could be reversed (Hübler & Hübler, 2006; Pinheiro & Visschers, 2015; Scicchitano et al., 2019). The change results from deducting the final effect of the risk level from the two groups. Thus, if the risk term were positive, it would assume that a hedonic wage exists, while a negative term would assume a reverse hedonic wage, withholding risky people as reference groups.[75] The equation chart below specifies how much each factor accounted for in both the explained and non-explained salary gaps. For some concentrated results, the equation could be written as follows.

$$wage_{gap} = \begin{cases} \textbf{\textit{Explained Gap}} \\ \textit{due to individual characteristics } \sum \beta_s(\bar{h}_s - \bar{h}_f) \\ \textit{due to job characteristics } \sum_2^n \gamma_s\,(\bar{z}_s - \bar{z}_f) \\ \textit{due to nitaqat risk } \rho_s(\bar{\bar{R}}_s - \bar{\bar{R}}_f) \\ \textbf{\textit{Unexplained Gap}} \\ \textit{due to individual characteristics} \sum(\beta_s - \beta_f)\,\bar{h}_f \\ \textit{due to job characteristics } \sum_2^n(\gamma_s - \gamma_f)\,\bar{z}_f \\ \textit{due to nitaqat risk } (\rho_{sb} - \rho_{fb})\bar{\bar{R}}_f \\ \textit{due to intercept } (\alpha_s - \alpha_f) \end{cases} \qquad 4\text{-}45$$

[75] On the contrary, if a non-risky environment is considered the reference group, a positive relationship denotes the reverse case, and the hedonic case would be negative.

This feature could be helpful for understanding the gap sources and explaining whether this gap existed because the workers differed or because the job characteristics varied. The Nitaqat risk could explain part of this gap, as well, concerning a pooled sample for both firms' localised statuses. Similar analyses conducted a separate estimation for the pay gap between Saudis and non-Saudis concerning the Saudisation status of firms. The main reason behind applying such an approach was to examine the wage gap between those two groups, whether the firms satisfied the Nitaqat criteria or not. The programme could be evaluated this way. Knowing the gap was widening or narrowing because of Nitaqat implied a change in the condition of the labour market. This was consistent with the economics of discrimination, which suggests that reducing the wage gap improves the competition state (Liu, 2004). Following the concept of the hiring quota, one could generalise whether this policy helped to enhance competition by using "reduce the gap" as an index. In this case, the hypothesis was that reducing the wage gap in all labour market units would improve competition between the two groups and vice versa. If the gap was narrowed, this could imply that Nitaqat helped increase the competition between the two groups through changing the job loss risk. Therefore, when the reduction was found, the question was about the source of this reduction. Was it increased foreign worker wages (spill-over effect) or decreased Saudi wages (welfare harmed)? These applications of the OB shed light on the Saudi labour market. To summarise, the approach first estimated wage functions. Second, it addressed sources of the gap between Saudi and non-Saudi workers' wages with respect to the hiring quota effect (risk). The estimation results were exploited to understand the wage gap and determine whether competition was fulfilled using a hiring quota.

4.4.2.2 Identification issue

This issue was recognised earlier in the literature when the contribution of a particular dummy variable constructed from an initial (multi-) categorical variable was variant when the reference group for the initial categorical variable changed. This issue was inevitable in the detailed decomposition (Oaxaca & Ransom, 1999). Accordingly, there were several solutions displayed (Nielsen, 2000; Gardeazabal & Ugidos, 2004; Yun, 2005; Gomes et al., 2020). All of these valuable studies contributed to solving the identification issue, using several methods. The drawback of these contributions is that all these methods require starting from scratch and ignoring the main Oaxaca decomposition. However, we propose a promising solution to this issue that could be added to the literature. This solution would require completion beyond the PhD thesis in terms of the slandered error. This approach could work as a post-estimate for the unexplained part if a researcher was not concerned about the z-score. Additionally, our approach benefitted from generating the fixed effects for each variable, corresponding to solving the identification issue. It respected two angles: looking at the average distribution for the summation of the unexplained part, excluding the coefficients of the continuous variables among dummy variable categories and normalising the coefficients as follows.

$$\tau = \beta_{iX1} - \frac{(\sum \beta_i)\, X1}{N(\beta_{X1})} + \frac{intercept + \sum \beta_i Xi}{N\beta \sum Xi + 1} \qquad 4\text{-}46$$

τ was the Bagazi coefficient, and β_{iX1} was the Oaxaca coefficient for a variable category, X1. N was the number of categories in the variable X1. Xi denoted all categorical variables in the regression. This could be expressed in words as follows.

$_{\text{Bagazi}}$ Coefficient = $_{\text{Oaxaca}}$ coefficient -

$$\sum \frac{all\ coefficients\ of\ a\ specific\ categorical\ variable}{number\ of\ Coefficients\ of\ a\ specific\ categorical\ variable} + \sum \frac{unexplained\ part\ excluded\ continuous\ variables}{number\ of\ al\ thoes\ coefficients} \qquad 4\text{-}47$$

However, the full work began from the basic stages of the decomposition, which enabled us to produce the statistical values required. We chose 2013 to apply this coefficient (see Table 4-1). Similarly, the calculation used in Table 4-2 provided the full results of the Bagazi coefficients.

Table 4-1: Application example for a selected variable to find the Bagazi coefficients.

	Oaxaca coefficients	Bagazi coefficients Calculations	Values
_ISIZE_1	0.019701	= 0.019701−0.0055091 +(−0.02108357)	−0.00689167
_ISIZE_2	0.0157199	= 0.0157199 −0.0055091 +(−0.02108357)	−0.01087277
_ISIZE_3	0	= 0 −0.0055091 + (0.02108357)	−0.02659267
_ISIZE_4	−0.0133845	=−0.0133845−0.0055091 +(−0.02108357)	−0.03997717
Average size = (0.019701 + 0.0157199 + −0.0133845) / 4 = 0.0220364 / 4 =	0.0055091	Sum of normalised size coefficients given by the two highlighted above = 0	
Average omitted category = all Oaxaca coefficients in Table 1-2, Column 2, unless age, age2 and age3 / 62 = −1.30718133 / 62	−0.02108357	The average of each variable equalled the average of the omitted variable values. However, the summation of the Bagazi coefficients provided a fixed value for any base categories, and the value constructed from the normalised Oaxaca coefficients + the average coefficient of these categories was omitted as a base.	

Table 4-2: The researcher's coefficients compared to coefficients of the OB

2013			**2017**		
	Bagazi coefficients	Oaxaca coefficient		Bagazi coefficients	Oaxaca coefficient
Total unexplained part	−1.196468	−1.196468	**Total unexplained part**	−0.4588490	−0.458849
age	3.810503	3.810503	**age**	2.285915	2.285915
age2	−6.902262	−6.902262	**age2**	−2.831601	−2.831601
age3	3.202472	3.202472	**age3**	1.060243	1.060243
_IQualifica_1	−0.030135	0.0016659	**Education_1**	−0.06602	0.0022221
_IQualifica_2	−0.0029269	0.0288738	**Education_2**	−0.04485	0.0233931
_IQualifica_3	−0.0290409	0.0027598	**Education_3**	−0.0572	0.0110413
_IQualifica_4	−0.0292939	0.0025068	**Education_4**	0.0073	0.0755492
_IQualifica_5	−0.031955	−0.000154	**Education_5**	0.362467	0.4307137
_IQualifica_6	−0.031806	−5.73E−06	**Education_6**	−0.03776	0.0304911
_IQualifica_7	−0.0259596	0.0058411	**Education_7**	−0.06809	0.0001522
_IQualifica_8	−0.0234705	0.0083302	**Education_8**	0.02137	0.0896195
_IQualifica_9	−0.0306909	0.0011098	**Education_9**	−0.06227	5.98E−03
_IQualifica_10	−0.0315094	0.0002913	**Education_10**	−0.06814	0.0001091
_IQualifica_11	0.071545026	0.1033457	**Education_11**	−0.068223	0.0000165
_IQualifica_12	−0.02432937	0.0074713	**Education_12**	−0.068246	
_IQualifica_13	−0.03177367	0.000027	**Qualifica_1**	−0.00896	0.0097503
_IQualifica_14	−0.02242197	0.0093787	**Qualifica_2**	−0.01245	0.0062308

_IQualifica_15	−0.03176797	0.0000327	Qualifica_3	−0.00542	0.0132593
_IQualifica_16	−0.03180067		Qualifica_4	−0.013796	0.0048824
Female	−0.01922077	0.0037256	Qualifica_5	−0.018668	9.59E−06
Male	−0.02294637		Qualifica_6	−0.01874	−0.0000635
_Iregions_1	−0.00052496	−0.0019848	Qualifica_7	−0.01888	−0.000205
_Iregions_2	−0.04253876	−0.0439986	Qualifica_8	−0.0146	0.0040719
_Iregions_3	−0.11529206	−0.1167519	Qualifica_9	−0.018495	0.0001824
_Iregions_4	0.001459844		Qualifica_10	−0.01516	0.0035131
_Iregions_5	0.000850044	−0.0006098	Qualifica_11	−0.01749	0.0011869
_Iregions_6	0.001171544	−0.0002883	Qualifica_12	0.0464096	0.0650871
_Iregions_7	0.007289344	0.0058295	Qualifica_13	−0.017628	0.0010498
_Icolour_1	−0.01711674	−0.0002048	Qualifica_14	−0.014777	0.0039004
_Icolour_2	−0.01784874	−0.0009368	Qualifica_15	−0.01835	0.0003245
_Icolour_3	−0.01764214	−0.0007302	Qualifica_16	−0.013495	0.0051822
_Icolour_4	−0.01466924	0.0022427	Qualifica_17	−0.01864	0.0000372
_Icolour_5	−0.01691194		Qualifica_18	−0.01918	−0.0005036
_Icolour_6	−0.0221242	−0.0052123	Qualifica_19	−0.018677	
_Icolour_7	−0.03520714	−0.0182952	female	0.029725	0.0843953
_Icolour_8	−0.03132004	−0.0144081		−0.05467	
_Icolour_9	−0.01691194		_Icolour_1	−0.019856	−0.0050074
_ISIZE_1	−0.00689167	0.019701	_Icolour_2	−0.01475	0.0000989
_ISIZE_2	−0.01087277	0.0157199	_Icolour_3	−0.032259	−0.0174077
_ISIZE_3	−0.02659267		_Icolour_4	−0.014878	−0.0000263
_ISIZE_4	−0.03997717	−0.0133845	_Icolour_5	−0.01485	
_Ioccupatio_1	0.034017763	0.0682252	_Icolour_6	−0.014402	0.0004494
_Ioccupatio_2	−0.033958	0.0002491	_Icolour_7	−0.012386	0.0024661
_Ioccupatio_3	−0.0285768	0.0056307	_Icolour_8	0.02607	0.0409241
_Ioccupatio_4	0.013015663	0.0472231	_Icolour_9	−0.014936	−0.0000843
_Ioccupatio_5	−0.02310124	0.0111062	_ISIZE1_1	0.0103499	0.0261519
_Ioccupatio_6	−0.04735364	−0.0131462	_ISIZE1_2	−0.01326025	0.0025418
_Ioccupatio_7	−0.03426014	−0.0000527	_ISIZE1_3	−0.01566115	0.0001409
_Ioccupatio_8	−0.03532804	−0.0011206	_ISIZE1_4	−0.01483575	0.0009663
_Ioccupatio_9	−0.03420744		_ISIZE1_5	−0.01580205	
_IActivitie_1	−0.01901218	−0.0026474	_ISIZE1_6	−0.02562515	−0.0098231
_IActivitie_2	−0.04209648	−0.0257317	_Iregions2_1	−0.00859075	−0.0023172
_IActivitie_3	−0.01636478		_Iregions2_2	−0.00792235	−0.0016488
_IActivitie_4	−0.00690478	0.00946	_Iregions2_3	−0.02851455	−0.022241
_IActivitie_5	−0.01653828	−0.0001735	_Iregions2_4	−0.03579275	−0.0295192
_IActivitie_6	−0.01116288	0.0052019	_Iregions2_5	−0.01159715	−0.0053236
_IActivitie_7	−0.01793618	−0.0015714	_Iregions2_6	−0.00855545	−0.0022819
_IActivitie_8	−0.02156448	−0.0051997	_Iregions2_7	−0.00700275	−0.0007292
_IActivitie_9	−0.01273078	0.003634	_Iregions2_8	−0.00678135	−0.0005078
_IActivitie_10	−0.01704228	−0.0006775	_Iregions2_9	−0.00821865	−0.0019451
_IActivitie_11	−0.05056618	−0.0342014	_Iregions2_10	−0.00794725	−0.0016737
Non-Saudi non-localised	−0.02052637	0.0011144	_Iregions2_11	−0.00627355	
others	−0.02164077		firm-age	0.0482546	0.0482546

Saudi non-localised	-0.01713452	0.0078981	firm-age2	-0.011395	-0.011395
Others	-0.02503262		_IActivitie_1	-0.01289275	-0.0000233
_cons		-1.384289	_IActivitie_2	0.002262049	0.0151315
			_IActivitie_3	-0.012869	
			_IActivitie_4	-0.01485915	-0.0019897
			_IActivitie_5	-0.01461055	-0.0017411
			_IActivitie_6	-0.01294145	-0.000072
			_IActivitie_7	-0.01331665	-0.0004472
			_IActivitie_8	-0.01700335	-0.0041339
			_IActivitie_9	-0.00547815	0.0073913
			_IActivitie_10	-0.01309985	-0.0002304
			_IActivitie_11	-0.02238725	-0.0095178
			_Ioccupatio_1	0.002383974	0.0382032
			_Ioccupatio_2	-0.02301883	0.0128004
			_Ioccupatio_3	-0.02289533	0.0129239
			_Ioccupatio_4	0.078130574	0.1139498
			_Ioccupatio_5	-0.019025	0.01679
			_Ioccupatio_6	-0.02851153	0.0073077
			_Ioccupatio_7	-0.03547853	0.0003407
			_Ioccupatio_8	-0.02801773	0.0078015
			_Ioccupatio_9	-0.03581923	
			Non-Saudi non-localised	0.025773335	0.0764915
				-0.05071816	
			_cons	0	-2.146027

4.5 Model specifications

4.5.1 The definition of the variables

Our model was inspired by the Mincer equation to estimate the earnings function; the first empirical aim followed the equation below.

$$\bar{w}_{ij} = \alpha_j + \sum \beta_j \bar{x}_{ij} + \epsilon_{ij}, \quad \mathrm{E}\,(\epsilon_{ij}, |X_i) = 0, \text{ where j} \in (\text{s, f}) \qquad 4\text{-}48$$

$\bar{w}_{ij}$, the natural logarithm of wage was the dependent variable in the standard earnings function by Mincer. We are aligned with the extensive literature in using the logarithmic form. We used monthly wage data consistent with Mahdi (2005); and Smith & Fernandez (2017). Unlike, Simón et al. (2008) interested to calculate the hourly wage from annual data. The α_j was the intercept, while ϵ_{ij} indicated the error terms. Moreover, the i subscript indicated individuals, and j denoted the main comparative groups, Saudis, and non-Saudis. Ignoring the j subscript provided the pooled sample of both groups. This equation was used for separate two years of cross-sectional data (2013- 2017). However, our two datasets could be pooled, and empirical analyses could be estimated. [76]

The $\sum \beta_j$ comprised the model parameters, which were interesting points for estimating the effect of $\bar{x}_{ij}$, which comprised the explanatory variables, including the observed worker characteristics, job characteristics and policy variables. We used educational qualification for both years and education for year 2017. From one angle, we are aligned with the literature that used those as a categorial variable and used an indicator categories IC to capture those who has unknown qualification or education (Lehmer and Ludsteck, 201; Luik et al., 2018). Moreover, our education classification somewhat like Agrawal (2014) study on India who used illiterate, primary education categories in his classification which did not used in other study in European counties. This can imply that the education level in Saudi Arabia is different than European countries and like Asian countries such as India. From another angle, our qualification variable is relatively like Luik et al. (2018) classification who use humanity, and health for example as a category and unlike Longhi et al. (2012) who set several qualification levels based on education. Accordingly, we are different than Kee (1995); and Fernandez (2017) who use a continues variable for education and Joona 2010 who use a dummy variable for qualification. Furthermore, we used continuous age as a demographic variable unlike (Luik & Steinhardt, 2016; Longhi 2012) and non-linear term unlike (Luik et al., 2018). Although we used a non-linear variable, we were different from most of the research that used squared terms such as Siddiqui et al. (1998). However, we are being aligned with Gottschalk (1978) study that used cubic age in the regression. We used gender variable as an independent variable in the regression to capture the heterogeneity between the two gender groups like Abdullah et al. (2020); and Smith & Fernandez (2017) which unlike some study that excluded female from the sample (Lehmer & Ludsteck, 2011; Longhi et al., 2012; Kee, 1995). In terms of workers' origin, we have available information in 2013 dataset only. We treat

[76] The dependent variable was found in the literature, as explained, as a response, predicted or regressed variable, while the independent variable was found as a control, explanatory, predictor or regressor variable. This also could be known as the covariate (Wooldridge, 2008).

this variable in two strategy: including origin as an independent variable in OLS regression combined with gender, and aligned with Frank et al. (2013) while we use separate origin estimation in the decomposition estimation following Simón et al. (2008). However, unlike other studies, our 2017 dataset has a dummy variable if Saudi equal 1 and 0 if not.

Work characteristics are considered an important dimension as well. We used a categorical variable firm size as most of the study did (Bílková, 2019; Longhi et al. (2012). Unlike Lehmer & Ludsteck (2011) who used the cumulative distribution to select firms' size, we follow the classification defined by the MLSD. Furthermore, firms' age was available only for the 2017 dataset. This variable was rarely used in the decomposition approach previously; however, we used this variable following Rand & Torm (2012) but we differ in our specification by using age and its square unlike their 2012 study that used age in? its logarithmic form. In line with majority of study we used occupation and economic activities following the Saudi classification that follow relatively closely the international occupation and industries classification only on the first digit; unlike Longhi et al, (2012). They used the third digit in their classification. we used occupation and economic activities variables in our regression in line with Hofer et al. (2017). Additionally, we used the region categories as a control variable in line with the literature (Lehmer & Ludsteck, 2011; (Hofer et al., 2017); however, we do not tend to use the dummy method used by Rand & Torm (2012) for region area. We used a categorical variable indicating firms' status according to their saudization percentage. This variable was in context of quota system unlike those who use female or citizen percentage out of the quota purposes (Rand & Torm, 2012; Lehmer & Ludsteck, 2011).

Table 4-3: Model specifications: Definitions and availability in the two datasets

Variable	Definition	2013	2017
Ln (w)	Continuous dependent variable denoting the natural logarithmic form of the monthly earnings in Saudi riyal for each individual for a cross-sectional dataset from 2013 or 2017.	✓	✓
Qualification	Series of dummy variables to indicate if the individual belonged to a particular qualification category. These variables were arranged in descending order*	✓	✓
Education	Series of dummy variables to indicate the individuals' education levels. These variables were arranged in ascending order.	-	✓
Age	The continuous variable measured in years.	✓	✓
Age-squared	The quadratic term of age.	✓	✓
Age-cubed	The cubic term of age.	✓	✓
Saudi	Dummy variable equalled 1 if Saudi and 0 otherwise.	✓	✓
Nationality	Categorical variable to allocate a number to each nationality, which involved a series of dummy variables.	✓	-
Female	Binary variable equal to 1 if female and 0 if male.	✓	✓
Size	A categorical variable to allocate a number to each firm size. However, the usable form involved a series of dummy variables in ascending order. The MLSD provided firm sizes according to Nitaqat criteria*	✓	✓
Firm age	The continuous variable measured by years. This variable informed how long the firms had been in the market.	-	✓
Occupations	Categorical variable to allocate a number to each occupation. This involved a series of dummy variables. This variable was linked to the Saudi classifications for giving job titles. This variable was in descending order and had a similar classification for both datasets.	✓	✓
Activities	Categorical variable to allocate a number to each activity and involved a series of dummy variables. This variable was ordered ascending.	✓	✓
Zone	Categorical variable to allocate a number to each region and involved a series of dummy variables. This variable denoted the administrative region distributed geographically. It was not systematically ordered.	✓	✓
Colour	Ordered categorical variable classified ascendingly that allocated a number to each colour band and involved a series of dummy variables. This variable represented the quota, where a colour band was the firm status according to the Nitaqat programme*	✓	✓
Localised firms	Binary variable equal to 1 if a firm was classified as green or above and equal to zero when a firm was red or yellow.	✓	✓
Interaction variable	Interaction variable between Saudis and the firm status used to explore the Nitaqat risk effects*. For 2013, detailed worker nationality was considered as well.	✓	✓

*Indicates that the variable had different categories for each dataset. For details (see appendix A, table 10-3).

4.5.2 Variable justifications

The coefficients of a categorical variable could have a similar effect on each other. In this case, joining two or more categories together might be economically and statistically justified because there would be no significant reduction in the goodness of fit. The F-test was performed in two steps: first, we examined the equality of the joint coefficients. Equal coefficients were the null hypothesis; when the null could not be rejected, combining the relevant categories could be better. This step ignored if the variable was continuous. Second, we examined two nested models to justify which was the best model – the restricted or the original model using the F-test (this required two nested models). In other words, we examined if a variable's coefficient was equal to zero or if the coefficients could be equal. This required the sum of squared residuals or R^2. The original model being equal to the restricted model was the basis for the null hypothesis. If the null were not rejected, the original model would be kept.[77] The specified variables were used in both the earnings function estimations and the Oaxaca decomposition.[78]

4.5.2.1 2013 variable justification

Size: Restricting firm size was essential, as the micro firms were omitted due to the collinearity between firm size and the colour bands. To solve this issue, small and micro firms were joined in one category, which means this variable had four categories: small firms, including micro firms, medium, big, and giant firms. SIZE was the new name of this variable. Merging the categories did not generate any difference between the two estimations in terms of the model fit statistical result. To ensure that was the correct step, the colour band was reduced, and the size was kept extended. The result showed that the merged size category generated a slightly better result than the merged colour band. The root mean square errors (RMSEs) were 0.4748 and 0.47482, respectively.

Age: The polynomial variable was examined between the quadratic and cubic terms. According to a significant F-test result, where the F-test was 21.45, and the p-value was zero, using the cubic age was better than using the quadratic age. In other words, the cubic age should be considered in the earnings function (which was not common in the prior literature).

Nationalities versus Saudi: With the 2013 data, we benefitted from the details on nationality. Thus, we looked at whether it was important to include a non-Saudi category to distinguish such individuals from their Saudi counterparts or use non-Saudi

[77] For the formula (see Appendix B, Equation 10-1).
[78] For details on these variable specifications, (see Appendix B, table 10-1 and 10-2).

as equal to zero. Two nested models were compared through the F-test, where the null hypothesis was that non-Saudi equalled zero. The F-calculated was approximately 277, and the F-critical was 1.326, and the p-value was 0.0000. The result supported rejecting the null hypothesis, which meant that non-Saudi did not equal zero. In other words, the extended variable was much better than the reduced one. Thus, we included nationality in our model.

Nationality: This variable contained roughly 35 categories from several countries. This variable was restricted to 15 categories, where the best compensation was renamed nation3. In this step, the F-test was exploited to know the equality between categories. Then, the restricted and unrestricted models were examined. The null hypothesis could not be rejected when the p-value equalled 0.1219. Therefore, nation3 was used in the earnings function estimation. Accordingly, this variable was supposed to be better than the simple dummy variable, Saudi. For confirmation, the F-test equalled 700.93, and the p-value equalled zero. This suggested that the null hypothesis was rejected, which implied that nation3 was better than the simple and extended variables.

Qualifications: This variable was formed by 25 categories for several educational qualifications. The F-test showed the possibility of some equal coefficients. The restricted specification involved 17 categories, renamed as a qualification. A comparison of the restricted and extended models confirmed that the null hypothesis could not be rejected when the F-calculated was equal to 0.512 and the p-value was equal to 0.8671. Therefore, the restricted model was at least as satisfactory as the extended model. Thus, the new variable was used in the estimation.

Zone versus regions: Similarly, this model was nested from the last model. However, the two nested restricted variables – regions with seven categories and regions23 with eight categories – were found equal to the main specification, which contained 13 zone categories. The two variables were nested from each other as well; the models were examined under the F-test null hypothesis. The p-value was equal to 0.8748, which did not allow us to reject the null, which indicated that regions was the most-reduced variable and was better than the less-reduced variable, regions23.

Finally, adding an interaction variable was much better than using nation3 alone. Several interactions could be generated: first, including each 2nationality with the localised status (nation3_status), second and third were the interactions of nation3 with firm status, which generated nation3_localised and nation3_non_localised. The difference between the three options was the interpretation of the model.

4.5.2.2 2017 variables justification

Age: This was a continuous variable; thus, we compared the nested model directly. The result of the F-test showed that the unrestricted model using cubic age was better than using the restricted model with just quadratic age. The F-calculated was 201.680, while the F-critical was 6.63, with a 1% significance level. Doubtlessly, cubic age needed to be considered in the estimation model. This was like the 2013 dataset.

Size: Firm size was also restricted in the first categories of small and micro to solve the collinearity with colour bands. This variable was renamed SIZE1. The point was taken that the priority might have been to solve a multicollinearity problem. However, the F-calculated was equal to 1.328, and the p-value was equal to 0.2492. This insignificant result means the null hypothesis could not be rejected. In other words, the two categories could have equal coefficients. Indeed, equal coefficients suggest effectively merging two size categories. This means using the restricted model was much better than the original model. The test result supported our priority point. Three medium levels – A, B and C – were defined in this dataset, unlike the previous dataset. This differential came from the change of firm size definitions under Nitaqat2.

Saudi: Adding a Saudi dummy variable to the last model generated better results than restricting the model by ignoring the Saudi variable. In terms of the nested model test, the F-calculated was equal to 3,937,152.517. This number and the p-value of 0.0000 implied the necessity of a null hypotheses rejection. Therefore, Saudi needed to be included in the estimation, regardless of the unavailability of detailed non-Saudi nationality data.

Qualifications: This variable was examined to see if any coefficients could be equal. Then, it was restricted to 19 categories and renamed Qualifications_3. The usage of the new variable was better than the original variable, so the null hypothesis could not be rejected according to the F-calculated result, which equalled 1.234, and the p-value equalled 0.29. However, these variable categories were not identical to the categories in the 2013 dataset; new categories had been introduced (in the raw data), such as the School of Economics and the School of Law.

Zone: Following similar steps, the zone was successfully restricted to 11 categories, compared to 13 in the original variable. It was renamed regions2. This new variable was used in the restricted model against the zone in the original model. The statistical result confirmed that the null hypothesis could not be rejected; the p-value was equal to 0.1772. This meant that using regions2 would generate satisfactory results as much as using the zone.

Interaction variable: Using an interaction variable between Saudi and the localised status was much better than excluding this interaction variable from the regression; the p-value was equal to 0.0000. This interaction variable would, therefore, be considered in the estimation.

4.6 Conclusion

In this chapter, the framework suggested explaining why this gap existed between natives and immigrants, which could fill the gap in the literature using Saudi Arabia as an example. The modern immigrant theory and the hedonic wage literature were exploited to construct this framework. The framework suggested that benefits would increase wages, while fees would decrease wages under the layoff risk resulting from the quota policy (Nitaqat), alongside consumption. However, the wage gap between Saudis and non-Saudis could mainly be explained by consumption if the government sector was ignored. These differences in government policies and consumption allowed the existence of a multi-supply side. Thus, the consumption and the heterogeneity of the policies could be new explanatory variables in the wage gap. Accordingly, we chose the OB methodology to be empirically used on the Mincer earnings function, and the OLS method was used in the estimation. The variables were carefully defined and justified. The next chapter covers the empirical analysis.

Chapter 5 Empirical Analysis

5.1 Introduction

Empirical analyses were applied to support the research hypothesis and answer the question. First, the logarithmic form of the earnings function from the labour market was estimated using the OLS method. Second, the wage differential was examined using the Oaxaca methodology, which depended on the OLS method for the analysis. Both analyses contributed to yielding inferences that helped understand the earnings structure, Nitaqat effect on the wage structure and the wage gap between Saudi and non-Saudi workers in several respects.

5.2 Earnings functions

As far as the earnings structure is concerned, the earnings function was estimated for the total labour market, using the OLS for each separate cross-section. Both datasets had different specifications; the 2013 dataset had the advantage of providing nationality information. However, the 2017 dataset had the advantage of variables such as firm age, in addition to the very large number of observations. However, we attempted to pool the two datasets to evaluate the changes in the earnings structure. This strategy only provided a tentative result because of the data differential.

5.2.1 Earnings function for 2013

The specifications from the last section were used to run an OLS regression to estimate the earnings function for 2013. As mentioned earlier in the model sub-heading, some assumptions needed to be met to have a BLUE estimator. The error term was exogenous; no correlation was found between the error term and the independent variables. Moreover, it had a mean of almost zero, following the OLS assumptions. Furthermore, the error term was roughly normally distributed according to the graphical approach, although this test failed in the statistical method.[79] The graphical approach was used because of the limitations of the statistical approach in terms of the number

[79] For test results (see Appendix B, Section 10.2.2).and graphs (see Appendix B, Section 10.2.2).

capacity of sample observations. Similarly, heteroscedasticity was found via the statistical method, while it was unrecognised via the graphical method. This could be explained because one potential solution to heteroscedasticity is to increase the sample size, and our sample was large. The dependent variable points were spread out, surrounded by the fitted value line. However, using robust standard errors was recommended (Arkes, 2019). When the interaction variable was included in the regression, using the robust standard error generated insignificant results for the interaction variable. However, the significant result for the variable's coefficients did not change. This means that using robust standard errors under multicollinearity could potentially have led to a misleading conclusion for an important variable, even though the coefficients were BLUE under the heteroscedasticity. Desai et al. (2013) supported this point of view, by stating that using the robust standard error with the presence of intra-cluster correlation would provide a type I error, which would lead to a misleading conclusion. Several estimations were performed to confirm that using robust standard errors could lead to a misleading conclusion, and the results were compared to confirm or reject this claim. The results confirmed that inconsistent conclusions could be reached when the robust standard error was considered in multicollinearity conditions. Therefore, we could state that the heteroscedasticity did not violate the BLUE estimation when multicollinearity was absent.

Table 5-1 display the relationship between age and female according to the earning function estimation; model1. **First**, the relationship between age and log-earnings was positive, even though all three polynomial variables were significant at the 1% significance level. This indicated that earnings increased by around 6% when workers worked for an additional year. Age was used as a proxy for experience (Hayfron, 2002). Predicted log-earnings initially increased at a decreased rate until a specific age, then began to increase at a growing rate when the age increased one year. Indeed, this could be found when the estimated equation was differentiated with respect to age.[80] This result provided an implication that each additional year could enhance a worker's experience, which would be reflected in their earnings. However, in advanced

[80] This age could be found by the second derivative of the (log) earnings function with respect to age, where age = 0.0022142 / 0.000050 – approximately 44 years.

economies, it would be more likely for the gain in log-earnings to be higher at young ages and for this to be reduced as further experience (age) was gained. **Second**, the mean of female earnings was lower than that for men by 0.13260, at the 1% significance level, which indicated a gender pay gap between men and women unexplained by roughly 12.5%,[81] given the semi-log specification.

Table 5-1: mean value and the coefficients of age and female (2013).

The natural logarithm of wage	Coefficient	Mean
Age	0.0601485***	49.62691
Age-squared	−0.001107***	2,511.506
Age-cubed	0.0000083***	129,440
Female	−0.132602***	0.007709

***significant at 1%.

Furthermore, the return on education measured by educational qualification was positive at the 1% significance level compared to those who did not register their educational qualification, except those with an educational qualification of veterinary training and animal production and secondary agricultural; these were insignificant (see Table 5-2).[82] This was consistent with the human capital theory; the earnings increased when the education increased, which could explain the earning differential between groups (Gottschalk, 1978; Purnagunawan, 2007; Longhi et al., 2012). For example, a Saudi worker in a localised firm would earn 10,377 SR if he had a secondary qualification, 15,521 SR if he held a degree in administrative science or translation and 8,538 SR if he had an unregistered qualification.[83] Therefore, predicted earnings increased in some educational qualifications compared to others, such as engineering. This required further investigation into the occupations filled by those with this qualification. From a policy perspective, this needed to be taken into consideration regarding the Saudisation plan. Admittedly, the government expenditure on education must be used effectively by engaging Saudis, for example, in engineering occupations, which cannot be filled by qualified workers where Saudi unemployment is recognised.

[81] Using the pooled option in the OB generated an unexplained part equal to the index coefficient value in the pooled model.
[82] Qualified veterinary surgeons in countries like the UK are very highly qualified and well-paid, unlike in Saudi Arabia. The reason for this is that they work, in truth, as shepherds or possibly in other jobs. The mismatch leads to such results.
[83] Concerning those specifications: aged 30, Riyadh, manager, education, platinum.

Table 5-2: the returns of educational qualification on wage (2013).

The natural logarithm of wage	model1	Mean
College of Literature	0.33115***	0.0028098
Languages and Translation & Administration Science	0.59770***	0.0110378
Colleges of Education & Colleges of Agriculture	0.26608***	0.0025553
Colleges of Science	0.69557***	0.0018661
Colleges of Pharmacy	0.74342***	0.0010073
Colleges of Medicine	0.85926***	0.0013784
Engineering Faculties	0.90833***	0.0081644
Architecture and Planning & Technical College	0.47143***	0.0036157
Colleges of Medical and Applied Sciences	0.61667***	0.0006256
Faculty of Computer Science and Information	0.50779***	0.0043367
Secondary School	0.19510***	0.0278331
Industrial & Trading School	0.27343***	0.0180995
Veterinary Training and Animal Production & Secondary Agricultural	−0.004445#	0.0014526
Institute including (Institute of Management, Technical Institute, Air Science, Institute of Professional Observers)	0.28283***	0.0115362
Health Institute	0.40692***	0.0008376

***significant at 1%, # insignificant.

Surprisingly, workers in Makkah earned more than workers in any other region at the 1% significance level (see Table 5-3). For example, a Saudi worker in Riyadh who had attributes identical to other workers in Makkah would earn less by approximately 8.3%, which would be around 354.45 SR less than the wages in Makkah.[84] Similarly, workers in Eastern Province, who had identical attributes to other workers in Makkah, could earn lower (by 0.17066 logarithmic points). This means that workers in Riyadh would earn higher than those in Eastern Province, but lower compared to the workers in Makkah. On the one hand, this could be a drawback of the sample structure as there were no observations of Saudis registered in a non-localised firm. On the other hand, it could be that Makkah was non-localised at the beginning of the Nitaqat period. Thus, firms paid extra to attract Saudis and satisfy the policy requirement. This could imply genuine Saudi employment rather than a dummy or temporary Saudisation. These results confirmed that the difference in earnings across the regions needed a different policy in each region (Burke et al., 2009).

[84] For similar characteristic in footnote 83.

Table 5-3: coefficients and mean for region as a control variable (2013).

	Coefficients	Means
Al-Jouf and Qassim	−0.30001***	0.0277589
Riyadh	−0.08662***	0.1719612
Eastern Province	−0.17066***	0.2917656
Tabuk and Najran	−0.12783***	0.0042094
Hail and Asir	−0.21636***	0.0483183
Others	−0.15456***	0.0877619

***significant at 1%.

Table 5-4 show that workers in basic engineering earned less than those in skilled and semi-skilled occupations, apart from service occupations. Those workers earned roughly 8% less than those in basic engineering. This percentage is worth around 65 SR. For example, for an Indian in a green2 firm in Makkah working in mining and quarry with a secondary school qualification, the predicted earnings would be around 800 SR in a service occupation compared to 865 SR in a basic engineering occupation. Workers in unskilled occupations, such as agriculture and industrial, earned less than basic engineering, like those works in service occupations. This could imply that service and unskilled occupations were not the best occupations for Saudis to be replaced with non-Saudis. Thus, when Nitaqat is modified, occupations will need to be taken into consideration.

Table 5-4: the coefficients and mean value for occupation.

	Coefficients	Mean
Managers, directors, and senior officials	0.632118***	0.0135084
Specialists	0.550457***	0.0990966
Technicians	0.207356***	0.0817181
Clerical occupations	0.195009***	0.0113453
Sales occupations	0.126054***	0.0792582
Services occupations	−0.07741***	0.4205721
Agriculture and animal husbandry	−0.07083***	0.0055348
Industrial and chemical processes	−0.04377***	0.0317987

***significant at 1%.

However, those lower-paid occupations formed over 50% of all economic activities. This means that economic activities (industrial sector category) could not influence the workers as much as the occupations. Frankly, all activities earned more than construction activities (the base case) whether in high or low values, apart from agriculture, which was insignificant. Education, manufacturing, mining, and quarrying and retail and household had higher earnings values compared to construction, as specialist and technical formed a good proportion of these activities (see Table 5-5). However, professional activity had an insignificant result, likely because workers in construction do not usually earn less than those in professional activities when they have similar attributes.

Table 5-5: the coefficients and mean of economic activities, a Nitaqat criterion (2013).

	Coefficients	Mean
Agriculture, Forestry and Fishing	−0.006463#	0.0143248
Manufacturing, Mining and Quarrying.	0.20273***	0.131277
Wholesale and Retail Trade	0.12899***	0.2173424
Repair of Motor Vehicles and Motorcycles	0.07791***	0.0405887
Transportation and Storage	0.04868***	0.0313428
Accommodation and Food Service	0.08147***	0.0403978
Professional, Scientific, and Technical	0.0013864#	0.0179298
Education, Human Health and Social Work	0.15029***	0.0393057
Other Personal Services	0.08708***	0.005344
Agriculture, Forestry and Fishing	0.05901***	0.070341

***significant at 1%.

In terms of firm size, it seems that big firms rewarded their employees better than other firm sizes, at the 1% significance level. A lower predicted log-earning was found in small and micro firms compared to big firms. In our example above, if the person were working in a small or micro firm, he would be predicted to earn less by approximately 19.42% (see Table 5-6). Indeed, this was an expected result; small firms have a lower ability to pay high salaries compared to the other firm sizes. However, the predicted earning would be lower by trivial amounts if medium and giant firms were taken into consideration. This implies that workers in medium, big and giant firms have a converged earning when their workers' attributes are similar, unlike small firms, which, to some extent, follow the literature intonations (Schmidt & Zimmermann, 1991; Bílková, 2019). This could mean that small firms could be less able to provide acceptable jobs for Saudis, which was one of the Nitaqat programme's strongest criteria. Small firms could run as self-employed firms, helping Saudis engage in the market; as noted in the descriptive statistics chapter, there was an increase in those firms. From another angle, lowering the Saudisation percentage in small firms could increase al-tasatur, and defeating these firms was one aim of the MLSD. However, this did not necessarily mean that firms would be legally owned by Saudis in other firm sizes. In this context, providing some criteria to ensure that firms existed legally would be helpful. Indeed, the earnings function would not help to determine those criteria.

Table 5-6: mean value and the coefficient of firms' size, a criterion of Nitaqat (2013).

	Coefficients	Mean
Small	−0.1759431***	0.3744804
Medium	−0.0552357***	0.3462656
Giant	−0.0379957***	0.0844643

***significant at 1%.

In all estimated models[85], the earnings function predicted that Saudis would earn more than other nationality groups, even if they had identical attributes, according to estimated models at the 1% significance level, with one exception: European3.[86] This group was insignificant, which indicated that this group could earn equal to or more than Saudis. This could be understood as this group included some high background nationalities, such as Greek, and immigrants usually accepted jobs if the pay were higher than they received in their countries.

Table 5-7: the mean and semi-log specification for each nationality (2013).

	Coefficients	Mean
Nepalese	−2.101352***	0.0007316
South African, Somali & Jordanian	−1.557937***	0.0112817
Mali	−2.182796***	0.0006892
Pakistani & Afghani	−1.837344***	0.1790546
Indian and Swiss	−1.956804***	0.3170116
Sudanese & Filipino	−1.576919***	0.0814318
Bangladeshi	−2.021013***	0.2093371
European1	−1.219735***	0.000721
European2	−1.607793***	0.000053
Syrian, Chinese & Turkish	−1.406029***	0.0096488
Other Asians	−1.776143***	0.0084401
African2	−1.723263***	0.0925757
European3	−0.2875063#	0.000106
Palestinian, Yemeni & Mauritanian	−1.684686***	0.0683158

***significant at 1%, # insignificant. The coefficients based on model1; Saudi is the base category.

85 See footnote 79 above.
86There were only ten individuals in the European3 category. Similarly, European2 had only five localised observations.

To clarify, if we took the accounting occupation as an example, we could obtain the disparities between Saudis and non-Saudis. In Greece, the average accounting monthly gross wage is worth around 16,159.80SR (Explorer, 2020), compared to the 10,400SR salary in Saudi Arabia (Explorer, 2020). However, the salary of this occupation in Yemen is worth around 1,593.61SR (Explorer, 2020).[87] Thus, an individual from Yemen would be expected to accept an accounting job with any higher wage, for example, 3,000SR, which would be unacceptable for a Saudi. This would create employment and wage gaps between Saudis and the other remaining groups and give those workers a competitive advantage, unlike Greeks, who, it seems, would not accept less than they would earn in their own country. Thus, we found that several workers from high background countries had rare qualifications in Saudi Arabia, unlike qualified workers from other lower background countries. From another angle, if this was the case for a qualified person from Yemen or other countries with similar or lesser economic backgrounds, we assumed that unskilled workers would accept any salary better than they could receive at home, maximising their future consumption – savings. This was unlike those from high background countries where the minimum wage was much better than the average wage in Saudi Arabia. For example, the minimum wage in the UK was around £6.31/hour in 2013 (GOV.UK, 2020). This is worth approximately 5,000SR per month, which is much higher than the occupation earnings considered the minimum wage level. Logically, workers from those countries would work in skilled occupations, exploiting their knowledge and benefitting from the rarity of their specialist skills. Incredibly, in the Saudi labour market, wages are determined according to the demand and supply sides, whether through the recruiting offices or bargaining power, ignoring all the variations between the backgrounds of Saudis and other workers.[88] According to this analysis, it is recommended to determine the occupations that should engage migrants by setting an effective salary scale for each occupation regardless of nationality. For example, an accountant might be rewarded 7,000SR, whether they were Saudi or not. In this case, firms would choose productive workers rather than low-cost workers. This way, firms could maximise the gains from the

[87] A currency conversion from 05/04/2020 was used for the relevant exchange rates. According to the website, this salary was inclusive of all benefits.
[88] The government sector has a clear salary scale, unlike the private sector.

recruitment process from two angles. **First,** recruits on low-salary scales, such as cleaners, would provide low-cost labour and guarantee Saudi welfare; also, the remittance expectation would be lower. **Second,** firms would recruit workers with high qualifications who would add to the economy even though they were highly paid.

Although under Nitaqat, firms were classified according to their Saudisation status in several colour bands, firms considered themselves localised or non-localised. The green1 category was the baseline of this policy and was used as the reference (colour category) group in the regression. Table 5-8 revealed that workers in green1 firms could earn less than those in other localised firms by trivial amounts compared to other green levels and by significant amounts compared to platinum and excluded firms. These results were significant at the 1% significance level. However, earnings compared to those working in non-localised firms were insignificantly different, although there was a positive difference apart from for yellow firms: employees in those firms could earn less than those in green1 at the 10% significance level in model1 and model2, but this was insignificant in model3 and model4. Those results imply that earnings were similar in non-localised firms and the early stages of localised firms. This could indicate that the higher the proportion of Saudis a firm employed, the higher the earnings could be expected. Exploring the reasons that enabled those firms to pay a higher salary compared to other firms would be essential. The statistical chapter displayed that capital formation changed during the Nitaqat application; thus, it could be assumed that the capital was an essential issue contributing to this variation. It was recommended to use firm capital as a criterion side-by-side with the number of workers to determine the firm size. This could be presented as the labour/capital intensity measurement.

Table 5-8: the coefficients and mean of firms' zone according to Nitaqat (2013).

	Coefficients	Mean
Red	0.0453482#	0.0044215
Red small A	0.011057#	0.0102214
Yellow	−0.060196*	0.0109
Green small A	0.04074***	0.0950356
Green2	0.04194***	0.374173
Green3	0.06749***	0.1660128
Platinum	0.17485***	0.0883027
Excluded	0.38743***	0.0002333

***significant at 1%, * significant at 10%, and # insignificant.

By looking at the policy variables – the interaction variable in each model – some information could be extracted. Both model1 and model2 provided similar conclusions on how much a group was expected to earn compared to a similar group working in a firm with different localisation status. For example, Saudis in non-localised firms could earn around 39% more than their Saudi peers. This implies that Saudis experienced higher wage offers if they applied to non-localised firms. This conclusion seems to be expected; firms try to attract Saudis to satisfy the policy percentage (see Table 5-9).

Table 5-9: interaction variable between nationality and firm's status.

	Coefficients	Mean
Saudi	−0.33069***	0.0257761
Nepalese	−0.083858#	0.0204746
South African, Somali & Jordanian	0.345695***	0.0006998
Mali	0.3264345#	0.0110378
Pakistani & Afghani	−0.055337#	0.0006786
Indian	−0.026477#	0.1736364
Sudanese & Filipino	−0.15448***	0.3091653
Bangladeshi	−0.08708***	0.0798838
European1	0.3580799#	0.2039931
European2	Omit	0.0007104
Syrian, Chinese & Turkish	−0.20128***	0.000053
Other Asians	−0.1572629#	0.0093944
African2	−0.0592559#	0.008281
European3	Omit	0.0896917
Palestinian, Yemeni & Mauritanian	Omit	0.000106

***significant at 1%, and # insignificant. Non-localized is the base category.

Additionally, Table 5-9 show that non-Saudis, who were not policy targets, were affected, as well. From one angle, some nationality groups were expected to earn more than their peers when they worked in a localised firm. For example, at a significance level of 1%, only workers from Jordan, South Africa and Somalia were expected to earn more if they worked at a localised firm, by approximately 41%. Although those workers were restricted from moving between firms, they had a lower layoff risk compared to their peers; this was the only group to behave opposite of the expectations of the hedonic wage model. The result was consistent with (Pinheiro & Visschers, 2015; Scicchitano et al., 2019). Under a layoff risk, Jordanians, South Africans, and Somalians were expected to earn less. From another angle, non-Saudi workers at non-localised firms (being under layoff risk but also free to move) were expected to earn more than their peers in localised firms, for example, Sudanese and Filipinos by around

17%, Bangladeshis by around 9% and Syrian, Turkish and Chinese workers by around 22%. Those findings followed the hedonic literature's prediction. Moreover, the remaining nationalities were insignificant.

We obtained broad implication of expected earnings for each group of Saudis or non-Saudis from model3. Although the interaction variable was reduced, Saudis had a similarly expected log-earning; in this model, it was approximately 9.24725 in a localised firm, worth approximately 10,376.01 SR, compared to a log-earning of 9.24734, equal to 10,376.95 SR, in the other models. This log-earning variation of 0.00009 was approximately 0.009%, which is a negligible amount for arguing that model3 generated biased results.[89] Moreover, the response to the policy was similar (see Table 5-10). However, model3 did not distinguish between the nationality groups on the interaction variable, and it provided a broad perspective regarding the expected earnings for non-Saudis based on firm status. It seems that non-Saudis, in general, exhibited results that were the opposite of the hedonic wage expectation; non-Saudis earned less when they were exposed to layoff risks. In 2013, this was expected; the programme was newly introduced, and the deportation levels were high. It was reasonable for workers in non-localised firms to accept lower wages rather than deportation, whether through negotiation for an existing job or by moving to a new job. Those non-Saudi workers in non-localised firms could be classified as double negative under the policy. The reduction in non-Saudi wages would allow adding a new Saudi, which could change the firm status. Although model3 was reduced in the interaction variable, it still provided a convergent result. For similar characteristics, a Bangladeshi in a localised firm would be predicted to earn around log 7.4706, worth 1,756 SR, for model3, compared to log 7.4699, worth roughly 1,754 SR in model1. Indeed, this small variation of approximately 0.009% could not significantly reduce model3's power.[90] The highest risk groups would earn less, and the lowest groups would earn more, unlike the hedonic wage assumption of the more the risk, the higher the earnings (see Table 5-10). Thus, we could state that the wages responded to Nitaqat opposite to the hedonic literature,

[89] Exp (0.00009) =1.00009 = 0.009%

[90] Exp (0.0009) = 1.00009

following (Hübler & Hübler, 2006; Pinheiro & Visschers, 2015; Scicchitano et al., 2019).

Table 5-10: The expected earnings according to model3

	Saudi	Non-Saudi
Localised	-	+
Non-localised	+	-

The interaction variable was estimated in four models.

Furthermore, model4 provided how much a nationality group earned compared to all workers in another localised status. Doubtless, the result based on model4 supported the view that, in general, Saudi workers earned more than other nationality groups, apart from the European3 group.[91] In a localised firm, at a 1% significance level, it was expected that Saudis would earn around 278% more compared to all workers in non-localised firms, except those in European3, who were expected to earn more than others by approximately 304%. In general, European3 would earn more than Saudis by 26% if in localised firms. After Saudis and European3, Euorpean1could earn more by roughly 128%, while Jordanians, South Africans and Somalis could earn more than workers in non-localised firms by approximately 60%. Moreover, Syrian, Turkish, and Chinese workers earned around 8% more than workers from other nationality groups, who earned significantly less than the above groups, at the 1% significance level. In non-localised firms, at a 1% significance level, it was expected that Saudis would earn approximately 439% more than all workers in localised firms, regardless of nationality, and European1 workers would earn around 59% more. Furthermore, at a 5% significance level, Syrian, Turkish and Chinese workers were predicted to earn more than others by 32%; similarly, Sudanese and Filipinos earned roughly 11% more. All remaining nationalities were expected to earn less than workers in localised firms. Therefore, it was deduced that some nationality groups earned more than other groups under a similar policy, which implies the heterogeneity between non-Saudi groups, which would exploit the recruiting policies.

[91] Only 10 observations were in this group (only just over 0.01% of the sample).

5.2.2 Earnings function for 2017

Unlike the 2013 dataset, the nationality variable was bounded due to the change in the MLSD privacy policy. This did restrict the results; however, the estimation still provided some essential implications for the research theme. The specified variables in the previous section were used to estimate the earnings function. The estimation satisfied the OLS assumptions, apart from homoscedasticity as the null hypothesis was rejected in the statistical method. However, a large observations' number was enough to ignore this issue, as shown by the graphical method. Although the model broke the homoscedasticity assumption, it was still able to provide BLUE coefficients. The heteroscedasticity issue could affect the standard error, providing inaccurate significance levels. Therefore, it was recommended to use a robust standard error (Arkes, 2019). The robust standard error could be used safely in this model because there was no multicollinearity issue between the categories in each variable. The multicollinearity that was present stemmed from the polynomial age variable, and when this variable was excluded, the mean of the VIF result, including the interaction terms, was around 1.76.

Although the relationship between wage and worker age was non-linear, it was positively linked; earnings increased around 4% with each additional year of work (see Table 5-11). This result was understandable because age was used as a proxy for experience (Hayfron, 2002). Therefore, when age increased, experience increased, which was reflected in the earnings increase. However, this increase had a non-linear growth rate; wage increased at a slower rate in the earlier years until around the age of 47, then it increased at a faster rate.[92] This result was consistent with the 2013 estimation, which means the earnings direction, according to age, did not change. Typically, men earned about 26% more than women, rather than the 14% in 2013, which pointed to the widening gender gap. This could result from the policy where Saudi women were extensively engaged in the market when women were allowed to

[92] To find a point of inflection of the cubic earning, $\frac{\Delta\, SALARY}{\Delta\, AGE}$= 0.03797 −0.0012856 age + 0.00001368 age^2, the solution was approximated to −9 and 47. Thus, we only considered 47. There was a similar result if we found the age of f (age) = 0. The inflection point = 0.0012856 / (2 * 0.00001368) = 47.

work in most of the minimum wage occupations. The gender gap needs to be investigated under Nitaqat in detail, which is not the interest of this research.

Table 5-11: earning function coefficients for age, female, and firms' age.

	Coefficients	VIF	Mean
Age	0.03797***	1,252.85	40.0667
Age-squared	−0.00064***	5,368.18	1,734.06
Age-cubed	4.56E−06 ***	1,526.42	80,007.38
Female	−0.23442***	1.57	0.145349
Firm age	0.00173***	6.24	12.58336
Firm age 2	−0.00002***	8.74	355.0301

***significant at a 1% significance level.

The existing literature has found a positive relationship between firm age and wage (Coad, 2018). However, controlling for worker characteristics could lead to a negative or insignificant relationship between earnings and firm age (Brown & Medoff, 2003). Surprisingly, our primary estimation of linear firm age was positive and insignificant due to controlling for firm size differently than the past literature Brown and Medoff (2003). However, using a firm age-squared variable generated a significant result and slightly increased the model fit (see Table 5-11 above).[93] This assumed a significant non-linear relationship between the firm age and worker earnings, even though it was a very small effect. Since the firm age coefficient was positive, and the quadratic term was negative, it implied a maximum turning point for firm age of around 35 years.[94] This means that for one additional year of a firm's age, earnings were predicted to increase until firms reached the age of 35, then the expected earnings would decrease. Accordingly, the percentage change would not be fixed for each additional year increase in a firm's age. This could be calculated by plugging the variable – firm age – into the first derivative equation. It assumed that the change in year 35 would be zero; less than 35 would be a positive value, and any year after 35 would be negative.[95] However, the maximum turning point could be expected as the second derivative was negative: −0.0000488. However, the linear relationship of firm age was examined among each

[93] This non-linear firm age was applied with respect to each firm's size, and we found that the power increased.

[94] We calculated the age according to the first differential equation; the result was approximated. Moreover, there were 209,159 firms aged 35 or older, forming 58%.

[95] Inflection point = $\frac{\Delta\, log-Salary}{\Delta\, firms'age}$ = 0.0017289 −0.0000488 (firm age) =35

firm size separately (see Table 5-12). The results displayed evidence of inconsistent responses among firm sizes. Surprisingly, small-micro, medium A and giant firms had similar coefficients exhibiting a small, negative linear impact of approximately 0.1% on wage when firm age increased by one year, compared to the positive effect on medium B, medium C and big firms. Accordingly, old, giant firms could struggle under the Nitaqat policy. Although they have earnings structures like small firms, they have a higher expectation of Saudisation percentage. Thus, firm size depending only on the number of workers could burden some firms. It was assumed this would need to be combined with another measurement, such as the capital/labour intensity.[96] Firms might share this key measurement, leading to a similar earnings structure between the giant and small firms.

Table 5-12: Firm age coefficients extracted from earnings regressions for each firm size

Firms age coefficient in each firm's size	Coefficients	Number of observations
Small and Micro	−0.00091***.	1,672,562
Medium A	−0.00056***	481,045
Medium B	0.00096***	391,441
Medium C	0.00053***	496,727
Big	0.00188***	712,485
Giant	−0.00054***	617,002

***significant at a 1% significance level. For the full estimation (see appendix B, Section 10.2.3).

If small and giant firms had a similar earnings structure, this did not necessarily mean they paid similarly; it seems that workers at giant firms earned more than any other size category. The workers in this category were expected to earn slightly more (around 9.28%) than big firms, unlike the rest of the categories (see Table 5-12). If a worker was predicted to earn 4,478SR at a big firm, he could earn approximately 415SR more if he worked at a giant firm. Also, he would earn 237SR less in medium C firms and 876SR less in small firms.[97] At the 1% significance level, firm size followed the common belief in the literature, where it is agreed that earnings increase when firm size increases (Schmidt & Zimmermann, 1991; Bílková, 2019). Admittedly, this dataset had a trending

[96] There were several financial measurements, such as sales, assets and revenue. The Small and Medium Enterprises General Authority in Saudi Arabia, which was established in 2016, combined both the number of employees and revenues as a measurement of firm size.
[97] This was our example for female, Saudi, non-registered education or qualification, platinum, clerical occupation, and education activity.

order in this dimension compared to the 2013 dataset (see Table 5-13). Moreover, at a 1% significance level, workers localised firms seemed to have a higher earnings expectation than those in non-localised firms, except for small green A. Even if firms were localised, it seems that workers in micro firms would earn less than workers in firms of other sizes. This could imply those very small firms had a lower ability to paid high salaries compared to firms of other sizes, which could be an obstacle to fulfilling the appropriate Nitaqat percentage and competing with other firms. However, the Nitaqat baseline was used as the reference group, which revealed that the wage expectation was higher when the localised status increased. Indirectly, this implied the existence of a wage gap between Saudis and non-Saudis as the percentage of employed Saudis increased when firms moved to a higher Nitaqat colour level.

Table 5-13: semi-log specification for firms' size and colour.

Variables	Coefficients	VIF	Mean
Size			
Small and Micro	−0.21779***	2.6	0.3826268
Medium A	−0.15103***	1.6	0.1100472
Medium B	−0.10326***	1.47	0.0895487
Medium C	−0.05426***	1.53	0.1136347
Giant	0.08871***	1.71	0.1411496
Colour			
Red	−0.02697***	1.39	0.006483
Red small A	−0.03059***	1.12	0.1155794
Yellow	−0.02623***	3.11	0.0438018
Green small A	−0.00388***	1.2	0.2281604
Green1	0.01383***	1.4	0.1430351
Green3	0.09981***	1.41	0.1509209
Platinum	0.16761***	1.52	0.0003109
Excluded	0.33817***	1.02	0.006483

***significant at a 1% significance level.

Another variable indicating the wage gap was a Saudi dummy variable where Saudis earned around 163% more than any non-Saudi by assuming both had similar attributes; however, in 2013, this variable amounted to 187%, which implied the possibility of approximately a 24% wage gap reduction between Saudis and non-Saudis.[98] However, it was assumed that Nitaqat would affect earnings through the layoff risk, which was indicative of a gap between the Saudi groups themselves, as well as non-Saudis. The

[98] The reduction was 1.375. The expected value in 2013 was around 6.488, where 21% = 1.375/6.488.

expected earnings could be seen for each group according to the localisation status by looking at the policy indicator – the interaction variable (see Table 5-14).

Table 5-14: semi-logarithmic specification of the interaction variable implication.

	Coefficients	VIF	Mean		Localised	Non-localised
Saudi	1.63180***	3.56	0.4344851	Saudi	+	-
Expat non-localised	0.02184***	3.31	0.0942536	**Expat**	-	+

This sign was extracted the above estimation using different interaction variable options, as demonstrated in 0.02184 a positive coefficients his sign wrote in column 5, and raw 2.

For example, Saudis in localised firms were expected to earn around 2.2% more than Saudis who were safe in non-localised firms. Saudis in localised firms experienced a layoff risk because a localised firm has such a high proportion of Saudi workers that it can afford to shift its Saudi/non-Saudi mix a bit in favour of the latter (who are cheaper) without the firm being re-graded as non-localised. It can also shift its qualified/unqualified Saudi mix in favour of unqualified to become a localised firm or move into a higher colour band. This relationship supported hedonic wage literature (Bloch, 1979; Hutchens, 1983).

Outwardly, a similar conclusion was found for non-Saudis; those who worked in non-localised firms earned more than those in localised firms by approximately 2.2%. Unlike 2013, this coefficient was around 32%, which contributed significantly more than all localised colour bands[99]. However, combining this interaction variable with the colour band would yield that non-Saudis followed the opposite of the hedonic literature in 2017. For example, non-Saudi workers in platinum firms could be predicted to earn around 2,210SR, while workers in yellow firms could be predicted to earn 1,861SR.[100] Seemingly, workers in small green A would earn more by the ignorable amount (around 1,862SR). Admittedly, in the earlier years of Nitaqat, the demand for Saudis increased in non-localised firms, which increased the earnings for this Saudi group. However, in localised firms, high-wage Saudi workers could be replaced by low-wage Saudi workers when firms aimed to improve their Nitaqat level, which reflected a decrease in the Saudi

[99] The interaction variable replaced with the other three interactions.
[100] Non-Saudi female, aged 30, big firm, Riyadh, unregistered qualification, bachelor's degree, clerical, education, firm age 20.

wage level in the 2013 dataset.[101] Similarly, non-Saudis were under direct deportation risk when firms with Nitaqat shock deported a range of non-Saudis to meet the percentage required in Nitaqat while also employing Saudis in lower occupations around the minimum wage of the quota. This way, the firms' costs could be balanced. However, localised firms who expected to keep non-Saudi workers could dismiss non-Saudis to move to a higher Nitaq. This reflected an indirect risk for non-Saudis in localised firms, who could show dissimilar responses among different non-Saudi groups. Therefore, some nationalities' earnings would increase while others would decrease, according to the indirect risk they experienced. The shock of the policy and firms' responses to it contributed to shaping the earnings structure, responding according to the reverse of the hedonic wage. However, approximately seven years after the start of Nitaqat, improvements to the earnings function were seen in response to the policy, as the hedonic wage assumed. Saudi workers who faced a layoff risk could receive more than their peers in firms whose status was safe, which was the expected relationship according to the hedonic theory.

In terms of the human capital variables, education, and educational qualification, supporting evidence was captured that education could enhance workers' earnings. Table 5-15 results demonstrated that illiterate workers would have the lowest-earning expectations. It might be that a high percentage of workers with unregistered education could hold a higher degree but preferred not to share this information because they worked in a job that was not under the educational qualification compulsory registrations. This allowed more flexibility for them to work in any occupation without restrictions. Another consequence might be that unregistered people could earn more than those holding a secondary degree. A similar conclusion could be drawn for the educational qualification. From one angle, some educational groups were expected to generate higher wages compared to an unregistered category, such as engineering, which was consistent with the 2013 dataset. From another angle, some educational

[101] Supporting this point of view, many newspapers reported the contract termination of several Saudis. Some writers linked the high termination of Saudi workers and law number 77, which allowed firms to dismiss any workers. Firms used this law to choose the best combination of Saudis and non-Saudis by either dismissal of qualified Saudis and replacement with unqualified Saudis or by dismissal of non-Saudi workers. However, these frequent incidents drew the MLSD’s attention to improving the criteria to be qualitative rather than (solely) quantitative.

categories were expected to earn less than the unregistered categories, such as the College of Literature or education. The occupation was another variable that captured higher returns on education (Luik et al., 2018).

Table 5-15: education and education coefficients based on the earning function.

Education level	Coefficients	VIF	Mean
Illiterate	−0.21474***	1.07	0.0336255
Reads and writes	−0.14376***	1.18	0.1186676
Primary degree	−0.16582***	1.1	0.0457637
Intermediate degree	−0.15165***	1.34	0.0610002
Secondary degree	−0.08529***	3.27	0.2563434
Diploma	0.01195***	1.48	0.0314156
Bachelor's degree	0.90278***	1.08	0.0002626
Master's degree	0.04565***	2.05	0.0507629
PhD degree	0.57423***	1.43	0.0167796
Higher diploma	0.67609***	1.04	0.0009137
Fellowship	0.74867***	1	0.0000327
Educational qualifications			
College of Literature	−0.01632***	1.16	0.0211193
Languages and Science and Institute of Management	0.09817***	1.13	0.0126227
Administration Science	0.23498***	1.31	0.0268726
Colleges of Education	−0.05816***	1.31	0.0127567
Colleges of Agriculture	0.01925***	1.02	0.0013143
Colleges of Pharmacy	0.31765***	1.05	0.0013765
Colleges of Medicine	0.39220***	1.15	0.0016478
Engineering Faculties	0.49717***	1.27	0.0121338
Architecture and Planning	0.28966***	1.01	0.000943
Dentistry, Computer & Management	0.21206***	1.16	0.0123635
Colleges of M and A Science and Technical Institute	0.19146***	1.09	0.0090939
High school	−0.00996***	1.83	0.1563855
Secondary trading	0.12808***	1.04	0.0060223
Industrial high School	0.16837***	1.09	0.0093916
Secondary agricultural	−0.03976***	1.01	0.0006927
Technical College	0.07286***	1.15	0.0093646
School of law	0.25701***	1.01	0.0005753
School of Economics	0.33754***	1.01	0.001061

*** significant at 1%.

As skilled and semi-skilled occupations required a specific educational qualification, workers in skilled and semi-skilled occupations were predicted to earn more than those in basic engineering, apart from the service occupations, at the 1% significance level (see Table 5-16). However, workers in basic engineering were expected to earn more than those working in service occupations and other unskilled occupations. Moreover, in this dataset, both specialists and managers earned more than workers in basic engineering by roughly 56% and 44%, respectively. It seems that the specialists earned more than managers by around 12%. This indicated a slight change in the earnings structure compared to the 2013 dataset.

Table 5-16: occupations coefficients where basic engineering is the base category.

	Coefficients	VIF	Mean
Managers, directors, and senior officers	0.36158***	1.31	0.0437135
Specialists	0.44300***	1.46	0.0910673
Technicians	0.16414***	1.42	0.096622
Clerical occupations	0.09943***	1.84	0.1285613
Sales occupations	0.07449***	1.59	0.112365
Service occupations	−0.09497***	1.54	0.1537375
Agriculture and animal husbandry professionals	−0.14658***	1.13	0.0060392
Industrial and chemical processes	−0.07341***	1.42	0.1206601

*** significant at 1%.

Workers in all activities, including agriculture and fishing, were expected to earn more than those in construction activities, at the 1% significance level, except for professional activities, which was insignificant. The activity of mining seemed to have the highest return compared to base categories, including construction by 35%, other activities by around 20% and educational activities by 18%. This result was slightly inconsistent with 2013 estimation results.

Table 5-17: activity contribution on wage, the construction base category.

	Coefficients	VIF	Mean
Agriculture, Forestry and Fishing	0.10382***	2.82	0.0151725
Mining and Quarrying, Manufacturing	0.29740***	1.33	0.1100346
Wholesale and Retail Trade	0.10747***	1.5	0.1944592
Repair of Motor Vehicles and Motorcycles	0.06056***	1.1	0.0341716
Transportation and Storage	0.09167***	1.08	0.0307335
Accommodation and Food Service Activities	0.10311***	1.15	0.0485112
Professional, Scientific and Technical	0.00068#	1.12	0.0329786
Education, Human Health and Social Workers	0.16236***	1.37	0.0720517
Others Personal Services	0.08269***	1.03	0.0060635
Other Activities	0.18553***	1.23	0.0934241

*** significant at 1%, #insignificant.

In terms of regions, earnings structures seemed to be different compared to 2013. In 2017, workers in the Eastern Province and Riyadh were expected to earn more than those working in Makkah (the base category), with significance levels of 1% and 10%, respectively, and the Eastern Province was the highest among the three regions. These earnings variations were negligible, at approximately 0.39% and 0.11%, respectively, compared to Makkah.[102] This could indicate that the Eastern Province and Riyadh had the highest response to the policy among other administrative regions, even though Nitaqat required an equal percentage for all regions. This confirms the importance of considering the regional dimension when a policy is set up (Burke et al., 2009)

Table 5-18: earning function coefficients according to the region.

	Coefficients	VIF	Mean
Al-Jouf	−0.17295***	1.03	0.0077026
Riyadh	0.00108*	1.65	0.3645931
Eastern Province	0.00386***	1.57	0.2138552
Qassim	−0.14895***	1.13	0.0357961
Madinah	−0.04723***	1.13	0.0401799
Tabuk	−0.03036***	1.04	0.0115051
Jazan	−0.05549***	1.05	0.0131468
Hail	−0.13388***	1.05	0.0121301
Asir	−0.09366***	1.12	0.0332

* significant at 10%, *** significant at 1%

5.2.3 The changes in the earnings structure

Applying a policy such as Nitaqat would leave an impact on the earnings structure during the application period. A pooled regression was used to address the change in the earnings structure (Meng & Kidd, 1997). This method required the data from 2013 and 2017 to be pooled in one sample. A dummy variable called year was generated, where 2017 was equal to one, and 2013 was equal to zero. Then, the interaction variables for each regressor with the dummy (year) were generated. The pooled regression included all variables, the year dummy and all interactions.[103] The results of the main variables were equal to the 2013 estimation results, and the interactions (our interest variables) explained the changes in the earnings structure between the two years. However, because of the differences in the variables between the two datasets, some arrangements were made. First, education and nationality were excluded from the pooled regression. Second, when the categories were slightly different, such as in qualifications, they had zero observations for the year that was not included. For example, the School of

[102] See similar example to footnote 83 above.
[103] For summary statistic (see Appendix B, table 10-15) and for detailed estimation result the (see table 10-16).

Economics, which was introduced in 2017, had zero observations in the 2013 dataset.[104] Finally, the medium levels for colour and size were combined into one category.

The main result that could be deduced was that the wage gap between Saudis and non-Saudis was significantly reduced at the 1% significance level by approximately 36%. This indicated the success of the Nitaqat policy, which aimed to reduce the gap between Saudi and non-Saudi wages. This result was consistent with Burger et al. (2016) finding that the gap would narrow because of affirmative action. However, the result was contradictory this study as the target group's wage reduced, unlike Burger et al. 's finding that Black wages increase through an increase in Black education. The expected earnings decreased in 2017 compared to 2013 in all colour bands apart from green3, where earnings increased slightly by 3%. However, the wages decreased sharply in non-localised firms; for example, earnings decreased in red small A firms by about 49%, compared to a 5% decrease for small green A. Similarly, for red and yellow firms, the reductions were 54% and 38%, respectively, while the expected earnings for workers in green2 decreased by 3% compared to 2013. Statistically insignificant earnings decreased in excluded and platinum firms by around 23% and 1%, respectively. However, when the groups were considered, it revealed that Saudis in localised firms could earn more in 2017 compared to 2013. In contrast, Saudis who worked in non-localised firms seemed to earn less than they did in 2013. In terms of non-Saudis, they were expected to earn more in 2013 if they were in a localised firm, while if they worked in a non-localised firm, they would earn more in 2017. When the colour effect was taken into consideration alongside the effect of the interaction variable, all non-Saudi workers were expected to earn less in 2013 unless they worked in yellow firms. They were expected to earn around 0.049% more (see Table 5-19). This means the response of non-Saudi workers to layoffs was accepting lower wages.[105] An earnings decrease could be deduced because of the quota system being applied for both groups. This result was not consistent with the literature, where a quota was expected to at least increase the target group's earnings (Burger et al., 2016). For example, Ransom and

[104] These categories were omitted automatically, and the comparison was on the matched categories only.
[105] This could be through moving to a localised firm if they were in non-localised firms or through internal negotiation.

Megdal (1993) found that women's wages increased even though the gender gap persisted. The economy and the workers' differences created this inconsistent result. Saudis were earning more, and they were the treated groups, unlike Blacks and women. Thus, workers responded to Nitaqat by accepting lower wages as though they were under unemployment risk.

Table 5-19: The earnings structure changes due to Nitaqat variables

	y-colour	Non-Saudi	Saudi
Red	−54%	−16%	−92%
Red small A	−49%	−11%	−87%
Yellow	−38%	0.5%	−76%
Green small A	−5%	−43%	33%
green2	−3%	−41%	35%
green3	3%	−35%	41%
Platinum	−1%	−39%	37%
Excluded	−30%	−68%	8%

Source: author's calculations from the pooled regression

In terms of the dummy variable of the year was insignificant, indicating no change in the earnings expectation of 2017 compared to 2013 when the worker attributes were similar. This was investigated in terms of mean values; the mean was higher in 2017 compared to 2013 in general because the number of Saudi observations was very small in the 2013 dataset, which allowed non-Saudi wages to pull the mean down. However, the mean of the subgroup decreased; for example, the Saudi mean wage in 2013 was higher than in 2017, as was the mean for non-Saudis. This could partly explain why the dummy coefficient was positive and insignificant (see Table 5-20 and Table 5-21).

Table 5-20: Summary statistic of log-salary 2013

	Obs.	Mean	Std. Dev.	Min	Max
Total	94,312	6.905299	0.674167	5.991465	11.51299
Saudi	1,943	8.697481	0.770318	7.313221	10.71442
Saudi localised	1,931	8.697033	0.768941	7.313221	10.71442
Saudi non-localised	12	8.769524	1.006792	7.824046	10.71442
non-Saudi	92,369	6.8676	0.618547	5.991465	11.51299
non-Saudi localised	89,950	6.869618	0.62088	5.991465	11.51299
non-Saudi non-localised	2,419	6.792555	0.519042	5.991465	10.34174

Table 5-21: Summary statistic of log-salary 2017

	Obs.	Mean	Std. Dev.	Min	Max
Total	4,371,262	7.41003	0.9792075	5.991465	11.51299
Saudi	1,899,248	8.305954	0.5794908	5.991465	11.45105
Saudi localised	1,704,446	8.332404	0.5959978	5.991465	11.45105
Saudi non-localised	194,802	8.074519	0.3261143	6.684612	10.71442
non-Saudi	2,472,014	6.721692	0.5890831	5.991465	11.51299
non-Saudi localised	2,060,007	6.738793	0.6048867	5.991465	11.51299
non-Saudi non-localised	412,007	6.636186	0.4938652	5.991465	11.51293

From another angle, earnings significantly increased in 2017 in all administrative areas apart from Najran, where earnings insignificantly decreased. Although some area increases involved a higher percentage compared to others, they all responded similarly. This was an expected result as the Nitaqat policy was equally applied to all areas. However, the variation in the percentage reflected how much the earnings structure changed according to the area. For example, in 2017, earnings were expected to increase by around 8% in Tabuk, compared to 21% in the Eastern Province. Furthermore, the relationship between earnings and firm size slightly changed; the earnings increased in giant firms by approximately 14%, which allowed the common trend presented in the 2017 estimations. Frankly, the worst change in the earnings structure was that the gender gap widened; it was around 4% in 2013, and it increased to 25%, which means the gap widened by 21% between the two years. In terms of age, it appears that the earnings function changed between the two years: the earnings decreased for young workers until the age of 21, then it started to increase. This trend continued until age 49. After this age, the earnings began to decrease. This indicated that the maximum and minimum points had an inflection point at 24 years.[106] However, this decrease did not change the total structure of the age–earning relationship by 2017; the expected wage increased at a growing rate after the age of 47 years in 2017, while it was after 44 years in 2013. The earnings decreased slightly compared to 2013 in the activities of repair of motor vehicles and wholesale and retail trade by 2% and 3%, respectively. The earnings increased in all other activities by a trivial amount, as well, except for other activities,

[106] The result was obtained by solving the first derivation equation in age. The equation was a quadratic: −0.000015 age2 + 0.0007146 age −0.0053831, whose solutions were when age equalled 21 and 49. The second derivative was 24, which indicated an inflection point between the minimum point at 21 and the maximum point at 49.

which rose by approximately 15%. It seems that Nitaqat reduced educational returns significantly, between 10% and 49% in all educational categories except architecture and planning, Colleges of Dentistry and secondary agricultural, which insignificantly decreased. Furthermore, among all types of occupations, service occupations increased significantly by only 2% compared to 2013. However, all other occupations were expected to earn negligibly less compared to 2013, apart from agriculture and fishing, which was lower by approximately 35%. This indicated that this activity was the most negatively affected by Nitaqat.

5.3 Oaxaca decompositions

It has been explained that the Saudi and non-Saudi wage gaps were associated with the employment gap. Therefore, Nitaqat was imposed, aiming to reduce this gap and enhance job creation for Saudis. According to the statistical descriptive, the number of employed Saudis slightly increased from when Nitaqat2 was imposed until 2016, then it sharply decreased and has remained steady at a lower number. This is consistent with Peck (2014) study, which revealed a slight increase in the micro-level data study that depended on 2014 data. This implied that Nitaqat success in engaging Saudis was less than in its success in decreasing in the wage gap, meaning that the increase in employment was little compared to the decrease in the wage gap. Following the expectation of Nitaqat, the increase in employment was associated with decreases in the wage gap during the application period: the gap was 1.82988 in 2013 and decreased to 1.58426 in 2017. Therefore, this section investigates this gap to understand the nature of the wage gap thoroughly and provide some recommendations for the authorities.

5.3.1 The Saudi and non-Saudi wage gap in general

Before the results are reported, it must be stated that splitting the indicators of non-Saudi nationalities and including them in the decomposition had important implications. **First**, for the omega or pooled option (in the Stata estimation), the earnings gap was explained fully, and the unexplained part had a trivial value; it had a p-value equal to one in the first option, and it was insignificant in the second option. These results could be because of the correlation between the subset and the index,[107] but they were only found when the interaction variable Saudi localised or non-Saudi localised was used; the correlation matrix indicated a high correlation between the index and the interaction variable by 0.9968 and −0.6586, respectively. Note that these specifications with suspicious results were not used. Fortunately, the correlation was 0.2464 for Saudi non-localised and 0.2828 for non-Saudi non-localised. More theoretical investigation is needed to assess the effect of the high correlation between the indicator and other variables on the explained part of the Oaxaca decomposition. **Second**, the pooled option proposed by Jann (2008a) would generate a lower explained part as the unexplained part is captured by the index value in the pooled model. This means the unexplained part would be higher when the heterogeneity was higher. This result supports our criticism of (Jann) as found in the literature review discussing. However, this criticism does not mean that the omega approach is better. Each approach would help in understanding the gap from a different perspective.

Explaining the wage gap between the Saudi and non-Saudi groups depended on the approach used. The gap was generated for different approaches: weighted group1, group2, pooled and omega.[108] The varied results relatively explained a reasonable amount of the gap when the weighted coefficient proposed by Oaxaca and Ransom (1988) was used, in line with the relative sample sizes of the two groups (see Table 5-22).

For the 2013 dataset, the explained, the results of all approaches indicated that Saudis had more statistically significant characteristics at the 1% significance level by 35%, 14%, 21%, and 15% for omega, pooled w (0), and w (1), respectively, compared to non-Saudis. A similar result was found for the 2017 dataset for omega and w (1). These results were unlike those for pooled and w (0), where the explained part suggested that non-Saudis had higher compositional characteristics. This did not match reality, as most non-Saudi workers have less education and come from low background countries.

[107] For correlation and decomposition results (see Appendix B, Section 10.4.1.2).

[108] Fortin's (2006) approach has been excluded as it generated a small biased result according to our manual calculations ((Abdullah et al., 2020). This approach was very useful for the gender pay gap. Our approach to improving the decomposition of the native–immigrant pay gap is not completed yet.

Moreover, this result contradicted most of the research findings where the gap was in favour of natives (Kee, 1995; Lehmer & Ludsteck, 2011; Longhi et al., 2012). However, for the two datasets, all approaches implied discrimination against non-Saudis as the unexplained part was negative. This discrimination result was usually found in the literature, except for Moroccans in the Netherlands, who had positive discrimination (Kee, 1995). Our results were consistent with the literature and other groups, such as Antilleans, Surinamese, and Turks, compared to the Netherlands (Dutch) in Kee's study.

The pooled results followed the pure discrimination approaches, w (0) and w (1), which means that although the pooled model was used, the result did not differ from the discriminatory approaches. Unlike omega approach that following the non-discriminatory assumption was able to weight the groups much better than the pooled approach, which explained around 35% in 2013. This result is consistent with findings by Lehmer and Ludsteck (2011). They found that less than 50% of the gap was explained, and they concluded that there was, therefore, substantial residual discrimination. However, in the 2017 results, around 71% of the gap was explained, which is inconsistent with that study but consistent with Hayfron (2002). He found the discrimination part to be around 32% or less in all weighted approaches.

Table 5-22: Explained and unexplained parts for Oaxaca decomposition.

		2013	2017
Wage gap		−1.829881***	−1.584262***
Number of observations		94,312	4,371,262
Omega	Explained	−0.633413***	−1.125414***
	Unexplained	−1.196468***	−0.4588485***
Pooled	Explained	−0.2584093***	0.0475387***
	Unexplained	−1.571471*	−1.631801***
W (0)	Explained	−0.3922593***	0.1653512***
	Unexplained	−1.437621***	−1.749613***
W (1)	Explained	−0.2666037***	−0.3212055***
	Unexplained	−1.563277***	−1.263057***

***significant at 1%. For details about the estimation according to the available data for each year, see Appendix B, Section 10.4.1.1)

Under the omega approach, the 2013 results included an explained part of around 12% of the existing gap because of higher Saudi attributes compared to non-Saudis, such as age and qualification (see Table 5-23). They also included 19% because of differences in job attributes, such as firm size and location, and only 3% because of the preferential effects of policies to Saudis. These were 40%, 22% and 9% in 2017. This means that adjusting the non-Saudi attributes to Saudi would increase non-Saudi wages from 34% (2013) to 71% (2017). 109 Consistent with (Longhi et al., 2012), the results indicated that the characteristics explained maybe over half of the gap. The policy coefficients at the explained part were negative for both years, which indicates that Nitaqat contributed directly to widening the gap for both groups by 3% and 9% respectively, according to Equation (5.44). This increase in the explaining percentage implies that the compositional differences where Saudi percentage increased with time. This could explain the higher average wage of Saudi.

Surprisingly, in 2017, the combined worker, job and policy attributes would yield positive discrimination towards non-Saudis, which contributed to narrowing the gap. In this case, the results followed the Moroccan unexplained gap (Kee, 1995). The results did not support the statement that immigrants could have unfavoured characteristics and supported the statement that discrimination could result from segregation, heterogeneity or different structures captured by the constant (Lehmer & Ludsteck, 2011). Accordingly, the results show that the structural differences led to the indirect effect of Nitaqat, i.e. narrowing the wage gap where Saudis were less preferred in some jobs, unlike in 2013. Thus, we can infer that Nitaqat2 had an indirect negative impact on the average wage of the Saudi compared to Nitaqat1. Saudis had higher starting wages, which is the main structural difference. Thus, the gap was still high even though non-Saudi worker attributes contributed to narrowing the gap. This precisely describes the Saudi labour market, where the preferred group has lower salaries. Accordingly, these structural differences are probably a core issue that policy needs to tackle. An argument could be raised to consider the w (1) approach as non-Saudis experienced around 1% negative discrimination. However, this argument is not strong because it suggested that the gap between the two groups was widening by 1%, which was not the case in the Saudi labour market.

[109] The explained part as a percentage of the total gap.

Table 5-23: the coefficients of Oaxaca decomposition omega approach.

			2013	2017
			Coefficient	**coefficient**
		Differences	−1.829881***	−1.584262***
Omega	**Explained**	**worker attributes**	**−0.22557*** **	**−0.62945*** **
		job attributes	**−0.34929*** **	**−0.35359*** **
		policy attributes	**−0.05856*** **	**−0.14237*** **
	Unexplained	**worker attributes**	**0.285913#**	**1.134706*** **
		job attributes	**−0.06956#**	**0.34928*** **
		policy attributes	**−0.02853#**	**0.092954*** **
		_cons	**−1.38429#**	**−2.03579*** **

*** significant at 1%, *at 10% #insignificant.

Attention must be paid to detail for the explained component of the wage gap under the omega approach: first, worker attributes. In 2013, it seems that age and female narrowed the gap between Saudis and non-Saudis by 1% and 0.01%, respectively; this was unlike 2017, where the gap widened between the two groups by 6% and 1%, respectively. **First**, age in 2013 was consistent with Hindu Indians in some quantiles, while in 2017, it was consistent with first- and second-generation Muslim Pakistanis compared to British workers (Longhi et al., 2012). The female result of 2013 was consistent with(Rand & Torm, 2012) and comparing developed country immigrants to Spain's native workers, while the 2017 result was consistent with comparing developing country immigrants in Spain (Simón et al., 2008). **Second**, non-Saudi female attributes contributed to reducing the gap between the two groups in 2013, while Saudi females contributed to widening the gap in 2017. It seems that there was a change in the gap structure in terms of qualifications between the two groups once Nitaqat was applied. Seemingly, Saudis earned more according to their qualifications by 13% and 6% in 2013 and 2017, respectively. However, in 2013, the gap narrowed if non-Saudis qualified for the Colleges of Medicine and the Health Institute, with a 1% significance level. This was similar for those who qualified with the Colleges of Medicine or Pharmacy or School of Law in 2017. This result was consistent with Longhi et al. (2012), who found that higher qualifications of British workers could explain the gap between them and first- and second-generation Muslim Indians and first-generation Pakistani Muslims. Saudi qualifications formed around 37% of the explained part or explained 13% of the wage gap in 2013, whereas these percentages were only 8.8% and 6%, respectively, in 2017. Although qualification formed a much lower percentage in

2017, education formed around 37% of the explained gap and 26% of the wage gap. Accordingly, the human capital theory was able to explain a substantial amount of the gap for both the 2013 and 2017 results (Becker, 2010; Collard, 1972). Second, regarding job attributes, in 2013, occupations formed approximately 50% of the explained gap. However, non-Saudi specialists narrowed the gap by around 6%, technical by 0.91% and sales by 0.31%; only the last had a 10% significance level.[110] This was unlike 2017, where occupation formed around 19% of the explained gap. As in 2013, specialists and technical contributed to reducing the gap by around 1.55% and 0.03%, respectively. The percentage variation between the two years implies that Saudis were more engaged in low occupations in 2017, which may have resulted in the gap reductions. The result was consistent with Lehmer and Ludsteck (2011), who found that the explained gap increased by 20–30% when controlling for occupation. This could imply that Saudi workers in higher attribute jobs had some similarities to native US workers compared to immigrants, as in (Smith & Fernandez, 2017). Surprisingly, economic activities formed only around 6.67% and 3.11% of the explained gap in 2017 and 2013, respectively. However, in 2013, activities assumed that the gap could be reduced by 1.16% according to non-Saudi job attributes as they worked in retail and wholesale; repairing, transport and storage; personal services and professional; scientific; and technical, which were statistically insignificant. However, approximately 0.28% of the gap in 2017 was reduced due to non-Saudi workers in mining, repairing and personal services. This result was consistent with Hayfron (2002), who found different earnings returns for natives and immigrants in different industries. Hayfron did not report the economic activities coefficient in his research even though it was included in the regression; this was similar to the study by Simón et al. (2008). Firm size was another job attribute factor; the gap existed in small and micro firms but was narrowed in medium and giant firms between the two groups, according to the 2013 results. This means wage inequality could be reduced when the firm size was larger. In contrast, in the 2017 results, firm size explained around 5.62% of the explained gap spread across all firm sizes apart from medium C firms. Surprisingly, in 2013, the gap was in favour of non-Saudis in Riyadh and the Eastern Province, although the negative gap in other

[110] Others were significant at the 1% level.

areas contributed to reducing the gap by roughly 1.54%. However, in 2017, Saudis were paid more in all areas, explaining 1.49% of the gap. This was consistent with the Hindu Indian effect (see Longhi et al. (2012).The policy formed around 9% and 12% of the explained gap in the two years, respectively.

The unexplained part provided a similar conclusion between the two datasets among the independent variables, except for activities and policy. However, apart from the region, those variables in the two datasets displayed positive coefficients, implying that non-Saudis could earn more: if Saudis lost the combination of attributes, they could earn less than non-Saudis, which is consistent with the Moroccan group compared to natives (Kee, 1995). For example, according to worker attributes, the positive coefficients of non-Saudis in all qualification categories reduced the gap by around 9.37% and 7.44% in 2013 and 2017, respectively, apart from some qualification categories that revealed a negative coefficient. In 2013, Saudis were the preferred group only if they held College of Medicine or Pharmacy qualifications, which had statistically insignificant effects, unlike in 2017, when both categories were significant, in addition to the School of Economics. Furthermore, non-Saudis were the (significantly) preferred group against Saudis in all education categories, which contributed to narrowing the gap. The preferences for non-Saudis according to their advancing ages were approximately 6.05% and 32.48% for 2013 and 2017, respectively, while females contributed by more modestly reducing the gap by 0.2% in 2013 and 5.32% in 2017. This implies that employers preferred an older non-Saudi to an otherwise identical Saudi, assuming that the immigrant workers were more experienced. Similarly, the non-Saudi females were preferred to Saudi females as they showed unlimited supply compared to the lower Saudi female labour market participation. The results did not support the double negative assumption for females in Saudi Arabia, unlike Boyd (1984); Hayfron (2002) findings. Turning to job attributes, regions contributed to widening the gap between the two groups by around 8.62% and 4.30% in 2013 and 2017, respectively. By contrast, occupations contributed to narrowing the gap between the groups by 6% in 2013 and 13% in 2017. Accordingly, Saudi workers experienced the glass ceiling issue in both years, and this issue was higher in 2017 by around 7%. However, both activities and policy implied discrimination against non-Saudis in 2013, unlike in 2017, when the discrimination was reversed towards Saudis. Regarding the policy in 2013, non-Saudis were preferred according to the interaction terms, while they significantly experienced discrimination in firms from the red, small red A and green3 bands. Otherwise, discrimination was insignificant. However, in 2017, Saudis experienced discrimination in most of the policy categories apart from red and yellow bands, as well as the excluded band. This discrimination against Saudis reduced the earnings gap between the two groups. Thus, we could say that Nitaqat indirectly narrowing the gap.

5.3.2 Saudi and non-Saudi according to their origin (2013)

The 2013 dataset is exploited in this section to explore the earnings gap differences according to worker origin. The workers were divided into six groups: European, American and Canadian, Saudi, Asian, African and Arabic. The first and second groups were merged into high background countries because they had a small number of observations – 85 in total: 29 for the first group and 56 for the second, which hindered the performance of the decomposition. For each origin, the earnings gap was decomposed (against Saudi) to understand if worker origin mattered. Table 5-24 demonstrates the decomposition results for each of these origin groups. The workers from high background countries had a positive insignificant gap. This means that Saudis earned less than those groups on average unlike Arabic, African, and Asians.

Table 5-24: Saudi and non-Saudi wage gaps according to their origins

		High background countries	**African**	**Arabic**	**Asian**
Omega	Differences	0.0830914#	−1.72886***	−1.42015***	−1.945074***
	Explained	0.32152***	−0.953263***	−0.6507155***	−0.8372177***
	Unexplained	−0.23843***	−0.775597***	−0.769435***	−1.107857***
Pooled	Explained	0.484104***	−0.55551***	−0.048245***	−0.3298121***
	Unexplained	−0.40101***	−1.173349***	−1.371905***	−1.615262***
W (0)	Explained	0.41892***	−0.567367***	−0.2831576***	−0.4228967***
	Unexplained	−0.33583***	−1.161493***	−1.136993***	−1.522178***
W (1)	Explained	−0.5589#	0.0529138#	−0.0168834#	−0.331946***
	Unexplained	0.64199#	−1.781774***	−1.403267***	−1.613128***
Number of observations		85	133	19,887	72,264

***significant at 1%, #insignificant. For more detailed results (see Appendix B, Section 10.4.2).

For workers from high background countries, there was a statistically insignificant log-earnings gap of around 8%. Employers discriminated in favour of Saudis (via a negative unexplained effect), narrowing this gap, which had an insignificant positive intercept, meaning that the reservation wage could be higher for the high background country group compared to Saudis. However, w (1) had insignificant explained and unexplained parts, suggesting (statistically very weakly) that Saudis had higher attributes, but they were treated as an unfavoured group. For the different conclusions according to the w (1) approach, qualification was the key variable that pushed the gap positively in favour of non-Saudis, which emphasised that human capital could explain a reasonable amount of the earnings gap. However, according to the omega approach, Saudi workers had higher attributes in all variables on average, apart from the occupation and qualifications, where non-Saudis were higher than Saudis in some qualifications, such as the College of Education, science, medicine, engineering, architecture and planning, computer, medical and applied sciences; only engineering was significant at the 1% significance level; otherwise, they were insignificant. In terms of occupations, non-Saudi origins were superior compared to Saudi origins in all occupations apart from clerical and sales occupations; the latter was not significant. Moreover, the coefficients of qualification, colour, occupation, and activities contributed to narrowing the earnings gap between Saudis and those who came from (non-Saudi) high background countries, unlike other variables. For example, the gap widened in some regions, such as Al-Jouf, Qassim, the Northern Borders, the Eastern Province and Riyadh. These results were somewhat expected since energy companies are in some of those areas.

Among the three disadvantaged non-Saudi groups, Arabic origin formed the lowest earning gap at a logarithmic level, with around 142%, followed by Africans with 173%, and then by Asians with 195%. According to the omega approach, in ordered approximately 46%, 55% and 43% of the gap was explained, of which roughly 14% and 12% were due to higher Saudi occupations and qualifications for Arabic. However, this percentage was 22% and 7.5% respectively, for African while it was 20% and 16%, respectively for Asian. However, non-Saudis from Arabic and African origin had a better endowment if they held a degree in medicine or engineering or from health institutions, and if it were not for these qualifications, the gap could have been bigger. Unlike Asian, the gap could negligibly increase if they were qualified with Colleges of Medicine, agriculture and veterinary, or the Health Institute or if they worked in a specialist or technical occupation. Saudis were superior in two occupations: manager and clerical occupations, which was enough to widen the gap between Saudis and non-Saudi workers. Furthermore, the gap existed in all regions; Saudis had, on an average, higher attributes by around 4%, 7.5%, 1% for Arabic, African and Asian in ordered. Although the gap was explained by roughly 3% on average for firms' colour for Arabic, 5% for African, and 3% for Asian. For Arabic, the gap was narrowed according to the non-Saudis' higher endowment in most firm colour bands, apart from platinum and green3. Moreover, 7%, 5%, and 1% of the gap was explained through firm size for Arabica, African, and Asian. Saudis had higher attributes in micro and small firms for all origin. Furthermore, 3% 2%, 1% of the gap was related to economic activity differences for these groups in ordered.

However, the unexplained 54% of the earnings gap implies that Arabic workers experienced discrimination; for Africans, this percentage is 45% and for Asians 44%. The Arabic group considered the preferred workers, according to all the variables apart from regions. Although the average of discrimination according to regions was 3%, Arabic workers seemed to be preferred in Al-Baha, Madinah and Jazan. Asian workers confronted discrimination in similar regions on an average of 7%. Unlike, employers preferred Africans on an average of 5%, while they discriminated against Africans in Tabuk, Najran, Hail and Asir. Generally, non-Saudis were the preferred workers if they qualified, with some exceptions. Saudis were preferred compared to Arabic people when Saudis had education related to pharmacy or medicine. The preferred qualifications were administration, science, languages, and translation for African workers, while Saudis were only preferred as pharmacists compared to African workers. Saudis with these qualifications earned more than non-Saudi workers. Moreover, Africans were the preferred group considering these occupations by approximately 5% on average.

Asian experience discrimination according to activities on average by 3% apart from workers in wholesale and retail trade, transportation and storage, and education, human health and social workers. However, Arabic and African seem to have been preferred according to occupation on average by 10% and 0.22%, respectively. Although Arabic workers preferred in all activities, employers discriminated against Africans in three activities: accommodation and food service activity, mining and quarrying and manufacturing, and other activities. Employers discriminated in favour of Saudis on average firms' colour by 3% or less than for all three origin groups. Arabic seems preferred groups in non-localized firms while African seems the preferred groups in localized firms.

For all origin, the intercept formed the highest percentage of the unexplained gap between the two groups. It was significant at 1% and 10% for Arabic and African, respectively while it was insignificant for Asian. This implying the differences in the starting wage which confirmed that the gap was mainly constructed by the differences in worker supply, not employers' discrimination against non-Saudi groups. The results from all groups indicate that the labour structure plays a big role in the wage gap. Thus, considering the structure differences between the two groups when the policy updated is recommended.

When the high background origins were decomposed, the omega approach explained a lower proportion of the gap compared to the pooled approach, which implies that omega approach does not transferred part of the unexplained gap to the explained part (Jann, 2008a). We went behind each group to justify the reason statistically and found that those groups had a constant variance according to the heteroscedasticity test of Breusch–Pagan/Cook–Weisberg; however, the results indicated that Asian, Arabic and African groups had a non-constant variance at a 1% significance level, apart from African (10% level). This supports our criticism to the pooled approach proposed by Jann. However, this needs to be proven theoretically.[111]

5.3.3 The wage gap between Saudis and non-Saudis by firm localisation status

It is interesting to know how the gap looks for firms from each localisation status: localised and non-localised. Thus, decompositions were performed on each status separately for each year. In 2013, the significant gap between Saudis and non-Saudis was slightly bigger in non-localised firms. Additionally, over 65% of the gap was unexplained in all decomposition approaches for both localised and non-localised firms

[111] For the detailed estimation and heteroscedasticity test (see Appendix B, Section 10.4.2.5).

(see Table 5-25). In contrast, for 2017, the gap was smaller in the non-localised firms (see Table 5-26).

Table 5-25: Saudi and non-Saudi wage gaps according to the firm status in 2013

	Localised		Non-localised	
	Explained	Unexplained	explained	unexplained
Omega	−0.5981***	−1.22932***	−0.25164*	−1.72532***
Pooled	−0.22687***	−1.6005***	−0.1996#	−1.77736***
W (0)	−0.3392***	−1.48821***	−0.22466#	−1.75231***
W (1)	−0.24136***	−1.58605***	−0.04203#	−1.9349***

***significant at a 1% level, #insignificant, *significant at 10%. For other approaches' details (see Appendix B, section 10.4.3, 2013).

Table 5-26: Comparing the wage gap between localised and non-localised firms (2017).

	Localised		Non-localised	
	explained	Unexplained	explained	unexplained
Omega	−1.07156***	−0.52206***	−1.13249***	0.06866***
Pooled	0.089***	−1.68261***	−0.30585***	−1.507***
W(0)	0.20145***	−1.79506***	−0.0613***	−0.34035***
W (1)	−0.29149***	−1.30212***	−1.37701***	−1.09798****

*** significant at 1% level. For other approaches' details (see Appendix B, section 10.4.3, 2017).

Since the unexplained gap was higher in all approaches, the two groups on the two firm types could have structures differences. At localised firms in 2013, Saudis could earn relatively less than non-Saudis if they were non-Saudi members due to worker attributes, while Saudis could earn relatively more compared non-Saudis due to job attributes. This implies that there were segregation issues where Saudi workers were preferred for high-quality jobs compared to non-Saudis. In both firm statuses 2017, according to omega approach, there was evidence of discrimination against Saudis on average, based on worker and job attributes, where Saudis could earn relatively less, which contributed to narrowing the earnings gap between the two groups; however, the unexplained gap indicated discrimination against non-Saudis. This implies a glass ceiling issue in both firm statuses in 2017. Non-Saudis experienced a high risk of deportation in the bottom wage categories.

The starting wage (via the negative constant-coefficient), which was responsible for widening the gap between Saudis and non-Saudis in both years, formed most of the unexplained gap. In 2013, the constant in non-localised firms was lower than in localised firms. This implies that Saudis asked for higher wages when a firm was required to meet the percentage even though they were safe from a layoff, which could imply a reversal of the hedonic wage theory (Theodossiou & Vasileiou, 2007; Pinheiro and Visschers, 2015). Similarly, non-Saudi could accept lower wages when moving to a more secure job, unlike Saudis. However, in 2017 Saudi workers in localised who were under indirect risk firms could earn more than Saudi who were safe in non-localized firms.[112] This because the unemployment risk for Saudis took a natural risk thread after applying for the SANED programme. Assuming positive relationship between risk and wages is the core of the hedonic wage theory (Hutchens, 1983).

5.3.4 The wage gap between workers by firm localisation status

As Nitaqat classified firms as localised or non-localised, this allowed an examination of the wage differential between the workers in each type of firm, following the method previously used for the wage differential between union and non-union workers (Oaxaca & Ransom, 1988). This approach provided some implications for understanding the effect of Nitaqat on wages for Saudis and non-Saudis due to the risk of layoff.

First, the full sample was taken into consideration across 2013 and 2017. The results indicated that workers in localised firms earned more than those in non-localised firms by 0.10571 and 0.36241 based on the log-earnings in 2013 and 2017, respectively (see Table 5-27). This could be because non-localised firms could not pay as much as localised firms. Of this gap, the explained part was (statistically significant) near 99% in 2013 and 88% in 2017, according to the omega approach.[113] This could imply that localised firms reduced the workers or job quality and increased the Saudi share in 2017 compared to 2013. The gap was explained by around 9% due to worker attributes, 65% due to job attributes and 24% due to employing Saudis in 2013, while they were rather different in 2017, with only 1% due to worker attributes, 26% due to job attributes and 61% due to employing Saudis. These results imply that localised firms rewarded higher wages compared to non-localised firms, not because they employed workers with higher

[112] Saudis were under indirect risk through the replacements between Saudis themselves.
[113] The result was similar for the pooled and w (0) approaches, while w (1) explained a lower percentage of around 82% in 2013 and 63% in 2017. Moreover, all four approaches had a similar conclusion, regardless of the percentage variation.

attributes but because they employed a greater percentage of Saudi workers. Although job attributes explained a good proportion of the gap between the two types of firm status, it decreased in 2017 compared to 2013, which implies that employing Saudis was then associated with the low job and qualification attributes. This assumed a limitation of the quota system's impact (Bertrand et al., 2014; 2019). Additionally, this implies that Saudis in 2017 had a lower job and worker attributes compared to 2013, which is consistent with other quota system findings that the quota provided lower job quality (Holzer & Neumark, 1999).

From another angle, the unexplained part of 2013 revealed that workers in non-localised firms received less than their peers in localised firms according to their attributes and job attributes. However, Saudis in non-localised firms earned more than Saudis in localised firms; this is unlike the assumptions of the hedonic theory. This resulted from the expected indirect layoff risk for Saudis in localised firms, which is consistent with the literature's finding that secure workers would have higher wages compared to insecure workers (Theodossiou & Vasileiou, 2007; Pinheiro & Visschers, 2015). In 2017, the result was slightly different; the unexplained part implied that Saudis who worked in localised firms could earn relatively less if they were working in non-localised firms, by roughly 27% (see Table 5-27).

Table 5-27: Wage gap between firm status according to the omega approach

		2013		2017	
		Coefficient	Percentage	Coefficient	Percentage
explained	Differences	−0.10571***	100%	−0.36241	100%
	Total	−0.10452***	98.87135	−0.31994	88.28049
	worker	−0.00961***	9.08757	−0.00404	1.113909
	job	−0.06915***	65.41274	−0.09643	26.60707
	Saudi	−0.02576	24.37104	−0.21947	60.55951
Unexplained	total	−0.00119	1.14154	−0.04247	11.71954
	worker	−0.67982	643.0953	−0.56198	155.0671
	job	−0.31051***	293.7359	−0.04237	11.69051
	Saudi	0.00165#	−1.55699	−0.10031	27.67912
	_cons	0.98749#	−934.146	0.66218	−182.717
Number of observations		94,312		4,371,262	

The percentage calculated from the total gap. *** significant at the 1% level, # insignificant, * significant at 10%. For other approaches' details (see Appendix B, section 10.4.4, A and D).

Indeed, this result indicates that Saudi workers under indirect risk received higher wages than those Saudis in safer job environments, which means the hedonic wage theories were applied. This supported the above findings. Other variable effects were like the 2013 results where they contributed to widening the gap between workers by firm localisation status, especially worker attributes. For Saudi, Table 5-28 shows an insignificant small gap of around 7% in favour of non-localised firms in 2013. This gap insignificantly explained by the higher worker and job attributes for Saudi workers in non-localised firms with approximately 10% and 166%, explaining around 176% of the log-earnings gap.[114] By contrast in 2017, a significant gap of around 26% was detected in favour of Saudi workers in localised firms which explained around 33% and 49%, due to worker attributes and job attributes, respectively.[115] These results supported the view that human capital could explain part of the gap (Becker, 2010; Collard, 1972). Similarly, non-Saudi had a significant gap at around 7% in 2013 and 10% in 2017 (see Table 5-29). In 2013, around 95% of non-Saudi earnings gap explained through worker characteristics and job attributes at around 12%, and 84%, respectively. This result implies that non-Saudi workers in non-localised firms in 2013 were heavily found in unskilled jobs. For 2017, 19% of the gap was explained through worker attributes and 48% through job attributes.

Table 5-28: Saudi wage gap according to firm status using the omega approach.

		2013		2017	
		Coefficient	Percentage	Coefficient	Percentage
	Differences	0.0724909#	100	−0.257885***	100
explained	Total	0.1275784#	−176%	−0.2107936***	82%
	Worker	0.00711#	−10%	−0.0840444***	33%
	Job	0.12047#	−166%	−0.1267492***	49%
Unexplained	Total	0.2000693#	276%	−0.0470914***	18%
	Worker	−135.4777***	−186,889%	0.5846929***	−227%
	Job	−2.97241***	−4,100%	−0.0532815***	21%
	_cons	138.6502***	191,266%	−0.5785028***	224%
Number of observations		1,943		1,899,248	

The percentage calculated from the total gap. *** significant at the 1% level, #insignificant, * significant at 10%. For other approaches' details (see Appendix B, section 10.4.4, B and E).

[114] There were only 12 observations for non-localised firms, which could make the estimation insignificant, unlike for 2017, which had 194,802 Saudis in non-localised firms.
[115] The sample size for 2017 was so large that its standard errors tended to be very small. Thus, it was more likely to be statistically significant compared to the 2013 data.

Table 5-29: non-Saudi wage gap according to firm status using the omega approach.

		2013		2017	
		Coefficient	Percentage	Coefficient	Percentage
explained	Differences	−0.07706***	100	−0.10261***	100
	Total	−0.07368***	96%	−0.06875***	67%
	Worker	−0.00933***	12%	−0.01937***	19%
	Job	−0.06438***	84%	−0.04938***	48%
Unexplained	Total	−0.00338#	4%	−0.03386***	33%
	Worker	−0.25546#	331%	0.05533#	−54%
	Job	−0.32227***	418%	−0.02648***	26%
	_cons	0.57435#	−745%	−0.06271#	61%
Number of observations		92,369		2,472,014	

The percentage calculated from the total gap. *** significant at 1% level, # insignificant, * significant at 10%. For other approaches' details (see Appendix B, section 10.4.4, C and F).

The positive unexplained part for Saudi in 2013 revealed that Saudi workers in localised firms could earn relatively less than workers in non-localised firms due to the starting wage, while they could earn relatively more due to their attributes and their job attributes. The lower starting wage (via intercept) in localised firms could imply that these firms depended on dummy Saudisation or temporary Saudisation for those who received 1,500SR or less to achieve the percentage required, along with real Saudis, which was then reflected in lower mean earnings for these types of firms, unlike 2017. For non-Saudi, in 2017, the log-earnings gap widened further because the localised workers asked for a higher starting wage compared to non-localized workers while they asked lower wage in 2013. For both groups, log-salary mean decreased between the two years in both firm types; non-localised firms had the largest decrease. Thus, the gap increased slightly between the two years in favour of workers in localised firms which implies that non-Saudis in non-localised firms, who experienced deportation risk, earned less than those with a lower risk of deportation. Thus, the final effect of the Nitaqat was to promote the reverse of the hedonic wage theory effect, unlike Saudi. In summary, non-Saudis in non-localised firms were affected negatively in both years compared to those in localised firms, which could imply a double negative effect from being non-Saudi in a non-localised firm. This term was used frequently to describe the sub-group effect, such as for female immigrants (Boyd, 1984; Hayfron, 2002).

From the description above, we deduce that there are no structural differences among worker status in both firms' localization status, where workers from both groups mixed smoothly. However, workers in localised firms were rewarded with a better salary compared to their peers in non-localised firms, which indicates the importance of searching behind the firms' ability for input substitution.

5.3.5 Consumption as a new explanatory variable

The wage gap between Saudis and non-Saudis is now examined, following the modern theory on the economics of immigration. The hypothesis is that the gap between the two groups can be explained through consumption (Massey et al., 1993). According to the framework, the wage could be determined by consumption habits for workers as the wage increased when autonomous consumption increased, and wage decreased when the consumption function's slope was steeper. Unfortunately, the datasets were poor in this respect. Thus, some assumptions were set for the slope of the consumption function.[116] According to Keynes (as noted in texts for introductory undergraduate economics, like McTaggart et al., 2015), consumption falls when wage increases, which implies that consumption is higher for lower-wage individuals compared to the high-wage individuals.[117] Considering that, a higher φ_1 (0.9) was applied for workers who received less than 4,000SR per month compared to those who received more than this amount (around 0.8), regardless of whether or not they were Saudi.[118] After the induced consumption was generated, it was added to the decomposition to see how much the induced consumption could explain the gap between the groups. The decomposition results supported the framework's hypothesis that the gap between the two groups is underpinned by consumption behaviour. This result implies that the modern theory of immigration could provide an explanation of the wage gap between Saudis and non-Saudis (see Table 5-30 and Table *5-31*). Indeed, this needs to be examined in other countries to clarify the explanatory validity of this theory among several economies. For 2013, 80% of the wage gap was explained according to the omega approach, 69% for the pooled approach, 35% when weighted to non-Saudi and 102% when weighted to

---

116 According to our framework, the autonomous consumption equalled the consumption slope multiplied by the wage. Moreover, consumption would equal double the autonomous consumption value. $W_i^* = \frac{\varphi_0}{\varphi_1}$; $\varphi_1 W_i^* = \varphi_0$; this leads to C = $2\varphi_0$ which indicates that consumption should decrease, and wages would decrease under Nitaqat.
117 The estimation of the consumption function was not the aim of the research. We used a simple possible value of consumption that we could achieve to include consumption in the decomposition.
118 3,000 SR is the baseline of Nitaqat as the minimum wage to be considered a quota. We assumed those workers had a high marginal propensity to consume.

Saudi. Similarly, for 2017, the percentages of the wage gap that were explained were around 85%, 33%, 21% and 75% according to the four approaches (respectively).

Table 5-30: Consumption as an explanatory variable for 2013

Differences −1.829881***		2013			
		Omega	Pooled	W (0)	W (1)
explained	Total	−1.46214***	−1.26125	−0.64403***	−1.85854***
	Worker	−0.09831***	−0.03663***	−0.04654***	−0.01543***
	Job	−0.04955***	−0.016583*	−0.09473***	−0.05714***
	Nitaqat	−0.02111***	−0.0166***	−0.02087***	−0.01372***
	consumption	−1.29317***	−1.19144***	−0.48189***	−1.77224***
Unexplained	Total	−0.36774***	−0.56863***	−1.18585***	0.02866
	worker	−0.97733*	−1.03901**	−1.0291**	−1.06021**
	Job	−0.11274***	−0.14571***	−0.06756***	−0.10515***
	Nitaqat	0.00964	0.00513*	0.009397	0.00225
	consumption	1.05214***	0.95041***	0.24086***	1.53121***
	cons	−0.33945	−0.33945	−0.33945	−0.33945

Significance level: *** 1%, ** 5%, * 10%, otherwise insignificant

Table 5-31: Consumption as an explanatory variable for 2017

Differences −1.584263 ***		2017			
		Omega	Pooled	W (0)	W (1)
Explained	Total	−1.34319***	−0.52505***	−0.33665***	−1.19165***
	worker	−0.53601***	−0.01582***	0.06497***	−0.13434***
	Job	−0.16276***	−0.06680***	−0.03273***	−0.07121***
	Nitaqat	−0.07115***	−0.01859***	−0.01244***	−0.01106***
	consumption	−0.57328***	−0.42383***	−0.35645***	−0.97504***
Unexplained	Total	−0.24107***	−1.05922***	−1.24761***	−0.39262***
	worker	−0.142336***	−0.6625***	−0.74331***	−0.54401***
	Job	0.1764742***	0.08052***	0.04645***	0.08493***
	Nitaqat	0.0563416***	0.00379***	−0.00237***	−0.00375***
	consumption	0.3936936***	0.24425***	0.17686***	0.79546***
	cons	−0.725245***	−0.72524***	−0.72524***	−0.72524***

significance level if *** 1%, ** 5%, * 10%, otherwise insignificant.

The explained gap improved significantly by adding the consumption to the decompositions for all approaches, and it formed a high percentage of this part. In 2013, according to the omega approach, around 71% of the wage gap was explained by Saudis' higher consumption, 5% by higher worker attributes, 3% by higher job attributes and 1% by the Nitaqat percentage. In 2017, these percentages were 36%, 34%, 10% and 4% respectively. Although the consumption explained substantially the gap, this did not change the fact that Nitaqat directly affected the widening of the gap (see Equation 5.44), while the indirect effect indirectly narrowed this gap. However, the results suggested that Saudis consumed more than non-Saudis, which pushed them to ask for higher salaries to maintain their needs. In another word, non-Saudi could accept wages equalizing subsistence wage for Saudi. This result was consistent with the assumption that workers would ask for different salaries according to their origin, which created several segments within the Saudi labour market (with workers from low background countries at the low end of the consumption spectrum and thus willing to accept lower wages than the higher-consumption Saudi workers).

Although adding consumption improved the explained gap, around 19% and 46% were unexplained in 2013 and 2017, respectively. It worth noticing that worker attributes contributed to a widening of the gap between the two groups in both datasets, although it was higher in 2017. This implies that employers tended to discriminate in favour of educated Saudi workers. This was unlike job attributes, on which employers apparently discriminated in favour of Saudis in 2013, while they discriminated against them in 2017. Accordingly, in 2017, Saudi workers seemed to experience the glass ceiling issue, while non-Saudis experienced a high deportation risk in the bottom wage categories. This confirms the previous results that Saudis engaged in lower occupations in 2017, which was reflected in lower average wage for Saudis that year (see heading 5.3.3 aboveTable 5-23). Moreover, the 2017 indication that policy creates discrimination against Saudis implies that removing Nitaqat could increase Saudis' average wages by improving their job attributes to reach the level before Nitaqat. In other words, it seems that Nitaqat has created a glass ceiling for Saudis because employers need to achieve the highest percentage (of Saudi workers) with the lowest possible cost. The starting wage (via intercept) widen the gap in lower power compared to the decomposition excluded consumption which could be attributed to omitted variables, such as the number of dependents, family financial background (wealth), working hours and the opportunity cost of unemployment.

In general, the results indicated that heterogeneity in the policy did not explain a substantial amount of the gap compared to the heterogeneity in the individual labour supply due to differences in consumption. Mahdi (2005) used some variables to indicate policy heterogeneity, such as worker mobility and security. However, those were not included in the decomposition because they were equal to the indicator. For example, the movement was a dummy variable equal to one if a worker was free to move between firms and zero if not.

For a sensitivity check, several $\varphi_1$values were assumed. First, it was assumed that Saudis had a higher consumption slope, if non-Saudis tend to appreciate saving, especially for remittance (to family members in their native country). Second, it was assumed than non-Saudis had a higher consumption rate as they had lower wages. Third, it was assumed that both groups had a similar slope for their respective consumption functions. Finally, it was assumed that Saudis and non-Saudis had a similar consumption function slope if they earned above the minimum wage of the quota (3,000SR), but if they earned less than that limit, Saudis would have a higher consumption function slope than non-Saudis. Among all those scenarios, the wage gap explained 71–83% in 2013 according to omega approach, while it was 80–85% in 2017. Moreover, the explained part seemed higher when the consumption slope for Saudi workers was higher than that for non-Saudi workers. This could indicate that the first scenario was more consistent with the nature of the workers who sought to maximise their utility by providing remittances to their family. Furthermore, the explained gap increased when the gap in the consumption slope increased. For example, when Saudis had a 0.9 slope and non-Saudis had 0.8, around 81% of the wage gap was explained, compared to 83% when the non-Saudi slope was 0.7.[119] Thus, the structure of the labour market and nationalities' behaviour could be a key point in Saudi Arabia.

5.4 Conclusion

The empirical results indicated that Nitaqat successfully reduced the gap between the two groups. This reduction came from decreasing both groups' average wages and decreasing the Saudi average wage by double. This supported the hypothesis that the quota system (Nitaqat) would reduce wages through the layoff risk; the relationship between the risk of layoff and wages seemed to be negative unless unemployment benefits existed. This implies that non-Saudis in non-localised firms experienced a double negative effect by Nitaqat. This response varied concerning the origin, which could implicate the importance of nationality in the labour market. Moreover, the modern research theory provided a new explanatory variable – consumption – to explain the gap between Saudis and non-Saudis. The results indicated that firms resisted Nitaqat

[119] For details (see Appendix B, Section 10.4.5).

by redistributing workers, creating a glass ceiling issue for Saudis and increasing the deportation risk at the bottom of the non-Saudi workers. This could affect Saudi welfare from two angles. The **first** angle is the decrease in the medium society layer among Saudis. The **second** angle is the deportation risk of non-Saudis in the bottom categories, which would increase labour costs and, thus, prices. Moreover, it seems that there was a substantial gap resulting from the different earnings structure of both groups, represented by the intercept.

Chapter 6 Inverse Probability Weighting IPW

6.1 Introduction

Handling missing data is a debatable issue. Some researchers argue that using the complete cases (CC) or (unweighted) method could generate unbiased result (consistent with other approaches) if it follows the MCAR or MAR mechanism (Enders, 2010; VanGeest et al., 2017). Unlike others expected, the CC method generates biased results as it ignores missing data. Therefore, they prefer using more advanced strategies such as Inverse probability weighting (IPW) or imputation methods. The former strategy is more commonly used to correct the bias generated. The CC method could be consistent in two cases: first, a representative subsample is given by the CC. Second, where the CC are a random sample of the population (Seaman & White, 2013). However, multiple imputations are more efficient than IPW, but they could cause extra bias. Inefficiency is the downside of IPW since, just like the CC, it possibly ignores useful data on the missingness. Moreover, it can be subject to a huge bias in finite-samples (Seaman & Vansteelandt, 2018).

Given the missing data discussed, Table 4.1 shows a possibility of bias results that could be generated from using CC for both OLS and Oaxaca decomposition. In this respect, we use the IPW method to address this issue. This method is applied in two stages: **first**, by performing logistic regression on the missing binary variable where the observed values are giving 1 and missing 0. There are limited variables which control our model choice. Thus, we included all variables that could possibly predict the missingness. However, the missingness in non-Saudi sub-sample initially comes from the wage registration policy which cannot be measured in this situation. For robustness checks, I used two model specifications to estimate the logistic regression, aiming to estimate the inverse probabilities (see Table 6-1). I chose these from the available variables, which determined the CC. Accordingly, these variables can contribute to model the probability of CC. None of these variables contributed to produce the probability of CC equal to 1. Moreover, these variables successfully contributed to calculating the missingness probability, where the mean value of the probability was 0.33863 for weight A and 0.35486 for weight B. We excluded the polynomial variable which could lead to an increase or decrease in the efficiency of the IPW outcome. **Second** stage is using the inverse of predicted probability (1/pr) to reweight the sample to then estimating the equation.

Table 6-1: summary of the inverse probability weight (1/pr).

	Observation	Mean	Std. Dev.	Min	Max
Weight A	9,584,165	3.147262	2.082781	1.000775	22.98084
Weight B	9,584,165	2.956895	1.870644	1.000957	63.94699

WeightA: logistic r age age2 age3 female ib5.colour ib5.SIZE1 ib11.regions2 firm-age firm_age2 ib3.Activities ib9.occupation. **WeightB** logistic r age female ib5.colour ib5.SIZE1 ib11.regions2 firm-age ib3.Activities ib9.occupation.

6.2 Ordinary Least Square (OLS)

Starting with the OLS estimation, we found in relative term (weighting coefficient - unweighted coefficient)/absolute unweighted coefficient *100) that unweighted OLS slightly overestimated compared to IPW weighted on average around 0.15% in weightA and 0.10% in weightB, when the Saudi dummy variable was excluded from the model. Similarly, when the Saudi indicator was included, CC was overestimate by 1.32 in weightA while it IPW was overestimate when weightB considered by 0.06%. However, most of the coefficients did not differ in terms of the sign and standard errors, although with some exceptions. Table 6-2 reported two predictors were dissimilar when Saudi dummy variable exclude from 79 predictors included on the model. Professional activity significance level was consistence in all regressions while coefficients was positive when (unweighted) used unlike weightA and weightB which were negative. Unlike, secondary trading was inconsistent for both coefficients and significance level.

Table 6-2: dissimilar estimates when Saudi excluded (sign and statistical significance).

	Unweighted	WeightA	WeightB
Secondary trading	0.005379 (1.27)	-0.02413 (-4.38)	-0.01045 (-2.13)
Professional	0.018864 (8.95)	-0.05191 (-22.23)	-0.022498 (-10.66)
Number of observations 4,371,262			
R-squared	0.5891	0.6045	0.6129
Root MSE	0.62769	0.60023	0.58771

The t-ratio between brackets.

Table 6-3 below shows the dissimilarity between CC and IPW when Saudi dummy variable included. We found that professional activities were negative and significant in both weights. High school of qualification was negative in unweighted while it was positive in IPW, these coefficients were significant in all regressions. Unlike, the consistent coefficients for collage of agriculture were insignificant in weightB compared to using CC and weightA. Concerning for small green category, we found only weightA generated positive insignificant coefficients compared to unweighted regression while weightB had negative insignificant coefficients. However, workers in Riyadh and eastern province significantly earn lower than those in Makkah (the base category) when weightA, and weightB considered.

It is noticeable that unweighted regression had the highest root MSE when compared to other weighted regressions when Saudi was excluded or included (see Table 6-2 and Table 6-3). Moreover, unweighted regression was inconsistent with other weight regressions between approximately 3% and 6% of the total coefficients in both excluded and included Saudi dummy variable.[120] Finding differences between the results is expected. However, we expected a relatively consistent outcome. Although weighted regression seems to generate better Statistically, we need rethinking on the weighted regressions that are generating unexpected relationships, such as the negative sign on Riyadh, East Provence compared to Makkah (the reference category), similarly professional activities compared to construction (the reference category).

Table 6-3: summary of dissimilar estimates when Saudi dummy included.

	Unweighted	WeightA	WeightB
High school	-0.009959 (-13.81)	0.0030261 (3.9)	0.0020866 (2.91)
Collage of Agriculture	0.0192535 (2.94)	-	0.0064681 (0.95)
green small	-0.0038797 (-3.74)	0.0000813 (0.07)	-0.000329 (-0.3)
Riyadh	0.0010838 (1.84)	-0.005594 (-9.15)	-0.006799 (-11.8)
Eastern province	0.0038593 (5.67)	-0.0121045 (-17.91)	-0.009157 (-14.3)
Professional	0.0006793 (0.43)	-0.067086 (-42.58)	-0.048893 (-33.04)
Number of observations 4,371,262			
R-squared	0.781	0.7967	0.7993
Root MSE	0.45827	0.43033	0.42321

The t-ratio between brackets.

[120] This is counted by dividing number of inconsistent categories on total categories multiplied by100.

6.3 Saudi and non-Saudi wage gap in general

To explore the wage gap using IPW, we would move to Oaxaca decomposition, which is expected to be affected as well. This methodology has an identification issue which causes different unexplained part once the reference group was changed in one categorical variable. Accordingly, we expected great relative differences in the unexplained part between unweighted and weighted regressions. However, we are keen to explore these differences in mean wage values and wage gap and the explained part. Oaxaca decomposition has four approaches, as discussed before. From the number of the observation in each model, we noticed that pooled and omega follow the double robust method combined between imputation and IPW, unlike w (1) and w (0) approaches that used actual observations performing IPW.

Using IPW decomposition generated changes in the average mean of logarithmic salaries for Saudi and non-Saudi, which was reflected in the wage gap (see Table 6-4Table 6-1). However, the significance level was consistent to unweighted decomposition for both weights where IPW generated significant results at 1% confidence level like CC.

Moreover, the coefficients show consistency in it is sign with relative changes. Saudi average wage coefficients relatively decreased by (0.23%) in weightA and (0.76%) in weightB. Similarly, non-Saudi wages coefficients relatively decreased by (1.13%) in weightA and (0.95%) weightB. This implies that weighting for wage missingness causes a decrease in the non-Saudi average mean by a higher percentage compared to the decrease in Saudi wages. The differences between the group's mean (the wage gap) relatively decreased, as well, by (3.60%) in weightA and (0.02%) in weightB.

Table 6-4: logarithmic salaries mean for each group and the wage gap.

	unweighted	weightA	weightB
Non-Saudi	6.721692 (1.80E+04)	6.64555 (1.90E+04)	6.657959 (1.90E+04)
Saudi	8.305954 (2.00E+04)	8.286735 (1.90E+04)	8.242581 (2.00E+04)
Wage gap	-1.58426 (-2812.98)	-1.64119 (-2889.63)	-1.58462 (-2899.74)

The t-ratio between brackets.

Indeed, these relative differences in the average wage gap among the weight predictions compared to unweighted results would influence the explanation percentage power of the explained and unexplained part from the wage gap for each approach. Table 6-5 shows that the explained part relatively increased in all approaches unless w(1) in weightB. However, the relatively changes of the explained part was higher in weightB for w(1) and omega approaches while w(0) and pooled approaches increased relatively higher in weightA. We would explore the effect of these changes in each approach on the next headings.

Table 6-5: the percentage changes of the explained and unexplained part compared to the unweighted approach.

		WeightA	WeightB
Omega	Explained	0.60%	2.57%
	Unexplained	-0.60%	-2.57%
Pooled	Explained	4.01%	0.19%
	Unexplained	-4.01%	-0.19%
W (0)	Explained	3.98%	-0.87%
	Unexplained	-3.98%	0.87%
W (1)	Explained	0.47%	-1.57%
	Unexplained	-0.47%	1.57%

The researcher calculation (unweighted percentage – weighted percentage) for both explained and unexplained part.

6.3.1 Omega approach

In the omega approach, the coefficients for all parts–explained and unexplained–and their attributes were consistent. They followed a similar sign and at a 1% significance level of confidence like the unweighted decomposition. However, comparing to CC the coefficients of the explained part decreased in weightA by (4.47%) and in weightB by (3.64%). Similarly, the unexplained part coefficient decreased by (1.46%) weightA while it increased relatively by 8.86% in weightB (see Table 6-6). In terms of the explained part, workers attributes relatively decreased in both weights compared to unweighted by (18.42%) in weightA and (27.02%) in weightB. This associated with an increase in the explanation percentage of these weights by 45.42% and 50.45%, respectively, compared to unweighted decomposition where workers attributes explained 39.73%. Unlike job and policy attributes, the coefficients relatively increased associated with an increase in the explanation percentage compared to unweighted. For job attribute, the coefficients increased by 17.52% in weightA and 31.26% in weightB. Job attribute explained 22.32% from the wage gap in unweighted regression while it explained lower percentage on the weighted decomposition around 17.77% in weightA and 15.34 % in weightB. Similarly, policy attribute coefficients increased by 2.64% in weightA and 13% in weightB explaining lower percentage around 8.45% and 7.82%, respectively compared to 8.99% when unweighted decomposition used.

Table 6-6: omega explained and unexplained part for all weighing approaches.

Explained	Unweighted	WeightA	WeightB
Explained	-1.125414 (-1761.42)	-1.175661 (-1704.6)	-1.166414 (-1805.74)
Workers' attributes	-0.6294533 (-913.48)	-0.7454177 (-1038.94)	-0.7995007 (-1152.45)
Jobs' attributes	-0.3535917 (-597.17)	-0.2916373 (-539.27)	-0.2430577 (-478.43)
Policy' attributes	-0.1423687 (-430.9)	-0.1386056 (-380.55)	-0.1238553 (-376.97)
Unexplained	-0.4588485 (-872.76)	-0.4655253 (-812.82)	-0.4182079 (-790.36)
Workers' attributes	1.386136 (35.58)	1.135958 (30.16)	1.28584 (34.42)
Jobs' attributes	0.2031385 (105.16)	0.1096699 (57.82)	0.0941181 (52.73)
Policy' attributes	0.0979042 (138.16)	0.095368 (126.65)	0.0904512 (130.17)
Constant	-2.146027 (-54.97)	-1.806522 (-47.84)	-1.888617 (-50.44)

The t-ratio between round brackets.

Regarding the unexplained part, workers attribute were positive coefficients in all weights and consistent with unweighted decomposition. This attribute was relatively lower in both weights compared to CC. Workers attribute coefficients decreased by (18.05%) in weightA, and (7.24%) in weightB. Accordingly, there was (87.49%) remained unexplained due to workers attribute on unweighted decomposition, and it increased by (18.28%) in weightA and (6.35%) in weightB which means higher percentage remained unexplained due workers attribute when IPW considered. A similar pattern was found in the job and policy attribute coefficients unlike the intercept where IPW had lower percentage remained unexplained due the intercept (see Table 6-7).

Table 6-7: the absolute differences of the unexplained part explanation percentage.

	Attribute	WeightA	WeightB
Absolute differences It equal the subtraction of the explanation percentage between unweighted and IPW.	Workers	18.28%	6.35%
	Job	6.14%	6.88%
	Policy	0.37%	0.47%
	Constant	-25.38%	-16.28%

6.3.2 Pooled approach

Table 6-8 shows that the coefficient of the explained gap was inconsistent in weight A when the polynomial specification was used, while it was consistent in weight B compared to the unweighted decomposition.

Table 6-8: coefficients of the decomposition for the pooled approach.

	Unweighted	WeightA	WeightB
Explained	0.0475387 (55.31)	-0.016549 (-19.8)	0.0444929 (52.24)
Workers' attributes	0.2094674 (285.17)	0.1273537 (180.01)	0.1583038 (216.02)
Jobs' attributes	-0.1291281 (-310.97)	-0.114828 (-305.7)	-0.087793 (-251.47)]
Policy' attributes	-0.0328007 (-142.46)	-0.029075 (-122.8)	-0.026018 (-118.26)
Unexplained	-1.631801 (-1668.73)	-1.624637 (-1685.65)	-1.629114 (-1721.89)
Workers' attributes	0.5472156 (14.05)	0.263187 (6.99)	0.3280351 (8.79)
Jobs' attributes	-0.0213251 (-11.25)	-0.067139 (-36.03)	-0.061147 (-34.73)
Policy' attributes	-0.0116638 (-17.48)	-0.014163 (-20.18)	-0.007386 (-11.33)
Intercept	-2.146027 (-54.97)	-1.806522 (-47.84)	-1.888617 (-50.44)

The t-ratio between brackets.

However, in both weights this coefficient was significant at 1% like unweighted. Using IPW for pooled approach did not provide one answer for the total explained part. Taking into consideration that using CC or IPW did not explained much of the wage gap due the high heterogeneity of the two groups and the limitation of the variable we have got. Although the total explained part shows inconsistency in weightA, the decomposition coefficients in all attributes displayed a coordinated coefficient and significance level with relative decrease compared to unweighted decomposition in both weights. It seems that the decrease in workers attribute in weightA was high compared to job and policy attribute which was sufficient to turn the explained part to negative unlike the decrease in weightB. Worker's attribute decreased in weightA by (39.20%) and in weightB by (24.43%). Explained workers attribute was significant at 1% level of confidence. Worker's attribute explained around 13.22% of the wage gap according to unweighted decomposition results while it was 7.76% in weightA, and 9.99% in weightB. However, the coefficients of job and policy attribute were coordinated and significant at 1% level of confidence for all weighted and unweighted decomposition, as well. However, both weights show a relative increase in job and policy attributes. Job attributes increased by 11.07% in weightA, and 32.01% in weightB. This explained 7.00% in weightA and 5.54% in weightB compared to 8.15% of the unweighted decomposition. Similarly, the coefficients of policy attribute relatively increased by 11.36% in weightA and 20.68% in weightB. Policy attribute explained 2.07% according to the unweighted decomposition and 1.77% in weightA and 1.64% in weightB. he coefficients of the unexplained gap- for both weights- were consistent with unweighted regression regardless of the slight relative differences. These coefficients relatively increased by 0.44% in weightA, and 0.16% in weightB. All coefficients were significant at 1% confidence level like unweighted. In term of the unexplained part components' coefficients, they were consistent with unweighted coefficients sign with relative decrease or increase and were significant at 1% like unweighted decomposition. The wage gap remained unexplained by lower percentage for the intercept of IPW and policy attribute in weightB compared to unweighted, otherwise the unexplained component explanation percentage increased in IPW compared to CC (see Table 6-9).

Table 6-9: relative and explanation percentages for the unexplained part.

		WeightA	WeightB
Relative differences	Workers	-51.90%	-40.05%
	Jobs	-214.84%	-186.74%
	Policy	-21.43%	36.68%
	Intercept	15.82%	11.99%
Absolute differences %	Workers	18.50%	13.84%
	Jobs	2.74%	2.51%
	Policy	0.13%	-0.27%
	Intercept	-25.38%	-16.28%

6.3.3 w (0) approach

Using IPW, we generated coordinated significant coefficients compared to the unweighted decomposition, although there were some relative changes. Total coefficient of the explained part relatively decreased in weightA by (35.89%) while it increased relatively in weightB by 8.31%. The coefficients and significance level of explained workers, job, and policy attribute were generated by IPW were consistent with unweighted decomposition with relative changes (see Table 6-10).

Table 6-10: coefficients of the decomposition for the w (0) approach.

	unweighted	weightA	weightB
Explained	0.165376 (162.63)	0.106018 (92.22)	0.179122 (155.17)
Workers' attributes	0.309487 (388.71)	0.250919 (269.02)	0.285616 (294.8)
Jobs' attributes	-0.11021 (-196.1)	-0.11307 (-200.36)	-0.08123 (-157.12)
Policy' attributes	-0.03393 (-176.99)	-0.03183 (-154.74)	-0.02526 (-141.48)
Unexplained	-1.74964 (-1659.01)	-1.7472 (-1391.44)	-1.76374 (-1420.45)
Workers' attributes	0.195766 (11.53)	0.139621 (3.7)	0.200723 (5.37)
Jobs' attributes	0.105896 (-21.87)	-0.0689 (-36.74)	-0.0677 (-37.7)
Policy' attributes	-0.01549 (-16.48)	-0.01141 (-18.76)	-0.00814 (-14.09)
_cons	-2.03579 (-55.26)	-1.80652 (-47.84)	-1.88862 (-50.44)

The t-ratio between brackets.

Worker's endowment explained roughly 19.54% of the wage gap using CC while this percentage was lower in IPW. It reached 15.29% in weightA and reached 18.02% in weightB because workers coefficients decreased in ordered by 18.93, and (7.72%). Similarly, job attribute coefficient was consistent comparing to CC and decreased relatively by (2.60%) in weightA whereas using weightB showed relative increased by 26.29%. Accordingly, due to job attribute IPW contributed to explained 6.89% in weightA, and 5.13% in weightB from the existing wage gap which was lower compared to the percentage provided by unweighted decomposition which was 6.96%. Likewise, policy attribute coefficients were consistent with relative differences. This coefficient increased by 6.18% in weightA, and 25.55% in weightB. Policy attribute contributed to explain the existing wage gap by 1.94% in weightA, and 1.59% in weightB compared to

2.14% in CC. However, using IPW generated coordinated significant unexplained coefficients compared to the unweighted decomposition with relative changes. The total unexplained part relatively increased in weightA by 0.14%, while it decreased by (0.81%) in weightB. Table 6-11 display both the explanation percentage and the relative changes of the unexplained part components.

Table 6-11: unexplained part relative and explanation percentage, w (0) approach.

		Unweighted	WeightA	WeightB
Relative differences	Workers	-	-68.78%	-55.11%
	Jobs	-	-71.19%	-68.23%
	Policy	-	-8.25%	22.70%
	Intercept	-	15.82%	11.99%
Unexplained part percentage from total gap	Workers	-28.23%	-8.51%	-12.67%
	Jobs	2.54%	4.20%	4.27%
	Policy	0.67%	0.69%	0.51%
	Intercept	135.46%	110.07%	119.18%

6.3.4 W (1) approach

Generally, the coefficients' sign and significance level of IPW were consistent compared to the unweighted decomposition for both weights with some relative differences (see Table 6-12). For total explained part, the coefficients decreased relatively in weightA by (6.03%) whereas it increased relatively in weightB by 7.78%. The coefficients of workers attribute relatively decreased in weightA by (25.67%) and weightB by (16.03%). This relative decrease in the weighted coefficients was associated with an increase in the explanation percentage of the workers attribute where it was explained around 12.77% in weightA, and 12.21% in weightB compared to 10.52% in unweighted decomposition. Unlike job attribute, which showed a relative increase by 14.06% in weightA, and 35.12% in weightB. However, this relative increase in both weights compared to the CC was associated with a decrease in the explained percentage where job attribute explained 7.98% when CC are considered while it explained slightly lower 6.62%, 5.18% in weightA, and 7.45%, in weightB. Policy attribute coefficients shows a relative increase in weightA and weightB show by 21.74%, and 27.18% in order. However, unweighted decomposition shows that policy attribute explained 1.67% from the existing gap, which is considered slightly higher compared to IPW where this percentage was roughly 1.26% in weightA and 1.22% in weightB.

Table 6-12: coefficients of the decomposition, w (1) approaches.

	unweighted	weightA	weightB
Explained	-0.3196617 (-128.78)	-0.3389316 (-135.68)	-0.2947779 (-104.24)
Workers' attributes	-0.1667281 (-72.14)	-0.2095349 (-86.88)	-0.19345 (-70)
Jobs' attributes	-0.1264728 (-139.76)	-0.1086879 (-156.29)	-0.0820585 (-134.01)
Policy' attributes	-0.0264608 (-61.26)	-0.0207088 (-51.37)	-0.0192693 (-49.37)
Unexplained	-1.2646 (-504.45)	-1.302254 (-513.38)	-1.289844 (-451.25)
Workers' attributes	0.9234111 (24.76)	0.6000756 (16.49)	0.6797889 (19.18)
Jobs' attributes	-0.0239804 (-11.46)	-0.0732794 (-36.82)	-0.0668811 (-36.02)
Policy' attributes	-0.0180036 (-20.93)	-0.0225288 (-27.3)	-0.0141348 (-18.09)
Intercept	-2.146027 (-55.26)	-1.806522 (-47.84)	-1.888617 (-50.44)

The t-ratio between brackets.

However, the total unexplained part coefficients and its components (workers, job, and policy attribute) were consistent with unweighted coefficients in both weights. Furthermore, considering unweighted decomposition, the explanation percentage for each unexplained part component increased compared to the percentage generated by the IPW unless the intercept (see Table 6-13).

Table 6-13: unexplained part relative and explanation percentage, w(1) approach.

		unweighted	WeightA	WeightB
Relative differences	Worker	-	-35.02%	-26.38%
	Jobs	-	-205.58%	-178.90%
	Policy	-	-25.13%	21.49%
	Intercept	-	15.82%	11.99%
Unexplained part %	Workers	-58.29%	-36.56%	-42.90%
	Jobs	1.51%	4.47%	4.22%
	Policy	1.14%	1.37%	0.89%
	Intercept	135.46%	110.07%	119.18%

6.4 The wage gap between Saudis and non-Saudis in firm localisation status

From the last section, we can notice that the CC and IPW were consistent for the entire sample. However, the pooled approach was sensitive to the weight. In this section, we will explore the wage gap between Saudis and non-Saudis within a finite sample, separated for localised and non-localised firms. The significance level was set at 1%

when CC or IPW were used. The coefficients provided a similar conclusion in terms of the wage gap between Saudi and non-Saudi had a lower gap in non-localized firms compared to localised firms for both weights like unweighted (see Table 6-14). In localised firms, there was a relative increase on the wage gap by 0.18% in weightB, compared to unweighted while there was a relative decrease in weightA by (3.5%). The wage gap in non-localised firms relatively decrease by (4.07%) in weightA and (2.40%) in weightB.

Table 6-14: the wage gap between Saudi and non-Saudis in firms' localised status.

	unweighted	WeightA	weightB
Localised	-1.593612 (-2564.94)	-1.649458 (-2642.93)	-1.590819 (-2645.6)
Non-localised	-1.43833 (-1348.35)	-1.49688 (-1268.88)	-1.472788 (-1317.1)

The t-ratio between brackets.

6.4.1 Localised firms

6.4.1.1 Omega approach

Table 6-15 show that using IPW produced consistency with CC in terms of the coefficients, and significance level. All coefficients were significant at 1% level of confidence. The coefficients of the total explained part relatively decreased in weightA by (5.30%), and in weightB by (4.73%). Accordingly, the unweighted explained part coefficients explained around 67.24% of the wage gap which is slightly lower compared to 68.41% in weightA, and 70.54% in weightB. The highest differences in the explanation power were in weightB by 3.30%. Unlike, total unexplained part coefficients increased relatively compared to unweighted coefficients by 0.19% in weightA, and 10.24% in weightB.

Unweighted decomposition coefficients explained the gap by workers characteristic lower than IPW for both weights where we found relative decrease of (18.23%) in weightA, and (25.89%) weightB. This increase in the coefficients associated with an increased in the explanation percentage where this coefficient explained 5.91% in weightA and 10.85% in weightB higher compared to CC. Unlike job characteristic coefficients which was higher when CC was used compared to IPW where weightA and weightB coefficients relatively increased by 15.62%, and 29.52%. Accordingly, job attribute of IPW explained lower percentage compared to CC by 4.74% in weightA and 7.55% in weightB.

Similarly, the unexplained part of workers characteristics, jobs and the intercept show relative decrease in both weights. Worker's attribute decreased by (14.71) in weightA and (3.85%) in weightB. Job characteristics of the unexplained part decreased relatively in IPW decomposition by (39.02%) in weightA, and (47.33%) in weightB. The intercept coefficients increased by 14.06% in weightA and 10.38% in weightB. Although all coefficients decreased in both weights compared to CC, the explanation percentage decreased only for the intercept by 23.36% in weightA and 14.07% in weightB, otherwise the explanation percentage increased. To clarify, workers attribute explained 15.67% in weightA and 3.28% higher in weightB higher compared to unweighted. Similarly, job attribute explained 6.52% in weightA and 7.50% in weightB higher compared to CC.

Table 6-15: omega approach for unweighted and IPW decomposition (localised firms).

		Unweighted	**weightA**	**weightB**
total	Explained	-1.071556 [-2564.94]	-1.128398 [-1499.8]	-1.122216 [-1595.08]
	Unexplained	-0.5220555 [-1527.15]	-0.52106 [-814.83]	-0.468603 [-795.05]
Explained	Workers	-0.6623225 [-875.2]	-0.7830842 [-1031.2]	-0.8337747 [-1144.57]
	Jobs	-0.4092336 [-902.72]	-0.3453137 [-542.3]	-0.2884414 [-481.76]
Unexplained	Workers	1.419315 [-591.63]	1.210567 [28.67]	1.364719 [32.68]
	Jobs	0.2528696 [32.32]	0.1542082 [75.91]	0.1331906 [69.86]
	Intercept	-2.19424 [121.59]	-1.885835 [-44.57]	-1.966512 [-47]

The t ratio between brackets.

6.4.1.2 Pooled approach

Table 6-15 shows that using IPW produced consistency with CC, in terms of the coefficients and significance level. The finding was statistically significant for all coefficients at 1% level of confidence. The coefficients of the explained part show high relative decrease of around (84.04%) in weightA, and (12.27%) in weightB. Unlike the coefficients of the unexplained part, relative changes were around 1% or less. The coefficients increase relatively in weightA, and weightB by 1.13%, and 0.81%, respectively. The coefficients of CC explained (in total) small percentage about (5.58%), similarly IPW did where it explained (0.86%) by weightA, and (4.91%) by weightB.

Table 6-16: pooled approach for unweighted and IPW decomposition (localised firms).

		Unweighted	weightA	weightB
Total	Explained	0.088998 [99.44]	0.0142033 [16.31]	0.0780783 [87.68]
	Unexplained	-1.68261 [-1637]	-1.663661 [-1.64E+03]	-1.668897 [-1674.13]
Explained	Workers	0.2320001 [296.25]	0.1444676 [192.03]	0.1790835 [229.62]
	Jobs	-0.1430021 [-307.17]	-0.1302643 [-307.3]1	-0.1010052 [-255.9]
Unexplained	Workers	0.5249922 [11.96]	0.2830149 [6.71]	0.3518607 [8.43]
	Jobs	-0.0133618 [-6.58]	-0.0608412 [-30.64]	-0.0542457 [-28.97]
	Intercept	-2.19424 [-49.88]	-1.885835 [-44.57]	-1.966512 [-47]

The t-ratio between square brackets.

By looking at detail of the explained part, we found that workers characteristic coefficient decreased by (37.73%) in weightA, and (22.811%) in weightB. Unlike, job characteristics coefficient increased by 8.91% in weightA, and 29.37% in weightB. However, workers attribute of IPW explained higher compared to unweighted by 5.80% in weightA and 3.30% in weightB while job attribute explained lower compared to CC by (1.08%) in weightA and (2.62%) in weightB. The coefficients of workers and job characteristics of the unexplained part relatively decreased on both weights. However, the workers remained unexplained by (32.94%) on unweighted decomposition while it increased by (15.79%) in weightA, and (10.83%) in weightB. Job coefficient on the unexplained part relatively decreased by over 300% in both weights cause an increase of the explanation percentage around 2% compared to CC. Around 0.84% remained unexplained due job characteristic on unweighted decomposition while it was 3.69% in weightA and 3.41% in weightB. Unlike the intercept coefficients increased by (14.06%) in weightA, and (10.11%) in weightB and explained lower percentage compared to CC. Around (137.69%) remained unexplained via the intercept on the unweighted decomposition decreased in IPW to (114.33%), and (123.62%) in both weights, respectively. However, the intercept formed higher percentage of the unexplained part in both CC and IPW.

6.4.1.3 W (0) approach

Table 6-17 shows that the coefficients of the explained and unexplained parts were consistent with unweighted decomposition, concerning the coefficients' sign and significance level. All coefficients were significant at 1% in both CC and IPW.

Table 6-17: w (0) approach for unweighted and IPW decomposition (localised firms).

		Unweighted	weightA	weightB
Total	Explained	0.2014461 [189.43]	0.133221 [111.13]	0.2073433 [172.5]
	Unexplained	-1.795058 [-1619.83]	-1.782679 [-1346.63]	-1.798163 [-1377.4]
Explained	Workers	0.3257728 [385.19]	0.2636276 [269.9]	0.3011816 [296.82]
	Jobs	-0.1243267 [-205.79]	-0.1304066 [-213.93]	-0.0938383 [-169.53]
Unexplained	Workers	0.4312195 [9.86]	0.1638549 [3.88]	0.2297625 [5.5]
	Jobs	-0.0320373 [-16.22]	-0.0606989 [-30.59]	-0.0614126 [-32.34]
	Intercept	-2.19424 [-50.11]	-1.885835 [-44.57]	-1.966512 [-47]

The t-ratio between brackets.

The explained part was relatively decreased by (33.87%) in weightA while it relatively increased by 2.93% in weightB. Unlike, the coefficients of the unexplained part which increased by a small proposition in weightA, around 0.69% while it decreased on weightB by (0.17%). This approach assumes a high proportion of the wage gap was unexplained around 112.64% and (12%) explained when CC were used. The CC approach was in the middle among all other weights decomposition results where the unexplained (explained) percentage were 108.08% (8.08%) for weightA, 113.03% (13.03) for weightB. The coefficients of workers characteristics decreased relatively by (19.08%) in weightA and (7.55%) in weightB. Accordingly, using CC explained around 20.44% of the wage gap by workers attributes which was lower compared to the IPW decomposition by around 1.10% in weightA and 1.51% in weightB. Similarly, job characteristics coefficients relatively decreased in weightA by (4.89%) while it increased relatively in weightB by 24.52%. Job attribute explained around 7.80% when

CC was used while using IPW decompositions increases this percentage in weightA by 1.67% unlike, it decreased by (1.85%) in weightB.

In terms of the unexplained parts, the coefficients of workers attributes show a relative decrease by (62.00%) in weightA and (46.72%) in weightB. Accordingly, around 27.06% remained unexplained by workers characteristics when unweighted decomposition was used while it was smaller using IPW explaining 9.93% in weightA, and 14.44% in weightB. Similarly, job attribute coefficient of IPW decreased by (89.46%) in weightA and (91.69%) in weightB. Although the relative decrease was significant, the explanation percentage slightly changed where 2.01% remained unexplained due to job attribute when unweighted decomposition was considered while it was 3.68% in weightA and 3.86%. The intercept increased relatively by 14.06% in weightA and 10.38% in weightB compared to unweighted decomposition. However, the intercept explained a higher percentage when unweighted decomposition used around 137.69%, similarly when IPW used, this percentage was 114.33% in weightA, and 123.62% weightB.

6.4.1.4 W (1) approach

Using IPW for the w (1) approach generated consistent results, in terms of the coefficients' sign and significance level with relative differences compared to CC. The explained coefficients relatively decreased in weightA by 9.06%, while it increased in weightB by 6.07%. The unexplained part decreased relatively by (2.26%) in weightA and (1.14%) in weightB. The explanation proportion when CC used was 4.26% in the middle compared to that proportion generated by IPW where it explained 5.02% in weightA and 3.73% in weightB (see Table 6-18).

Workers attribute coefficients decreased relatively by (27.42%) in weightA and (16.56%) in weightB when the IPW was considered. Although the coefficients relatively decreased the explanation percentage increased for this coefficient. Around 10.00% of the gap explained by workers attributes when unweighted decomposition was used while this percentage increased to 12.31% in weightA and 11.67% in weightB. Unlike job attribute coefficients which increased relatively when IPW decomposition was used by 13.07% in weightA and 33.36% in weightB. This relative increase in all coefficients associated with a decrease in the explanation percentage for both weights. Using CC explained 8.29% by job attribute while weightA and weightB had lower percentage 6.97%, and 5.54%, respectively.

Table 6-18: w (1) approach for unweighted and IPW decomposition (localised firms).

		Unweighted	weightA	weightB
Total	Explained	-0.2914937 [-109.46]	-0.3178981 [-119.88]	-0.273792 [-90.17]
	Unexplained	-1.302118 [-483.86]	-1.33156 [-494.24]	-1.317027 [-429.17]
Explained	Workers	-0.1593207 ([-63.45]	-0.2030039 [-78.49]	-0.1857064 [-62.13]
	Jobs	-0.132173 ([-134.14]	-0.1148942 ([-152.48]	-0.0880857 ([-131.64]
Unexplained	Workers	0.916313 [21.77]	0.6304865 [15.45]	0.7166505 [18.07]
	Jobs	-0.0241909 [-10.67]	-0.0762113 [-35.79]	-0.0671652 [-33.87]
	Intercept	-2.19424 [-50.11]	-1.885835 [-44.57]	-1.966512 [-47]

The t-ratio between square brackets.

However, workers attribute coefficients on the unexplained part were relatively decreased relatively compared to unweighted decomposition by (31.19%) in weightA, and (21.79%) in weightB. Accordingly, workers attribute explained (57.50%) when CC was used while it increased in weightA and weightB by (19.28%), (12.45%), respectively. Job attribute coefficients were significantly decreased by (215.04%) in weightA and (177.65%) in weightB. Although the relative decrease were high, the explanation percentage absolute change was small at around 3%. The unweighted approach suggested that the wage gap remained unexplained due to job attribute by around 1.52% while weightA, and weightB had higher percentage around 4.62%, and 4.22%. The coefficients of the intercept was relatively increased compared to CC by 14.06% in weightA, 10.38% in weightB. Although unweighted had the highest unexplained part due to the intercept, IPW shared this feature, as well. Accordingly, unweighted intercept contributed by around 137.69%, compared to 114.33% in weightA and 123.62% in weightB.

6.4.2 *Non-localised firms*

6.4.2.1 Omega approach

The coefficients and significance level of the omega approach showed consistency between CC and IPW (see Table 6-19). Generally, IPW coefficients of the explained part decreased relatively to CC in both weights: a decrease by 1.54% in weight A and 4.29% in weight B. However, this coefficient explained 78.74% of the gap with the application of the unweighted decomposition, which was lower than the percentage explained by weight A (76.82%). On the contrary, weightB explained 80.20% of the gap, which was higher compared to the unweighted decomposition. The coefficients of the unexplained part decreased by 13.44% in weight A, while it increased by 4.63% in weightB. However, 21.26% of the gap remained unexplained when the CC approach was used, while this percentage was higher in weight A, reaching 23.18%. As for the unexplained part of weightB, the percentage was 19.80%.

Table 6-19: omega approach for CC and IPW decomposition (non-localised firms).

		Unweighted	weightA	weightB
Explained	Total	-1.132487 [-1348.35]	-1.149933 [-650.5]	-1.181113 [-727.53]
	Workers	-0.9294387 [-239.61]	-1.004692 [-526]	-1.065411 [-595.59]
	Jobs	-0.2030488 [-546.3]	-0.1452408 [-131.69]	-0.1157024 [-116.58]
Unexplained	Total	-0.3058455 [-782.92]	-0.3469484 [-213.47]	-0.2916749 [-201.36]
	Workers	1.221743 [-161.45]	0.9119108 [10.95]	1.021086 [12.34]
	Jobs	0.1508555 [15.02]	0.0803779 [14.22]	0.0681067 [12.98]
	Intercept	-1.678444 [28.73]	-1.339237 [-16.02]	-1.380867 [-16.62]

The t-ratio between brackets.

On the explained part, the IPW coefficients of worker attributes decreased relatively to the unweighted decomposition in both weights, while the job attribute increased. Although the unweighted decomposition had higher coefficients of worker attributes, it explained a lower percentage compared to the IPW decomposition (64.62% by CC *versus* 67.12% in weight A and 72.34% in weightB). Moreover, 14.12% of the wage gap could be explained through job attributes when CC was used, while it decreased when IPW was applied (9.70%, 7.86%, and 12.87% in weight A, and weightB, respectively). The coefficients of the worker and job attributes for the unexplained part decreased in both weights, unlike the intercept. Higher percentages remained unexplained due to the worker and job attributes when IPW was compared to the unweighted decomposition, while the remaining unexplained percentage via intercept was lower for IPW.

6.4.2.2 Pooled approach

The coefficients' sign and the significance level of the overall explained and unexplained parts via IPW and unweighted decomposition were consistent, while they were inconsistent when the components of the explained and unexplained parts were taken into consideration (see Table 6-20).

Table 6-20: pooled approach for CC and IPW decomposition (non-localised firms).

		Unweighted	weightA	weightB
Explained	Total	-0.0613225 [-27.46]	-0.0750869 [-31.52]	-0.0549546 [-23.11]
	Workers	0.0072699 [3.35]	-0.0243177 [-10.73]	-0.021048 [-9.28]
	Jobs	-0.0685924 [-76.05]	-0.0507692 [-61.87]	-0.0339066 [-45.14]
Unexplained	Total	-1.37701 [-541.84]	-1.421795 [-512]	-1.417834 [-532.86]
	Workers	0.2850348 [3.51]	-0.0684639 [-0.82]	-0.0232774 [-0.28]
	Jobs	0.0163991 [3.16]	-0.0140937 [-2.51]	-0.0136891 [-2.62]
	Intercept	-1.678444 [-20.57]	-1.339237 [-16.02]	-1.380867 [-16.62]

The t-ratio between square brackets.

All explained and unexplained part coefficients were negative and significant at 1% level of confidence for IPW and CC. The explained part decreased for IPW compared to CC by (22.45) in weightA while it increased relatively by 10.38% in weightB. Additionally, total unexplained part coefficients decreased relatively in both weights by (3.25%) in weightA and (2.91%) in weightB. The explained workers attribute was inconsistent in term of the coefficient sign while it was significant at 1% level in both CC and IPW. This coefficient decreased relatively by over 300% results in a negative coefficient of the IPW compared to the positive coefficient for CC. However, job attribute had negative and significant coefficient at 1% level of confidence in both CC and IPW. Both workers and job attribute unexplained coefficients were negative for IPW and positive at CC. IPW show in significance level for workers attribute and significance level of 1% for job attribute while CC had significance level for both coefficients at 1%. The unweighted intercept was the only component of the unexplained part consistent with the IPW in both weights with relative decreased by (20.21%) in weightA and (17.73%) in weightB.

6.4.2.3 W (0) approach

IPW generated consistent coefficients and significance levels in the entirety of the explained and unexplained parts. However, unexplained worker attributes had negative coefficients in weight A, contrary to the CC. Otherwise, the remaining coefficients were consistent. This coefficient was significant in weightA while it was insignificant in weightB (see Table 6-21).

Table 6-21: w (0) approach for CC and IPW decomposition (non-localised firms).

		Unweighted	weightA	weightB
Explained	Total	0.0686643 [23.85]	4.61E-02 [10.37]	0.0632205 [15.81]
	Workers	0.1098498 [46.37]	0.0812233 [22.14]	0.085793 [24.24]
	Jobs	-0.0411855 [-25.07]	-0.0351181 [-14.94]	-0.0225725 [-13]
Unexplained	Total	-1.506997 [-500.9]	-1.542987 [-324.34]	-1.536009 [-363]
	Workers	0.1824549 [2.28]	-0.1740049 [-2.09]	-0.1301184 [-1.57]
	Jobs	-0.0110078 [-2.23]	-0.0297448 [-4.91]	-0.0250232 [-4.57]
	Intercept	-1.678444 [-20.91]	-1.339237 [-16.02]	-1.380867 [-16.62]

The t-ratio between square brackets.

Using IPW, accordingly, did not generate a consistent result compared to unweighted on one component of the unexplained part; otherwise, it would have had a similar conclusion. For example, the job attribute of the unexplained gap explained around 0.77% when unweighted decomposition was respected, and it explained around 1.99%, and 1.70% for weightA, and weightB, respectively. Similarly, the intercept was explained around 116.69% when CC was used and around 89.47% in weightA, and 93.76%, in weightB. Moreover, using CC explained around (4.77%) while 104.77%

remained unexplained. This percentage middled between IPW results where weightA explained lower percentage around (3.08%), and 103.08 remained unexplained while weightB explained higher percentage around (4.29%) compared to unweighted. Likewise, the explained part component had a similar conclusion. Around (7.64%) explained the wage gap due to workers attributes when CC was used while the percentages were (5.43%), and (5.83%), for both weights in order. Similarly, it explained around 2.86% due to job attribute when unweighted decomposition was used and where this percentage was 2.35%, and 1.53% for both weights in order.

6.4.2.4 W (1) approach

Table 6-22 illustrates that the coefficients and significance level were consistent with CC decomposition when IPW was used, except for the unexplained worker attributes in weight A, where the significance level reached 10%. The total explained part coefficient relatively increased by 3.87% in weightA, and 14.00% in weightB. Accordingly, around 23.66% of the gap was explained using unweighted decomposition while IPW explained slightly lower 21.86% in weightA, and 19.87% in weightB.

Table 6-22: w (1) approach for CC and IPW decomposition (non-localised firms).

		Unweighted	weightA	weightB
Explained	Total	-0.3403504 [-47.61]	-0.3271666 [-42.6]	-0.2926935 [-36.29]
	Total	-1.097983 [-152.78]	-1.169715 [-150.64]	-1.180095 [-144.87]
	Workers	-0.2302336 [-33.74]	-0.2432444 [-32.33]	-0.2322636 [-29.16]
	Jobs	-0.1101168 [-47.09]	-0.0839222 [-45.26]	-0.0604299 [-38.72]
Unexplained	Workers	0.5225383 [6.83]	0.1504628 [1.89]	0.1879382 [2.42]
	Jobs	0.0579235 [10.74]	0.0190593 [3.23]	0.0128341 [2.36]
	Intercept	-1.678444 [-20.91]	-1.339237 [-16.02]	-1.380867 [-16.62]

Standard error between round brackets and t ratio between square brackets.

Although, the coefficients of workers attribute decreased in both weights of IPW, the coefficient explanation percentage of unweighted middled between the percentage generated by IPW. Around 16.01% of the gap was explained due to workers attributes on the unweighted decomposition while this percentage increased in IPW to reach

16.25% in weightA while it decreased in weightB to 15.77%. Unlike, job attribute coefficients significantly increased in a relative term by around 23.79% in weightA and 45.12% in weightB. however, the explanation percentage of these coefficients decreased with less than 5%. Approximately 7.66% of which was explained by CC while it decreased to 5.61% in weightA, and 4.10% in weightB.

Despite of the total coefficient of the unexplained part decreased relatively using IPW by (6.53%) in weightA and (7.48%) in weightB compared to unweighted decomposition, IPW had higher percentage remained unexplained at 78.14% in weightA and 80.13% in weightB compared to 76.34% when unweighted decomposition used. However, all unexplained part components decreased relatively compared to the unweighted decomposition unless the intercept. This decrease was associated with an increase in the explanation percentage of workers attributes. For clarification, (36.33%) of the unweighted gap remained unexplained due to workers attributes while this percentage was higher in absolute term by (26.28%) and (23.57%) in both weights, respectively. The gap remained unexplained due to job attribute when unweighted decomposition used by (4.03%) which were lower compared to IPW percentage by (2.75%) in weightA and (3.16%), in weightB. Similarly, the intercept which remained unexplained by around 116.69% when CC was used while this percentage decreased to 89.47%, and 93.76% when IPW was used for all weights in ordered.

In this section, we found that IPW was consistent with CC in all the approaches and weights for the localised sub-sample, which forms 86% of the dataset. For the non-localised sub-sample, both the omega and w (1) approaches showed similar findings. This indicates that IPW was sensitive to the approaches and weights used.

6.5 The wage gap between workers according to firm status

6.5.1 Saudi and non-Saudi included in the sample.

The average wage gap was in favour of workers in localised firms at a 1% level of confidence whether we use CC or IPW. The coefficients of the average wages and wage gap were significant at 1% significance level (see Table 6-23). This gap relatively increased in both weight of IPW (in weightA by 2.73% and weightB by 7.89%) compared to unweighted decomposition. The average wage relatively decreased for workers in localised and non-localised workers in weightA and weightB compared to CC.

Table 6-23: average wage gap for all workers (firms' status the reference).

	unweighted	weightA	weightB
Localised	7.460339 [15000.00]	7.293744 [13000.00]	7.375556 [1.3000.00]
Non-localised	7.09793 [6854.48]	6.941245 [6081.11]	7.04175 [5733.75]
difference	-0.36241 [-313.62	-0.3525 [-275.78]	-0.3338063 [-247.94]

The t-ratio between square brackets.

6.5.1.1 Omega approach

This approach indicates that the coefficients, the standard error of both the explained and the unexplained parts, and their components were consistent with CC when IPW was used in both weights (see Table 6-24). The explained part increased relatively by 2.82% in weightA and 7.68% in weightB. The wage gap was explained by 88.28% when unweighted decomposition was used which middled among IPW percentage, approximately 88.21% in weightA, while it explained 88.49% in weightB. Although the coefficients changed in relative term, those coefficients had similar explanation power in both weights of the IPW decomposition compared to CC. By that, I meant being Saudi has the highest explanation percentage over 60.56% then job attribute explained the gap by 26.61%, while 1.11% is explained by the workers' attribute when CC used. Similarly, in IPW, in weightA around 1.50% was explained by workers attribute, 22.01% by job attribute, and 0.55%, 22.21%, and 65.72% in weightB.

The total unexplained part increased relatively by 2.12% and 9.53 in weightA, and weightB in order. Workers attribute coefficients decreased by (16.11%) in weightA compared to unweighted decomposition while it increased in weightB by 54.21%. Unweighted unexplained percentage were in the middle in between IPW percentage for workers attribute, Saudi, and the intercept. For example, 185.11% of the gap was unexplained due to workers attribute when weightA was used, while this percentage was 155.07% when CC was used. The percentage decreased further and reached 135.00% in weightB. Unlike, 11.69% was unexplained due job attribute when CC used while this percentage was slightly lower when IPW used 11.21% in weightA, and 10.92% in weightB.

Table 6-24: the wage gap between workers by firms' status (omega approach).

	unweighted	weightA	weightB
Explained	-0.31994 [-295.73]	-0.31093 [-262.49]	-0.2953793 [-235.67]
Workers	-0.00404 [-12.74]	-0.00527 [-20.6]	-0.0018483 [-6.56]
Job	-0.09643 [-266.77]	-0.07759 [-224.24]	-0.0741426 [-222.04]
Saudi	-0.21947 [-200.66]	-0.22806 [-203.47]	-0.2193884 [-165.57]
Unexplained	-0.04247 [-77.45]	-0.04157 [-69.94]	-0.038427 [-69.25]
Workers	-0.56198 [-28.57]	-0.6525 [-31.17]	-0.4635699 [-22.77]
Job	-0.04237 [-14.74]	-0.03953 [-12.82]	-0.036457 [-12.73]
Saudi	-0.10031 [-106.56]	-0.06036 [-77.87]	-0.0799686 [-86.78]
_cons	0.662183 [33.03]	0.710815 [33.5]	0.5415685 [26.1]

The t-ratio in parentheses.

6.5.1.2 Pooled approach

Both weights of IPW generated explained and unexplained gap coefficients; sign and significance level with consistent to the unweighted decomposition, in total and with respect of each part compared to the CC (see **Error! Not a valid bookmark self-reference.**). The coefficients of the total explained gap increased relatively by 2.75% in weightA and 7.54% weightB. This decrease associated with an increase in the explanation percentage to 87.54% in weightB and a slightly decreased to 87.20% in weightA compared to 87.21% for unweighted decomposition. From the other side, the coefficients of the unexplained part increased relatively in weightA and weightB by 2.65%, and 10.28%, respectively. The explained part components coefficients of the CC were middled in between the coefficients produced when IPW where it relatively decreased in weight A by (31.55%) and increased in weightB by 54%. Similarly, being Saudi where the coefficient relatively decreases in weightA by (3.89%) while it increases in weightB by 0.04%. Unlike, job attribute coefficients were higher in both weights of IPW by 19.89% and 23.23%, respectively. This applied on the unexplained part components as well. The coefficients of workers attribute and being Saudi was middle between the coefficients produced by IPW. The workers attribute on the unexplained part relatively decreased by (16.10%) in weightA while it increased weightB by 17.51%, unlike, the intercept relatively increased by 7.34% in weightA while it decreased in weightB by 18.21%. However, the coefficients of job attribute and being Saudi relatively increased in both weights of IPW compared to unweighted. The changes in the explanation percentage of worker attribute on the unexplained part and the intercept were high compared to other components.

Table 6-25: the wage gap between workers by firms' status (pooled approach)

	unweighted	weightA	weightB
Explained	-0.31605 [-292.21]	-0.30737 [-259.66]	-0.2922179 [-233.3]
Workers	-0.00395 [-12.48]	-0.00519 [-20.32]	-0.0017935 [-6.38]
Job	-0.09295 [-255.4]	-0.07449 [-213.61]	-0.0713597 [-212.21]
Saudi	-0.21916 [-200.65]	-0.22768 [-203.47]	-0.2190647 [-165.57]
Unexplained	-0.04636 [-78.09	-0.04513 [-70.41]	-0.0415884 [-69.7]
Workers	-0.56207 [-28.57]	-0.65258 [-31.18]	-0.4636247 [-22.77]
Job	-0.04584 [-15.94]	-0.04263 [-13.84]	-0.03924 [-13.7]
Saudi	-0.10063 [-106.93]	-0.06074 [[-78.4]	-0.0802923 [-87.16]
_cons	0.662183 [33.03]	0.710815 [33.5]	0.5415685 [26.1]

the t-ratio in parentheses.

6.5.1.3 W (0) approach

The IPW decomposition produced consistent coefficients and significance levels compared to CC in weight A. However, it showed inconsistent coefficients only for the explained worker attributes. However, the rest of the coefficients reached the significance level of 1% in both IPW and CC (see Table 6-26).

Total coefficients of the explained part increased in weightA by 3.65% and in weightB by 8.07%. This associated with decrease in the explanation percentage compared to unweighted. Approximately 88.19% of the gap was explained when CC were used while this percentage decreased to 87.36% in weightA, and 88.02% in weightB. The inconsistent and insignificant coefficient implies that workers attribute did not explained the gap. Unweighted suggested 0.60% of the gap explained through workers attributes while it was 0.93% in weightA. Similarly, unexplained workers attribute, Saudi dummy variable and the intercept of the unweighted decomposition middle in between of the IPW while unexplained gap due job attribute was higher when CC used by less than 2%. The explanation percentage of the unexplained part was high in both weight for workers attribute and the intercept, otherwise it the change was small compared to unweighted.

Table 6-26 : the wage gap between workers by firms' status w (0) approach.

	unweighted	weightA	weightB
explained	-0.3196 [-293.43]	-0.30794 [-259.32]	-0.2938004 [-234.3]3
workers	-0.00218 [-6.69]	-0.00326 [-12.37]	0.0000832 [0.28]
job	-0.09574 [-253.38]	-0.07508 [-205.5]	-0.072764 [-207.01]
Saudi	-0.22168 [-200.98]	-0.2296 [-203.31]	-0.2211196 [-165.49]
unexplained	-0.04281 (0.000613) [-69.81]	-0.04456 (0.000654) [-68.16]	-0.0400059 (0.0006024) [-66.4]1
workers	-0.56383 [-25.88]	-0.65451 [-31.27]	-0.4655014 [-22.87]
job	-0.04306 [-16.16]	-0.04204 [-13.57]	-0.0378357 [-13.14]
saudi	-0.09811 [-124.04]	-0.05882 [-79.42]	-0.0782373 [-87.86]
_cons	0.662183 [29.9]	0.710815 [33.5]	0.5415685 [26.1]

The t-ratio in parentheses.

6.5.1.4 W (1) approach

IPW and CC produced consistent coefficients and significance level when the w (1) decomposition was considered for all weights. Indeed, there were changes (see Table 6-27). Regardless of the changes, all coefficients were significant at the level of 1%.The coefficients of the total explained part decreased relatively by (7.34%) in weightA, while it increased by 0.32% in weightB. This was associated with a slight increase on the explanation percentage to 68.99% in weightA, and 67. 66% in weightB, compared to 62.52% when unweighted decomposition was considered. The Saudi component, which form most of the explained part, explained 50.06% using unweighted decomposition, which was lower compared to IPW percentage, where this percentage were 55.77% in weightA, and 56.21% in weightB. However, workers and job coefficients for unweighted explained the middle percentage compared to IPW. Approximately 5.50% of the gap was explained through workers attribute when unweighted decomposition was considered while this percentage increased to 6.05% in weightA, unlike it decreased to 5.42% in weightB. Approximately 6.96% of the gap was explained through job attribute when unweighted decomposition was considered, while this percentage increased to 7.18% in weightA unlike it decreased to 6.03%, and 6.81% in weightB. The coefficients of unexplained part (in total) were lower when IPW used compared to unweighted decomposition where around 37.48% of the gap remained unexplained. This percentage decreased to 31.01% in weightA, and 32.34% in weightB. the intercept follows similar pattern to workers attribute while job attribute and Saudi on the unexplained part share lower percentage both weights of IPW compared to unweighted. However, using CC yielded 150.68% unexplained share due to workers attribute, 31.34% due to job attribute, 38.18% due to Saudi status, and (182.72%) due to intercept. In weightA the percentage was 180.55%, 26.05%, 26.05%, and (201.65%). In weightB, it was 134.01%, 27.10%, 33.47%, and (162.24%) in order.

Table 6-27: the wage gap between workers by firms' status w (1) approach

	unweighted	**weightA**	**weightB**
explained	-0.22658 [-190.29]	-0.24321 [-183.39]	-0.2258553 [-163.37]
workers	-0.01994 [-69.78]	-0.02133 [-59.74]	-0.0181007 [-55.44]
job	-0.02522 [-38.09]	-0.02529 [-35.34]	-0.0201277 [-30.92]
Saudi	-0.18142 [-192.05]	-0.19658 [-190.38]	-0.1876268 [-158.93]
unexplained	-0.13583 [-157.34]	-0.10929 [-117.73]	-0.107951 [-125.3]
workers	-0.54607 [-25.04]	-0.63644 [-30.37]	-0.4473175 [-21.95]
job	-0.11357 [-43.69]	-0.09183 [-30.58]	-0.090472 [-32.39]
Saudi	-0.13837 [-127.17]	-0.09184 [-80.86]	-0.1117301 [-89.67]
_cons	0.662183 [29.9]	0.710815 [33.5]	0.5415685 [26.1]

The t-ratio in parentheses.

6.5.2 *Non-Saudi sample only*

When the localised status was the decomposition reference, the total sample showed sensitivity when weightB used, where the polynomial specification excluded). Otherwise, the two methods were consistent. However, we found this result when the estimation was limited to the non-Saudi sample. The coefficients of the non-Saudi average wage in localised and non-localised firms were consistent compared to unweighted and the t ratio showed significantly at 1% level of confidence for both weights in IPW and unweighted decomposition. This also applied to the wage differences; the wage gap (see Table 6-28). However, IPW reduced the average wage for non-Saudi in both firm's status compared to using CC. Non-Saudi in non-localised firms average wage reduced relatively by (2.45%) in weightA, and (0.98%) in weightB. Similarly, in localised firms the average wage reduced by (1.26%) in weightA, and (0.97%) in weightB. This decrease in average wage associated with high reduction on the wage gap between non-Saudi according to firms' status by (75.83%) in weightA while the gap increased relatively in weightB by 0.61% compared to unweighted.

Table 6-28:non-Saudi average wage in each firms' status

	unweighted	weightA	weightB
Localised	6.738793 (16000)	6.65403 (595.46)	6.673454 (15000)
Non-localised	6.636186 (8625.09)	6.473614 (134.34)	6.571474 (8061.1)
Difference	-0.10261 (-116.96)	-0.18042 (-3.65)	-0.1019797 (-109.73)

The t-ratio between round brackets.

6.5.2.1 Omega approach

The coefficients' sign of the total explained and unexplained parts of IPW in both weights were consistent with the unweighted decomposition. Meanwhile, at least one result of IPW was inconsistent with CC, for the components of either the explained or the unexplained parts (see Table 6-29). Unweighted coefficient of total explained part was significantly lower compared to weightA by (124.46%), and slightly lower compared to weightB by (3.48%). This relative decrease associated with an increase in the explanation percentage by 18.62% in weightA and 2.76% in weightB compared to unweighted, which was 67.00%. Similarly, the coefficients of the explained part components were consistent and significant with CC in weightB, unlike weightA, although the coefficients sign was consistent with CC it was insignificant. Unlike the unexplained part components coefficients sign and significance level of unweighted were inconsistent with IPW in both weights. The changes on the unexplained part is expected however, the significance level is matter. Only unexplained worker attribute has insignificance level of confident in both IPW and CC. However, this coefficient was similar with CC in weightA. Job attribute neither the sign nor the significance level were consistent between the two methods. The intercept of unweighted was consistence on the sign with weightA.

Table 6-29: the decomposition coefficients between non-Saudi for omega approach.

	unweighted	weightA	weightB
explained	-0.06875 (-114.29)	-0.15446 (-3.21)	-0.0711367 (-118.14)
Workers	-0.0193696 (-60.81)	-0.0435897 (-1.75)	-0.0182673 (-60.39)
Job	-0.049376 (-109.61)	-0.1108731 (-4.09)	-0.0528695 (-115.68)
unexplained	-0.03386 (-48.58)	-0.02595 (-1.99)	-0.030843 (-40.67)
Workers	0.0553304 (0.63)	0.4045074 (1.48)	-0.1043959 (-0.97)
Job	-0.026484 (-6.79)	0.0486353 (1.79)	0.0077962 (1.85)
Intercept	-0.0627072 (-0.71)	-0.4790958 (-1.7)	0.0657567 (0.61)

The t-ration displayed between brackets.

6.5.2.2 Pooled approach

Unweighted decomposition coefficients of both the explained and the unexplained parts were consistent and significant at 1% like the IPW for both weights. The components' coefficients of the explained part were consistent on the sign between IPW and CC, while worker attributes showed inconsistency in terms of the significance level. Unlike, at least one component of the unexplained part was inconsistent with either the sign or the significance level. The intercept sign was consistent in weight A and CC unlike, while it was insignificant in CC and unlike weight A which was significant at 10%. Job attribute via CC was inconsistent with IPW for both the sign and the significance level (see Table 6-30).

The explained part coefficients increased with moderate percentage unless weightA it decreased by (119.27%). Accordingly, weightA explained the gap by around 79.88% in total, around 22.21% due to workers attributes and 57.67% due to job attributes while these percentage was 67.22%, 17.83%, and 49.38%, respectively in weightB compared to 64.06% for unweighted 18.81% due to workers attributes and 45.24% due job attributes. Total unexplained part coefficients of unweighted increased relatively when IPW used by 1.57% in weightA and 9.35% in weightB. The changes on the unexplained part were over 100% in both weights compared to CC.

Table 6-30: the decomposition coefficients between non-Saudi for pooled approach.

	unweighted	weightA	weightB
Explained	-0.06573 (-108.61)	-0.14412 (-3.13)	-0.0685463 (-113.13)
Workers	-0.0193 (-60.67)	-0.04008 (-1.69)	-0.0181866 (-60.19)
Job	-0.04642 (-101.96)	-0.10404 (-4.01)	-0.0503597 (-109)
Unexplained	-0.03688 (-48.7)	-0.0363 (-2.29)	-0.0334334 (-40.74)
Workers	0.055264 (0.63)	0.400995 (1.47)	-0.1044766 (-0.97)
Job	-0.02944 (-7.55)	0.0418 (1.57)	0.0052865 (1.25)
Intercept	-0.06271 (-0.71)	-0.4791 (-1.7)	0.0657567 (0.61)

The t-ratio between parentheses.

6.5.2.3 W (0) approach

The coefficients of the explained and unexplained parts were consistent and significant in both the unweighted and weighted decomposition, except for weight A where the total unexplained part was insignificant (see Table 6-31). Similarly, explained workers attributed coefficients was insignificant in weightA compared to unweighted while explained job attribute was significant and consistent to unweighted. Unexplained job attribute was insignificant in weightA and weightB unlike unweighted.

The total coefficients of the explained part decreased relatively by (88.43%) in weightA, and (2.49%) in weightB. The coefficients of the unweighted explained 64.87% of the gap, which was lower compared to IPW where it increased to 69.52% in weightA, and 66.89% in weightB. For CC, 18.44% of the gap was explained due to workers attribute and 46.42% due job attributes. This percentage decreased for workers attributes to 16.70% in weightA and 17.96% in weightB while for job attribute this percentage increased to reach 52.82% in weightA and 49.46% in weightB. The total unexplained part coefficient explained 35.13% of the gap for unweighted while it was lower when IPW used. In other words, 30.48% of the gap remained unexplained in weightA and 33.95% in weightB.

Table 6-31: the decomposition coefficients between non-Saudi for w (0) approach.

	CC	weightA	weightB
Explained	-0.066559 (-107.38)	-0.125418 (-2.73)	-0.0682175 (-109.44)
Workers	-0.018925 (-59.85)	-0.030125 (-1.1)	-0.0177793 (-58.66)
Job	-0.047634 (-100.88)	-0.095293 (-3.99)	-0.0504382 (-104.09)
unexplained	-0.036047 (-46.47)	-0.054999 (-1.36)	-0.0337622 (-41.11)
Workers	0.0548861 (0.63)	0.3910426 (1.51)	-0.1048839 (-0.98)
Job	-0.028226 (-8.15)	0.0330547 (0.88)	0.005365 (1.27)
Intercept	-0.062707 (-0.72)	-0.479096 (-1.7)	0.0657567 (0.61)

The t-ratio between round brackets.

6.5.2.4 W (1) approach

It seems that the coefficients of the explained and the unexplained parts of the unweighted decomposition were consistent and significant at 1%, while it was insignificant for weight A. At least one component of the unexplained part was inconsistent between the weighted and the unweighted method. In terms of significance level, the components of the explained and the unexplained parts were consistent with CC, unlike for weight A (see In this section, we notice that all the decomposition approaches applied on the unexplained part had at least one coefficient sign inconsistent between CC and IPW. In terms of significance level, they show consistency. This implies that the finite sample, including the missingness sample, could produce inaccurate results for the unexplained part.

Table 6-32). However, the relative decrease was high especially in weightA where the total explained coefficients decrease relatively by (201.78%) compared to unweighted, and weightB by 27.15%. Unlike, the coefficients of the total unexplained part increased relatively in both weighs. The lowest relative increase was in weightA by 0.08%, while it was 17.35% in weightB. The coefficients of the explained part for unweighted decomposition explained 37.61% of the wage gap among non-Saudis while this percentage increased to 64.55% in weightA, and 48.11% in weightB. This increase was associated with higher explanation percentage of workers attribute in weightA to 29.67% compared to 20.55% on unweighted while it slightly reduced to 19.75% in weightB. Accordingly, the source of the explanation percentage increase are job attributes where it increased in theses weights to 34.87%and 28.37% compared to 17.06% for unweighted.

In this section, we notice that all the decomposition approaches applied on the unexplained part had at least one coefficient sign inconsistent between CC and IPW. In terms of significance level, they show consistency. This implies that the finite sample, including the missingness sample, could produce inaccurate results for the unexplained part.

Table 6-32: the decomposition coefficients between non-Saudi for w (1) approach.

	CC	weightA	WeightB
Explained	-0.0385888 (-107.38)	-0.116453 (-2.73)	-0.0490661 (-46.7)
Unexplained	-0.0640175 (-46.47)	-0.063963 (-1.36)	-0.0529136 (-41.79)
Workers	-0.021082 (-59.85)	-0.053538 (-1.1)	-0.0201381 (-44.22)
Job	-0.0175061 (-100.88)	-0.062915 (-3.99)	-0.0289279 (-31.94)
Workers	0.0570436 (0.63)	0.4144556 (1.51)	-0.102525 (-0.96)
Job	-0.0583539 (-8.15)	0.0006771 (0.88)	-0.0161453 (-3.92)
Intercept	-0.0627072 (-0.72)	-0.479096 (-1.7)	0.0657567 (0.61)

The t-ration between round brackets.

6.6 Conclusion

We re-estimated Oaxaca decomposition using IPW to weight for missing wages. According to the result above, using CC could yield a consistent result with IPW when missingness follows the MAR mechanism, like Enders, 2010; VanGeest et al. (2017) suggested. This consistency is conditional to the model specified for logistic regression. Using the total sample provided consistent result for CC and IPW with some dissimilarity. This implies that non-Saudi missingness seems to be a subsample from total labour markers. This result is consistent with Seaman & White (2013) when the total sample was respected. Notice this constant was sensitive to the weight used. To clarify, when the total sample was used to decompose the wage gap between firm's localization status, we found that that workers attribute did not explained the gap in weightB for w (0) approaches only. However, Seaman & Vansteelandt's (2018) suggested that IPW in a finite-sample could produce inaccurate result, we found that omega approach and w (1) produced had consistent result when IPW used compared to CC for localized and non-localized samples. However, on the non-Saudi sample w (1) was more accurate considering weightB compared to omega. Therefore, further research on this respect might be required.

To sum up, using complete cases under MAR machinery could produce valid tentative results consistent with IPW. However, with a finite sample we were aligned with other studies that IPW would reduce biases generated by CC, which ignores missingness; but IPW can generate inaccurate results because it sensitive to the weights and approach chosen. Thus, great caution should be exercised when IPW is applied.

Chapter 7 General Conclusions

7.1 Overview

The research combined three literatures – affirmative action Nitaqat quotas, Oaxaca decomposition, and the earning functions – to address the impact of Nitaqat on the wages and wage gap between Saudi and Non-Saudi. Unlike other policies, Nitaqat is designed to enhance the employment of Saudis who received double wages compared to non-Saudis on average and who suffered from a high unemployment rate. This distinguishing feature of the Saudi labour market could lead the quota policy (Nitaqat) to produce an undesirable outcome for the target group (Saudis).

This policy was evaluated in the literature and gained huge attention because it restricted employers' choices of labour, increasing their operating costs. Keep in mind that employers, according to the rewards and penalties associated with Nitaqat classification of firms' status (localised and non-localised), will choose the best combination of workers' groups and capital to minimise their costs and satisfy the required quota percentage. Accordingly, the workers in both groups are under direct or indirect layoff risk according to the status of the firms they belong to. Non-Saudis would be under direct layoff risk at non-localised firms, unlike their peers in localised firms who are exposed to indirect layoff risk. Similarly, Saudis are exposed to indirect risk in both firms' statuses because employers reallocate Saudis themselves to satisfy the Nitaqat criteria, which would influence their wages.

Therefore, unlike other studies, we evaluated the effect of Nitaqat on wages through the interaction of the employees with this policy through our simple framework. We expect a negative relationship between wage and layoff risk, unlike the hedonic wage literature assumption, although especially that literature provided evidence of the possibility of reverse hedonic wages (Theodossiou & Vasileiou, 2007). Unlike the finding of the previous studies, Nitaqat could successfully decrease the average wage gap between the two groups because of the decrease of one group's wages at least or both groups if the layoff risks were appreciated. This reduction of the wage gap could harm at least one group's welfare. Thus, the success of this policy is associated with an increase in the target group's (Saudis) welfare not reducing the wage gap between the two groups. This is empirically applied using Oaxaca decomposition as a standard tool to evaluate the wage gap between the respective means of two groups group's mean. We used two separate cross-sections, 2013 and 2017.

7.2 Summary of finding

We summarise our findings on a chapter-by-chapter basis. In the first chapter, we displayed some important figures on the Saudi labour market regarding Nitaqat. We find that Nitaqat can lead to undesirable results for at least one group's wages because of workers' response to the policy and their distribution among the occupation categories. Therefore, there were several questions that needed to be answered empirically to understand the effect of Nitaqat on wages.

In chapter 2, we find that Nitaqat applied for Saudis who had lower unemployment and high wages, unlike the quota principle, where disadvantaged groups suffered from lower wages and unemployment. Additionally, Nitaqat is associated with other policies that could affect workers' layoff risk and distribution. Therefore, we expected that Nitaqat quotas could have perverse outcomes on wages, given that a quota is usually associated in the literature with an increase in the wages of the disadvantaged group (such as women) or both groups but a lower percentage for the original group (men). Accordingly, the quota policy is associated with a decrease in the wage gap. Thus, if the gap was decreased due to Nitaqat, this will be associated with a decrease in Saudi worker's welfare. Moreover, we found explaining the wage gap between native and immigrants in Saudi Arabia was not given attention in the literature – both theoretically and empirically.

In chapter 3, although we acquired huge datasets, the privacy restriction banned us from having more variables and tracking individual's IDs. Indeed, we appreciated the effort put into collecting the data by the MLSD. However, we found that the datasets need to be improved, especially where it contains large omissions. Fortunately, the data was following the MAR mechanism, which was expected where the missingness was found when it was not mandatory to provide them, especially for non-Saudis. Additionally, the data reflected the differences in the wage distribution regarding the research scope (Saudis and non-Saudis) where non-Saudis were distributed intensively below the lowest category of Saudi wages. As expected, we found the sample was heterogonous between the two groups and among non-Saudis.

In chapter 4, the effect of Nitaqat is examined theoretically, through our frameworks. We started our analysis from the utility function for workers which could be maximised by wage and consumption conditional on the layoff risk resulting from Nitaqat.
This contrasted with other work in the literature where the researcher investigates firms' size through cost minimisation or profit maximisation. Our framework exploited two theories which were the modern research theory and the hedonic wage theory. We found that the modern research theory (through consumption) and the hedonic wage theory (through layoff risk) can explain the wage gap between Saudis and non-Saudis. Then we used Oaxaca decomposition methodology to empirically measure the effect Nitaqat and consumption have in explaining the wage gap. We then exploited Oaxaca decomposition to address the effect of layoff risk (Nitaqat variable) on explaining the compositional differences (see Equation 5.44). We assumed that the negative sign of the coefficient of interest implied that the Saudi worker had a higher average wage because of the direct effect of Nitaqat. On the contrary, the unexplained part implies an indirect effect of Nitaqat on average wages. We display the model specification we used, as well. Moreover, we developed a new strategy to fix the identification issue of the unexplained part of the Oaxaca decomposition. This strategy is a calculation method depending on distributing the constant on categorial coefficients after we considered the omitted category. However, this finding needs to be enhanced in terms of finding the associated standard errors.

Chapter 5 provided the empirical finding that the wage gap was explained by the higher Saudi characteristics in educational qualifications, occupations, quota policies, and consumptions in both years. This result is consistent with theoretical and empirical views, such as human capital theory (Becker, 2010; Collard, 1972), and empirically, such as in (Longhi et al., 2012). Moreover, the gap resulted from segregations in occupations which was supported by some empirical evidence (Lehmer & Ludsteck, 2011; Smith & Fernandez, 2017). The occupations had less power to explain the gap in 2017, which indicates a glass ceiling issue for Saudis under Nitaqat2. Furthermore, the modern immigrant theory was able to explain the gap (through consumption) by over 20% in both years. We then exploited Oaxaca decomposition to address the effect of layoff risk (Nitaqat variable) on explaining the compositional differences (see Equation 5.44). We assumed that the negative sign of the coefficient of interest implied that the Saudi worker had a higher average wage because of the direct effect of Nitaqat. On the contrary, the unexplained part implies an indirect effect of Nitaqat on average wages. Similarly, the hedonic wage theory contributed successfully to explaining the compositional differences of the gap by 3% in 2013 and 9% in 2017 through Nitaqat variables. This implies that Nitaqat contributed to the increase of the compositional differences of the existing gap in 2017 with an increase of 6% compared to 2013. Occupation explained the gap by 6% and 13% in 2013 and 2017, respectively. This higher percentage implies a limited direct effect of the quota on the gap, similar to (Bertrand et al., 2014; 2019) 's findings. However, this indirect Nitaqat effect came from firms' resistance in both years through redistributing Saudi workers among occupations. We found that Nitaqat successfully narrowed the gap, reflecting a decrease in Saudi average wages. Accordingly, Saudi welfare decreased because of Nitaqat. This result supported by the pooled earnings function, as well, where Saudi wages decreased by around 30% between the two years. This implies that Saudis had an advantage from the segregations that existed in 2013, while Saudi workers could be redistributed to lower layers because of Nitaqat.

However, the unexplained part formed a substantial percentage of the gap in both years, which was consistent with some research findings (Hayfron, 2002; Lehmer & Ludsteck, 2011). It was noticed that the gap could be narrowed by increasing non-Saudi worker attributes. However, the gap was not closed because of the heterogeneity of the wage structure between both groups (via intercept), implying that the differences of the starting wage for both groups were due to the unobserved variables, such as the wages in the sending countries. Thus, the result was different concerning the origins of non-Saudis; the constant of workers from high background countries was higher compared to Saudis, unlike other origins. This result was supported by other research findings (Kee, 1995; Lehmer & Ludsteck, 2011; Longhi et al., 2012).

Additionally, workers in localised firms earned more than their peers in non-localised firms by 11% and 36% in 2013 and 2017, respectively. However, the gap was larger between Saudis and non-Saudis in non-localised firms compared to localised firms in 2013, but it was smaller in 2017. This implies that the two groups responded differently due to the heterogeneity of the layoff risk according to firm status. However, the gap between Saudis revealed heterogeneity in their responses between the two years because of the heterogeneity in the policies between the two years, such as the introduction of SANED (the unemployment benefit).

In 2017, they followed the hedonic wage response, unlike in 2013, which followed the theoretically expected result (Pinheiro & Visschers, 2015). However, non-Saudis' responses were not substantially different; non-Saudis in localised firms earned more by 7% and 10% in the two years, respectively, compared to their peers in non-localised firms. This could imply a double negative for non-Saudis in non-localised firms, which contradicts the use of this term found in some research, where it has been used for female immigrant groups (Boyd, 1984; Hayfron, 2002). It seems that Nitaqat caused heterogeneity in the layoff risk, resulting in different effects, both following and against the hedonic response for both groups. The larger gap in non-localised firms implies that Saudis asked for higher wages in 2013, even though they were not under the layoff risk. This result is supported in the literature as the reverse of the hedonic wage (Theodossiou & Vasileiou, 2007). However, non-Saudis responded opposite to the hedonic wage in both years, as they earned more when the deportation risk was lower in localised firms, and they earned less if they worked in non-localised firms where this risk was high.

In chapter 6, we found that using complete cases and using inverse probability weighting (IPW) provided consistent results generally in the total sample unlike the finite sample where the explained part shows substantial dissimilarity. Moreover, we found that with IPW, the results were sensitive to the weights which were used. Using IPW for Oaxaca decomposition was sensitive to the approach used where the results could be consistent with the complete cases in one approach and inconsistent with another. Omega approaches were the highest, consistent with IPW in all cases. We concluded that under the MAR mechanism Oaxaca decomposition could produce a similar conclusion for both CC and IPW in some IPW weights and some Oaxaca approaches.

7.3 Implications and recommendations

According to our findings, we can provide some recommendations. We recommend improving the data collection strategy that could provide predictions via other methodological tools and in any new scenarios. The Wages Protection Program, which was announced in 2017, seems a promising program to improve the data quality and quantity. Moreover, we recommend using the data to direct the inspection tours arranged by the MLSD. Generally, Nitaqat inefficiently contributed to reducing wage gap as the reduction stemmed from reducing the average wage of both groups, and the Saudi average wage was reduced by double. Nitaqat provided low-quality jobs concentrated around the minimum monthly wage of 3,000SR. Thus, we do not recommend linking Nitaqat to a specific wage as much as to occupations, according to the occupation's structure in each activity to avoid replacements among Saudis (worker redistribution among Saudis). It seemed that Nitaqat affect Saudi welfare on average and had a limited effect on Saudi employment, which could be because of the redistribution of workers and firms' Size. Although using a similar percentage in all administrative areas was a strong criterion, using the number of employees to select firm size caused firm size redistribution. Thus, we recommend combining other firm size measurements or using a fixed percentage to avoid firm redistribution.

However, varying the percentage among regions could redistribute the non-Saudi population and occupations toward the cities that required the lowest percentage. Non-Saudi females did not experience a double negative issue which implies that Nitaqat provided low-quality jobs for Saudis in general and Saudi women, particularly. Thus, further research on the gender gap is recommended in the light of the Nitaqat. Moreover, engaging non-Saudis in the programme percentage could also redistribute them as an alternative policy of entry quota as used in some other countries. This would require some details on those occupations, such as extensively considering the relationships between them (complements versus substitutes) for capital versus labour and among labour (qualified versus less qualified). In other words, understanding the labour market’s structural differences among the workers in firms and the automation possibility is recommended.

Moreover, rethinking the recruitment policies and fee structures is advised to treat the dumping effect that coincides with a systematic replacement policy. For example, suspending new visas in specific occupations based on the number of Saudis seeking jobs. The quickstep would be to link the fees to these target occupations so that they are at least equal to the Saudi payment to the GOSI. This would increase the non-Saudi labour cost in selected occupations. Frankly, there is no clear wage scale in the private sector, and the employers followed non-linear pricing according to the nationalities that were profitable. Thus, we recommend exploiting this distribution of non-linear pricing in the complementary occupations and increasing the non-Saudi workers' costs in the substituted occupations. This needs further information before any step is taken. Indeed, using the remittances as an indicator to detect al-tasatur could be helpful in some cases. The national anti-al-tasatur law needs to be offered some ideas that could break the relationships between Saudi sponsors and illegal non-Saudi merchants. Legalising those firms would reduce money leaking out through remittance as the merchant's life cycle end in another country; it would also guarantee legal control for those firms.

Theoretically, we recommended assessing the effect of the correlation between a variable and the index on the explained part of the Oaxaca decomposition. Additionally, investigate why a lower explained part associated with the high heterogeneity of the reference group (the index) when the pooled approach is used. Additionally, we suggest more research using IPW with Oaxaca decomposition. We recommend the use of Monte Carlo Simulation in future research.

7.4 The research contributions

This project contributes to the literature in several aspects. **First**, unique cross-sections were used. This is the first study that used individual-level wage data in Saudi Arabia using secondary data. This dataset was protected by the Privacy Policy, and it is rare to obtain permission to view it. Fortunately, the research benefitted from this access to such a unique dataset, which is restricted from public use (Mahdi, 2005). Thus, this study is the first to address an economic issue with micro-level data using this type of dataset in Saudi Arabia. **Second**, this study exploited the modern theory of immigrants and the hedonic wage theory to construct a simple framework representing the source of the wage gap between Saudis and non-Saudis. This framework considered the quota policy (represented by Nitaqat) as a source of job loss risk to explain the source of the wage gap. Additionally, the framework assumed a multi-supply for both natives and immigrants, which stemmed from the differences in consumption. **Third,** the research addressed whether the hiring quota eliminated or reduced the wage gap between Saudis and non-Saudis, considering that the primary problem (of low wages demanded by foreign workers) caused a preference for foreign workers relative to Saudi workers. Frankly, this was one clear aim of Nitaqat. This helped evaluate the programme's ability to fulfil its objectives. **Fourth,** it contributed to understanding the wage gap according to firm localisation status, which had not yet been researched in the literature; other studies have focused on the wage gap in terms of firm size, ownership, and firm age. **Fifth**, we got the opportunity to find out the effect of using IPW on the Oaxaca decomposition. **Finally**, the study introduced a simple solution for the categorical explanatory variable identification issue in the Oaxaca decomposition. This solution provided a fixed value for each categorical variable.

Chapter 8 Bibliography

1. Abdullah, B., Theodossiou, I. & Zangelidis, A. (2020) Native-immigrant wage differentials in Malaysia. *Journal of Population and Social Studies*, 28(3), 232–249.

2. Acemoglu, D. (2002) Technical change, inequality, and the labor market. *Journal of Economic Literature*, 40(1), 7–72.

3. Acock, A. C. (2014) *Gentle introduction to Stata*, 4th edition. College Station, TX: Stata Press

4. Aggarwal, C. C. (2016) *Outlier analysis,* 2ed edition. New York city, united states: Springer International Publishing

5. Agrawal, T. (2014) Gender and caste-based wage discrimination in India: some recent evidence. *Journal for Labour Market Research*, 47(4), 329–340.

6. Ahern, K. R. & Dittmar, A. K. (2012) The changing of the boards: the impact on firm valuation of mandated female board representation. *The Quarterly Journal of Economics*, 127(1), 137–197.

7. Al-Ahsa Chamber (2014) *Nitaqat guide*. Available online: http://www.hcci.org.sa/Arabic/Users/Circulars/Documents/2014/%D8%AF%D9%84%D9%8A%D9%84%20%D9%86%D8%B7%D8%A7%D9%82%D8%A7%D8%AA%20%D8%A7%D9%84%D8%B5%D8%A7%D8%AF%D8%B1%20%D9%85%D9%86%20%D9%88%D8%B2%D8%A7%D8%B1%D8%A9%20%D8%A7%D9%84%D8%B9%D9%85%D9%84.pdf [Accessed 15/06/2016].

8. Al-Ali, S. S. A. (2015) *The role of the Nitaqat Program in the localization of educational jobs in private schools From the point of view of the owners and teachers of the city of Mecca: Development proposals*. Master's thesis. Umm Al Qura University. Available online: http://b7oth.com/wp-content/uploads/2015/07/%D8%AF%D9%88%D8%B1-%D8%A8%D8%B1%D9%86%D8%A7%D9%85%D8%AC-%D9%86%D8%B7%D8%A7%D9%82%D8%A7%D8%AA-%D9%81%D9%8A-%D8%AA%D9%88%D8%B7%D9%8A%D9%86-%D8%A7%D9%84%D9%88%D8%B8%D8%A7%D8%A6%D9%81-%D8%A7%D9%84%D8%AA%D8%B9%D9%84%D9%8A%D9%85%D9%8A%D8%A9-%D9%81%D9%8A-%D8%A7%D9%84%D9%85%D8%AF%D8%A7%D8%B1%D8%B3-%D8%A7%D9%84%D8%A7%D9%87%D9%84%D9%8A%D8%A9-%D9%85%D9%86-%D9%88%D8%AC%D9%87%D8%A9-%D9%86%D8%B8%D8%B1-%D8%A7%D9%84%D9%85%D9%84%D8%A7%D9%83-%D9%88%D8%A7%D9%84%D9%85%D8%B9%D9%84%D9%85%D9%8A%D9%86-%D8%A8%D9%85%D8%AF%D9%8A%D9%86%D8%A9-%D9%85%D9%83%D8%A9-%D8%A7%D9%84%D9%85%D9%83%D8%B1%D9%85%D8%A9.-%D9%85%D9%82%D8%AA%D8%B1%D8%AD%D8%A7%D8%AA-%D9%84%D9%84%D8%AA%D8%B7%D9%88%D9%8A%D8%B1.pdf [Accessed 12/12/2019].

9. Albrecht, J., Van Vuuren, A. & Vroman, S. (2004) Decomposing the gender wage gap in the Netherlands with sample selection adjustments. Labour Economics, 16(4), 383–396.

10. Alfarhan, U. (2015) The public earnings premium in a mineral rich economy: new piece in the puzzle. *Journal of Economic & Financial Studies*, 3(01), pp.10-23.

11. Al-Farhan, U. F. & Al-Busaidi, S. (2019) Migrant earnings gaps in Gulf Cooperation Council countries: employers' perceptions or opportunity costs? *International Labour Review*, 158(2), 273–295.

12. Allanson, P. (2003) *CAP reform and the distribution of farming income in* Dundee, *Scotland*. Discussion Papers in Economics. Working Paper No. 147

13. Allanson, P. (2006) The redistributive effects of agricultural policy on Scottish farm incomes. *Journal of Agricultural Economics*, 57(1), 117–128.

14. Allanson, P. & Atkins, J. P. (2005) The evolution of the racial wage hierarchy in post-Apartheid South Africa. *The Journal of Development Studies*, 41(6), 1023–1050.

15. Allanson, P., Atkins, J. & Hinks, T. (2000) A multilateral decomposition of racial wage differentials in the 1994 South African labour market. *The Journal of Development Studies*, 37(1), 93–120.

16. AlShik, H. (2008) *Natural resources and environmental economics*. Riyadh: Obekan Library.

17. Angrist, J. D. & Keueger, A. B. (1991) Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4), 979–1014.

18. Appasani, K. & Visscher, P. M. (2016) *Genome-wide association studies: from polymorphism to personalized medicine*. Cambridge: Cambridge University Press.

19. Arcidiacono, P. (2004) Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1–2), 343–375.

20. Aristei, D. (2013) A Blinder–Oaxaca decomposition for double-hurdle models with an application to migrants' remittance behaviour. *Applied Economics Letters*, 20(18), 1665–1672.

21. Arkes, J. (2019) *Regression analysis: a practical introduction*. Routledge an imprint of the taylor and Francis group, an Informa business.

22. Arrow, K. J. (1971) *Some models of racial discrimination in the labor market*. Santa Monica : The Rand corporation.

23. Arrow, K. (1973) The theory of discrimination. *Discrimination in Labor Markets*, 3(10), 3–33.

24. Ashenfelter, O. C. & Heckman, J. J. (1974) Measuring the effect of an anti-discrimination program. Cambridge, MA: National Bureau of Economic Research.

25. Avery, T. E. & Burkhart, H. E. (2015) *Forest measurements*, 5th edition. Long grove, Illinois, USA: Waveland Press.

26. Aydemir, A. & Skuterud, M. (2008) The immigrant wage differential within and across establishments. *ILR Review*, 61(3), 334–352.

27. Barbezat, D. A. (2002) History of pay equity studies. *New Directions for Institutional Research*, 2002(115), 9–40.

28. Barrett, A., Bergin, A., Kelly, E. & McGuinness, S. (2016) Ireland's recession and the immigrant-native earnings gap. In Kahanec, M. & Zimmermann, K. (eds.) Labor migration, EU enlargement, and the great recession. Berlin: Springer-Verlag, 103–122.

29. Barrow, M. (2013) *Statistics for economics, accounting and business studies*, 6th edition. Harlow, UK: Pearson

30. Barsky, R., Bound, J., Charles, K. K. & Lupton, J. P. (2002) Accounting for the black–white wealth gap: a nonparametric approach. *Journal of the American statistical Association*, 97(459), 663–673.

31. Barth, E., Kerr, S. P. & Olivetti, C. (2017) *The dynamics of gender earnings differentials: evidence from establishment data*. Available online: https://www.nber.org/papers/w23381.pdf [Accessed 12/10/2019].

32. Bauer, T. K. (2002) Educational mismatch and wages: a panel analysis. *Economics of Education Review*, 21(3), 221–229.

33. Bauer, T. K. & Sinning, M. (2008) An extension of the Blinder–Oaxaca decomposition to nonlinear models. *AStA Advances in Statistical Analysis*, 92(2), 197–206.

34. Bauer, T. K. & Sinning, M. (2010) Blinder–Oaxaca decomposition for tobit models. *Applied Economics*, 42(12), 1569–1575.

35. Bauer, T. K. & Sinning, M. G. (2011) The savings behavior of temporary and permanent migrants in Germany. *Journal of Population Economics*, 24(2), 421–449.

36. Bauer, T., Göhlmann, S. & Sinning, M. (2007) Gender differences in smoking behavior. *Health Economics*, 16(9), 895–909.

37. Bazen, S. (2011) *Econometric methods for labour economics*. Oxford: Oxford University Press.

38. Bazen, S., Joutard, X. & Magdalou, B. (2017) An Oaxaca decomposition for nonlinear models. *Journal of Economic and Social Measurement*, 42(2), 101–121.

39. Beaurain, G. & Masclet, D. (2016) Does affirmative action reduce gender discrimination and enhance efficiency? New experimental evidence. *European Economic Review*, 90, 350–362.

40. Becker, G. S. (1971) *The economics of discrimination*. Chicago: University of Chicago Press.

41. Becker, G. S. (2010) *The economics of discrimination*, 2nd edition. Chicago: University of Chicago Press.

42. Belman, D. & Heywood, J. S. (1990) Application of the 'Oaxaca decomposition' to probit estimates: the case of unions and fringe benefit provision. *Economics Letters*, 32(1), 101–104.

43. Beng, C. S. (2017) *Labour economics and public policy: managing the labour markets for competitiveness*. World Scientific Publishing Company Pte Limited.

44. Bertrand, M., Black, S.E., Jensen, S. and Lleras-Muney, A., 2019. Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway. The Review of Economic Studies, 86(1), pp.191-239.

45. Bhattacharya, S (2009) *Compensation management*. New Delhi, Delhi India: Excel Books,

46. Bílková, D. (2019) Higher earnings in large firms? Employer size-wage relation in the Czech Republic. *Industrial Relations: A Journal of Economy and Society*, 31(2), 229–249.

47. Bin-Obaid, A. (2003) Determinants of government labour demand in the GCC states. *King Saud University*, 15(2), 67–81.

48. Bin-Obaid, A. (2003) Determinants of private sector labor demand in the GCC states. *Social Scientific Journal Kuwait*, 31(1).

49. Bin-Obaid, A. S. (2003) Determinants of employees in the GCC states. *Social Science Journal*, 7.

50. Black, D., Haviland, A., Sanders, S. & Taylor, L. (2006) Why do minority men earn less? A study of wage differentials among the highly educated. *The Review of Economics and Statistics*, 88(2), 300–313.

51. Blaug, M. (1997) *Economic theory in retrospect*. Cambridge: Cambridge University Press.

52. Blinder, A. S. (1973) Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8(4), 436–455.

53. Bloch, F. E. (1979) Labor turnover in US manufacturing industries. *The Journal of Human Resources*, 14(2), 236–246.

54. Bøhren, Ø. & Staubo, S. (2014) Does mandatory gender balance work? Changing organizational form to avoid board upheaval. *Journal of Corporate Finance*, 28, 152–168.

55. Books, V. (2015) *Intermediate I year economics (English medium) test papers: model papers, question bank, practice papers*. India: Vikram Publishers Pvt Ltd.

56. Borisov, G. V. & Pissarides, C. A. (2019) Intergenerational earnings mobility in post‐Soviet Russia. Economica, 87(345), pp.1-27.

57. Bouza-Herrera, C. N. (2013) *Handling missing data in ranked set sampling*. New York Dordrecht London: Springer Briefs in Statistics.

58. Boyd, M. (1984) At a disadvantage: the occupational attainments of foreign born women in Canada. *International Migration Review*, 18(4), pp.1091-1119.

59. Brenzel, H. & Reichelt, M. (2017) Job mobility as a new explanation for the immigrant-native wage gap: a longitudinal analysis of the German labor market. *International Migration Review*. Available online: https://onlinelibrary.wiley.com/doi/pdf/10.1111/imre.12313 [Accessed 12/09/2019].

60. Brooks, C. (2019) *Introductory econometrics for finance*. Cambridge: Cambridge University Press.

61. Brown, J. D. (2019) *Advanced statistics for the behavioral sciences: a computational approach with R*. Switzerland: Springer International Publishing.

62. Brown, C. & Medoff, J. L. (2003) Firm age and wages. *Journal of Labor Economics*, 21(3), 677–697.

63. Brown, R. S., Moon, M. & Zoloth, B. S. (1980) Incorporating occupational attainment in studies of male-female earnings differentials. Journal of Human Resources, 15(01) 3–28.

64. Burger, R. & Jafta, R. (2006) Returns to race: labour market discrimination in post-Apartheid South Africa. *Department of Economics, University of Stellenbosch*. Available online: https://ideas.repec.org/p/sza/wpaper/wpapers18.html [Accessed 12/12/2017].

65. Burger, R. & Jafta, R. (2010) *Affirmative action in South Africa: an empirical assessment of the impact on labour market outcomes*. CRISE (Centre for Research on Inequality, Human Security and Ethnicity) Working Paper No. 76, 09–36. Available online: https://d1wqtxts1xzle7.cloudfront.net/30687097/workingpaper76.pdf?1361918001=&response-content-disposition=inline%3B+filename%3DAffirmative_Action_in_South_Africa_an_em.pdf&Expires=1600299847&Signature=clhdq1wcW1qDFpicvxZ0V9Rg7OKy50gn4uKvPHXRz3SnAp3eT6vB~SYkGiSwQS1HEdFqD-laihqfR7zbyy3HIq~9wxgqleRG-4kPZ6tzgJhIV3ct3w5br4gnUSwmVd9sUHnavpam5IWpTFhpTXfInCWAxKUl0eCigM9t2Xt2YgJwecabxoH4cmU0rS2ZNLExpr28Tv5EGdK2QaUXu37djOXUwQv7SFJVN8HrlnitJ~KPlDoOFjeQD-XV-N~IlfRWBxxR7Xkx0TY6WCRyQz8aZVtoqz40-nDAOKE2qyobUA9IHddFhjrDoklzMw8telbcR4NU53rDikodixIZE076Tg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA [Accessed 12/12/2017].

66. Burger, R., Jafta, R. & von Fintel, D. (2016) *Affirmative action policies and the evolution of post-apartheid South Africa's racial wage gap*. [Lecture]. WIDER Working Paper. Unpublished.

67. Burke, A. E., Fitzroy, F. R. & Nolan, M. A. (2009) Is there a North–South divide in self-employment in England? *Regional Studies*, 43(4), 529–544.

68. Butcher, K. F. & DiNardo, J. (1998) *The immigrant and native-born wage distributions: evidence from United States censuses*. Available online: https://www.nber.org/papers/w6630.pdf [Accessed 19/09/2019].

69. Butcher, K. F. & DiNardo, J. (2002) The immigrant and native-born wage distributions: evidence from United States censuses. *ILR Review*, 56(1), 97–121.

70. Cahuc, P. & Zylberberg, A. (2004) *Labor economics*. Cambridge, Massachusetts. London, England: Massachusetts Institute of Technology (MIT) Press.

71. Cahuc, P., Carcillo, S., Zylberberg, A. & McCuaig, W. (2014) *Labor economics*, 2[ed] edition. Cambridge, Massachusetts. London, England: Massachusetts Institute of Technology (MIT) Press.

72. Cain, G. G. (1986) The economic analysis of labor market discrimination: a survey. *Handbook of Labor Minnesota: University of Wisconsin-Madison, Institute for Research on Poverty*.

73. Callea, A., Lo Presti, A., Mauno, S. & Urbini, F. (2019) The associations of quantitative/qualitative job insecurity and well-being: the role of self-esteem. *International Journal of Stress Management*, 26(1), p.46. available online: https://jyx.jyu.fi/bitstream/handle/123456789/64271/1/2017callealoprestimaunourbini.pdf [Accessed 02/10/2019].

74. Chugh, S. K. (2015) *Modern macroeconomics*. Cambridge, Massachusetts. London, England: Massachusetts Institute of Technology (MIT) Press.

75. Coad, A. (2018) Firm age: a survey. *Journal of Evolutionary Economics*, 28(1), 13–43.

76. Coate, S. & Loury, G. C. (1993) Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, 83(5) 1220–1240.

77. Cobb, J. A., Lin, K.-H. & Gabriel, P. (2016) *Growing apart: the changing firm-size wage effect and its inequality consequences*. Available online : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2785868 [Accessed 12/10/2019].

78. Cohen, J. (2013) *Statistical power analysis for the behavioral sciences, 2ed edition*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

79. Collard, D. (1972) The Economics of Discrimination. by G. S. Becker. *The Economic Journal*, 82(326), 788–790. Available online: http://www.jstor.org/stable/pdf/2230052.pdf [Accessed 21/05/2017].

80. Cook, R. D. (1977) Detection of influential observation in linear regression. *Technometrics*, 19(1), 15–18.

81. Cook, R. D. (1979) Influential observations in linear regression. *Journal of the American Statistical Association*, 74(365), 169–174.

82. Cordesman, A. H. & Obaid, N. E. (2005) *National security in Saudi Arabia: threats, responses, and challenges*. Washington: Greenwood Publishing Group.

83. Corns, A. & Schotter, A. (1999) Can affirmative action be cost effective? An experimental examination of price-preference auctions. *The American Economic Review*, 89(1), 291–305.

84. Cotton, J. (1988) On the decomposition of wage differentials. *The Review of Economics and Statistics*, 70(2), 236–243.

85. Damas de Matos, A. (2017) Firm heterogeneity and immigrant wage assimilation. *Applied Economics Letters*, 24(9), 653–657.

86. Daneshvary, N. & Weber, W. L. (1991) Sources of wage differentials between native and immigrant workers: a regional analysis. *The Review of Regional Studies*, 21(2), 119–135.

87. Delgado, M. S., Ozabaci, D., Sun, Y. & Kumbhakar, S. C. (2020) Smooth coefficient models with endogenous environmental variables. *Econometric Reviews*, 39(2), 158–180.

88. Denes, T. A. (1997) Do small business set-asides increase the cost of government contracting? *Public Administration Review*, 57(5), 441–444.

89. Desai, M., Bryson, S. W. & Robinson, T. (2013) On the use of robust estimators for standard errors in the presence of clustering when clustering membership is misspecified. *Contemporary Clinical Trials*, 34(2), 248–256.

90. Dougherty, C. (2016b) Introduction to econometrics, Fifth edition [eBook]. Oxford, United Kingdom: Oxford University Press

91. Dunlop, J. & Segrave, M. (2016) *The theory of wage determination*. New York: Springer.

92. Durlauf, S. N. & Blume, L. E. (2016) The new Palgrave dictionary of economics, 2ed edition. Palgrave: Macmillan publisher Ltd.

93. Dwivedi, D. N. (2002) *Microeconomics: theory and applications*. Delhi: Pearson Education India.

94. Ehrenberg, R. G. & Smith, R. S. (2016) *Modern labor economics: theory and public policy*. Abingdon-on-Thames: Routledge, parent company Taylor & Francis.

95. Elder, T. E., Goddeeris, J. H. & Haider, S. J. (2010) Unexplained gaps and Oaxaca–Blinder decompositions. *Labour Economics*, 17(1), 284–290.

96. Enders, C. K. (2010) *Applied missing data analysis*. New York, London: Guilford Publications.

97. Even, W. E. & Macpherson, D. A. (1990) Plant size and the decline of unionism. *Economics Letters*, 32(4), 393–398.

98. Explorer, S. (2020) *The scale of salaries and wages in Yemen 2020*. [Video]. Available online: http://www.salaryexplorer.com/ar/salary-survey.php?loc=241&loctype=1 [Accessed 05/04/20].

99. Fairlie, R. W. (1999) The absence of the African-American owned business: an analysis of the dynamics of self-employment. *Journal of Labor Economics*, 17(1), 80–108.

100. Fairlie, R. W. (2006) An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement*, 30 (4), 305-316.

101. Figart, D. M. & Marangos, J. (2013) *Living standards and social well-being*. Abingdon-on-Thames: Routledge, parent company Taylor & Francis.

102. Fleischhauer, K.-J. (2007) A review of human capital theory: microeconomics. *University of St. Gallen, Department of Economics Discussion Paper* (2007–01). Available online: http://ux-tauri.unisg.ch/RePEc/usg/dp2007/DP01_Fl.pdf [Accessed date 30/09/2019].

103. Fortin, N. M. (2006) *Greed, altruism, and the gender wage gap*. Unpublished manuscript, Department of Economics, University of British Columbia.

104. Fortin, N., Lemieux, T. & Firpo, S. (2010) *Decomposition methods in economics*. *National Bureau of Economic Research* 4, 1–102.

105. Fortin, N., Lemieux, T. & Firpo, S. (2011) Decomposition methods in economics. NBER Working Paper 16045 (eds) *Handbook of labor economics*. Cambridge: National Bureau of Economic Research NBER, 1–102. Available online: https://www.nber.org/papers/w16045.pdf [Accessed date 12/09/2019].

106. Fox, J. (ed) (2015) *Applied regression analysis and generalized linear models*,3rd edition. S New York: AGE Publications.

107. Frank, K., Phythian, K., Walters, D. & Anisef, P. (2013) Understanding the economic integration of immigrants: a wage decomposition of the earnings disparities between native-born Canadians and recent immigrant cohorts. *Social Sciences*, 2(2), 40–61.

108. Gardeazabal, J. & Ugidos, A. (2004) More on identification in detailed wage decompositions. *Review of Economics and Statistics*, 86(4), 1034–1036.

109. GaStat, General authority statistic (2014) *Annual economics survey for firms. Riyadh: General authority statistic GaStat*

110. GaStat, General Authority for Statistics (2018) Establishments Economic Survey. Saudi Arabia: General Authority for Statistics. Available online: https://www.stats.gov.sa/en/64 [Accessed 12/12/2019]

111. GaStat, General Authority for Statistics (2018). *Enterprise economic surveys 2005-2018*. Available online: https://www.stats.gov.sa/ar/64 [Accessed 12/12/2019].

112. GOSI, General Organization of Social and Insurance (2018) *Annual reports 2003-2018.* Available online: https://www.gosi.gov.sa/GOSIOnline/Open_Data_Library [Accessed 10/11/2019].

113. Goldstein, M. & Smith, R. S. (1975) Estimated impact of the antidiscrimination program aimed at federal contractors*. ILR Review*, 29(4), 523-543.

114. Gomes, M. R., de Souza, S. d. C. I., Mantovani, G. G. & de Paiva, V. F. (2020) Wage gap decomposition models: a methodological contribution. *Brazilian Review of Econometrics*, 39(2), 243–267.

115. Gomulka, J. & Stern, N. (1990) The employment of married women in the United Kingdom 1970-83. *Economica*, 57(226), 171–199.

116. Gottschalk, P. T. (1978) A comparison of marginal productivity and earnings by occupation. *ILR Review*, 31(3), 368–378.

117. gov.uk (2020) *Employ someone: step by step*. Available online: https://www.gov.uk/national-minimum-wage-rates [Accessed 05/04/2020].

118. gov.uk (2020) *National minimum wage and national living wage rates*. Available online: https://www.gov.uk/national-minimum-wage-rates [Accessed 05/04/2020].

119. Griffin, P. (1992) The impact of affirmative action on labor demand: a test of some implications of the Le Chatelier principle. *The Review of Economics and Statistics*, 74(2), 251–260.

120. Groenwold, R. H., White, I. R., Donders, A. R. T., Carpenter, J. R., Altman, D. G. & Moons, K. G. (2012) Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis. The Canadian Medical Association Journal (*Cma)j*, 184(11), 1265–1269. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3414599/ [Accessed 21/08/2019].

121. Groshen, E. L. (1991) The structure of the female/male wage differential: is it who you are, what you do, or where you work? *Journal of Human Resources*, 26(3), 457–472.

122. Hahs-Vaughn, D. L. (2016) Applied multivariate statistical concepts. Available online: https://books.google.co.uk/books?id=5UAlDwAAQBAJ [Accessed 28/06/2019].

123. Hanson, G. H. (2009) *The economics and policy of illegal immigration in the United States*. Washington DC: Migration Policy Institute.

124. Harmon, C. & Walker, I. (1995) Estimates of the economic return to schooling for the United Kingdom. *The American Economic Review*, 85(5), 1278–1286.

125. Hasebe, T. (2016) Estimating the variance of decomposition effects. *Applied Economics*, 48(20), 1902–1913.

126. Hayfron, J. E. (2002) Panel estimates of the earnings gap in Norway: do female immigrants experience a double earnings penalty? *Applied Economics*, 34(11), 1441–1452.

127. Heckman, J. J. & Wolpin, K. I. (1976) Does the contract compliance program work? An analysis of Chicago data. *ILR Review*, 29(4) 544–564.

128. Heckman, J. J., Lochner, L. J. & Todd, P. E. (2003) *Fifty years of Mincer earnings regressions*. Available online: https://www.nber.org/papers/w9732.pdf [Accessed 12/10/2019]

129. Hellerstein, J. K., Neumark, D. & Troske, K. R. (1999) Wages, productivity, and worker characteristics: evidence from plant-level production functions and wage equations. *Journal of Labor Economics*, 17(3), 409–446.

130. Henry, A. J., Hevelone, N. D., Lipsitz, S. & Nguyen, L. L. (2013) Comparative methods for handling missing data in large databases. *Journal of Vascular Surgery*, 58(5), 1353–1359. Available online: https://reader.elsevier.com/reader/sd/pii/S0741521413009452?token=DDBBC845F103094A03780BC0B7957BBF751B71F222E0E24B3E95F86DAAB5690ECB4C67382C794072C06C9B8249B6B652 [Accessed 21/08/2019].

131. Hill, R. C., Griffiths, W. E. & Lim, G. C. (2018) *Principles of econometrics*. Hoboken, New Jersey: John Wiley & Sons.

132. Himmler, O. & Jaeckle, R. (2017) Literacy and the migrant–native wage gap. *Review of Income and Wealth*, 64(3), 592-625.

133. Hinz, T. & Lechmann, D. S. (2018) *Does low job satisfaction lead to higher job mobility? Investigating the role of local labor market conditions*. Available online: https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.585042.de/038_soep2018_ps1_abstract_hinz-lechmann.pdf [Accessed 02/10/2019]

134. Hofer, H., Titelbach, G., Winter-Ebmer, R. & Ahammer, A. (2017) Wage discrimination against immigrants in Austria? *Labour*, 31(2), 105–126.

135. Holzer, H. & Neumark, D. (1999) Are affirmative action hires less qualified? Evidence from employer-employee data on new hires. *Journal of Labor Economics*, 17(3), 534–569. Available online: https://www.journals.uchicago.edu/doi/abs/10.1086/209930 [Accessed 05/08/2019].

136. Holzer, H. & Neumark, D. (2000) Assessing affirmative action. *Journal of Economic Literature*, 38(3), 483–568.

137. Holzer, H. J. & Neumark, D. (2004) *The economics of affirmative action*. California: Edward Elgar Pub.

138. Horrace, W. C. & Oaxaca, R. L. (2001) Inter-industry wage differentials and the gender wage gap: an identification problem. *ILR Review*, 54(3), 611–618.

139. Houser, J. (2009) *Nursing research: reading, using, and creating evidence*. Sudbury: Jones & Bartlett Learning.

140. Howard, L. L. & Prakash, N. (2012) Do employment quotas explain the occupational choices of disadvantaged minorities in India? *International Review of Applied Economics*, 26(4), 489–513.

141. Salary Explorer (2020) The scale of salaries and wages in Yemen 2020. Avalable online : http://www.salaryexplorer.com/ar/salary-survey.php?loc=241&loctype=1 [Accessed 04/05/2020].

142. Salary Explorer (2020) The scale of salaries and wages in Greece 2020. Available online : http://www.salaryexplorer.com/salary-survey.php?loc=84&loctype=1 [Accessed 04/05/2020].

143. Salary Explorer (2020) The scale of salaries and wages in Saudi Arabia 2020. Available online: http://www.salaryexplorer.com/salary-survey.php?loc=191&loctype=1 [Accessed 04/05/2020].

144. Hübler, D. & Hübler, O. (2006) Is there a trade-off between job security and wages in Germany and the UK? Discussion paper No. 2241. Zukunft der Arbeit: Institute for the Study of Labor. Available online: http://ftp.iza.org/dp2241.pdf [Accessed 10/10/2019]

145. Hutchens, R. M. (1983) Layoffs and labor supply. *International Economic Review*, 24(1), 37–55.

146. Jann, B. (2008) A Stata implementation of the Blinder-Oaxaca decomposition. *Stata Journal*, 8(4), 453–479. Available online: https://www.jstor.org/stable/pdf/2526114.pdf?casa_token=5T-BVAjaOpwAAAAA:D5Vhh25N4uc-77pXHpA7xB5Rfpxn-iYFg073CfLBdJrR3e0PIzuil125bA5BMR_08jUZEhgS59eZ20Ybnv1fs-v0sUdumOw6YhWCb_8mK6DSjJy5i20Z [Accessed 12/10/2019].

147. Jann, B. (2008) The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4), 453–479.

148. Javied, Z. & Hyder, A. (2009) Impact of training on earnings: evidence from Pakistani industries. Asian Social Science, 5(11), 76–85.

149. Haltiwanger, J & Lane, J & Spletzer, J (1999) Productivity differences across employers: the roles of employer size, age, and human capital. *The American Economic Review*, 89(2), 94–98.

150. Jones, F. L. (1983) On decomposing the wage gap: a critical comment on Blinder's method. *The Journal of Human Resources*, 18(1), 126–130.

151. Jones, F. L. & Kelley, J. (1984) Decomposing differences between groups: a cautionary note on measuring discrimination. *Sociological Methods & Research*, 12(3), 323–343.

152. Joona, P. A. (2010) Exits from self-employment: is there a native-immigrant difference in Sweden? *International Migration Review*, 44(3), 539–559.

153. Kabe, S., Ushiyama, R., Kinkyo, T. & Hamori, S. (2016) *Moving up the ladder: development challenges for low and middle-income Asia*. Toh Tuck Link, Singapore: World Scientific Publishing Company.

154. Kabli, S. (2015) *Unemployment in Saudi Arabia: assessment of the Saudization Program and the imperatives leading to youth unemployment*. Master's thesis. The American University in Cairo. Available online: http://dar.aucegypt.edu/bitstream/handle/10526/4273/Saud%20Kabli%20-%20Thesis.pdf?sequence=1 [Accessed 12/10/2017].

155. Kaletski, E. & Prakash, N. (2016) *Affirmative action policy in developing countries*, [Lecture]. Unpublished.

156. Kee, P. (1995) Native-immigrant wage differentials in the Netherlands: discrimination? *Oxford Economic Papers*, 47(2), 302–317.

157. Kline, P. (2014) A note on variance estimation for the Oaxaca estimator of average treatment effects. *Economics Letters*, 122(3), 428–431.

158. Kurtulus, F. A. (2012) Affirmative action and the occupational advancement of minorities and women during 1973–2003. *Industrial Relations: A Journal of Economy and Society*, 51(2), 213–246.

159. Kurtulus, F. A. (2016) The impact of affirmative action on the employment of minorities and women: a longitudinal analysis using three decades of EEO-1 filings. *35.* Journal of Policy Analysis and Management, 35(1), 34–66.

160. Lehmer, F. & Ludsteck, J. (2011) The immigrant wage gap in Germany: are East Europeans worse off? *International Migration Review*, 45(4), 872–906.

161. Leonard, J. S. (1983) *Anti-discrimination or reverse discrimination: the impact of changing demographics, Title VII, and affirmative action on productivity*. Cambridge, MA: National Bureau of Economic Research.

162. Leonard, J. S. (1984a) *Affirmative action as earnings redistribution: the targeting of compliance reviews*. Cambridge, MA: National Bureau of Economic Research.

163. Leonard, J. S. (1984b) *Splitting blacks?: affirmative action and earnings inequality within and across races*. Cambridge, MA: National Bureau of Economic Research.

164. Leonard, J. S. (1989) Women and affirmative action. *The Journal of Economic Perspectives*, 3(1), 61–75.

165. Lesik, S. A. (2018) *Applied statistical inference with MINITAB®*, 2nd edition. Boca Raton, USA: CRC Press, part of Taylor & Francis Group .

166. Li, C. (2013) Little's test of missing completely at random. College Station, TX: Stata Press.*stata Journal*, 13(4), 795–809.

167. Lin, L., Yuan, G. X., Wang, H. & Xie, J. (2019) The stochastic incentive effect of venture capital in partnership systems with the asymmetric bistable Cobb–Douglas utility. *Communications in Nonlinear Science and Numerical Simulation*, 66, 109–128.

168. Liu, A. Y. (2004) Gender wage gap in Vietnam: 1993 to 1998. *Journal of Comparative Economics*, 32(3), 586–596.

169. Liu, P.-W., Zhang, J. & Chong, S.-C. (2004) Occupational segregation and wage differentials between natives and immigrants: evidence from Hong Kong. *Journal of Development Economics*, 73(1), 395–413.

170. Lokanathan, V. (2018) *A history of economic thought*, 10th edition. New Delhi: S CHAND & Company Limited.

171. Longhi, S., Nicoletti, C. & Platt, L. (2012) Explained and unexplained wage gaps across the main ethno-religious groups in Great Britain. *Oxford Economic Papers*, 65(2), 471–493.

172. Looney, R. (2004) Can Saudi Arabia reform its economy in time to head off disaster? *Strategic Insights*, 3(1).

173. Luik, M.-A. & Steinhardt, M. F. (2016) Immigrant-native differences in stockholding – The role of cognitive and non-cognitive skills. *Journal of Empirical Finance*, 38, 103–119.

174. Luik, M.-A., Emilsson, H. & Bevelander, P. (2018) The male immigrant–native employment gap in Sweden: migrant admission categories and human capital. *Journal of Population Research*, 35(4), 363–398.

175. Lumley, T., Diehr, P., Emerson, S. & Chen, L. (2002) The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, 23(1), 151–169.

176. Machado, J. A. & Mata, J. (2005) Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4), 445–465.

177. Machina, M. & Viscusi, W. K. (2013) *Handbook of the economics of risk and uncertainty*. North-Holland: Elsevier Science.

178. MacLaury, J. (2010) *President Kennedy's EO 10925: seedbed of affirmative action. Federal History online, 42-57. Available online:* http://fd.valenciacollege.edu/file/ftua/History%20of%20EO%2010925%20and%20Affirmative%20Action.pdf [Accessed 17/07/2017]

179. Madley-Dowd, P., Hughes, R., Tilling, K. & Heron, J. (2019) The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology*, 110, 63–73.

180. Mahdi, S. E. T. (2005) The structure of employee compensation in Saudi Arabia: the case of chemical and petrochemical industries. PhD thesis. University of Hertfordshire. Available online: https://uhra.herts.ac.uk/handle/2299/14235 [Accessed 12/06/2018].

181. Manacorda, M., Manning, A. & Wadsworth, J. (2006) The impact of immigration on the structure of male wages: theory and evidence from Britain. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=937942 [Accessed 22/08/2019].

182. Manpower Council (2002) *Guide of activities, occupation, and jobs and the occupation in which work is restricted to Saudi and gradual replacement of national labour. 2ed print.* Riyadh: Studies and Research Department, Workforce Council, General Secretariat. Available online (Arabic): https://www.google.com/search?q=d8afd984d98ad984-d8a7d984d985d987d986-d8a7d984d985d982d8aad8b5d8b1d8a9-d8b9d984d989-d8a7d984d8b3d8b9d988d8afd98ad98ad986&rlz=1C1ASUC_enGB780GB780&oq=d8afd984d98ad984-d8a7d984d985d987d986-d8a7d984d985d982d8aad8b5d8b1d8a9-d8b9d984d989-d8a7d984d8b3d8b9d988d8afd98ad98ad986&aqs=chrome..69i57.786j0j15&sourceid=chrome&ie=UTF-8 [Accessed 12/02/2018]

183. Marchon, C. & Toledo, H. (2014) Re-thinking employment quotas in the UAE. *The International Journal of Human Resource Management*, 25(16), 2253–2274.

184. Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A. & Taylor, J. E. (1993) Theories of international migration: a review and appraisal. *Population and Development Review*, 19(3), 431–466.

185. Mazaheri, A. & Mazumdar, D. (2005) *The African manufacturing firm: an analysis based on firm studies in Sub-Saharan Africa*. e-library London and New York: Rutledge, Taylor & Francis group.

186. McTaggart, D., Findlay, C. & Parkin, M. (2015) *Economics*. Frenches forest: Pearson Australia (a division of Pearson Australia group Pty Ltd).

187. Meng, X. & Kidd, M. P. (1997) Labor market reform and the changing structure of wage determination in China's state sector during the 1980s. *Journal of Comparative Economics*, 25(3), 403–421.

188. Miller, C. C. M. (2014) *Women on the board: quotas have limited success.* Available online: https://www.nytimes.com/2014/06/20/upshot/women-on-the-board-quotas-have-limited-success.html#commentsContainer [Accessed 12/07/2017].

189. Ministry of Labour and social developments MLSD (2013) *Annual tactics book.* Riyadh: MLSD Ministry of labour.

190. Ministry of Labour and social developments MLSD (2014a) *Nitaqat guides* 1. Riyadh: MLSD.

191. Ministry of labour and social developments MLSD (2014b) *Saudi Arabia labour market report. Riyadh: MLSD.*

192. Montgomery, M. & Powell, I. (2003) Does an advanced degree reduce the gender wage gap? Evidence from MBAs. *Industrial Relations: A Journal of Economy and Society*, 42(3), 396–418.

193. Mori, Y. & Sakamoto, N. (2018) Economic consequences of employment quota system for disabled people: evidence from a regression discontinuity design in Japan. *Journal of the Japanese and International Economies*, 48, 1–14. Available online: http://www.sciencedirect.com/science/article/pii/S0889158317300035 [Accessed 01/06/2018].

194. Murphy, K. M. & Welch, F. (1990) Empirical age-earnings profiles. *Journal of Labor Economics*, 8(2), 202–229.

195. Neumark, D. (1988) Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources*, 23(3), 279–295.

196. Newton, L. H. (1973) Reverse discrimination as unjustified. *Ethics*, 83(4), 308–312.

197. Nielsen, H. S. (2000) Wage discrimination in Zambia: an extension of the Oaxaca-Blinder decomposition. *Applied Economics Letters*, 7(6), 405–408.

198. Nielsen, H. S., Rosholm, M., Smith, N. & Husted, L. (2004) Qualifications, discrimination, or assimilation? An extended framework for analysing immigrant wage gaps. *Empirical Economics*, 29(4), 855–883.

199. Ñopo, H. (2008) An extension of the Blinder–Oaxaca decomposition to a continuum of comparison groups. *Economics Letters*, 100(2), 292–296.

200. Nuri, A.-S. H. a. E. s. (2012) *The oil curse and labor markets: the case of Saudi Arabia*. [Lecture]. Unpublished.

201. Oaxaca, R. (1973) Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3), 693–709.

202. Oaxaca, R. L. & Ransom, M. R. (1988) Searching for the effect of unionism on the wages of union and nonunion workers. *Journal of Labor Research*, 9(2), 139–148.

203. Oaxaca, R. L. & Ransom, M. R. (1994) On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61(1), 5–21.

204. Oaxaca, R. L. & Ransom, M. (1998) Calculation of approximate variances for wage decomposition differentials. *Journal of Economic and Social Measurement*, 24(1), 55–61.

205. Oaxaca, R. L. & Ransom, M. R. (1999) Identification in detailed wage decompositions. *Review of Economics and Statistics*, 81(1), 154–157.

206. O'Brien, D. P. (2004) *The classical economists revisited*. Princeton and oxford: Princeton University Press.

207. OECD (2015) *OECD compendium of productivity indicators 2015*. Paris: OECD Publishing.

208. O'Sullivan, E., Rassel, G., Maureen, B. & Taliaferro, J. D. V. (2016) *Research methods for public administrators*. 6th edition. New York and London: Routledge, parent company Taylor & Francis.

209. Pampaka, M., Hutcheson, G. & Williams, J. (2016) Handling missing data: analysis of a challenging data set using multiple imputation. *International Journal of Research & Method in Education*, 39(1), 19–37.

210. Pande, R. and Ford, D., 2011. *Gender quotas and female leadership: a review*. World Development Report on Gender. Available online http://www.tinyurl.com/yxhp8v6z [Accessed 17/07/2017]

211. Parodi, G., Pastore, F., Nicodemo, C. & Ramos, R. (2012) Wage differentials between native and immigrant women in Spain: accounting for differences in support. *International Journal of Manpower*, 33(1), 118–136.

212. Patrinos, H. A. & Sakellariou, C. N. (1992) North American Indians in the Canadian labour market: a decomposition of wage differentials. *Economics of Education Review*, 11(3), 257–266.

213. Peck, J. R. (2014) *Can hiring quotas work? The effect of the Nitaqat Program on the Saudi private sector*. American Economic Journal: Economic Policy, 9(2), pp.316-47.

214. Pepinsky, T. B. (2018) A note on listwise deletion versus multiple imputation. *Political Analysis*, 26(4), 480–488.

215. Pinheiro, R. & Visschers, L. (2015) Unemployment risk and wage differentials. *Journal of Economic Theory*, 157, 397–424.

216. Purdie, N., Frigo, T., Stone, A. & Dick, W. (2006) Enhancing employment opportunities for Indigenous Victorians: a review of the literature. Indigenous Education Australian Council for Educational Research. Available online: https://research.acer.edu.au/cgi/viewcontent.cgi?article=1018&context=indigenous_education [Accessed 17/07/2017].

217. Purnagunawan, R. M. (2007) Earning motivation and the conventional earning function.avaliable online: https://melbourneinstitute.unimelb.edu.au/assets/documents/hilda-bibliography/student-essays-dissertations/RPurnagunawan.pdf [Accessed 03/10/2019]

218. Radchenko, S. I. & Yun, M.-S. (2003) A Bayesian approach to decomposing wage differentials. *Economics Letters*, 78(3), 431–436.

219. Rand, J. & Torm, N. (2012) The informal sector wage gap among Vietnamese micro-firms. *Journal of the Asia Pacific Economy*, 17(4), 560–577.

220. Ransom, M. R. & Megdal, S. B. (1993) Sex differences in the academic labor market in the affirmative action era. *Economics of Education Review*, 12(1), 21–43.

221. Razzolini, T., D'Ambrosio, A. & Leombruni, R. (2017) Native-migrant differences in trading off wages and workplace safety. Available online: https://www.econstor.eu/bitstream/10419/161146/1/dp10523.pdf [Accessed 17/07/2017].

222. Reilly, K. T. & Wirjanto, T. S. (1999) Does more mean less? The male/female wage gap and the proportion of females at the establishment level. *Canadian Journal of Economics*, 32(4), 906–929. Available online: https://www.jstor.org/stable/pdf/136410.pdf?refreqid=excelsior%3A3f5d6f6a33a9e60f9f1e6291299ec4a2 [Accessed 06/06/2018].

223. Reimers, C. W. (1983) Labor market discrimination against Hispanic and black men. *The Review of Economics and Statistics*, 65(4), 570–579.

224. Rios-Avila, F. (2019) A semi-parametric approach to the Oaxaca–Blinder decomposition with continuous group variable and self-selection. Econometrics, 7(2), 1-28.

225. Riyadh Economic Forum (December 2013) Labor policies and localization of jobs in the private sector. Riyadh. Available online: http://webcache.googleusercontent.com/search?q=cache:thROiRKPfWMJ:www.riyadhef.com/wp-content/uploads/2015/05/Saudization-Final-WEB-21.pdf+&cd=1&hl=en&ct=clnk&gl=uk [Accessed 12/12/2019].

226. Robinson, H. (2003) Are you experienced? British evidence on age-earnings profiles. *Applied Economics*, 35(9), 1101–1115.

227. Rodgers, W. M. (2009) *Handbook on the economics of discrimination*. Northampton: Edward Elgar Publishing Incorporated.

228. Royalty, A. B., Center, E. G. & New Haven, C. (1993) *Does job matching differ by sex?* Economic Growth Center No. 689, Yale University. Available online: https://www.econstor.eu/handle/10419/160611 [Accessed 12/12/2017].

229. Rubb, S. (2003) Overeducation in the labor market: a comment and re-analysis of a meta-analysis. *Economics of Education Review*, 22(6), 621–629.

230. Ruppert, E. (1999) Managing foreign labor in Singapore and Malaysia: are there lessons for GCC countries? Washington, DC 20433: World Bank. Available online: http://documents1.worldbank.org/curated/en/599931468758957274/109509322_20041117144524/additional/multi-page.pdf [Accessed 17/07/2017].

231. Samuele, C. & Jeffrey, S. R. (2017) Semiparametric varying coefficient models with endogenous covariates. *Annals of Economics and Statistics*, 128, 261–295.

232. Sapsford, D. & Tzannatos, Z. (1993) *The economics of the labour market*. New York: Macmillan Education Limited.

233. Scheffer, J. (2002) Dealing with missing data. *Research Letters in the Information and Mathematical Sciences*, 3, 153–160.

234. Schmidt, C. M. & Zimmermann, K. F. (1991) Work characteristics, firm size and wages. *The Review of Economics and Statistics*, 73(4), 705–710.

235. Scicchitano, S., Biagetti, M. & Chirumbolo, A. (2019) More insecure and less paid? The effect of perceived job insecurity on wage distribution. INAPP Working Paper n. 4. Roma: National Institute for Public Policies Analysis (INAPP). Available online: http://oa.inapp.org/bitstream/handle/123456789/383/Inapp_Scicchitano_Biagetti_Chirumbolo_More_Insecure_less_%20paid_2019.pdf?sequence=1&isAllowed=y [Accessed 02/10/2019].

236. Seaman, S. R. & Vansteelandt, S. (2018) Introduction to Double Robust Methods for Incomplete Data. Statist. Sci., 33(2), 184-19.

237. Seaman, S. R. & White, I. R. (2013) Review of inverse probability weighting for dealing with missing data. Statistical methods in medical research, 22(3), 278-295.

238. Sharma, R. C. (2016) *Industrial relations and labour legislation*. Delhi: PHI Learning Pvt Ltd.

239. Siddiqui, R. & Akhtar, M. R. (1998) A decomposition of male-female earnings differentials [with comments]. The Pakistan Development Review, 37(4) 885–898

240. Simón, H., Sanromá, E. & Ramos, R. (2008) Labour segregation and immigrant and native-born wage distributions in Spain: an analysis using matched employer–employee data. *Spanish Economic Review*, 10(2), 135–168.

241. Smith, W. C. & Fernandez, F. (2017) Education, skills, and wage gaps in Canada and the United States. *International Migration*, 55(3), 57–73.

242. Smith, J. P. & Welch, F. (1984) Affirmative action and labor markets. *Journal of Labor Economics*, 2(2), 269–301.

243. Sokoloff, N. J. (2014) *Black women and white women in the professions: occupational segregation by race and gender, 1960-1980*. New York, London: Routledge, part of Taylor & Francis group.

244. Sousa-Poza, A. & Henneberger, F. (2004) Analyzing job mobility with job turnover intentions: an international comparative study. *Journal of Economic Issues*, 38(1), 113–137.

245. Stabile, D. (1996) *Work and welfare: the social costs of labor in the history of economic thought*. London: Greenwood Press.

246. Steeh, C. & Krysan, M. (1996) Trends: affirmative action and the public, 1970-1995. *The Public Opinion Quarterly*, 60(1), 128–158. Available online: http://www.jstor.org/stable/2749501 [Accessed 18/06/2018].

247. Sung, Y. W., Zhang, J. & Chan, C. S. (2001) Gender wage differentials and occupational segregation in Hong Kong, 1981–1996. *Pacific Economic Review*, 6(3), 345–359.

248. Tachibanaki, T. (1996) *Wage determination and distribution in Japan*. Oxford: Clarendon Press.

249. Theodossiou, I. & Vasileiou, E. (2007) Making the risk of job loss a way of life: does it affect job satisfaction? *Research in Economics*, 61(2), 71–83.

250. Thode, H. C. (2002) *Testing for normality*. New York: CRC Press part of Taylor & Francis group Ltd.

251. Tverdostup, M. & Paas, T. (2019) Immigrant–native wage gap in Europe: the role of cognitive skills and their use at work. *International Journal of Manpower*, 40(1), 154–175.

252. Tyson, S. F., Crow, J. L., Connell, L., Winward, C. & Hillier, S. (2013) Sensory impairments of the lower limb after stroke: a pooled analysis of individual patient data. *Topics in Stroke Rehabilitation*, 20(5), 441–449.

253. Ulrick, S. W. (2012) The Oaxaca decomposition generalized to a continuous group variable. *Economics Letters*, 115(1), 35–37.

254. Urbanaviciute, I., Lazauskaite-Zabielske, J., Vander Elst, T., Bagdziuniene, D. & De Witte, H. (2015) Qualitative job insecurity, job satisfaction, and organizational commitment: the mediating role of control perceptions. *Human Resources Psychology/Psihologia Resurselor Umane*, 13(2), 217-231 .

255. VanGeest, J. B., Johnson, T. P. & Alemagno, S. A. (2017) *Research methods in the study of substance abuse*. Chicago: Springer International Publishing.

256. Vieira, E. T. (2017) *Introduction to real world statistics: with step-by-step SPSS instructions*. New York and London: Routledge, part of Taylor & Francis group.

257. Vignoles, A. (1999) Higher education, employment, and earnings in Britain. *Education Economics*, 7(3), 277.

258. Vittinghoff, E., Glidden, D. V., Shiboski, S. C. & McCulloch, C. E. (2012) Regression methods in biostatistics: linear, logistic, survival, and repeated measures models. Available online: https://books.google.co.uk/books?id=boMUM6N8V4wC [Accessed 30/07/2019].

259. Wang, M. & Kelan, E. (2013) The gender quota and female leadership: effects of the Norwegian gender quota on board chairs and CEOs. *Journal of Business Ethics*, 117(3), 449–466.

260. Weisberg, H. F. (2009) *The total survey error approach: a guide to the new science of survey research*. Chicago: University of Chicago Press.

261. Wheeler, A. R., Coleman Gallagher, V., Brouer, R. L. & Sablynski, C. J. (2007) When person-organization (mis) fit and (dis) satisfaction lead to turnover: the moderating role of perceived job mobility. *Journal of Managerial Psychology*, 22(2), 203–219.

262. Wood, J. C. (1991) *David Ricardo: critical assessments*. *Wiltshire:* Routledge, part of Taylor & Francis.

263. Wooldridge, J. M. (2008) *Introductory econometrics: a modern approach 4th edition*. Boston: Cengage Learning.

264. Wooldridge, J. M. (2015) *Introductory econometrics: a modern approach*, 6th edition. Boston: Cengage Learning.

265. Yang, K.-E. (2017) Employment gap between immigrant and native women in South Korea: discrimination in the labor market pre-entry stage. *OMNES: The Journal of Multicultural Society*, 7(2), 99–127.

266. Yin, Y. P. (2002) Skilled-unskilled wage/employment disparity - a CGE simulation analysis, *In: Procs of the International Conference on Policy Modelling, EcoMod2002*. Available online:

https://uhra.herts.ac.uk/bitstream/handle/2299/1792/901993.pdf?sequence=1 [Accessed 12/12/2019].

267. Yun, M. S. (2004) Decomposing differences in the first moment. *Economics Letters*, 82(2), 275–280.

268. Yun, M. S. (2005) A simple solution to the identification problem in detailed wage decompositions. *Economic Inquiry*, 43(4), 766–772.

269. Yun, M. S. (2007) An extension of the Oaxaca decomposition using generalized residuals. *Journal of Economic and Social Measurement*, 32(1), 15–22.

270. Zhang, Z. & Wu, X. (2012) Registration status, occupational segregation, and rural migrants in Urban china. Population Studies Centre Research Report 13-809. Michigan: population study centre, university of Michigan, instate for social research.

Chapter 9 Appendix A

9.1 average wage by gender in 2017 GaStat

		Saudi			Non-Saudi		
	Occupations	M	F	T	M	F	T
1	Directors	13,762	5,246	11,368	13,174	8,967	13,046
2	Specialists	15,983	8,479	13,631	14,683	12,463	14,514
3	Technicians	8,720	6,053	7,765	5,173	5,520	5,210
4	Clerical	7,734	4,490	6,389	7,204	5,476	7,023
5	Sales	4,946	3,761	4,396	4,035	5,205	4,063
6	Services	6,221	4,334	5,961	1,926	2,691	1,945
7	Agriculture & Fishing	5,369	3,648	5,145	1,184	1,433	1,185
8	Industrial, Chemical Operations and Food	9,470	4,023	8,197	2,342	1,927	2,328
9	Basic Engineering	6,745	4,006	6,520	1,712	2,670	1,713
	Total	8,388	4,939	7,372	2,679	4,737	2,731

9.2 Data opening and setup

At this point, we will share our experience in dealing with the secondary data provided by the MLSD. This was the most time-consuming process and very important as none of the estimations could be performed if this step was not managed.

9.2.1 Data opening

Fortunately, the research benefitted from having access to a unique dataset that is restricted from public use (Mahdi, 2005). The dataset consisted of two cross-sections of annual data for 2013 and 2017, which were provided by the MLSD. The first dataset contained around 100,000 observations, while the second provided over nine million observations. Both files required a password to be opened. The file for the 2013 data was an Excel spreadsheet, while the 2017 data was provided as a comma-separated values (CSV) file. The latter file was problematic as the password connected to the main server in the MLSD if opened through the recommended program, Microsoft Access. Thus, opening the data file was complicated; the server could not be reached from outside the MLSD. The only software that could open the data with the free connection required was SPSS - Statistical Package for the Social Sciences. Although the data was reached, the SPSS programme did not respond and collapsed during the analysis. Reading the data and understanding the contents took days. Moreover, as the file was in CSV format, there was a problem with reading the data from one economic activity.

This was recognised when tabulating the Saudi flag column; three classifications were given: Saudi, non-Saudi and poultry. This confirmed that the poultry-agriculture was divided into two columns when it should have been one column. As a result, all problematic rows had to be extracted and placed in a new file. The two data files were then merged into one new file. The final step was to resave the data as a CSV file to be opened easily in Stata.

9.2.2 Data setup.

In the beginning, the numeric variables were provided in English numbers using a string format, which could be converted to a numerical value using a Stata command. This could be done, for example, for wages and year of birth. While this was straightforward in the 2013 dataset, in the 2017 dataset, the data was complicated. Even though it was provided in string format, the complex variable was the date variable. It was provided as a string value using two different string formats, for example, "firms insert date", which made transferring them more difficult than expected. Once the date variable was encoded, missing data was given for the other format type on the same variable, which was time-consuming as well. Commands were applied to generate a new variable containing information for each individual. While the years were kept, details such as the day and month were missed. Fortunately, this did not make a huge difference as the year was still usable. Thus, the calculation of each group needed to be taken into consideration. To solve this issue, we generated a new variable exploiting the " if " command to calculate workers' ages.

Unlike the numerical data, the categorical variables were provided in the Arabic language, so translation into English was imperative to complete the research. Unfortunately, the auto-feature did not work because of the string variables, so Stata commands, including recode, label value and label define, were used to translate the data. We recoded the main string variables, such as colour, size, nationality, region, gender, occupation, activity and qualification. This strategy allowed us to recode the categorical variables while we completed the translation, which meant that each variable was recorded in order, where possible, according to variables such as colour or size, going from small to large. However, there was some difficulty finding the correct category name for occupations and activities given that a precise translation was not always possible. These two variables were time-consuming, as some jobs were not found in the dictionary. For example, "moaagip" is an occupation in which an employee is responsible for processing paperwork for government departments. They handle things like renewing passports, issuing licenses and sending and receiving mail. This occupation was translated as expeditor or pursuer, which needed an explanation. Thus, we avoided translating each category based on these two variables. Instead, we translated only the main categories to which they belonged. Therefore, we linked the observations one-by-one to the Saudi occupation classification published by the GaStat; this classification was also certified by the MLSD. We then recorded the data in English. Similar processes were necessary for economic activities. Unlike occupations, this variable was much quicker and was clarified in two steps. First, we linked the activities to the Saudi classification, which followed the international classification. However, the occupation variable differed in that respect. The standard Saudi classification for occupations was noticeably different from the international classifications and standard Arabic classifications. The main differences were after the fourth category – clerical occupation. Therefore, there was a high level of effort needed from the authorities to link the Saudi classification to the international classification. However, the Saudi classification is specifically suited to the Saudi labour market. Although they are different, these categories contain semi-skilled and unskilled workers in both Saudi and international classifications. This could prevent the research from precisely classifying those two-skill types. However, the first three categories – managers, specialists and technical – are roughly similar and considered skilled workers in both classification documents. Second, we exploited Stata for the translation. Unlike the problems raised above, education and qualification were numerical string variables, which required us to contact the MLSD to provide further information to prepare these variables. After this stage, the variables were ready for the analysis. After the variables were processed, the difficulty was in handling the sizable 2017 dataset through Stata. Thus, we used Viper to obtain some implications on the data.[121] Therefore, we had access to the data and found missing observations that need to be examined carefully.

[121] Viper is a software to help remotely access the high-performance computer.

9.3 Little's MCAR test

The null hypothesis of this test assumes missing values have similar means; otherwise, the missingness could be following one of other mechanisms.

The test value obtains as

$$d2 = \sum_{J=1}^{J} n_i (y_{0j} - \mu_{0j})^T \tilde{\Sigma}_{0j}^{-1} (y_{0j} - \mu_{0j}) \qquad 9\text{-}1$$

where y_{0j} the average of the observed sample (OS). Based on the null hypothesis, μ and $\sum$ are the estimator of maximum likelihood. The statistical value of chi-square is *d2*.

Table 9-1: Little's MCAR results.

	2013		2017			
	Wage	Quail.	Wage (1)	Wage (2)	Quail.	Educations
Observations #	100000	94312	9777328	9777328	4542744	4542744
Chi-square distance	0.4015	7277.866	529801.400	5016675.0000	1114860.0000	1114860.00
Degrees of freedom	1	2	1	9	18	18
Prob > chi-square	0.5263	0.0000	0.0000	0.0000	0.0000	0.0000

9.4 z-score test

This variation could be captured statistically by testing the difference between two proportions, applying a z-score test using the how far the conditional distribution in each category (Barrow, 2013); written as

$$t = \frac{(p_1 - p_2) - d}{\sqrt{\frac{\pi(1-\pi)}{n_1} + \frac{\pi(1-\pi)}{n_2}}} \qquad 9\text{-}2$$

Where d is the difference between the two proportions under the null hypothesis (zero when testing whether the proportions are equal). Both n_1, n_2 are sample sizes; and p_1, p_2 the proportions which are calculated as sample size divided by total sample[122]. The π is the weighted proportion for both samples[123].

[122] The proportion is equivalent (here) to the conditional distribution.
[123] Giving that $\pi = \frac{n_1 . p_1 + n_2 p_2}{n_1 + n_2}$ same source p196.

columns 3; illustrates the z-score for each category compared to the total apart from this category. The null hypothesis was rejected in all occupation categories apart from technicians – meaning missing education is different in all occupation categories regardless of occupations status, which confirms the relationship between missing and occupations.

9.5 Detecting outlier tests.

- Cook's distance value is given as:

$$C_i = \frac{\widehat{\varepsilon}_i^2}{p\,.S^2}\left[\frac{li}{(1-li)^2}\right] \qquad 9\text{-}3$$

Where $\widehat{\varepsilon}_i^2$ is the squared residual for the ith observation, p is the total number of coefficients in the regression, S^2 is the mean squared error for the model estimated, and l_i is the leverage for each observation (Lesik, 2018). C_i must be far from one, otherwise the result indicates that there is at least one influential observation.

- The studentized residual E_i^* given as:

$$E_i^* = \frac{E_i}{S_{E(-i)}\sqrt{1-l_i}} \qquad 9\text{-}4$$

Where $S_{E(-i)}$ is the standard error based on the regression excluding the outliers. E_i is the residual of the model. This value does not equal the variance. $E_i = \sigma^2(1 - l_i)$.

9.6 Sample test

- Shapiro-Wilk test

when the probability value is less than 0.05 then the null hypothesis should be rejected at the 5% significance level, which means the data is not normally distributed. It is calculated via a numerator (b^2) which is the square of the estimated value of weighted observation using generalized least-square; divided by (S^2) the sample variance multiplied by the sample size *n* minus 1:

$$W = \frac{b^2}{(n-1)\,S^2)} \qquad 9\text{-}5$$

This means that the test rejects respectively both the normal and lognormal assumption, while we got from the graphical methods that 2013 dataset was lognormal distributed.

- the Shapiro-Farmcia test statistic is expressed as:

$$W' = (a^*x)^2 / ((n-1)\, S^2) \qquad \text{9-6}$$

- Hotelling's T^2 test

by taking the square of pooled standard error for both group. , and rearranging equation 1-7, the T^2 is obtained as[124]:

$$T^2 = \frac{n_s + n_{ns}}{n_s\, n_s} (\mu_s - \mu_{ns})^T (S_p^2)^{-1} (\mu_s - \mu_{ns}) \qquad \text{9-7}$$

However, this test follows F-statistic as critical value unlike t-test for two groups. The result was significant for both datasets leading to rejection of the null hypothesis of equal means for both groups.

- Levene's test is generally written as:

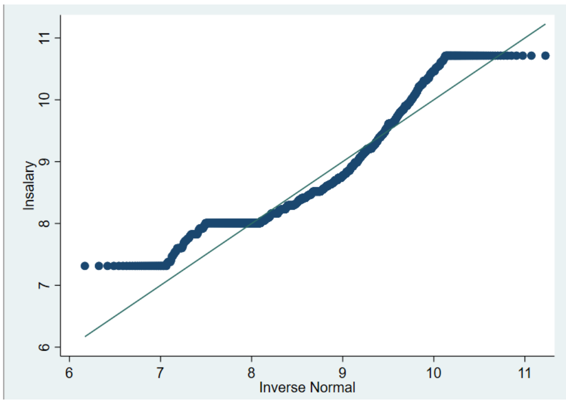
$$w_0 = \frac{\sum j\, n_j \left(Z_j - Z\right)^2 / \; g-1}{\sum j \sum i \left(\left(Z_{ij} - z_j\right)^2 / \; \sum_i (n_j - 1)\right.} \qquad \text{9-8}$$

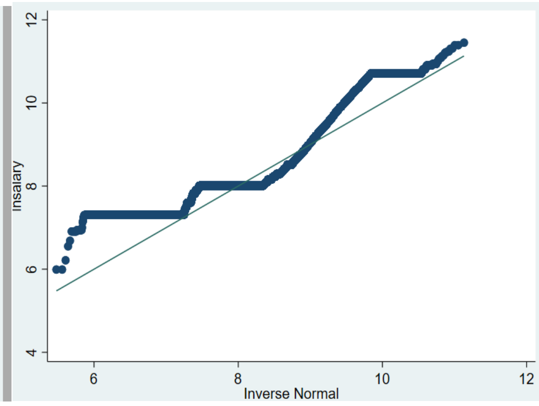
where $z_{ij} = (X_{ij} - \overline{X}_j)$ and X_{ij} is the i[th] observation of j[th] groups, $\overline{X}_j$ the mean of X in a group. The n_j is the number of observations in group j, and g represented groups enumeration- in our sample we have two groups. Replacing the mean value $\overline{X}_j$ with median or trimmed means; when Z is calculated, this results in w_{50}, and w_{10} respectively.

Figure 9-1: Q-Q plot ln(wage) for 2013 and 2017, respectively.

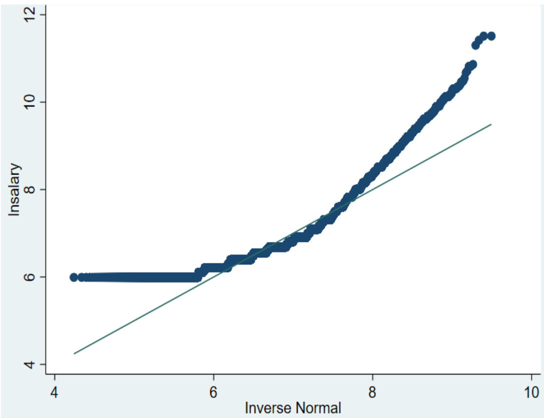
- *For Saudi*

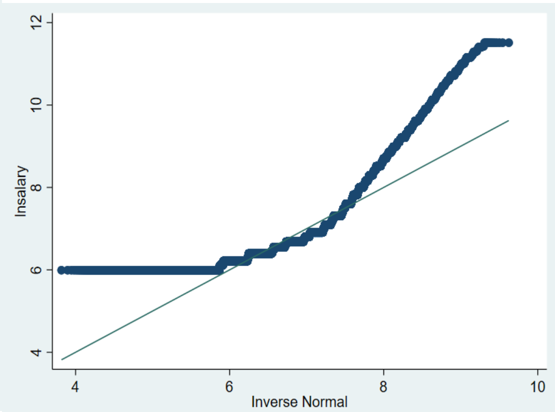
[124] At the equation, T notation denote to means covariance matrix in some literature.





- *For non-Saudi.*





9.7 Normality test STATA output

Shapiro-Wilk W test						
Variable		#	W	V	z	Prob>z
Log Salary	2013	94,312	0.89038	3215.244	22.614	0.00000
	2013	4,371,262	0.94867	8420.020	25.469	0.00000
Wage	2013	94,312	0.32562	2.0e	27.702	0.00000
	2017	4,371,262	0.52673	7.8e	31.729	0.00000
Shapiro-Francia W' test						
Variable		Observation	W	V	z	Prob>z
Log Salary	2013	94,312	0.89053	5146.767	25.771	0.00001
	2017	4,371,262	0.94868	1.0e	43.303	0.00001
Log salary for Saudi	2013	1,943	0.63198	450.082	14.601	0.00001
	2017	1,899,248	0.54469	3.9e	46.180	0.00001
Log salary for non-Saudi	2013	92,369	0.39578	2.8e	30.825	0.00001
	2017	2,472,014	0.32784	7.5e	49.238	0.00001
Wage	2013	94,312	0.32558	3.2e	31.254	0.00001
	2017	4,371,262	0.52673	9.3e	51.654	0.00001

Table 9-2: One-sample variance-comparison test for log salary (2013 & 2017).

Variable	Mean	Stoddard Deviation	Frequency

Saudi	2013	6.8676	.61854676	92,369
	2017	8.3059536	.57949079	1,899,248
Non-Saudi	2013	8.6974808	.77031805	1,943
	2017	6.7216916	.58908314	2,472,014
W0	2013	240.48595	df(1, 94310)	Pr > F = 0.000
	2017	560.77789	df(1, 4371260)	Pr > F = 0.000
W50	2013	170.21124	df(1, 94310)	Pr > F = 0.000
	2017	5253.99380	df(1, 4371260)	Pr > F = 0.000
W10	2013	191.93703	df(1, 94310)	Pr > F = 0.000
	2017	262.19913	df(1, 4371260)	Pr > F = 0.000

Table 9-3: Hotelling-test (2013 & 2017)

Log salary		**Observation**	**mean**	**Standard deviation**	**Min**	**Max**
Non-Saudi	2013	92,369	6.8676	.6185468	5.991465	11.51299
	2017	2,472,014	6.721692	.5890831	5.991465	11.51299
Saudi	2013	1,943	8.697481	.770318	7.313221	10.71442
	2017	1,899,248	8.305954	.5794908	5.991465	11.45105
group Hotelling's T-squared	2013	16467.716	F (1,94310)=	1.65e		
	2017	7878889.5	F(1,4371260)	7.88e		
F test statistic	2013	16467.716	Prob > F(1,94310)	0.0000		
	2017	7878889.5	Prob > F(1,4371260)	0.0000		

Table 9-4: t-test log salary average (2013 & 2017).

		Obs	**Mean**	**Standard Error**	**Standard Devotion**	**[95% Conf. Interval]**	
0	2013	92,369	6.8676	0.002035	0.618547	6.863611	6.871589
	2017	2472014	6.721692	.0003747	.5890831	6.720957	6.722426
1	2013	1,943	8.697481	0.017476	0.770318	8.663208	8.731754
	2017	1899248	8.305954	.0004205	.5794908	8.305129	8.306778
combined	2013	94,312	6.905299	0.002195	0.674167	6.900996	6.909602
	2017	4371262	7.41003	.0004684	.9792075	7.409112	7.410948
diff	2013		-1.82988	0.01426		-1.85783	-1.80193
	2017		-1.584262	.0005644		-1.585368	-1.583156
Ha		diff <=0	diff !=0	diff <=0	2013 t= -1.3e 2017 t= -2.8e =t		
Pr(T<t)		0	0	0			

Table 9-5: Bartlett's test result (2013 & 2017).

Source of Variance Analysis	**SS**	**df**	**MS**	**F**	**Prob> F**

Between groups	2013	6372.028	1	6372.028	16467.72	0.00
	2017	2695751.32	1	2695751.32	7.9e	0.00
Within groups	2013	36492.37	94310	0.386941		
	2017	1495620	.66437126	.342148639		
chi2(1)	2013	211.5974***				
	2017	578.1375***				

Table 9-6: multivariate analysis of variance (2013 &2017).

		Source	Statistic	df	F(df1,	df2)= F		Prob>F	
Saudi	2013	Wilks' lambda	0.791	1	2	94309	12462.26	0.000	e
	2017		0.2778	1	2	4.4e	5.7e	0.000	e
	2013	Pillai's trace	0.209		2	94309	12462.26	0.000	e
	2017		0.7222		2	4.4e	5.7e	0.000	e
	2013	Lawley-	0.2643		2	94309	12462.26	0.000	e
	2017	Hotelling	2.5996		2	4.4e	5.7e	0.000	e
	2013	Roy's largest	0.2643		2	94309	12462.26	0.000	e
	2017	root	2.5996		2	4.4e	5.7e	0.000	e
			2013	2017	e = exact,				
Residual			94310	4371260					
Total			94311	4371261					
Number of observations			94,312	4,371,262					

Table 9-7:equal groups mean assume homogeneous (2013 &2017)

		Statistic	F(df1,	df2)	F	Prob>F	
Wilks' lambda	2013	0.8513	1	94310	16467.72	0.000	e
	2017	0.3568	1	4.4e	7.9e	0. 0000	e
Pillai's trace	2013	0.1487	1	94310	16467.72	0.000	e
	2017	0.6432	1	4.4e	7.9e	0. 0000	e
Lawley-Hotelling	2013	0.1746	1	94310	16467.72	0.000	e
	2017	1.8024	1	4.4e	7.9e	0. 0000	e
Roy's largest root	2013	0.1746	1	94310	16467.72	0.000	e
	2017	1.8024	1	4.4e	7.9e	0. 0000	e

e = exact

Table 9-8: equal groups mean assume homogeneous.

		MNV	Prob > F
F(1,1995)	2013	10817.52	0.000
F(1,4117931.1)	2017	7.91e	0.000

Chapter 10 Appendix B

Some of the variables can be classified in a specific base. For example, the zone can be classified as they distributed geographically: south, north, west, east and central while this way would achieve a specific conclusion but could be not statistical the statistical term which would reflect on the estimation quality. Thus, the F-ratio test used to choose the best specifications according to the statistical measurements. The test requires nested models to be compared. The calculation F formula follows

$$F= \frac{SSR_R - SSR_{UR}/(df_R - df_{UR})}{SSR_{UR}/df_{UR}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..10\text{-}1$$

SSR is the sum of square residuals, df freedom degree, the notation R restricted models and UR the unrestricted model, See (NA et al., 2016).

10.1 Variable description.

Table 10-1 and Table 10-2 below are showing the result of the F- test for several nested models to find out the best specification of the estimation function. Critical variables were only reported. Other variables such as economic activities, occupation and educations were not reported because their coefficients tests were significant. This means there is no need to restrict them. Moreover, reporting occupation and activities were interesting to be reported as it was classified.

Table 10-1: F test result for nested models; 2013 dataset.

Null hypothesis	F-test	F-critical			P-value
		0.1	0.05	0.01	
Merge size (micro= small)	0	0	0	0	0
age3 =0	21.45	2.706	3.842	6.635	0.0000
nationality=0	277	1.326	1.436	1.66	0.0000
nationality= nation3	1.375	1.421	1.571	1.879	0.1219
Qualifications = Qualification	0.512	1.632	1.88	2.408	0.8671
zone = regions	1.001	1.774	2.099	2.802	0.4228
zone = regions23	1.196	1.847	2.214	3.017	0.3083
Original = reduced	1.089	1.316	1.423	1.639	0.330
Rigions = rigion23	0.025	2.706	3.842	6.635	0.8748
nation3 = nation status	3.412	1.546	1.752	2.185	0.0001
nation3_nonlocalizes =0	3.411	1.546	1.752	2.185	0.0001
nation3 = nation3_locslized	3.411	1.546	1.752	2.185	0.0001

Table 10-2: F test result for nested models; 2017 dataset.

null hypothesis	F-test	F-critical			P-value
		0.10	0.05	0.01	
age3 = 0	201.680	2.71	3.84	6.63	0.0000
micro= small	1.328	2.71	3.84	6.63	0.2492
Saudi=0	3937152.517	2.71	3.84	6.63	0.0000
Qualifications_3= qualifications	1.234	1.85	2.21	3.02	0.2900
zone=regions2	1.731	2.30	3.00	4.61	0.1772
Interaction variable=0	255.765	2.71	3.84	6.63	0.0000

Table 10-3: categorical variables in details.

	2013	2017
Qualifications	colleges of literature=1 Dentistry and administration science and Languages and translation =2 Colleges of Education and Colleges of Agriculture =3 *Colleges of Science=4 Colleges of Pharmacy =5 Colleges of Medicine=6 Engineering Faculties=7 Architecture and Planning & technical College=8 Colleges of Medical and Applied Science=9 Faculty of Computer Science and Inform=10 high school=11 industrial high School and trading School=12 Secondary agricultural and Veterinary Training and Animal *Product =13 Institute =14 Including (Institute of Management, Technical Institute, air science, Institute of Professional Observers) Health Institute=15 Unregistered=16	college of literature=1 Languages and translation and Colleges of Science and Institute of Management=2 administration science =3 Colleges of Education =4 Colleges of Agriculture=5 Colleges of Pharmacy =6 Colleges of Medicine=7 Engineering Faculties =8 Architecture and Planning =9 Colleges of Dentistry and Faculty of Computer Science and Inform and Management=10 Technical Institute and Colleges of Medical and Applied Science=11 High school =12 Secondary trading=13 industrial high School=14 Secondary agricultural =15 technical College=16 School of law =17 School of Economics 18 Unregistered=19
Educations		Illiterate=1 reads and writes=2 Primary degree =3 Intermediate degree=4 secondary degree=5 Diploma=6 Bachelor’s degree =7 Master’s degree =8 PhD degree=9 Higher diploma =10 Fellowship=11 unregistered =12

Firm size	Small and micro =1 Medium =2 Big =3 Giant =4	Small Medium A Medium B Medium C Big Giant
Zone	Al-Jouf & Qassim=1 Riyadh=2 East borders including (North Bordered and Eastern province) =3 Makkah =4 Tabuk and Najran=5 Hail and Asir=6 Others=7 Including (AL-Baha, Madinah, and Jazan)	AL-Baha and Northern Borders and Najran=1 AL-Jouf=2 Riyadh=3 Easter province=4 Qassim=5 Madinah =6 Tabuk =7 Jazan =8 Hail =9 Asir =10 Makkah=11
Nationalities	Saudi=1 Nepalese=2 South African and Somali and Jordanian=3 Mali=4 Pakistani and Afghanistan =5 Indian and sued=6 Sudanese and Philippine =7 Bangladesh =8 Euroupean1 =9 Including (Portuguese, British, and CAD) European2 =10 Including (American, Brazilian, and Croatian) Syrian and China and Turkish =11 Other Asians=12 Including (Indonesian, Thai, and Sri Lanka) African 2=13 Including (Ethiopian, Algerian, and Egyptian) Others =14 Includes (Slovakian, Venezuelan, Cyprus, and Greek) Palestinian and Yemeni and Mauritanian=15	-
Origin	European=1 Includes (Brazilian, Portuguese, British, Slovakian, Swiss, Venezuelan, Cyprus, Croatian, Greek) CAD and USA =2 Saudi =3 Asian =4 Afghanistan, Indonesian, China National, Pakistani, Bangladesh, Turkish, Sri Lanka, Philippine, Nepalese, Indian African =5	-

	Ethiopian, South African, Mali and Somali Arabic =6 Jordanien, Algerian, Soudanaise, Syrien, A Palestinien document, Egyptien, Mauritanien, Yémen	

10.2 Earning functions.

10.2.1 Some test for the earning function

Table 10-4: correlation matrix 2013.

	e	Log - earning	age	age2	age3	Qualification	female
E	1.0000						
Log -earning	0.6721	1.0000					
age	-0.0000	0.1371	1.0000				
age2	-0.0000	0.1499	0.9954	1.0000			
age3	-0.0000	0.1595	0.9835	0.9963	1.0000		
Qualification	0.0000	-0.3797	0.0050	0.0003	-0.0036	1.0000	
female	-0.0000	0.0434	0.0112	0.0128	0.0141	-0.0321	1.0000
regions	-0.0000	-0.0345	0.0339	0.0316	0.0297	0.0418	0.0021
colour	-0.0000	0.1709	0.0112	0.0120	0.0123	-0.1025	0.0382
SIZE	0.0000	0.1308	- 0.0026	- 0.0019	-0.0016	-0.0667	0.0589
Occupations	0.0000	-0.3895	0.0025	- 0.0008	-0.0040	0.2284	- 0.0719
Activities	0.0000	0.0131	- 0.0235	- 0.0209	-0.0184	-0.0325	0.1428
Nationalities/ localized	-0.0000	0.0353	0.0496	0.0497	0.0500	-0.0117	0.0035
Nationalities/ Non-localized	0.0000	-0.0160	0.0136	0.0145	0.0153	0.0213	- 0.0102
	regions	colour	SIZE	SIZE	Occupation	nation	nation
regions	1.0000						
colour	-0.1256	1.0000					
SIZE	-0.2182	0.2857	1.0000				
occupations	0.0110	-0.0796	- 0.0400	1.0000			
Activities	0.0241	-0.0960	- 0.0150	- 0.0761	1.0000		
nation3_lo~d	0.0624	0.1223	- 0.0495	- 0.1506	0.0163	1.0000	
nation3_no~d	0.0170	-0.4305	- 0.0815	0.0077	-0.0012	-0.3303	1.0000

Table 10-5: correlation matrix 2017.

	Log salary	age	age2	age3	Educations	Qualifications3
Log salary	1					
Age	-0.3493	1				
age2	-0.3194	0.9893	1			
age3	-0.2862	0.9622	0.9915	1		
Education	-0.1325	0.2424	0.2209	0.1989	1	
Qualification	-0.4452	0.2974	0.2754	0.2501	0.1506	1
female	0.2769	-0.1898	-0.1836	-0.1734	-0.1591	-0.3344
colour	0.3121	-0.1173	-0.1157	-0.1111	-0.0386	-0.109
SIZE1	0.2967	-0.1169	-0.123	-0.1243	-0.0401	-0.0234
regions2	-0.0684	0.0716	0.0724	0.0718	0.0884	0.0203
firm age	0.0317	0.0543	0.051	0.0475	0.0054	0.0412
firm_age2	-0.0017	0.0209	0.0195	0.0179	-0.0087	0.0158
Activities	0.1326	-0.0531	-0.0535	-0.052	0.0075	-0.149
occupations	-0.4354	0.1731	0.164	0.1514	0.022	0.3348
Saudi	0.802	-0.5436	-0.5018	-0.456	-0.2981	-0.4603
Non-Saudi non-localize	-0.2549	0.1555	0.1447	0.1327	0.0859	0.1434
e	0.468	0	0	0	0	0
	female	colour	SIZE1	regions2	firm age	firm2
female	1					
colour	0.0281	1				
SIZE1	-0.0682	0.3037	1			
regions2	-0.0076	-0.0869	-0.1231	1		
Firm age	-0.0759	0.1477	0.2738	-0.0958	1	
firm_age2	-0.0287	0.1239	0.1163	-0.0565	0.8871	1
Activities	0.1455	-0.0245	0.0223	-0.017	-0.1425	-0.1336
occupations	-0.2616	-0.1099	-0.0375	0.0148	0.0316	0.0155
Saudi	0.4349	0.2364	0.1652	-0.0629	-0.0236	-0.0134
non-Saudi non-localized	-0.1288	-0.5765	-0.1449	0.0447	-0.0607	-0.0423
e	0	0	0	0	0	0
	Activities	occupations	Saudi	non-Saudi non localized	e	
Activities	1					
occupations	-0.1825	1				
Saudi	0.1151	-0.3395	1			
non-Saudi non-localize	0.0114	0.1026	-0.2828	1		
e	0	0	0	0	1	

Table 10-6: Shapiro-Wilk W test

Error term (e)	Observation	W	V	Z	Probability
2013	94,312	0.97412	759.208	18.572	0.00000
2017	4,371,262	0.95609	7202.445	25.029	0.00000

Table 10-7: Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

fitted values of log earning	2013	2017
Ho: model has no omitted variables		
chi2(1)	15542.88	81743.46
Prob > chi2	0.0000	0.0000

Table 10-8: sum of error term.

Error term (e)	Observation	Mean	Standard deviation	Min	Max
2013	94,312	-6.49e-11	.4530971	-2.442881	4.92734
2017	4,371,262	-1.20e-11	.4582675	-2.71781	5.013273

Table 10-9: link test estimation result.

Log-earning		Coef.	Standard error	t	P>|t|	[95%Confident Interval]	
_hat	2013	0.482102	0.046978	10.26	0	0.390025	0.57418
_hat-square	2013	0.0352	0.003187	11.05	0	0.028954	0.041446
_cons	2013	1.889031	0.172233	10.97	0	1.551457	2.226606
_hat	2017	0.191968	0.0057057	33.65	0	0.1807851	0.203151
_hat-square	2017	0.0534823	0.0003773	141.76	0	0.0527428	0.0542217
_cons	2017	3.010859	0.021323	141.2	0	2.969067	3.052651

Table 10-10: Ramsey RESET test using powers of the fitted values of log earning.

fitted values of log earning	2013	2017
Ho: model has no omitted variables		
F(3, 94228)	20725.55	535.18
Prob > F =	0.0000	0.0000

Table 10-11: variance inflation factor (using both interaction variable) 2013.

Variable	VIF	1/VIF	occupation	VIF	1/VIF
age	4064.39	0.000246	4	1.09	0.921132
age2	17862.96	0.000056	5	1.38	0.72224
age3	5035.42	0.000199	6	1.72	0.582175
Qualification			7	1.13	0.885457
1	1.01	0.994858	8	1.15	0.871509
2	1.07	1	Activities		
4	1.02	0.982508	1	1.14	0.875867
5	1.05	0.948311	2	1.22	0.818594
6	1.01	0.988679	4	1.47	0.680067
7	1.04	0.957253	5	1.14	0.879631
8	1.05	0.951185	6	1.08	0.928595

9	1.01	0.98818	7	1.12	0.893644
10	1.03	0.975486	8	1.09	0.918871
11	1.02	0.980601	9	1.27	0.785009
12	1.07	0.936804	10	1.03	0.973565
13	1.01	0.987605	11	1.14	0.87459
14	1	0.998148	nation3_lo~d		
15	1.02	0.983644	1	1.6	0.624044
16	1.05	0.950588	2	1.02	0.981524
female	1.21	0.829749	3	1.19	0.839713
regions			4	1.01	0.988714
1	1.07	0.930399	5	3.24	0.308472
2	1.27	0.784862	6	4.42	0.226104
3	1.47	0.680338	7	2.2	0.453596
5	1.01	0.988043	8	3.68	0.271885
6	1.14	0.877043	9	1.02	0.981724
7	1.21	0.827251	10	1	0.99561
colour			11	1.15	0.867771
1	3.31	0.302068	12	1.14	0.880603
2	6.3	0.158813	13	2.33	0.429524
3	6.3	0.15883	14	1	0.996326
4	1.45	0.688045	nation3_no~d		
6	1.61	0.622048	1	1.07	0.934452
7	1.45	0.688293	2	1.02	0.976791
8	1.3	0.766808	3	1.13	0.882343
9	1.19	0.843346	4	1.01	0.993465
SIZE			5	3.85	0.260005
1	2.38	0.420374	6	5.17	0.193475
2	1.87	0.535839	7	1.82	0.549589
4	1.39	0.718929	8	3.83	0.261245
occupations			9	1.01	0.993413
1	1.18	0.84817	11	1.13	0.881085
2	1.57	0.637917	12	1.09	0.919543
3	1.29	0.776537	13	2.53	0.395323
Mean VIF = 338.6					

Table 10-12: variance inflation factor (using one interaction variable)2013.

Variable	**VIF**	**1/VIF**		**VIF**	**1/VIF**
age	4064.39	0.000246	4	1.09	0.921132
age2	17862.96	0.000056	5	1.38	0.72224
age3	5035.42	0.000199	6	1.72	0.582175
Qualification			7	1.13	0.885457
1	1.01	0.994858	8	1.15	0.871509
2	1.07	0.931238	Activities		

4	1.02	0.982508	1	1.14	0.875867
5	1.05	0.948311	2	1.22	0.818594
6	1.01	0.988679	4	1.47	0.680067
7	1.04	0.957253	5	1.14	0.879631
8	1.05	0.951185	6	1.08	0.928595
9	1.01	0.98818	7	1.12	0.893644
10	1.03	0.975486	8	1.09	0.918871
11	1.02	0.980601	9	1.27	0.785009
12	1.07	0.936804	10	1.03	0.973565
13	1.01	0.987605	11	1.14	0.87459
14	1	0.998148	nation3		
15	1.02	0.983644	2	1.1	0.910995
16	1.05	0.950588	3	1.73	0.579146
female	1.21	0.829749	4	1.06	0.942144
regions			5	10.47	0.095499
1	1.07	0.930399	6	14.65	0.06827
2	1.27	0.784862	7	5.81	0.172142
3	1.47	0.680338	8	11.5	0.086939
5	1.01	0.988043	9	1.06	0.941103
6	1.14	0.877043	10	1.01	0.994076
7	1.21	0.827251	11	1.63	0.612872
colour			12	1.56	0.642752
1	3.31	0.302068	13	6.46	0.154816
2	6.3	0.158813	14	1.01	0.992302
3	6.3	0.15883	15	5.09	0.196633
4	1.45	0.688045	Nationality Non-localized		
6	1.61	0.622048	1	1.08	0.928625
7	1.45	0.688293	2	1.07	0.934767
8	1.3	0.766808	3	1.16	0.863098
9	1.19	0.843346	4	1.02	0.978166
SIZE			5	3.95	0.253105
1	2.38	0.420374	6	5.3	0.188757
2	1.87	0.535839	7	1.86	0.537355
4	1.39	0.718929	8	3.92	0.254815
occupations			9	1.02	0.978743
1	1.18	0.84817	11	1.17	0.857969
2	1.57	0.637917	12	1.11	0.901592
3	1.29	0.776537	13	2.6	0.384445
Mean VIF 339.08					

10.2.2 graphs

Figure 10-1:the relationship between pronominal age and log earning (2013).

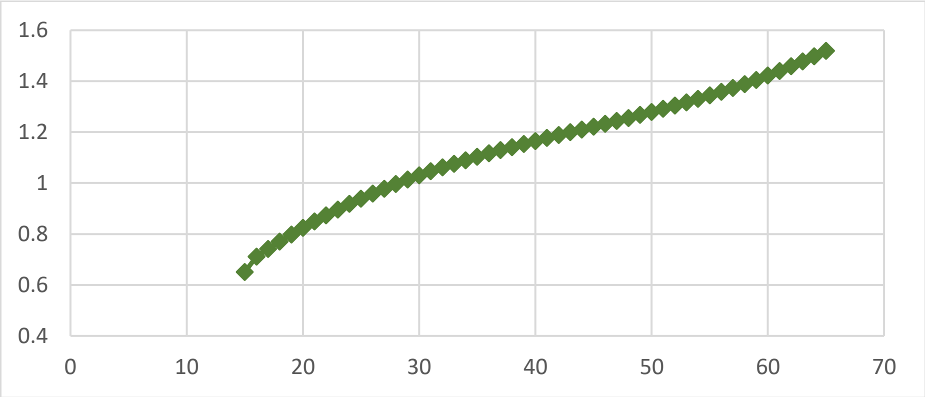


Figure 10-2: the growth rate of earning taking U shape (2013)

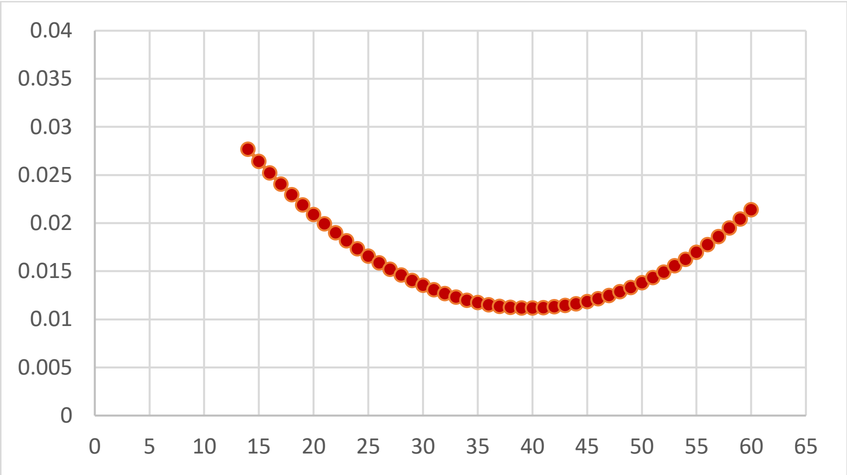


Figure 10-3: kernel density of the error term 2013

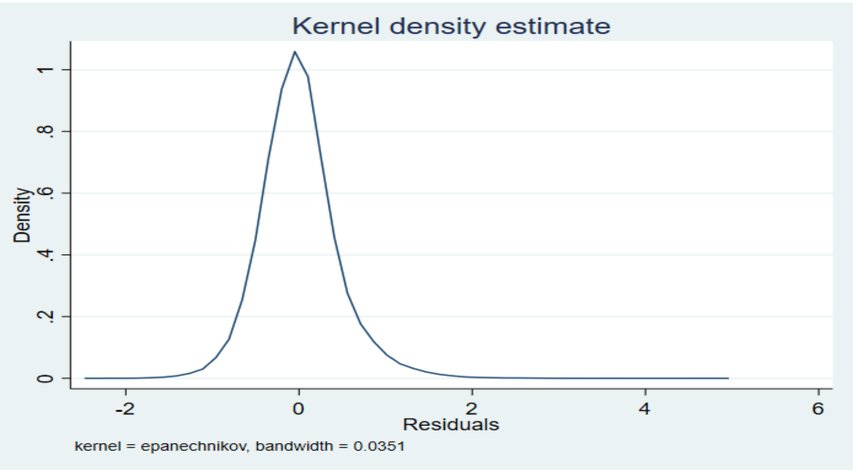


Figure 10-4: quantile–normal plots 2013

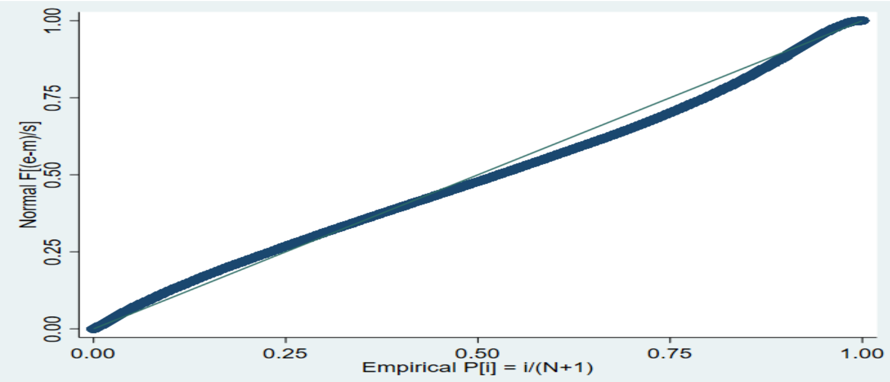


Figure 10-5:comparing the residual distribution to the normal distribution 2013.

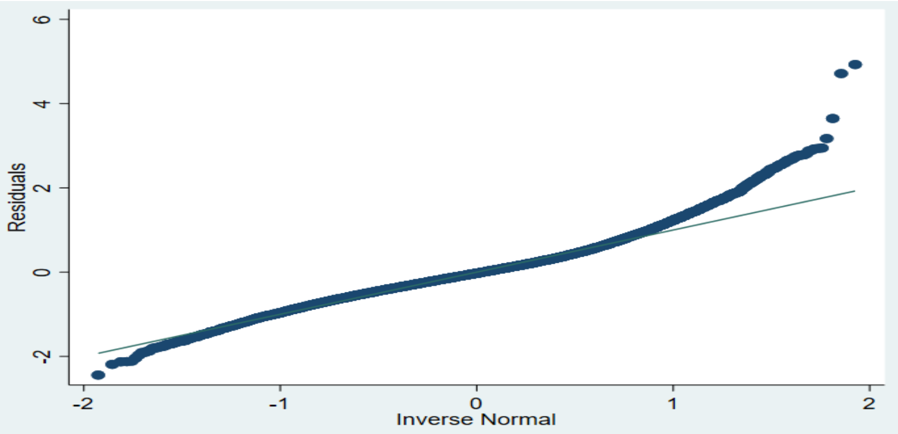


Figure 10-6: plot fitted value against residual 2013.

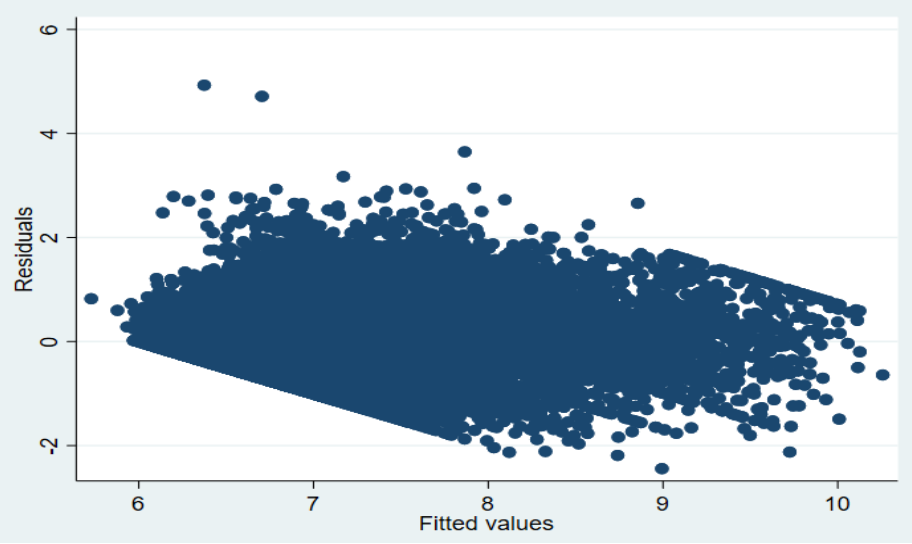


Figure 10-7: plot of the dependent variable (log salary) and the fitted value 2013.

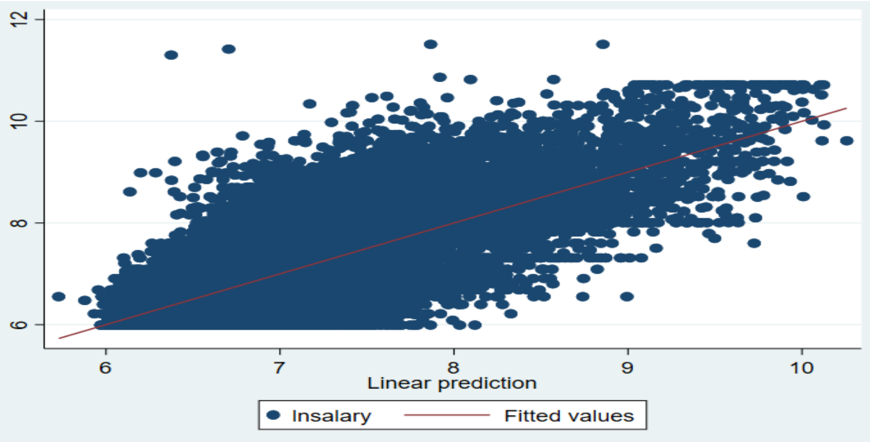


Figure 10-8: the relationship between pronominal age and log earning (2017)

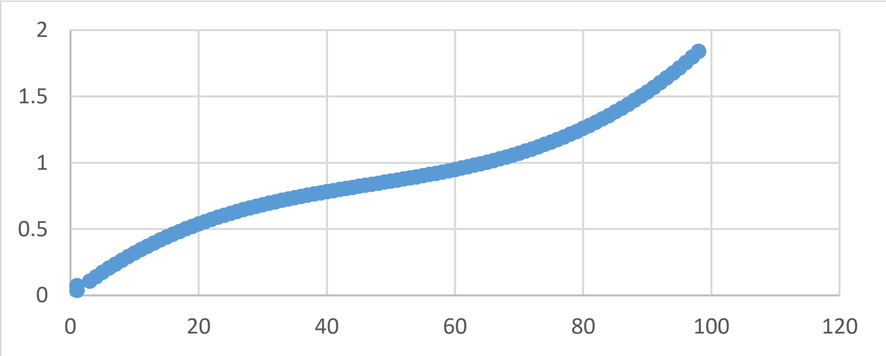


Figure 10-9: the growth rate of earning taking U shape (2017)

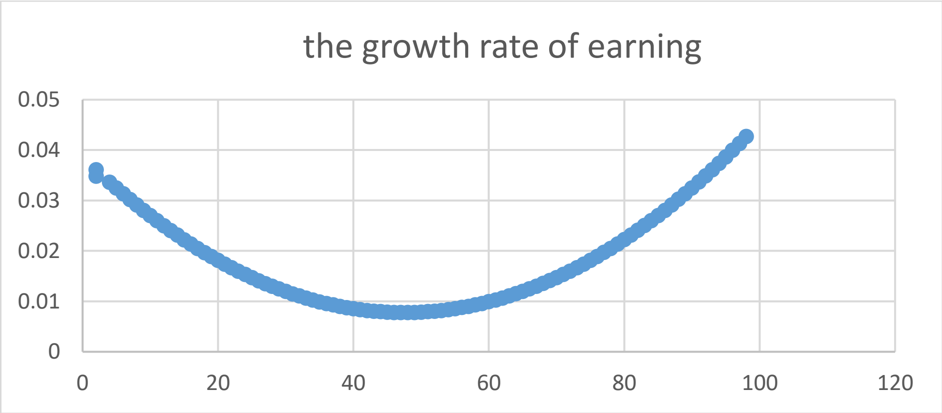


Figure 10-10: kernel density of the error term 2017

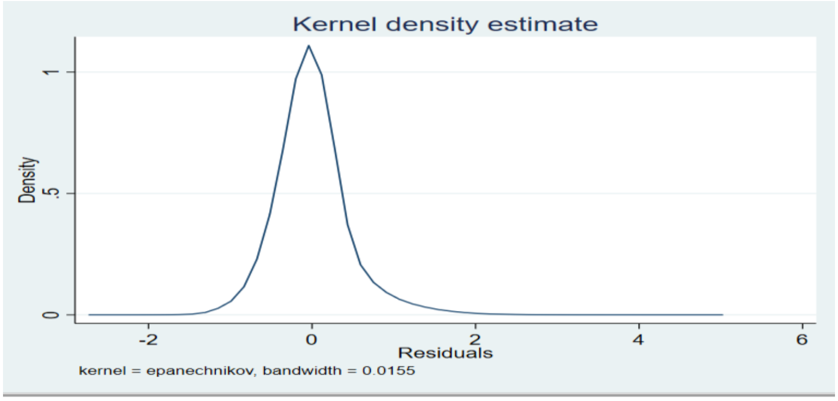


Figure 10-11: quantile–normal plots 2017

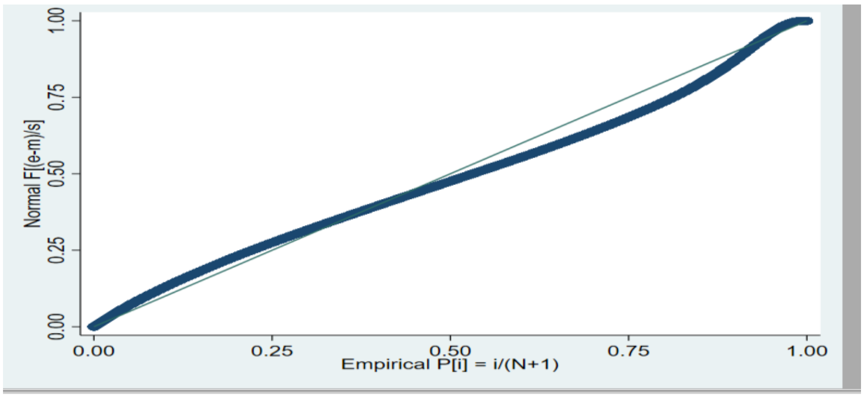


Figure 10-12: comparing the residual distribution to the normal distribution 2017

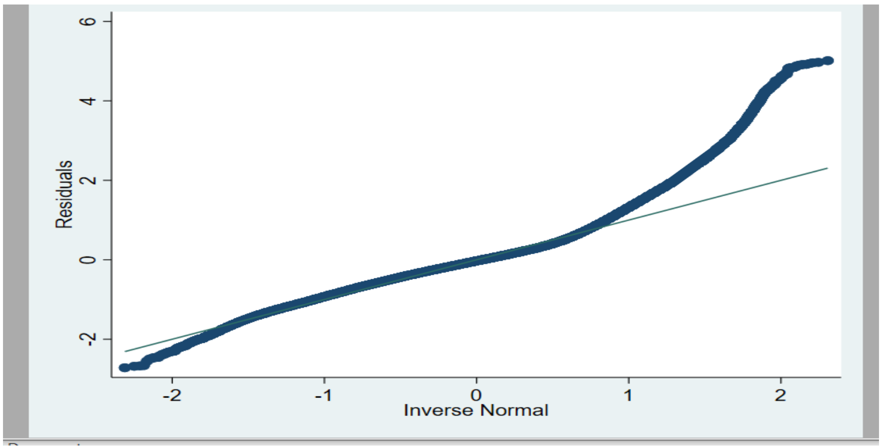


Figure 10-13: plot fitted value against residual 2017

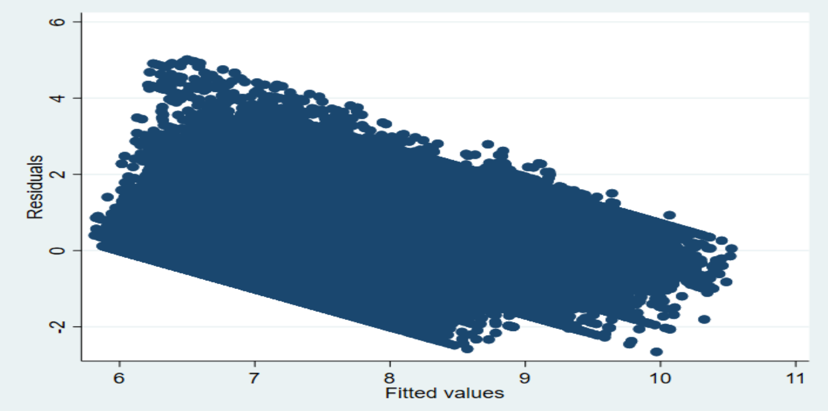
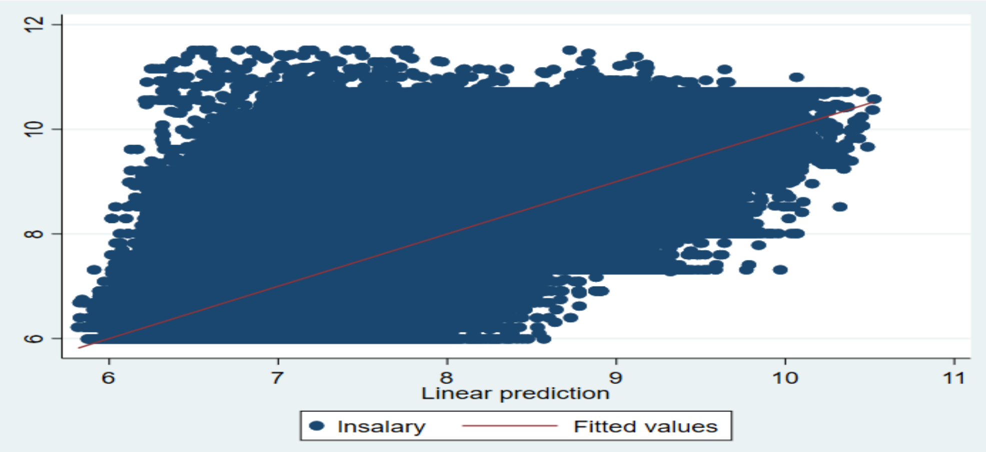


Figure 10-14: plot of the dependent variable (log salary) and the fitted value 2017.



10.2.3 Firm's age in each firms' size

Table 10-13: earning function in small, medium, and medium B.

	Small	MediumA	MediumB
Log earning	Coefficient	Coefficient	Coefficient
Age	0.016229***	0.0280***	0.0417***
age2	-0.0003979***	-0.0006***	-0.0008***
age3	3.85E-06***	5.17E-06***	6.68e-06 ***
Education			
Illiterate	-0.1194914***	-0.1360***	-0.1762***
reads and writes	-0.0826912***	-0.0988***	-0.1290***
Primary degree	-0.0354444***	-0.0452***	-0.0876***
Intermediate degree	0.0246227***	0.0017***	-0.0194***
secondary degree	0.0412545***	0.0361***	0.0180***
Diploma	0.0981546***	0.1065***	0.1098***

Bachelor's degree	0.7423692***	0.9054***	0.5750***
Master's degree	0.1082287***	0.1160***	0.1379***
PhD degree	0.678188***	0.5848***	0.5226***
Higher diploma	0.7776469***	0.7886***	0.8126***
Fellowship	0.4855487***	0.8940***	0.9009***
Qualifications_3			
college of literate.	0.0245176***	0.0269***	0.0238***
Languages and translation	0.0857308***	0.1374***	0.1625***
administration sci..	0.1963539***	0.2659***	0.3332***
Colleges of Education	0.0281178***	0.0125***	0.0053#
Colleges of Agriculture	-0.0421745***	0.0014***	0.0911***
Colleges of Pharmacy	0.3217499***	0.3405***	0.2926***
Colleges of Medicine	0.2180129***	0.3169***	0.1891***
Engineering faculty	0.5110932***	0.5944***	0.5708***
Architecture and Planning	0.2020871***	0.2641***	0.3817***
Dentistry & Computer science	0.1783281***	0.2369***	0.2651***
Technical Institute.	0.1372341***	0.1921***	0.2307***
High school	0.0417808***	0.0480***	0.0454***
Secondary trading	0.103038***	0.1214***	0.1449***
industrial high School	0.1211314***	0.1855***	0.1875***
Secondary agriculture	0.0155064***	0.0864***	0.0464#
technical College	0.1840459***	0.2373***	0.2267***
School of law	0.1270291***	0.2425***	0.2686***
School of Economics	0.0752293***	0.0555***	0.3192***
Female	-0.127405***	-0.1947***	-0.2388***
Colour			
Red	-0.0077689***	-0.0242***	0.0167***
red small A	-0.0291415***		
Yellow	-0.0072887***	-0.0221***	-0.0179***
green small A	-0.001767***		
green2	0.0162014***	0.0240***	0.0432***
green3	0.0430541***	0.0685***	0.0889***
Platinum	0.0975498***	0.1805***	0.2121***
Excluded	0.2013362***	0.3101***	0.4993***
Region			
AL-Baha & Northern border	-0.0928067***	-0.1118***	-0.1685***
AL-Jouf	-0.1961479***	-0.1766***	-0.1478***
Riyadh	0.0093223***	0.0027#	0.0018#
Easte province	-0.0414359***	-0.0247***	-0.0319***
Qassim	-0.1410866***	-0.1495***	-0.2012***
Madinah	-0.0333707***	-0.0451***	-0.0924***
Tabuk	-0.057785***	-0.0590***	-0.1502***
Jazan	-0.0515275***	-0.0635***	-0.1079***
Hail	-0.133029***	-0.1499***	-0.1914***

Asir	-0.0825993***	-0.1016***	-0.1307***
Activities			
Agriculture, and forest	0.0401173***	0.0664***	0.0105#
Mining and Quarrying	0.0637095***	0.1803***	0.1854***
Wholesale and Retailed	0.0392436***	0.1293***	0.1554***
Repair Of Motor Vehicle	-0.0069836***	0.0491***	0.1289***
Transportation	0.0326034***	0.0864***	0.1306***
Accommodation	0.0076244***	0.0585***	0.0949***
Professional, Scie	0.1542964***	0.2151***	0.1406***
Education, Human H..	0.0599654***	0.1615***	0.1811***
Other Personal Ser..	-0.0058937***	0.1071***	0.0298#
Other Activities	0.0951834***	0.2170***	0.2444***
Occupations			
Managers, Director	0.2242771***	0.3419***	0.5560***
Specialists	0.3150674***	0.3756***	0.4444***
Technicians	0.1313014***	0.1637***	0.2177***
The clerical occupation	0.0842203***	0.0907***	0.1186***
sales occupations	0.0789105***	0.0668***	0.1006***
Services occupation	-0.0046904***	-0.0205***	-0.0466***
Agriculture and animal	-0.0759986***	-0.1147***	-0.2032***
Industrial and chemical	-0.0369147***	-0.0394***	-0.0615***
Saudi	1.378765***	1.4303***	1.4633***
Non-Saudi non-localize	-0.0010585#	-0.0075***	-0.0205***
Firm age	-0.0009062***	-0.0006***	0.0010***
_cons	6.281887***	6.0425***	5.7850***
Sample size	1,672,562#	481,045#	391,441#
F-TEST	83446.87***	24608.60***	18099.38***
f-test	0.7822#	0.7817#	0.7640#

Table 10-14: earning function in small, medium, and medium B

	MediumC	**Big**	**Giant**
Log earning	Coefficient	Coefficient	Coefficient
Age	0.0525***	0.0729***	0.0439***
age2	-0.0010***	-0.0014***	-0.0004***
age3	7.22e-06 ***	.0000101***	1.71e-06***
Education			
Illiterate	-0.2295***	-0.2959***	-0.4109***
reads and writes	-0.1421***	-0.1946***	-0.2368***
Primary degree	-0.1318***	-0.2415***	-0.3297***
Intermediate degree	-0.0560***	-0.1753***	-0.2958***
secondary degree	0.0133***	-0.0712***	-0.2069***
Diploma	0.0982***	0.0384***	-0.1380***
Bachelor’s degree	1.0442***	1.1037***	0.9550***
Master’s degree	0.1411***	0.0768***	-0.0161***

PhD degree	0.5118***	0.4798***	0.5706***
Higher diploma	0.5860***	0.5871***	0.4657***
Fellowship	0.9019***	0.7515***	0.7889***
Qualifications_3			
college of literate.	0.0197***	-0.0067#	-0.0753***
Languages and translation	0.1717***	0.1555***	0.0680***
administration sci..	0.3116***	0.2881***	0.1861***
Colleges of Education	-0.0168***	-0.1126***	-0.0846***
Colleges of Agriculture	0.0431***	-0.0102#	-0.0508***
Colleges of Pharmacy	0.4323***	0.2586***	0.0673***
Colleges of Medicine	0.3539***	0.4228***	0.4802***
Engineering faculty	0.5647***	0.5064**	0.2018***
Architecture and Planning	0.2830***	0.3469**	0.3651***
Dentistry & Computer science	0.2719***	0.2266**	0.1390***
Technical Institute.	0.2385***	0.1915**	0.1275***
High school	0.0319***	-0.0038***	-0.1011***
Secondary trading	0.1713***	0.1410***	0.0466***
industrial high School	0.1760***	0.1999***	0.1726***
Secondary agriculture	-0.0257***	-0.0362*	-0.1535***
technical College	0.1769***	0.1160***	-0.0670***
School of law	0.3167***	0.2249***	0.1740***
School of Economics	0.4556***	0.5895***	0.3707***
Female	-0.2744***	-0.2333***	-0.2186***
Colour			
Red	-0.0736***	-0.0764***	-0.1323***
red small A			
Yellow	-0.0379***	-0.0335***	-0.0801***
green small A			
green2	0.0547***	0.0443***	-0.0167***
green3	0.0863***	0.0922***	0.0998***
Platinum	0.2119***	0.1806***	0.0564***
Excluded	0.3533***		
Region		-0.1944***	-0.1528***
AL-Baha & Northern border	-0.2264***	-0.0896***	-0.2767***
AL-Jouf	-0.3303***	-0.0301***	0.0289***
Riyadh	-0.0218***	0.0272***	0.0228***
Easte province	-0.0333***	-0.2628***	-0.1195***
Qassim	-0.2290***	0.0078***	-0.0864***
Madinah	-0.1353***	0.1676***	0.1653***
Tabuk	-0.1004***	-0.1854***	-0.0967***
Jazan	-0.1306***	-0.2344***	-0.0768***
Hail	-0.2519***	-0.1615***	-0.1808***
Asir	-0.1676***		
Activities		-0.2241***	0.1351***

Agriculture, and forest	0.0581***	0.2761***	0.5780***
Mining and Quarrying	0.2182***	0.1257***	0.1130***
Wholesale and Retailed	0.1927***	0.1235***	0.0961***
Repair Of Motor Vehicle	0.2247***	0.1125***	0.0488***
Transportation	0.1479***	0.1556***	0.2452***
Accommodation	0.1701***	0.0649***	-0.0463***
Professional, Scie	0.1344***	0.1800***	0.5163***
Education, Human H..	0.1611***	-0.0182#	0.2477***
Other Personal Ser..	0.2956***	0.1594***	
Other Activities	0.2105***		0.6482***
Occupations		0.8586***	0.5098***
Managers, Director	0.6957***	0.5728***	0.0683***
Specialists	0.5112***	0.2749***	0.0476***
Technicians	0.2378***	0.1558***	0.0670***
The clerical occupation	0.1311***	0.1286***	-0.1732***
sales occupations	0.0896***	-0.1003***	-0.2049***
Services occupation	-0.0726***	-0.2352***	-0.1821***
Agriculture and animal	-0.2271***	-0.0425***	1.9188***
Industrial and chemical	-0.0772***	1.6263***	-0.0952***
Saudi	1.5308***	0.0349***	-0.0005***
Non-Saudi non-localize	0.0360***	0.0019***	5.5012***
Firm age	0.0005***	5.2829***	-0.0167***
_cons	5.5748***	0.0443***	0.0998***
Observations number	496,727#	712,485#	617,002#
F(70, 496656)	21188.04***	31241.56***	44803.11***
Adj R-squared	0.7491#	0.7516#	0.8316#

10.3 Earning structure change, pooled sample (2013 and 2017).

Table 10-15: summary statistic of pooled sample.

Variable	Observations	Mean	Standard deviation	Min	Max
Log earning	4,465,574	7.39937	0.9764539	5.991465	11.51299
Log earning for Saudi	1,901,191	8.306354	.5798526	5.991465	11.45105
Log earning for non-Saudi	2,564,383	6.726947	.5907958	5.991465	11.51299
Age	4,465,574	40.26861	11.35424	15	65
Age-square	4,465,574	1750.48	940.2387	225	4225
Age-cubic	4,465,574	81051.39	62633.3	3375	274625
Female	4,465,574	0.1424422	0.3495031	0	1
Qualification	4,465,574	20.34286	6.369779	1	28
college of literature	4,465,574	0.0207326	0.1424878	0	1
Languages and translation	4,465,574	0.0030552	0.0551889	0	1
administration science	4,465,574	0.0265234	0.1606856	0	1
Colleges of Education	4,465,574	0.0125278	0.1112245	0	1
Colleges of agriculture	4,465,574	0.0012999	0.0360313	0	1
Colleges of Pharmacy	4,465,574	0.0013687	0.0369705	0	1

Colleges of Medicine	4,465,574	0.0016421	0.0404898	0	1
Colleges of Science	4,465,574	0.0068296	0.0823586	0	1
Engineering Faculties	4,465,574	0.01205	0.109109	0	1
Architecture and Planning	4,465,574	0.0009313	0.0305038	0	1
Colleges of Dentistry	4,465,574	0.0004949	0.0222408	0	1
Colleges of Medical and Applied Science	4,465,574	0.0030825	0.0554344	0	1
Faculty of Computer Science and Inform	4,465,574	0.0102417	0.1006817	0	1
High school	4,465,574	0.1536705	0.3606326	0	1
Secondary trading	4,465,574	0.0060122	0.0773051	0	1
industrial high School	4,465,574	0.0094584	0.0967931	0	1
Secondary agricultural	4,465,574	0.0006935	0.0263258	0	1
technical College	4,465,574	0.0092349	0.0956535	0	1
Institute of Management	4,465,574	0.0025446	0.0503796	0	1
Technical Institute	4,465,574	0.0060346	0.077448	0	1
School of Law	4,465,574	0.0005632	0.0237251	0	1
Management	4,465,574	0.0014627	0.0382179	0	1
School of Economics	4,465,574	0.0010386	0.0322108	0	1
Health Institute	4,465,574	0.0000177	0.004206	0	1
air science	4,465,574	2.46E-06	0.0015695	0	1
Veterinary Training and Animal Production Centre	4,465,574	0.0000152	0.0039022	0	1
Institute of Professional Observers	4,465,574	0.0000148	0.0038444	0	1
Colour	4465574	5.612248	1.611917	1	9
Red	4,465,574	0.0161905	0.1262078	0	1
Red small A	4,465,574	0.006562	0.0807398	0	1
Yellow	4,465,574	0.1133686	0.3170429	0	1
Green small A	4,465,574	0.0448838	0.207049	0	1
Green2	4,465,574	0.2312442	0.4216282	0	1
Green3	4,465,574	0.1435204	0.3506028	0	1
Platinum	4,465,574	0.1495985	0.3566774	0	1
Excluded	4,465,574	0.0003093	0.0175829	0	1
Size	4465574	2.061115	1.049854	1	4
Small	4,465,574	0.3824548	0.4859868	0	1
Medium	4,465,574	0.3139283	0.4640877	0	1
Giant	4,465,574	0.1399524	0.3469377	0	1
Zone	4465574	6.880256	3.449908	1	13
AL-Baha	4,465,574	0.0051942	0.0718833	0	1
AL-Jouf	4,465,574	0.0075643	0.0866435	0	1
North Border	4,465,574	0.0049259	0.0700117	0	1
Riyadh	4,465,574	0.3605248	0.4801528	0	1
Prov.	4,465,574	0.2154811	0.4111558	0	1
Qassim	4,465,574	0.0356019	0.1852955	0	1
Madinah	4,465,574	0.0410745	0.1984625	0	1
Tabuk	4,465,574	0.0113132	0.1057602	0	1
Jazan	4,465,574	0.0129515	0.1130654	0	1
Hail	4,465,574	0.0119087	0.1084751	0	1
Asir	4,465,574	0.0334846	0.1798983	0	1
Najran	4,465,574	0.0103277	0.101099	0	1
Activities	4465574	4.785657	2.82748	1	11
Agriculture, Forestry, And fishing	4,465,574	0.0151546	0.1221677	0	1

Mining and Quarrying, Manufacturing	4,465,574	0.1104832	0.3134911	0	1
Wholesale and Retail Trade	4,465,574	0.1949425	0.3961565	0	1
Repair of Motor Vehicles and Motorcycles	4,465,574	0.0343071	0.1820169	0	1
Transportation and Storage	4,465,574	0.0307463	0.1726297	0	1
Accommodation and Food Service Activities	4,465,574	0.0483398	0.2144833	0	1
Professional, Scientific and Technical	4,465,574	0.0326608	0.1777471	0	1
Education, Human Health and Social Workers	4,465,574	0.0713601	0.2574255	0	1
Others Personal Services Activity	4,465,574	0.0060483	0.0775351	0	1
Other Activities	4,465,574	0.0929365	0.2903435	0	1
Occupations	4465574	5.752843	2.581961	1	9
Managers, Directors, and senior officer	4,465,574	0.0430755	0.2030272	0	1
Specialists	4,465,574	0.0912369	0.2879457	0	1
Technicians	4,465,574	0.0963072	0.2950121	0	1
The clerical occupations	4,465,574	0.1260857	0.331946	0	1
sales occupations	4,465,574	0.1116658	0.3149549	0	1
Services occupations	4,465,574	0.159373	0.3660236	0	1
Agriculture and animal husbandry professional	4,465,574	0.0060286	0.0774094	0	1
Industrial and chemical processes	4,465,574	0.1187834	0.3235335	0	1
Year	4,465,574	0.9788802	0.1437837	0	1
Saudi	4,465,574	0.4257439	0.4944553	0	1
Saudi non-localized	4,465,574	0.3821182	0.4859053	0	1
Saudi non-localized	4,465,574	0.0436257	0.204261	0	1
Non-Saudi localized	4,465,574	0.0928046	0.2901585	0	1
Non-Saudi non-localized	4,465,574	0.4814514	0.4996559	0	1

Table 10-16: OLS result of the pooled sample of 2013 and 2017.

	Coefficients	Interaction coefficients
Age	0.0460842***	-0.0053831#
Age-square	-0.001022***	0.0003573#
Age-cubic	9.59E-06***	-5.04E-06***
Female	-0.042334***	-0.2065351***
college of literature	0.3795741***	-0.328508***
Languages and translation	0.6882234***	-0.4550356***
administration science	0.6289093***	-0.2406872***
Colleges of Education	0.3342939***	-0.2343726***
Colleges of agriculture	0.2606035***	-0.1582596***
Colleges of Pharmacy	0.826754***	-0.1608628***
Colleges of Medicine	0.8464669***	-0.0723679*
Colleges of Science	0.7126682***	-0.4873652***
Engineering Faculties	0.9638068***	-0.2206984***
Architecture and Planning	0.5110182***	-0.0818578#
Colleges of Dentistry	0.5946997***	-0.0809241#
Colleges of Medical and Applied Science	0.6066305***	-0.2766565***

Faculty of Computer Science and Inform	0.5237066***	-0.1514028***
High school	0.2093192***	-0.2019555***
Secondary trading	0.3003643***	-0.1513354***
industrial high School	0.2871958***	-0.1252778***
Secondary agricultural	0.0243769#	-0.0628661#
technical College	0.4682598***	-0.3172263***
Institute of Management	0.3562632***	-0.1733076***
Technical Institute	0.3044667***	-0.0969424***
School of Law	0.3255699***	
Management	0.2859245***	
School of Economics	0.3593561***	
Health Institute	0.3925872***	
air science	0.3939337***	
Veterinary Training and Animal Production Centre	0.015737#	
Institute of Professional Observers	0.2306014***	
Red	0.4015204***	-0.4313726***
Red small A	0.3726713***	-0.4010071***
Yellow	0.2900479***	-0.3185556***
Green small A	0.0474602***	-0.0474005***
Green2	0.0476163***	-0.0311063***
Green3	0.071519***	0.0334566***
Platinum	0.1832211***	-0.010089#
Excluded	0.7106212***	-0.2621012#
Small	-0.157173***	-0.4010071***
Medium	-0.047132***	-0.3185556***
Giant	-0.04479***	-0.0474005***
AL-Baha	-0.193024***	0.0917938***
AL-Jouf	-0.348794***	0.1564787***
North Border	-0.215392***	0.1075692***
Riyadh	-0.105976***	0.1009732***
Prov.	-0.216372***	0.2114976***
Qassim	-0.3341162***	0.1652734***
Madinah	-0.1575319***	0.1031735***
Tabuk	-0.1124791***	0.0779795***
Jazan	-0.1597722***	0.0913834***
Hail	-0.2641832***	0.1172273***
Asir	-0.2488606***	0.1472091***
Najran	-0.2488606***	-0.010735#
Agriculture, Forestry, And fishing	-0.0085806#	0.0255335*
Mining and Quarrying, Manufacturing	0.2038004***	0.1192002***
Wholesale and Retail Trade	0.1499563***	-0.0305614***
Repair of Motor Vehicles and Motorcycles	0.0893628***	-0.0248712***
Transportation and Storage	0.0527436***	0.0447752***
Accommodation and Food Service Activities	0.0730835***	0.0367921***
Professional, Scientific, and Technical	-0.0263682***	0.0326229***
Education, Human Health and Social Workers	0.176515***	0.0174226**
Others Personal Services Activity	0.040955**	0.051913***
Other Activities	0.0505667***	0.149038***
Managers, Directors, and senior officer	0.7376351***	-0.3514269***
Specialists	0.6637556***	-0.1634008***
Technicians	0.2545635***	-0.0680433***

The clerical occupations	0.2286566***	-0.1107579***
sales occupations	0.1879493***	-0.1028099***
Services occupations	-0.1147129***	0.0220507***
Agriculture and animal husbandry professional	-0.1181359***	-0.0491065***
Industrial and chemical processes	-0.0709331***	-0.0060582#
Year	0.2231111#	
Saudi	1.871159***	-0.3095805***
Non-Saudi localized	0.3008518***	-0.3208183***
Constant	5.575211***	

Table 10-17: age function shape

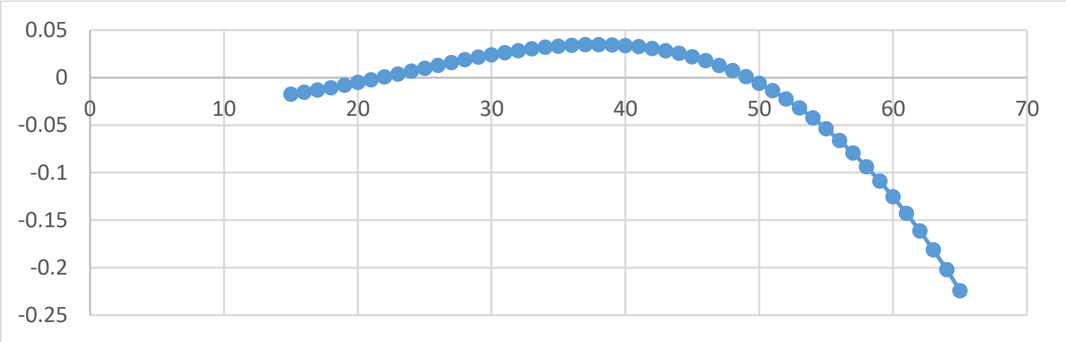


Table 10-18: summaries the error term.

Variable	Observation	Mean	Standard deviation	Min	Max
Error term (e)	4,465,574	9.12e-12	.4686642	-2.476634	5.073446

Table 10-19: variance inflation factor for pooled sample

Variable	VIF	1/VIF	Variable	VIF	1/VIF
Age	509111.4	0.000002	28	1	0.997265
age2	1.56E+06	0.000001	year	24866.72	0.00004
age3	335018.3	0.000003	y_age	630210.6	0.000002
Zone			y_age2	1.64E+06	0.000001
1	185.87	0.00538	y_age3	338547.7	0.000003
2	310.66	0.003219	y_female	899.02	0.001112
3	252.88	0.003954	y_zone		
4	96.08	0.010408	1	321.27	0.003113
5	52.08	0.019201	2	507.83	0.001969
6	66.88	0.014951	3	381.42	0.002622
7	29.17	0.034284	4	6168.28	0.000162
8	221.2	0.004521	5	4426.48	0.000226

9	157.5	0.006349	6	959.09	0.001043
10	341.49	0.002928	7	1026.6	0.000974
11	38.97	0.025664	8	515.07	0.001941
13	272.05	0.003676	9	492.73	0.00203
Colour			10	650.71	0.001537
1	6186.27	0.000162	11	868	0.001152
2	2477.15	0.000404	12	4899.48	0.000204
3	37736.71	0.000026	y_colour		
4	34.3	0.029155	1	3604.47	0.000277
6	57.75	0.017315	2	1397.67	0.000715
7	61.04	0.016383	3	22180.56	0.000045
8	97.36	0.010271	4	24399.05	0.000041
9	182.61	0.005476	5	122059.5	0.000008
Female	898.35	0.001113	6	103013.1	0.00001
Activities			7	71518.11	0.000014
1	57.02	0.017539	8	74779.05	0.000013
2	49.54	0.020185	y_SIZE		
4	63.07	0.015856	1	199.64	0.005009
5	45.64	0.021909	2	162.9	0.006139
6	49.94	0.020025	3	115.99	0.008621
7	62.62	0.015969	y_occupation		
8	91.98	0.010872	1	170.6	0.005861
9	105.77	0.009455	2	63.77	0.015681
10	54.93	0.018204	3	70.24	0.014236
11	69.68	0.014352	4	505.27	0.001979
Occupations			5	87.01	0.011493
1	170.37	0.005869	6	42.78	0.023374
2	63.66	0.015709	7	58.1	0.017211
3	69.96	0.014295	8	185.18	0.0054
4	504.25	0.001983	y_Qualific~d		
5	86.56	0.011552	1	344.73	0.002901
6	43.22	0.023139	2	324.58	0.003081
7	58.1	0.017212	3	127.75	0.007828
8	184.66	0.005415	4	310.84	0.003217
SIZE			5	96.94	0.010316
1	112.79	0.008866	6	64.88	0.015412
2	83.96	0.011911	7	58.8	0.017008
4	102.42	0.009763	8	181.33	0.005515
Qualification			9	72.99	0.0137
1	344.6	0.002902	10	112.48	0.008891
2	324.56	0.003081	11	93.34	0.010714
3	127.66	0.007833	12	238.63	0.004191

4	310.68	0.003219	13	113.34	0.008823
5	96.93	0.010317	14	243.58	0.004105
6	64.88	0.015413	15	51.83	0.019295
7	58.81	0.017003	16	36.08	0.027713
8	181.34	0.005514	17	44.97	0.022235
9	72.94	0.01371	18	136.32	0.007335
10	112.47	0.008891	19	104.96	0.009527
11	93.34	0.010713	20	30.38	0.032915
12	238.62	0.004191	y_Activities		
13	113.3	0.008826	1	65.3	0.015314
14	242.91	0.004117	2	105.63	0.009467
15	51.82	0.019297	3	192.57	0.005193
16	36.08	0.027719	4	150.22	0.006657
17	44.97	0.022235	5	64.51	0.015501
18	136.28	0.007338	6	67.64	0.014784
19	104.96	0.009528	7	87.71	0.011401
20	30.38	0.032911	8	108.8	0.009191
21	1	0.998931	9	142.09	0.007038
22	1.02	0.982317	10	58.15	0.017196
23	1	0.997348	Saudi	91846.09	0.000011
25	1.05	0.949607	Non Saudi localise	94238.38	0.000011
26	1	0.998044	y_non_saud..	93813	0.000011
27	1	0.998456	y_saudi	91832.17	0.000011
Mean VIF 41636.57					

10.4 Oaxaca decomposition

10.4.1 Oaxaca decomposition between Saudi and non-Saudi

10.4.1.1 Oaxaca decomposition result

Table 10-20: the total effect of each variable in 2013, omega approach.

	Explained	%	Unexplained	%
age	0.01	-1.46	0.11	-6.05
Female	0.00	-0.02	0.00	-0.20
Qualification	-0.23	37.10	0.17	-9.37
Regions	0.01	-1.54	-0.16	8.62
Size	-0.02	3.94	0.02	-1.20
Occupations	-0.31	49.63	0.12	-6.45
Activities	-0.02	3.11	-0.05	2.84
interaction	0.00	0.38	0.01	-0.49
colour	-0.06	8.87	-0.04	2.05

Table 10-21: the total effect of each variable in 2017, omega approach.

	Explained	%	Unexplained	%
age	-0.09	7.85	0.51	-32.48
female	-0.02	1.99	0.08	-5.33
education	-0.42	37.22	0.67	-42.25
Qualification	-0.10	8.88	0.12	-7.44
regions	-0.02	1.49	-0.07	4.30
size	-0.06	5.62	0.02	-1.26
occupations	-0.21	18.70	0.21	-13.26
Activities	-0.08	6.67	0.00	-0.28
interaction	-0.08	6.78	0.08	-4.83
colour	-0.07	5.87	0.02	-1.35

Table 10-22:the detail decomposition of all approaches 2013.

	pooled	omega	w(0)	w(1)
	Coef.	Coef.	Coef.	Coef.
explained	-0.2584***	-0.6334***	-0.3923***	-0.2666***
age	0.4336***	-1.6753***	-0.4752#	0.0152#
age2	-0.8507***	2.5244***	1.9745#	-0.1075#
age3	0.5445***	-0.8398***	-1.4350**	0.2110#
Qualification				
1	-0.0043***	-0.0056***	-0.0040***	-0.0042***
2	-0.0542***	-0.0696***	-0.0422***	-0.0551***
3	-0.0008#	-0.0010#	0.0005#	-0.0009#
4	-0.0050***	-0.0055***	-0.0034***	-0.0052***
5	0.0004#	0.0004#	0.0006#	0.0004#
6	0.0012***	0.0012***		0.0012***
7	-0.0163***	-0.0174***	-0.0130***	-0.0163***
8	-0.0074***	-0.0091***	-0.0021#	-0.0081***
9	-0.0003#	-0.0003#	0.0002#	-0.0003#
10	-0.0059***	-0.0068***	-0.0063***	-0.0058***
11	-0.0524***	-0.1096***	-0.0106#	-0.0588***
12	-0.0041***	-0.0045***	-0.0010#	-0.0042***
13				
14	-0.0064***	-0.0074***	-0.0010#	-0.0066***
15	0.0003***	0.0003***		0.0003***
female	0.0003#	0.0002#	0.0019*	0.0002#
1	-0.0058***	-0.0058***	-0.0027#	-0.0058***
2	0.0022***	0.0023***	-0.0025*	0.0023***
3	0.0367***	0.0327***	-0.0074#	0.0377***
5	-0.0002#	-0.0002#	0.0002#	-0.0002#
6	-0.0107***	-0.0109***	-0.0089#	-0.0109***
7	-0.0085***	-0.0084***	-0.0177***	-0.0085***
Colour				
1	-0.0014***	-0.0011***		-0.0014***
2	-0.0030***	-0.0025***		-0.0029***
3	-0.0027***	-0.0024***	-0.0019#	-0.0026***
4	0.0023***	0.0027***	-0.0004#	0.0023***
6	0.0034***	0.0040***	0.0045#	0.0034***
7	-0.0031***	-0.0037***	-0.0068***	-0.0029***

8	-0.0398***	-0.0532***	-0.0594***	-0.0384***
9				
Size				
1	-0.0277***	-0.0292***	-0.0428***	-0.0272***
2	0.0018***	0.0020***	0.0034***	0.0017***
4	0.0027***	0.0023***	-0.0030#	0.0029***
Occupation				
1	-0.1682***	-0.2737***	-0.2039***	-0.1633***
2	0.0403***	0.0386***	0.0426***	0.0403***
3	0.0061***	0.0058***	0.0040***	0.0062***
4	-0.0300***	-0.0680***	-0.0216***	-0.0343***
5	0.0020*	0.0019*	0.0003#	0.0020*
6	-0.0177***	-0.0171***	-0.0104#	-0.0176***
7	-0.0005***	-0.0005***	-0.0003#	-0.0005***
8	-0.0012***	-0.0013***	0.0001#	-0.0012***
Activity				
1	-0.0001#		0.0029***	-0.0001#
2	-0.0158***	-0.0175***	-0.0258***	-0.0154***
4	0.0043***	0.0044***	0.0029***	0.0043***
5	0.0017***	0.0018***	0.0018#	0.0017***
6	-0.0021***	-0.0037***	-0.0001#	-0.0022***
7	0.0009***	0.0008***	0.0016#	0.0009***
8	0.0004#	0.0002	-0.0021*	0.0005***
9	-0.0005	-0.0006	-0.0003#	-0.0005#
10	0.0002**	0.0002***	0.0031*	0.0002*
11	-0.0035***	-0.0053***	-0.0213***	-0.0029***
non_saudi_no~d	0.0107***	0.0094***		0.0105***
saudi_non_lo~d	-0.0044***	-0.0117***	-0.0038*	
unexplained	-1.5715***	-1.1965***	-1.4376***	-1.5633***
age	1.7016#	3.8105#	2.6104#	2.1200#
age2	-3.5272#	-6.9023***	-6.3524#	-4.2703#
age3	1.8181*	3.2025***	3.7977***	2.1516***
Qualification				
1	0.0003#	0.0017#		0.0002#
2	0.0134***	0.0289***	0.0014***	0.0143***
3	0.0026***	0.0028***	0.0013***	0.0027***
4	0.0020*	0.0025***	0.0004*	0.0021#
5	-0.0002#	-0.0002#	-0.0003#	-0.0002#
6	0.0000***		0.0012***	
7	0.0047*	0.0058***	0.0015***	0.0048***
8	0.0066***	0.0083***	0.0013***	0.0073***
9	0.0011#	0.0011#	0.0007***	0.0011#
10	-0.0006#	0.0003#	-0.0002#	-0.0007#
11	0.0461***	0.1033***	0.0044***	0.0526***
12	0.0071***	0.0075***	0.0040***	0.0072***
13				
14	0.0083***	0.0094***	0.0030***	0.0085***
15	0.0000***	0.0000***	0.0003***	
female	0.0036***	0.0037***	0.0019***	0.0037***
Region				
1	-0.002#	-0.0020#	-0.0051#	-0.0020#
2	-0.0439***	-0.0440***	-0.0392***	-0.0440***
3	-0.1208***	-0.1168***	-0.0766***	-0.1218***
5	-0.0006#	-0.0006#	-0.0010#	-0.0006#
6	-0.0004#	-0.0003#	-0.0022#	-0.0003#

7	0.0060**	0.0058#	0.0152***	0.0060***
Colour				
1		-0.0002***	-0.0014***	
2	-0.0004#	-0.0009***	-0.0034***	-0.0005#
3	-0.0004#	-0.0007#	-0.0012#	-0.0005#
4	0.0027#	0.0022#	0.0054#	0.0027#
6	-0.0046#	-0.0052#	-0.0057#	-0.0046#
7	-0.0189***	-0.0183***	-0.0151***	-0.0190***
8	-0.0278***	-0.0144#	-0.0081***	-0.0292***
9				
Size				
1	0.0182***	0.0197***	0.0333***	0.0178***
2	0.0159#	0.0157#	0.0143#	0.0160#
4	-0.0138***	-0.0134***	-0.0081***	-0.0140***
Occupation				
1	-0.0373***	0.0682***	-0.0016***	-0.0423***
2	-0.0015#	0.0002#	-0.0037#	-0.0015#
3	0.0053#	0.0056*	0.0074#	0.0053#
4	0.0092*	0.0472***	0.0008**	0.0135**
5	0.0111***	0.0111***	0.0127***	0.0110***
6	-0.0126#	-0.0131#	-0.0198#	-0.0126#
7	-0.0001#	-0.0001#	-0.0003#	-0.0001#
8	-0.0012#	-0.0011#	-0.0025#	-0.0012#
Activities				
1	-0.0026***	-0.0026***	-0.0056***	-0.0026***
2	-0.0274***	-0.0257***	-0.0174***	-0.0278***
4	0.0095#	0.0095#	0.0110#	0.0095#
5	-0.0001#	-0.0002#	-0.0001#	-0.0001#
6	0.0036#	0.0052*	0.0016#	0.0037#
7	-0.0017#	-0.0016#	-0.0024#	-0.0017#
8	-0.0054#	-0.0052*	-0.0029***	-0.0055***
9	0.0036#	0.0036#	0.0034#	0.0036#
10	-0.0007#	-0.0007	-0.0036*	-0.0007#
11	-0.0360***	-0.0342***	-0.0182***	-0.0366***
Non-Saudi non-localize-	-0.0003#	0.0011***	0.0105***	
Saudi non-localize	0.0005#	0.0079***		-0.0038*
_cons	-1.3843#	-1.3843#	-1.3843*	-1.3843#

Table 10-23: the detail decomposition of all approaches 2017

	Pooled	omega	w(0)	w(1)
	Coef.	Coef.	Coef.	Coef.
explained	0.0475	-1.1254***	0.1654***	-0.3197***
age	0.4724***	-0.6386***	-0.3928***	0.1654***
age2	-0.6105***	0.6128***	1.1949***	-0.3153***
age3	0.2619***	-0.0625***	-0.7102***	0.2242***
Education				
1	-0.0064***	-0.0033***	-0.0127***	-0.0053***
2	-0.0250***	-0.0409***	-0.0357***	-0.0194***
3	-0.0026***	0.0012***	-0.0072***	-0.0013***
4	0.0116***	-0.0341***	0.0304***	0.0001***
5	0.0421***	-0.2621***	0.1494***	-0.0780***

6	-0.0004***	-0.0130***	0.0080***	-0.0073***
7	0.0004***	0.0003***	-0.0003**	0.0004***
8	-0.0051***	-0.0744***	0.0141***	-0.0486***
9	0.0129***	0.0070***	0.0063***	0.0120***
10	0.0004***	0.0004***	0.0003***	0.0004***
11	0.00002***	0.00001***		0.000044***
Qualification				
1	0.0007***	-0.0077***	0.0017***	-0.0022***
2	-0.0022***	-0.0064***	-0.0010***	-0.0072***
3	-0.0109***	-0.0214***	-0.0090***	-0.0153***
4	0.0013***	-0.0036***	0.0014***	0.0024***
5	0.0000***	-0.0001***	0.0000***	-0.0001***
6	0.0000***	0.0000***	0.0000***	0.000025***
7	0.0001***	0.0002***	0.0001***	0.0001***
8	-0.0038***	-0.0057***	-0.0033***	-0.0047***
9	-0.0001***	-0.0002***	-0.0001***	-0.0001***
10	-0.0029***	-0.0054***	-0.0023***	-0.0034***
11	-0.0003***	- 0.0001****	-0.0001***	-0.0003***
12	0.0031***	-0.0420***	0.0188***	-0.0378***
13	-0.0003***	0.00001	0.0001***	-0.0002***
14	0.0014***	-0.0027***	-0.0004***	0.0007***
15	0.0000***	-0.0005***	-0.0001***	-0.0001***
16	-0.0011***	-0.0043***	-0.0001***	-0.0049***
17	0.0001***	0.0000***	0.0000***	0.0000451***
18	-0.0001***	-0.0001***	-0.0001***	0.0000194***
female	0.0725***	-0.0223***	0.0592***	-0.0209***
Colour				
1	-0.0003***	0.0038***	0.0005***	-0.0004***
2	0.0000***	0.0001***	0.00002***	
3	-0.0013***	0.0151***	-0.0001***	-0.0009***
4			-2.57e-08#	-5.70e-08#
5	0.0003***	0.0010***	0.0002***	0.0004***
7	-0.0072***	-0.0161***	-0.0076***	-0.0037***
8	-0.0280***	-0.0702***	-0.0270***	-0.0220***
9	0.0001***	0.0002***	0.0000***	0.0001***
Size				
1	-0.0243***	-0.0350***	-0.0257***	-0.0213***
2	-0.0044***	-0.0064***	-0.0043***	-0.0042***
3	-0.0010***	-0.0017***	-0.0009***	-0.0010***
4	0.0004***	0.0008***	0.0004***	0.0003***
6	-0.0084***	-0.0210***	-0.0135***	0.0028***
Region				
1	-0.0004***	-0.0004***	-0.0001***	-0.0006***
2	-0.0007***	-0.0008***	0.00004***	-0.0011***
3	-0.0001*	0.0005***	-0.0015***	0.0013***
4	-0.0001***	-0.0020***	-0.0024***	0.0024***
5	-0.0009***	-0.0009***	-0.0004***	-0.0013***
6	-0.0004***	-0.0002***	0.0000***	-0.0005***
7	-8.48e-06***	- 5.20e06***	0.00004***	0.000014***
8	-0.0001***	0.0000***	-0.0001***	-0.0001***
9	-0.0004***	-0.0004***	-0.0001***	-0.0006***
10	-0.0006***	-0.0004***	-0.0004***	-0.0007***
firm_age	0.0012***	0.0015***	-0.0009***	0.0017***

firm_age2	-0.0009***	-0.0016***	0.0002***	-0.0011***
Activities				
1	0.0005***	0.0009***	0.0003***	0.0005***
2	-0.0245***	-0.0472***	-0.0244***	-0.0159***
4	0.0000	-0.0001#	0.00003#	0.0000334***
5	0.0011***	0.0022***	0.0011***	0.0008***
6	-0.0003***	-0.0007***	-0.0003***	-0.0002***
7	-0.0004***	-0.0005***	-0.0004***	-0.0003***
8	-.00004#	-0.0001***	-0.0006***	0.0003***
9	-0.0091***	-0.0152***	-0.0083***	-0.0088***
10	0.0000***	0.00004***	0.00005***	0.000028***
11	-0.0087***	-0.0145***	-0.0117***	-0.0038***
Occupation				
1	-0.0250***	-0.0574***	-0.0214***	-0.0327***
2	0.0183***	0.0174***	0.0137***	.0199872 ***
3	0.0004***	0.0004***	0.0002***	0.00046***
4	-0.0272***	-0.1295***	-0.0167***	-0.0475***
5	-0.0038***	-0.0104***	-0.0011***	-0.0054***
6	0.0044***	-0.0052***	0.0044***	0.0052***
7	-0.0011***	-0.0014***	-0.0010***	-0.0010***
8	-0.0126***	-0.0245***	0.0056***	-0.0140***
Non-Saudi non-localize	0.0036***	-0.0763***		0.0002#
unexplained	-1.6318***	-0.4588***	-1.7496***	-1.2646***
age	1.1749***	2.2859***	2.0401***	1.4819***
age2	-1.6083***	-2.8316***	-3.4137***	-1.9035***
age3	0.7358***	1.0602***	1.7079***	0.7735***
Education				
1	0.0053***	0.0022***	0.0117***	0.0042***
2	0.0075***	0.0234***	0.0182***	0.0019***
3	0.0149***	0.0110***	0.0195***	0.0136***
4	0.0298***	0.0755***	0.0110***	0.0413***
5	0.1265***	0.4307***	0.0192***	0.2466***
6	0.0178***	0.0305***	0.0094***	0.0248***
7	-6.78e-07#	0.0002***	0.0007***	1.57e-06#
8	0.0203***	0.0896***	0.0011***	0.0638***
9	0.00004#	0.0060***	0.0067***	0.001016***
10	0.0001***	0.0001***	0.0002***	0.0001***
11	1.45e-06**	0.00002***	0.00004***	
Qualification				
1	0.0013***	0.0098***	0.0003***	0.0043***
2	0.0020***	0.0062***	0.0008***	0.0071***
3	0.0028***	0.0133***	0.0009***	0.0072***
4	-0.0001*	0.0049***	-0.0001***	-0.0011***
5	-0.0001***	9.59e-06#	-0.0001***	0.0000396***
6	0.0000***	-0.0001***	-0.0001***	0.000048***
7	-0.0002***	-0.0002***	-0.0002***	-0.0001***
8	0.0022***	0.0041***	0.0017***	0.0031***
9	0.0001***	0.0002***	0.0001***	0.0001***
10	0.0011***	0.0035***	0.0005***	0.0016***
11	0.0014***	0.0012***	0.0011***	0.0013***
12	0.0200***	0.0651***	0.0043***	0.0609***
13	0.0013***	0.0010***	0.0009***	0.0013***
14	-0.0002***	0.0039***	0.0016***	0.0006***
15	-0.0001***	0.0003***	-0.0001***	0.000023***
16	0.0020***	0.0052***	0.0010***	0.0058***

17	9.16e-06#	0.00004***	0.00005***	0.00003***
18	-0.0005#	-0.0005***	-0.0005***	-0.0006***
female	-0.0105***	0.0844***	0.0028***	0.0829***
Colour				
1	-0.0009***	-0.0050***	-0.0017***	-0.0008***
2	0.0003***	0.0001***	0.0003***	0.0002***
3	-0.0009***	-0.0174***	-0.0021***	-0.0013***
4	0.00002#	0.00003#	0.00002***	0.000022#
6	0.0011***	0.0004#	0.0012***	0.0011***
7	-0.0064***	0.0025***	-0.0060***	-0.0099***
8	-0.0013***	0.0409***	-0.0023***	-0.0073***
9	-3.41e-07#	-0.0001***	0.000037***	0.0000***
Size				
1	0.0154***	0.0262***	0.0168***	0.0125***
2	0.0005***	0.0025***	0.0004#	0.0003*
3	-0.0005***	0.0001#	-0.0006***	-0.0006***
4	0.0014***	0.0010***	0.0014***	0.0015***
6	-0.0224***	-0.0098***	-0.0172***	-0.0336***
Region				
1	-0.0023***	-0.0023***	-0.0026***	-0.0021***
2	-0.0018***	-0.0016***	-0.0024***	-0.0014***
3	-0.0217***	-0.0222***	-0.0202***	-0.0230***
4	-0.0314***	-0.0295***	-0.0292***	-0.0340***
5	-0.0053***	-0.0053***	-0.0059***	-0.0049***
6	-0.0022***	-0.0023***	-0.0025***	-0.0020***
7	-0.0007***	-0.0007***	-0.0007***	-0.0007***
8	-0.0004***	-0.0005***	-0.0005***	-0.0004***
9	-0.0019***	-0.0019***	-0.0022***	-0.0017***
10	-0.0015***	-0.0017***	-0.0017***	-0.0014***
firmage	0.0486***	0.0483***	0.0507***	0.0480***
firm_age2	-0.0120***	-0.0114***	-0.0132***	-0.0118***
Activity				
1	0.0004***	0.00002#	0.0006***	0.0004***
2	-0.0076***	0.0151***	-0.0077***	-0.0162***
4	-0.0020***	-0.0020***	-0.0020***	-0.0020***
5	-0.0007***	-0.0017***	-0.0007***	-0.0004***
6	-0.0005***	-0.0001#	-0.0005***	-0.0005***
7	-0.0005***	-0.0004***	-0.0005***	-0.0006***
8	-0.0043***	-0.0041***	-0.0037***	-0.0045***
9	0.0013***	0.0074***	0.0005***	0.00103***
10	-0.0002***	-0.0002***	-0.0002***	-0.00022***
11	-0.0153***	-0.0095***	-0.0123***	-0.0202***
Occupation				
1	0.0058***	0.0382***	0.0022***	0.0136***
2	0.0119***	0.0128***	0.0165***	0.0102***
3	0.0129***	0.0129***	0.0131***	0.0128***
4	0.0116***	0.1139***	0.0011***	0.0319***
5	0.0102***	0.0168***	0.0076***	0.0118***
6	-0.0023***	0.0073***	-0.0023***	-0.0030***
7	0.0001***	0.0003***	0.00004#	-7.80e-06 #
8	-0.0041***	0.0078***	-0.0222***	-0.0027***
Non-Saudi non-localize	-0.0034***	0.0765***	0.0002#	
_cons	-2.1460***	-2.1460***	-2.1460***	-2.1460***

10.4.1.2 Correlation issue

Although there was no correlation between Saudi and nationalities categories, the result was like the result from the interaction variable when the correlation was high. Thus, we assumed that the result of Oaxaca decomposition with including the index as a category would provide a suspicious result. Thus, theoretical prove needed. See the example below. We notice that any variable interacting with the index would give high explained gap.

Table 10-24: correlation between the nationalities (nation3) category and the index (Saudi) 2013.

	Saudi	natio~15	natio~14	natio~13	natio~12	natio~11
Saudi	1.0000					
nation_or~15	-0.0393	1.0000				
nation_or~14	-0.0015	-0.0028	1.0000			
nation_or~13	-0.0463	-0.0865	-0.0033	1.0000		
nation_or~12	-0.0134	-0.0250	-0.0010	-0.0295	1.0000	
nation_or~11	-0.0143	-0.0267	-0.0010	-0.0315	-0.0091	1.0000
nation_or~10	-0.0011	-0.0020	-0.0001	-0.0023	-0.0007	-0.0007
nation_ori~9	-0.0039	-0.0073	-0.0003	-0.0086	-0.0025	-0.0027
nation_ori~8	-0.0746	-0.1393	-0.0053	-0.1644	-0.0475	-0.0508
nation_ori~7	-0.0432	-0.0806	-0.0031	-0.0951	-0.0275	-0.0294
nation_ori~6	-0.0988	-0.1845	-0.0070	-0.2176	-0.0629	-0.0672
nation_or~n5	-0.0677	-0.1265	-0.0048	-0.1492	-0.0431	-0.0461
nation_or~n4	-0.0038	-0.0071	-0.0003	-0.0084	-0.0024	-0.0026
nation_or~n3	-0.0155	-0.0289	-0.0011	-0.0341	-0.0099	-0.0105
nation_or~n2	-0.0039	-0.0073	-0.0003	-0.0086	-0.0025	-0.0027
nation_or~n1	1.0000	-0.0393	-0.0015	-0.0463	-0.0134	-0.0143
	natio~10	nation~9	nation~8	nation~7	nation~6	natio~n5
natio~10	1.0000					
nation_ori~9	-0.0002	1.0000				
nation_ori~8	-0.0037	-0.0138	1.0000			
nation_ori~7	-0.0022	-0.0080	-0.1532	1.0000		
nation_ori~6	-0.0050	-0.0183	-0.3506	-0.2028	1.0000	
nation_or~n5	-0.0034	-0.0125	-0.2403	-0.1391	-0.3182	1.0000
nation_or~n4	-0.0002	-0.0007	-0.0135	-0.0078	-0.0179	-0.0123
nation_or~n3	-0.0008	-0.0029	-0.0550	-0.0318	-0.0728	-0.0499
nation_or~n2	-0.0002	-0.0007	-0.0139	-0.0081	-0.0184	-0.0126
nation_or~n1	-0.0011	-0.0039	-0.0746	-0.0432	-0.0988	-0.0677
	natio~n4	na~gion3	na~gion2	na~gion1		
natio~n4	1.0000					
na~gion3	-0.0028	1.0000				
nation_or~n2	-0.0007	-0.0029	1.0000			
nation_or~n1	-0.0038	-0.0155	-0.0039	1.0000		

Table 10-25: high correlation between the index and some interaction variable 2013.

	Saudi	Saudi localise	Saudi non-localise	Non-Saudi localise	Non-Saudi non-localise
Saudi	1.0000				
Saudi localise	0.9968	1.0000			
Saudi non-localise	0.0778	-0.0016	1.0000		
Non-Saudi localise	-0.6586	-0.6565	-0.0512	1.0000	
Non-Saudi non-localise	-0.0235	-0.0235	-0.0018	-0.7368	1.0000

Table 10-26: high correlation between the index and some interaction variable 2017.

	Saudi	Saudi localise	Saudi non-localise	Non-Saudi localise	Non-Saudi non-localise
Saudi	1.0000				
Saudi localise	0.9121	1.0000			
Saudi non-localise	0.2464	-0.1727	1.0000		
Non-Saudi localise	-0.875	-0.7548	-0.2039	1.0000	
Non-Saudi non-localise	-.2828	-0.2579	-0.0697	-0.3045	1.0000

Table 10-27: decomposition result including nationalities' categories in the regression 2013.

	pooled		omega		w(0)		w(1)	
explained	-1.871263	0.000	-1.82988	0.000	-0.379	0.000	-1.9	0.000
unexplained	0.041382	0.8	9.11E-12	1.000	-1.451	0.000	0.07	0.6

Table 10-28: decomposition result including high correlation 3 interaction variables 2013.

	-1.829881		omega	
explained	-1.829881	0.839	-1.829881	0.000
unexplained	-7.18e-12	1.000	-7.18e-12	1.000

The interactions variables were Saudi localise, non-Saudi localise, non-Saudi non-localize

10.4.2 Oaxaca decomposition according to the origins

High background countries				
	omega	pooled	w(1)	w(0)
explained	0.4206***	-0.5082***	-0.1161	-0.5086***
age	0.1233***	0.1744***	0.2073#	0.1674***
_IQualifica_1	-0.0008#	-0.0008#	-0.0007#	-0.0008#
_IQualifica_2	-0.0064#	-0.0061#	-0.0020#	-0.0064#
_IQualifica_3	0.0008#	0.0012#		0.0011#
_IQualifica_4	0.0157#	0.0168#	0.0066#	0.0188#
_IQualifica_5	-0.0006#	-0.0006#		-0.0006#

_IQualifica_6	0.0170#	0.0199#	0.0246#	
_IQualifica_7	0.0589***	0.0593***	-0.0036#	0.0720#
_IQualifica_8	0.0012#	0.0012#	-0.0059#	0.0017#
_IQualifica_9	0.0004#	0.0005#		0.0005#
_IQualifica_10	0.0046#	0.0042#	0.0019#	0.0045#
_IQualifica_11	-0.0113#	-0.0064#		-0.0103#
_IQualifica_12	-0.0032#	-0.0025#		-0.0029#
_IQualifica_13				
_IQualifica_14	-0.0019#	-0.0012#		-0.0017#
_IQualifica_15				
female	-0.0025#	-0.0023#	-0.0002#	-0.0025#
_Iregions_1	0.0013#	0.0016#		0.0018#
_Iregions_2	0.0099#	0.0099#	0.0145#	0.0107#
_Iregions_3	-0.0280***	-0.0208*	-0.1258#	-0.0215#
_Iregions_5	-0.0002#	-0.0002#		-0.0002#
_Iregions_6	0.0014*	0.0015*		0.0014#
_Iregions_7	-0.0001#	-0.0001#	-0.0002#	-0.0001#
_Icolour_1				
_Icolour_2	0.0026#	0.0029#	-0.0063#	
_Icolour_3				0.0013#
_Icolour_4	0.0001#	-0.0001#	0.0039#	-0.0002#
_Icolour_6	0.0064#	0.0062#	0.0210#	0.0056#
_Icolour_7	-0.0148***	-0.0142***	0.0128#	-0.0151***
_Icolour_8	-0.0186#	-0.0178#	0.0015#	-0.0183#
_Icolour_9				
_ISIZE_1	-0.0087#	-0.0084#	-0.0031#	-0.0084#
_ISIZE_2	-0.0040#	-0.0037#	-0.0017#	-0.0036#
_ISIZE_4	-0.0010#	-0.0008#	-0.0061#	-0.0007#
_Ioccupatio_1	0.0518#	0.0523#	-0.0104#	0.0519#
_Ioccupatio_2	0.2646***	0.3548***	-0.2241#	0.3708***
_Ioccupatio_3	0.0020#	0.0024#	-0.0042#	0.0020#
_Ioccupatio_4	-0.0228***	-0.0227***		-0.0223***
_Ioccupatio_5	-0.0014#	-0.0016#	0.0092#	-0.0008#
_Ioccupatio_6	0.0164**	0.0178***		0.0165#
_Ioccupatio_7	0.000016#	0.000035#		0.00004#
_Ioccupatio_8	0.0004#	0.0006#		0.0001#
_IActivitie_1	0.0013#	0.0014#	-0.0055#	0.0019#
_IActivitie_2	-0.0217*	-0.0210002*	0.0031#	-0.0220#
_IActivitie_4	0.0098#	0.0098#	0.0042#	0.0106#
_IActivitie_5	-0.0010#	-0.0009#	0.0024#	-0.0010#
_IActivitie_6	-0.0010#	0.0007#		-0.0004#
_IActivitie_7	-0.0035#	-0.0031#		-0.0034#
_IActivitie_8	-0.0015#	-0.0014#	-0.0018#	-0.0013#
_IActivitie_9	0.0017#	0.0016#	-0.0263#	0.0030#

_IActivitie_10	-0.0008#	-0.0008#		-0.0008#
_IActivitie_11	-0.0028#	-0.0027#	-0.0014#	-0.0027#
Saudi non-localized	-0.0022#	-0.0021#		-0.0038#
Non Saudi non-localize	-0.0103#	-0.0076#		
unexplained	-0.3375***	0.5913***	0.1992#	0.5917***
age	0.1902#	0.1391#	0.1062#	0.1460#
_IQualifica_1	-0.0003#	-0.0003#	-0.0003#	-0.0003#
_IQualifica_2	-0.0288#	-0.0291#	-0.0332#	-0.0287#
_IQualifica_3	0.0003#	0.00005#	0.0011#	
_IQualifica_4	-0.0119#	-0.0130#	-0.0028#	-0.0149#
_IQualifica_5	2.09e-06 #	-2.82e-06#	-0.0006#	
_IQualifica_6	0.0076#	0.0047#		0.0246#
_IQualifica_7	-0.0825***	-0.0830***	-0.0201***	-0.0957***
_IQualifica_8	-0.0161#	-0.0161#	-0.0090#	-0.0166#
_IQualifica_9	0.00002#	0.00003#	0.0005#	
_IQualifica_10	-0.0076#	-0.0072	-0.0049#	-0.0075#
_IQualifica_11	0.0011#	-0.0039***	-0.0103#	
_IQualifica_12	0.0004#	-0.0004***	-0.0029#	
_IQualifica_13				
_IQualifica_14	0.0003#	-0.0004***	-0.0017#	
_IQualifica_15				
female	0.0060#	0.0058#	0.0037#	0.0060#
_Iregions_1	0.0005#	0.0001#	0.0018#	
_Iregions_2	0.0129#	0.0129#	0.0082#	0.0121#
_Iregions_3	0.0294#	0.0222#	0.1272#	0.0229#
_Iregions_5	-.0000379#	0.0000308#	-0.0002#	
_Iregions_6	-.0000484#	-0.0000965#	0.0014#	
_Iregions_7	-0.0089#	-0.0089#	-0.0088#	-0.0089#
_Icolour_1				
_Icolour_2	-0.0098#	-0.0101#	-0.0009#	-0.0072#
_Icolour_3	0.0013#	0.0013#	0.0013#	
_Icolour_4	0.0127#	0.0129#	0.0089#	0.0130#
_Icolour_6	0.0705#	0.0707#	0.0559#	0.0712#
_Icolour_7	-0.0368#	-0.0374#	-0.0644#	-0.0364#
_Icolour_8	-0.0707#	-0.0715#	-0.0908#	-0.0709#
_Icolour_9				
_ISIZE_1	0.0370#	0.0367#	0.0314#	0.0366#
_ISIZE_2	0.0205#	0.0202#	0.0182#	0.0201#
_ISIZE_4	0.0477#	0.0476#	0.0528#	0.0474#
_Ioccupatio_1	-0.3194***	-0.3198***	-0.2571#	-0.3194#
_Ioccupatio_2	-0.5327*	-0.6229***	-0.0439#	-0.6388#
_Ioccupatio_3	-0.0355#	-0.0359#	-0.0293#	-0.0356#
_Ioccupatio_4	0.0005#	0.0004#	-0.0223#	
_Ioccupatio_5	-0.0099#	-0.0098#	-0.0206#	-0.0106#

_Ioccupatio_6	0.0002#	-0.0012#	0.0165#	
_Ioccupatio_7	0.00002#	0.000035#	0.000035#	
_Ioccupatio_8	-0.0005#	-0.0008#	-0.0001#	-0.0002#
_IActivitie_1	-0.0165#	-0.0166#	-0.0097#	-0.0170#
_IActivitie_2	-0.0539#	-0.0546#	-0.0787#	-0.0535#
_IActivitie_4	-0.0170#	-0.0169#	-0.0114#	-0.0178#
_IActivitie_5	-0.0040#	-0.0042#	-0.0075#	-0.0041#
_IActivitie_6	0.0006#	-0.0011***	-0.0004#	
_IActivitie_7	0.0001#	-0.0003#	-0.0034#	
_IActivitie_8	0.0014#	0.0012#	0.0016#	0.0011#
_IActivitie_9	-0.0586#	-0.0585#	-0.0307#	-0.0600#
_IActivitie_10	4.87e-06#	0.000023#	-0.0008#	
_IActivitie_11	-0.0198#	-0.0198#	-0.0212#	-0.0198#
Saudi non-localized	-0.0016#	-0.0018#	-0.0038#	
Non-Saudi non-localized	0.0103#	0.0076#		
_cons	0.5542#	0.5542#	0.5542#	0.5542#
Arabic				
	omega	pooled	w(0)	w(1)
explained	-0.6507***	-0.0482***	-0.2832***	-0.0169
Age	-0.9428***	0.8204***	-0.5413#	0.5075#
age2	1.1331#	-1.2537***	2.3122#	-0.6399#
age3	-0.2025#	0.6151***	-1.7292**	0.3187#
_IQualifica_1	-0.0047***	-0.0034***	-0.0030***	-0.0034***
_IQualifica_2	-0.0424***	-0.0314***	-0.0329***	-0.0303***
_IQualifica_3	0.0006#	0.0005#	-0.0005#	0.0006#
_IQualifica_4	-0.0041***	-0.0033***	-0.0030***	-0.0034***
_IQualifica_5	0.0025***	0.0027***	0.00401**	0.0027***
_IQualifica_6	0.0036***	0.0036***		0.0034***
_IQualifica_7	-0.0078***	-0.0070***	-0.0070***	-0.0068***
_IQualifica_8	-0.0076***	-0.0039***	-0.0021#	-0.0047***
_IQualifica_9	-0.0002#	-0.0001#	0.0002#	-0.0002#
_IQualifica_10	-0.0032***	-0.0026***	-0.0036***	-0.0024***
_IQualifica_11	-0.0952***	-0.0116***	-0.0103#	-0.0135***
_IQualifica_12	-0.0016***	-0.0011***	-0.0007#	-0.0011**
_IQualifica_13	-0.0003#	-0.0002#		-0.0002#
_IQualifica_14	-0.0087***	-0.0042***	-0.0012#	-0.0051***
_IQualifica_15	0.0001#	0.0003#		0.0002#
female	-0.0005#	0.0008#	0.0027***	-0.0003#
_Iregions_1	-0.0055***	-0.0057***	-0.0022#	-0.0059***
_Iregions_2	0.000041#	0.0003#	-0.0023*	0.0005#
_Iregions_3	-0.0208***	0.0304***	-0.0135#	0.0367***
_Iregions_5	-0.0009***	-0.0006***	0.0007#	-0.0007***
_Iregions_6	-0.0125***	-0.0127***	-0.0096#	-0.0130***
_Iregions_7	-0.0160***	-0.0166***	-0.0252#	-0.0167***

_Icolour_1	0.0053***	0.0019***		
_Icolour_2	0.0146***	0.0053***		0.0000202#
_Icolour_3	0.0093***	0.0027#	-0.0024#	-0.0011*
_Icolour_4	0.0092***	0.0070***	-0.0006#	0.0071***
_Icolour_6	0.0070***	0.0053***	0.0040#	0.0054***
_Icolour_7	-0.0101***	-0.0076***	-0.0090***	-0.0072***
_Icolour_8	-0.0813***	-0.0517***	-0.0572***	-0.0491***
_Icolour_9	0.0002#	0.0001#		0.00005#
_ISIZE_1	-0.0917***	-0.0716***	-0.0630***	-0.0716***
_ISIZE_2	0.0081***	0.0057***	0.0046***	0.0056***
_ISIZE_4	-0.0159***	-0.0101***	-0.0046#	-0.0113***
_Ioccupatio_1	-0.2440***	-0.1591***	-0.1916***	-0.1384***
_Ioccupatio_2	0.1034***	0.1332***	0.1635***	0.1343***
_Ioccupatio_3	0.0037***	0.0067***	0.0054***	0.0068***
_Ioccupatio_4	-0.0669***	-0.0243***	-0.0211***	-0.0318***
_Ioccupatio_5	0.0117***	0.0198***	0.0030#	0.0208***
_Ioccupatio_6	-0.0001#	-0.0002#	-0.0002#	-0.0002#
_Ioccupatio_7	0.000014#	-6.70e-06#	-0.0001#	8.15e-06#
_Ioccupatio_8	0.0003#	0.0003#	-6.70e-06#	0.0003#
_IActivitie_1	-0.0013***	-0.0010***	0.0035***	-0.0012***
_IActivitie_2	-0.0357***	-0.0279***	-0.0370***	-0.0259***
_IActivitie_4	0.0245***	0.0254***	0.0157***	0.0258***
_IActivitie_5	0.0011***	0.0012***	0.0011#	0.0012***
_IActivitie_6	-0.0059***	-0.0001#	-0.0001#	-0.0004#
_IActivitie_7	0.0004#	0.0004#	0.0008#	0.0004#
_IActivitie_8	-0.0034***	-0.0022***	-0.0018#	-0.0023***
_IActivitie_9	0.0045***	0.0044***	0.0018#	0.0045***
_IActivitie_10	0.00005#	0.0001#	0.0003#	0.000024#
_IActivitie_11	-0.0259***	-0.0182***	-0.0265***	-0.0160***
Saudi non-localize			-0.0038*	
Non-Saudi non-localize	-0.0347***	-0.0098*		0.0042***
unexplained	-0.7694***	-1.3719	-1.1370#	-1.4033***
Age	5.4308#	3.6677#	5.0293#	3.9806#
age2	-6.9436***	-4.5568#	-8.1227*	-5.1706*
age3	2.7428***	1.9252#	4.2695***	2.2216***
_IQualifica_1	0.0020#	0.0007#	0.0003#	0.0007#
_IQualifica_2	0.0084#	-0.0026#	-0.0011#	-0.0037#
_IQualifica_3	0.0021*	0.0022*	0.0032***	0.0022*
_IQualifica_4	0.0013#	0.0005#	0.0002#	0.0006#
_IQualifica_5	-0.0001#	-0.0002#	-0.0015#	-0.0002#
_IQualifica_6	-0.0002***	-0.0001***	0.0034***	
_IQualifica_7	0.0005#	-0.0003#	-0.0003#	-0.0004#
_IQualifica_8	0.0060***	0.0023#	0.0005#	0.0031#
_IQualifica_9	0.0009#	0.0009#	0.0006*	0.0009#

_IQualifica_10	-0.0020#	-0.0026#	-0.0016#	-0.0028#
_IQualifica_11	0.0852***	0.0016#	0.0004#	0.0035#
_IQualifica_12	0.0020#	0.0015#	0.0011#	0.0015#
_IQualifica_13	0.0001*	0.0000#	-0.0002#	
_IQualifica_14	0.0087***	0.0042*	0.0012*	0.00504*
_IQualifica_15	0.0001***	-0.0001***	0.0002#	
female	0.0047***	0.0035***	0.0015***	0.0045***
_Iregions_1	-0.0032***	-0.0030**	-0.0064***	-0.0028*
_Iregions_2	-0.0272***	-0.0276***	-0.0249***	-0.0278***
_Iregions_3	-0.0172#	-0.0684***	-0.0246***	-0.0748***
_Iregions_5	-0.0005#	-0.0007#	-0.0021#	-0.0007#
_Iregions_6	-0.0009#	-0.0008#	-0.0039#	-0.0005#
_Iregions_7	0.0032#	0.0039#	0.0124#	0.0039#
_Icolour_1	-0.0053***	-0.0019***		
_Icolour_2	-0.0145***	-0.0052***	0.00002#	2.89e-06#
_Icolour_3	-0.0097***	-0.0031#	0.00201#	0.0007#
_Icolour_4	0.0032#	0.0055#	0.0131#	0.0053#
_Icolour_6	0.0049#	0.0067#	0.0080#	0.0066#
_Icolour_7	-0.0038#	-0.0063#	-0.0049#	-0.0067#
_Icolour_8	0.0205*	-0.0091#	-0.0036#	-0.0117#
_Icolour_9	-0.0001#	-0.0001#	0.00005#	
_ISIZE_1	0.0134#	-0.0067#	-0.0152#	-0.0067#
_ISIZE_2	-0.0100#	-0.0076#	-0.0065#	-0.0075#
_ISIZE_4	0.0149***	0.0092#	0.0037#	0.0104#
_Ioccupatio_1	0.0468***	-0.0381***	-0.0056***	-0.0588***
_Ioccupatio_2	0.0259***	-0.0040#	-0.0342*	-0.0050*
_Ioccupatio_3	0.0056#	0.0026#	0.0039#	0.0025#
_Ioccupatio_4	0.0468***	0.0042#	0.0010#	0.0117#
_Ioccupatio_5	0.0219***	0.0139***	0.0306***	0.0128***
_Ioccupatio_6	0.0064#	0.0065#	0.0065#	0.0065#
_Ioccupatio_7	0.0001#	0.0001#	0.0002#	0.0001#
_Ioccupatio_8	-0.0016#	-0.0016#	-0.0013#	-0.0017#
_IActivitie_1	-0.0032***	-0.0036***	-0.0081***	-0.0034***
_IActivitie_2	-0.0109#	-0.0186***	-0.0096***	-0.0206***
_IActivitie_4	0.0136#	0.0126#	0.0224#	0.0122#
_IActivitie_5	0.0003#	0.0003#	0.0004#	0.0002#
_IActivitie_6	0.0059#	0.0002#	0.0002#	0.0005#
_IActivitie_7	-0.0020#	-0.0020#	-0.0023#	-0.0019#
_IActivitie_8	0.0022#	0.0011#	0.0007#	0.0011#
_IActivitie_9	0.0060#	0.0061*	0.0087***	0.0060**
_IActivitie_10	-0.0006#	-0.0006#	-0.0009#	-0.0006#
_IActivitie_11	-0.0068#	-0.0146***	-0.0063***	-0.0167***
Saudi non-localize	-0.0038***	-0.0038***#		-0.0038*
Non-Saudi non-localize	0.0389***	0.0140**	0.0042***	

_cons	-2.2786***	-2.2786***	-2.2786***	-2.2786***
African				
	omega	Pooled	w(0)	W(1)
Explained	-0.9533#	-0.5555#	-0.5674***	0.0529138#
Age	1.3266***	0.9107***	0.9722***	0.7911274#
age2	-1.4283***	-0.8551***	-0.9299***	-0.6231944#
_IQualifica_1	-0.0061***	-0.0050***	-0.0050***	
_IQualifica_2	-0.0479***	-0.0433***	-0.0431***	0.0719644#
_IQualifica_3	0.00004#	-0.0002#	-0.0004#	0.0038489#
_IQualifica_4	-0.0048***	-0.0043***	-0.0043***	
_IQualifica_5	-0.0006#	-0.0006#	-0.0006#	
_IQualifica_6				
_IQualifica_7	-0.0200***	-0.0193***	-0.0191***	
_IQualifica_8	-0.0029#	-0.0024#	-0.0024#	
_IQualifica_9	0.0004#	0.0005#	0.0005#	
_IQualifica_10	-0.0093***	-0.0086***	-0.0085***	
_IQualifica_11	-0.0346***	-0.0127*	-0.0120#	-0.1363292*
_IQualifica_12	-0.0002#	-0.0002#	-0.0001#	-0.000492#
_IQualifica_13				
_IQualifica_14	-0.0034#	-0.0018#	-0.0017#	
_IQualifica_15				
female	-0.0002#	-0.0002#	-0.0002#	-0.0006299#
_Iregions_1	-0.0003#	0.0016#	0.0016#	
_Iregions_2	-0.0221***	-0.0135***	-0.0134***	-0.0115787#
_Iregions_3	-0.0886***	-0.0209#	-0.0195#	0.2183383#
_Iregions_5	0.0006#	0.0006#	0.0007#	-0.0014042#
_Iregions_6	-0.0015#	-0.0019#	-0.0018#	-0.0035034#
_Iregions_7	-0.0169#	-0.0167**	-0.0180**	-0.0078253#
_Icolour_1				
_Icolour_2	0.0015#	0.0019#		
_Icolour_3	-0.0037#	-0.0045#	-0.0008#	-0.0041359#
_Icolour_4	-0.0006#	-0.0005#	-0.0004#	-0.013156#
_Icolour_6	0.0081#	0.0049#	0.0049#	0.0032948#
_Icolour_7	-0.0189***	-0.0144***	-0.0152***	0.0146424#
_Icolour_8	-0.0701***	-0.0569***	-0.0579***	-0.0307959#
_Icolour_9				
_ISIZE_1	-0.0767***	-0.0576***	-0.0533***	-0.068032*
_ISIZE_2	-0.0018#	-0.0013#	-0.0012#	-0.0028698#
_ISIZE_4	-0.0076*	-0.0061#	-0.0054#	-0.0213839#
_Ioccupatio_1	-0.2412***	-0.2106***	-0.2117***	
_Ioccupatio_2	-0.0245***	-0.0219***	-0.0225***	-0.0005032#
_Ioccupatio_3	-0.0063**	-0.0063***	-0.0059#	-0.0469687***
_Ioccupatio_4	-0.0328***	-0.0220***	-0.0227***	
_Ioccupatio_5	-0.0045*	-0.0012#	-0.0014#	-0.0127363#

_Ioccupatio_6	-0.0626***	-0.0367***	-0.0325*	-0.0467071#
_Ioccupatio_7	0.0001#	0.0001#	0.0001#	
_Ioccupatio_8	-0.0016#	-0.0001#	5.31e-06#	-0.0005541#
_IActivitie_1	-0.0029***	-0.0025***	-0.0025**	
_IActivitie_2	-0.0204*	-0.0186*	-0.0190*	-0.0082786#
_IActivitie_4	0.0060#	0.0044#	0.0037#	0.008229#
_IActivitie_5	0.0002#	0.0001#	0.0001#	0.0004696#
_IActivitie_6	-0.0021#	-0.0002#	-0.0001#	-0.0025337#
_IActivitie_7	0.0055#	0.0039#	0.0053#	-0.0065296#
_IActivitie_8	-0.0085***	-0.0047#	-0.0045*	
_IActivitie_9	-0.0050*	-0.0025#	-0.0025#	-0.0127815#
_IActivitie_10	-0.0008#	-0.0007#	-0.0007#	
_IActivitie_11	-0.0138#	-0.0123#	-0.0125#	-0.0093285#
Saudi non-localize	-0.0109***	-0.0116***	-0.0038*	
Non-Saudi non-localize	0.0031#	0.0153#		0.013251#
Unexplained	-0.7756***	-1.1733***	-1.1615***	-1.781774***
age	-1.1497#	-0.7339#	-0.7954#	-0.6142472#
age2	1.2786#	0.7053#	0.7802#	0.473445#
_IQualifica_1	0.0011***	0.00001#		-0.0049846***
_IQualifica_2	-0.0051#	-0.0097#	-0.0099#	-0.1249334***
_IQualifica_3	0.0132#	0.0134#	0.0136#	0.0093152*
_IQualifica_4	0.0005**	0.00002#		-0.0042792***
_IQualifica_5	-5.97e-06#	-4.53e-06#		-0.0005701#
_IQualifica_6				
_IQualifica_7	0.0009#	0.0002#		-0.0191459***
_IQualifica_8	0.0005#	-9.54e-06#		-0.0023894#
_IQualifica_9	0.000042#	0.00001#		0.0004502#
_IQualifica_10	0.0008*	0.0001#		-0.0084886***
_IQualifica_11	0.0298***	0.0079#	0.0072#	0.131547#
_IQualifica_12	0.0060#	0.0060#	0.0059#	0.0062672#
_IQualifica_13				
_IQualifica_14	0.0017***	0.0001#		-0.0017014#
_IQualifica_15				
female	-0.0108#	-0.0109#	-0.0108#	-0.0103897#
_Iregions_1	0.002***	0.00004#		0.0016314#
_Iregions_2	0.0073#	-0.0013#	-0.0014#	-0.0032722#
_Iregions_3	0.0651***	-0.0026#	-0.0040#	-0.2418517#
_Iregions_5	-0.0031#	-0.0031#	-0.0032#	-0.0010922#
_Iregions_6	-0.0032#	-0.0028#	-0.0030#	-0.0012208#
_Iregions_7	0.0156#	0.0153#	0.0167#	0.0064698#
_Icolour_1				
_Icolour_2	-0.0015#	-0.0019#		
_Icolour_3	-0.0057*	-0.0050#	-0.0086#	-0.0053191#
_Icolour_4	-0.0196#	-0.0198#	-0.0199#	-0.0071123#

_Icolour_6	-0.0109#	-0.0077#	-0.0076#	-0.0060682#
_Icolour_7	-0.0316#	-0.0361**	-0.0353#	-0.0651188***
_Icolour_8	0.0006#	-0.0126#	-0.0116#	-0.0386401#
_Icolour_9				
_ISIZE_1	-0.0049#	-0.0240#	-0.0283#	-0.0135509#
_ISIZE_2	-0.0457#	-0.0463#	-0.0463#	-0.0447098#
_ISIZE_4	0.0078#	0.0063#	0.0056#	0.0216371#
_Ioccupatio_1	0.0295***	-0.0012#		-0.2117402***
_Ioccupatio_2	-0.0031#	-0.0057#	-0.0051#	-0.0270966#
_Ioccupatio_3	0.0264#	0.0264#	0.0260#	0.0670431***
_Ioccupatio_4	0.0101***	-0.0008#		-0.0227183***
_Ioccupatio_5	0.0086#	0.0053#	0.0055#	0.0168535#
_Ioccupatio_6	0.0077#	-0.0182#	-0.0224#	-0.0081936#
_Ioccupatio_7				0.0001025#
_Ioccupatio_8	0.0007#	-0.0008#	-0.0009#	-0.0003899#
_IActivitie_1	0.0005*	0.000012#		-0.0024674**
_IActivitie_2	-0.0268#	-0.0286#	-0.0282#	-0.0389313#
_IActivitie_4	0.0259#	0.0275#	0.0282#	0.0236869#
_IActivitie_5	0.0090#	0.0091#	0.0091#	0.008715#
_IActivitie_6	0.0066#	0.0047#	0.0046#	0.0070432#
_IActivitie_7	-0.0205#	-0.0190#	-0.0204#	-0.0085205#
_IActivitie_8	0.0041***	0.0002#		-0.004462**
_IActivitie_9	0.0082***	0.0057#	0.0057#	0.0159528#
_IActivitie_10	0.0001#	0.00002#		-0.000724#
_IActivitie_11	-0.0062#	-0.0077#	-0.0075#	-0.0106885#
Saudi non-localized	0.0072***	0.0078***		-0.0037554***
Non-Saudi non-localized	0.0102#	-0.0020#	0.0133#	
_cons	-1.0132#	-1.0132#	-1.0132#	-1.013161#

Asian

	omega	pooled	w(1)	w(0)
explained	-0.8372***	-0.3298***	-0.3319***	-0.4229***
age	-1.7060***	0.5429***	0.0703	-0.4564
age2	2.4983***	-1.0136***	-0.1909	1.8787
age3	-0.8227***	0.5808***	0.2188*	-1.3516**
_IQualifica_1	-0.0057***	-0.0034***	-0.0029***	-0.0043***
_IQualifica_2	-0.0992***	-0.0681***	-0.0734***	-0.0448**
_IQualifica_3	-0.0011***	-0.0006*	-0.0008***	0.0007
_IQualifica_4	-0.0064***	-0.0057***	-0.0059***	-0.0036***
_IQualifica_5	-0.0002#	-0.0002#	-0.0001	-0.0004
_IQualifica_6	0.0003***	0.0004***	0.0004***	
_IQualifica_7	-0.0222***	-0.0204***	-0.0206***	-0.0148***
_IQualifica_8	-0.0104***	-0.0085***	-0.0093***	-0.0021
_IQualifica_9	-0.0003#	-0.0003#	-0.0003	0.0002

_IQualifica_10	-0.0095***	-0.0079***	-0.0077***	-0.0071***
_IQualifica_11	-0.1348***	-0.0635***	-0.0742***	-0.0107
_IQualifica_12	-0.0058***	-0.0053***	-0.0054***	-0.0011
_IQualifica_13	0.00002#	0.0001#	0.0001	
_IQualifica_14	-0.0078***	-0.0067***	-0.0069***	-0.0010
_IQualifica_15	0.0003***	0.0003***	0.0003***	
female	-0.0003#	-0.0002#	-0.0003	0.0017*
_Iregions_1	-0.0054***	-0.0053***	-0.0054***	-0.0028
_Iregions_2	0.0025***	0.0023***	0.0024***	-0.0025
_Iregions_3	0.0214***	0.0243***	0.0249***	-0.0057
_Iregions_5	-0.0001#	-0.0001#	-0.0001	0.0001
_Iregions_6	-0.0097***	-0.0094***	-0.0096***	-0.0088
_Iregions_7	-0.0061***	-0.0062***	-0.0061***	-0.0157***
_Icolour_1	-0.0010***	-0.0013***	-0.0012***	
_Icolour_2	-0.0022***	-0.0028***	-0.0027***	
_Icolour_3	-0.0020***	-0.0024***	-0.0022***	-0.0018
_Icolour_4	0.0015***	0.0011***	0.0011***	-0.0003
_Icolour_6	0.0032***	0.0025***	0.0026***	0.0046
_Icolour_7	-0.0028***	-0.0021***	-0.0019***	-0.0062***
_Icolour_8	-0.0521***	-0.0361***	-0.0340***	-0.0601***
_Icolour_9				
_ISIZE_1	-0.0240***	-0.0227***	-0.0222***	-0.0372***
_ISIZE_2	0.0016***	0.0015***	0.0014***	0.0030***
_ISIZE_4	0.0027***	0.0030***	0.0031***	-0.0026
_Ioccupatio_1	-0.3197***	-0.1485***	-0.1248***	-0.2076***
_Ioccupatio_2	0.0066***	0.0070***	0.0070***	0.0090***
_Ioccupatio_3	0.0047***	0.0051***	0.0051***	0.0036***
_Ioccupatio_4	-0.0722***	-0.0231***	-0.0282***	-0.0218***
_Ioccupatio_5	-0.0010***	-0.0009***	-0.0009***	-0.0004
_Ioccupatio_6	-0.0232***	-0.0237***	-0.0237***	-0.0132**
_Ioccupatio_7	-0.0007***	-0.0006***	-0.0006***	-0.0003
_Ioccupatio_8	-0.0014***	-0.0012***	-0.0012***	0.0002
_IActivitie_1	0.000014#	-0.000033#	-0.0001	0.0028***
_IActivitie_2	-0.0157***	-0.01401***	-0.0136***	-0.0228***
_IActivitie_4	-0.0009#	-0.0008#	-0.0008	-0.0007
_IActivitie_5	0.0018***	0.0017***	0.0016***	0.0020
_IActivitie_6	-0.0042***	-0.0024***	-0.0026***	-0.0001
_IActivitie_7	0.0009***	0.0010***	0.0010***	0.0018
_IActivitie_8	0.0010***	0.0012***	0.0013***	-0.0021*
_IActivitie_9	-0.0012**	-.000998**	-0.0010**	-0.0008
_IActivitie_10	0.0002***	0.0003***	0.0002**	0.0038*
_IActivitie_11	-0.0037***	-0.0018***	-0.00103***	-0.0199***
Saudi non-localized	-0.0109***	-0.0041***		-0.0038*

Non-Saudi non-localize	0.0079***	0.0097***	0.0091***	
unexplained	-1.1079***	-1.6153***	-1.6131***	-1.5222***
age	4.1472#	1.8983#	2.3709#	2.8976#
age2	-7.1504***	-3.6385#	-4.4612#	-6.5308#
age3	3.2209***	1.8174#	2.1794***	3.7498***
_IQualifica_1	0.0012#	-0.0011#	-0.0016#	-0.0002#
_IQualifica_2	0.0557***	0.0246***	0.0299***	0.0013***
_IQualifica_3	0.0023**	0.0017*	0.0019*	0.0004***
_IQualifica_4	0.0032***	0.0025***	0.0028***	0.0005***
_IQualifica_5	-0.0003#	-0.0004#	-0.0004#	-0.0001#
_IQualifica_6	0.000022*	-2.75e-06#		0.0004***
_IQualifica_7	0.0091***	0.0073***	0.0074***	0.0017***
_IQualifica_8	0.0098***	0.0079***	0.0088***	0.0015***
_IQualifica_9	0.0012	0.0012#	0.0012#	0.0007***
_IQualifica_10	0.0025#	0.0009#	0.0008#	0.0001#
_IQualifica_11	0.1294***	0.0581***	0.0688***	0.0053***
_IQualifica_12	0.0093***	0.0088***	0.0090***	0.0046***
_IQualifica_13	0.0000***	0.0000***		0.0001#
_IQualifica_14	0.0104***	0.0093***	0.0096***	0.0036***
_IQualifica_15	0.0000**	-9.32e-06***		0.0003***
female	0.0047***	0.0046***	0.0047***	0.0027***
_Iregions_1	-0.0015#	-0.0015#	-0.0015	-0.0040#
_Iregions_2	-0.0449***	-0.0447***	-0.0449***	-0.0399***
_Iregions_3	-0.1037***	-0.1065***	-0.1072***	-0.0765***
_Iregions_5	-0.0007#	-0.0007#	-0.0007#	-0.0008#
_Iregions_6	-8.48e-06#	-0.0003#	-0.0001#	-0.0009#
_Iregions_7	0.0070***	0.0070***	0.0070***	0.0166***
_Icolour_1	-0.0002#	0.0001#		-0.0012***
_Icolour_2	-0.0009***	-0.0003#	-0.0005#	-0.0032***
_Icolour_3	-0.0006#	-0.0002#	-0.0003#	-0.0008#
_Icolour_4	0.0013#	0.0017#	0.0017#	0.0032#
_Icolour_6	-0.0092#	-0.0085#	-0.0086#	-0.0106#
_Icolour_7	-0.0220***	-0.0226***	-0.0228***	-0.0185***
_Icolour_8	-0.0177#	-0.0337***	-0.0358***	-0.0097***
_Icolour_9				
_ISIZE_1	0.0215***	0.0202***	0.0197***	0.0348***
_ISIZE_2	0.0176#	0.0177#	0.0178#	0.0161#
_ISIZE_4	-0.0155***	-0.0157***	-0.0159***	-0.0102***
_Ioccupatio_1	0.1104***	-0.0608***	-0.0845***	-0.0017***
_Ioccupatio_2	-0.0059*	-0.0063**	-0.0063***	-0.0083***
_Ioccupatio_3	0.0043#	0.0040#	0.0040#	0.0055#
_Ioccupatio_4	0.0508***	0.0017#	0.0068#	0.0004#
_Ioccupatio_5	0.0028#	0.0027#	0.0027#	0.0022#
_Ioccupatio_6	-0.0148#	-0.0144*	-0.0144#	-0.0248#

_Ioccupatio_7	0.00003#	-0.0001#	-0.0001#	-0.0004#
_Ioccupatio_8	-0.0008#	-0.0010#	-0.0009#	-0.0023#
_IActivitie_1	-0.0027***	-0.0026***	-0.0026***	-0.0055***
_IActivitie_2	-0.0257***	-0.0273***	-0.0277***	-0.0186***
_IActivitie_4	0.0046#	0.0046#	0.0046#	0.0044#
_IActivitie_5	-0.0005#	-0.0004#	-0.0003#	-0.0007#
_IActivitie_6	0.0063**	0.0045#	0.0047#	0.0021#
_IActivitie_7	-0.0014#	-0.0016#	-0.0016#	-0.0024#
_IActivitie_8	-0.0069***	-0.0071***	-0.0072***	-0.0037***
_IActivitie_9	0.0009#	0.0008#	0.0008#	0.0006#
_IActivitie_10	-0.0007#	-0.0007#	-0.0007#	-0.0043*
_IActivitie_11	-0.0375***	-0.0393***	-0.0401***	-0.0213***
Saudi nonlocalized	0.0070***	0.0003#	-0.0038*	
Non-Saudi non-localize	0.0012#	-0.0006#		0.0091***
_cons	-1.4863#	-1.4863#	-1.4863#	-1.4863#

10.4.2.5 Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Table 10-29: heteroskedasticity for each group.

fitted values of log earning	high background	African	Asian	Arabic
Ho: model has no omitted variables				
chi2(1)	0.37	18314.00	10815.29	1591.02
Prob > chi2	0.5418	0.0000	0.0000	0.000

10.4.3 Saudi and non-Saudi wage gap according to firm's status

2013					
Localized firm		Omega	Pooled	W(0)	W(1)
total	explained	-0.5981***	-0.22687***	-0.3392048***	-0.241361***
	unexplained	-1.22932***	-1.60054***	-1.48821***	-1.58605***
explained	worker	-0.23821***	-0.03037***	-0.0263535#	-0.047178***
	job	-0.35989***	-0.1965***	-0.3128513***	-0.194183***
unexplained	worker	0.402835#	0.194998#	0.1909828#	0.211807#
	job	-0.04514#	-0.20853***	-0.0921787**	-0.210847***
	_cons	-1.58701*	-1.58701*	-1.587014*	-1.587014*
non-localized firms					
Total	explained	-0.251645*	-0.199612	-0.2246598	-0.0420333
	unexplained	-1.72532***	-1.77736***	-1.75231***	-1.93494***
	worker	-0.110102	-0.06437	0.5114431***	0.035791

explained	job	-0.14154***	-0.13524***	-0.7361029**	-0.0778244
unexplained	worker	133.6198***	133.574***	5.107935***	5.583587***
	job	3.499805***	3.493499***	-1.278907***	-1.93719***
	_cons	-138.845***	-138.845***	-5.581338***	-5.58134***
2017					
Localized firm					
Total	Explained	-1.07156***	0.088998***	0.201446***	-0.29149***
	unexplained	-0.522056***	-1.68261***	-1.79506***	-1.30212***
Explained	worker	-0.662323***	0.232000***	0.325773***	-0.15932***
	job	-0.409234***	-0.143002***	-0.12433***	-0.1322***
unexplained	worker	1.41932***	0.5249921***	0.43122***	0.91631***
	job	0.25287***	-0.013362***	-0.03204***	-0.02419***
	_cons	-2.19424***	-2.19424***	-2.19424***	-2.19424***
Localized firm					
Total	explained	-1.13249***	-0.0613***	0.06866***	-0.3404***
	unexplained	-0.30585***	-1.3770***	-1.507***	-1.098***
explained	Worker	-0.92944**	0.00727***	0.1098***	-0.2302***
	Job	-0.20305***	-0.0686***	-0.0412***	-0.1101***
unexplained	Worker	1.2217***	0.28503***	0.1825***	0.5225**
	Job	0.15086***	0.0164***	-0.011***	0.0579***
	_cons	-1.6784***	-1.6784***	-1.6784***	-1.6784***

10.4.4 The wage gap by firm status between the group members.

2013				
A	Total dataset	Pooled	W(0)	W(1)
Total	Explained	-.1044984***	-.1051826***	-.087321***
	unexplained	-.0012119#	-.0005277 #	-.0183894#
Explained	worker	-.0096053***	-.009713***	-.0002004#
	job	-.0691306***	-.0697328***	-.056107***
	Saudi	-.0257626***	-.0257368***	-.031014***
unexplained	worker	-.6798193#	-.6797116#	-.6892242#
	job	-.3105265***	-.3099243***	-.323551***
	Saudi	.0016457#	.00162*	.006897***
	_cons	.9874882#	.9874882#	.9874882#
B	Saudi			
Total	Explained	-.1289455#	-.1278626#	.473501#

	unexplained	.2014365#	.2003535#	-.40101***
Explained	worker	-.0072698#	-.0072766#	.2983408#
	job	-.1216758#	-.120586#	.17516#
unexplained	worker	-135.4776***	-135.4776***	-135.78***
	job	-2.971199***	-2.972288***	-3.26803***
	_cons	138.6502***	138.6502***	138.6502***
C	Non-Saudi			
Total	Explained	-.0736313***	-.074339***	-.053629***
	unexplained	-.0034323 #	-.0027246#	-.0234345**
Explained	worker	-.0092968***	-.0094387***	.00106
	job	-.0643345***	-.0649002***	-.054689***
unexplained	worker	-.2554612#	-.2553192#	-.2658179 #
	job	-.3223218***	-.3217561***	-.331967***
	_cons	.5743508#	.5743508#	.5743508#
2017				
D	Total dataset			
Total	Explained	-.3160532***	-.3195993***	-.226581***
	unexplained	-.0463553***	-.0428092***	-.135828***
Explained	worker	-.0039468***	-.0021843***	-.01994***
	job	-.0929495***	-.0957363***	-.0252238***
	Saudi	-.2191569***	-.2216788***	-.18142***
unexplained	worker	-.5620665***	-.5638291***	-.54607***
	job	-.0458442***	-.0430574***	-.11357***
	Saudi	-.1006274***	-.0981055***	-.13837***
	_cons	.6621828***	.6621828***	.66218***
E	Saudi			
Total	Explained	-.206842***	-.2124794***	-.06374***
	unexplained	-.0510431***	-.0454056***	-.194145***
Explained	worker	-.0836028***	-.0865097***	-.038215***
	job	-.1232391***	-.1259698***	-.025525***
unexplained	worker	.5842513***	.5871581***	.538863***
	job	-.056792***	-.0540609***	-.154506***
	_cons	-.5785028***	-.5785028***	-.578503***
F	non-Saudi			
Total	Explained	-.0657257***	-.0665592***	-.038589***
	unexplained	-.0368806***	-.0360472***	-.064018***
Explained	worker	-.0193029***	-.0189253***	-.021083***
	job	-.0464228#	-.0476339***	-.017506***
unexplained	worker	.0552638***	.0548861#	.0570436#
	job	-.0294372 #	-.0282261***	-.058354***
	_cons	-.0657257***	-.0627072#	-.0627072#

10.4.5 Consumption's slop-assumption

10.4.5.1 2013

Table 10-30: the percentage of explained part 2013.

	Consumption slop	Omega	Pooled	W(0)	W(1)
1	4000>	80%	69%	35%	102%
2	Saudi .9 Non-Saudi .7	83%	72%	36%	130%
3	Saudi .7 Non-Saudi .9	71%	59%	34%	74%
4	Saudi 0.77 Non-Saudi 0.85	76%	64%	34%	89%
5	Saudi 0.85 Non-Saudi 0.77	81%	69%	35%	110%
6	Saudi 0.9 Non-Saudi 0.8	80%	70%	35%	113%
7	Saudi 0.8 Non-Saudi 0.9	75%	63%	34%	87%
8	Both .9	79%	67%	35%	99%
9	Saudi < 3000=0.9 non-Saudi < 3000=0.85 + 3000=.8 +10000 =0.75	81%	71%	36%	36%
10	Wage > 3000 = .8 Wage <3000=.9	79%	68%	35%	101%

Table 10-31: coefficients of OB

Differences -1.829881***		omega	Pooled	W(0)	W(1)
assumption 2					
explained	Total	-1.5175***	-1.3160***	-0.6555	-2.3814***
	Worker	-0.0942***	-0.0411***	-0.0455	-0.0176***
	Job	-0.0558***	-0.0308***	-0.0898	-0.0635***
	Nitaqat	-0.0234***	-0.0198***	-0.0197	-0.0146***
	consumption	-1.3441***	-1.2242***	-0.5004	-2.2857***
Unexplained	Total	-0.3124***	-0.5139***	-1.1744	0.5516***
	Worker	-0.9580*	-1.0111*	-1.0067**	-1.0347**
	Job	-0.1048***	-0.1297***	-0.0708***	-0.0970***
	Nitaqat	0.0156	0.0120	0.0120	0.0069
	consumption	1.0716***	0.9517***	0.2279***	2.0132***
	Cons	-0.3368	-0.3368	-0.3368	-0.3368
assumption 3					
explained	Total	-1.3082***	-1.0706***	-0.6138#	-1.3632***
	Worker	-0.1185***	-0.0317***	-0.0455#	-0.0176***
	Job	-0.0736***	-0.0212***	-0.0898#	-0.0635***
	Nitaqat	-0.0233***	-0.0163***	-0.0197#	-0.0146***
	consumption	-1.0928***	-1.0013***	-0.4587#	-1.2675***
Unexplained	Total	-0.5216***	-0.7593***	-1.2161#	-0.4667***
	worker	-0.9337*	-1.0205***	-1.0067**	-1.0347**
	Job	-0.08697***	-0.1394***	-0.0708***	-0.0970***
	Nitaqat	0.01557#	0.0086#	0.0120#	0.0069#
	consumption	0.8203***	0.7288***	0.1862***	0.9949***

	cons	-0.3367964#	-0.3368	-0.3368#	-0.3368#
assumption 4					
explained	Total	-1.3922***	-1.1704***	-0.6287***	-1.6200***
	Worker	-0.1082***	-0.0350***	-0.0455***	-0.0176***
	Job	-0.0621***	-0.0209***	-0.0898***	-0.0635***
	Nitaqat	-0.0227***	-0.0170***	-0.0197***	-0.0146***
	consumption	-1.1992***	-1.0974***	-0.4737***	-1.5243***
Unexplained	Total	-0.4377***	-0.6595***	-1.2011***	-0.2099#
	worker	-0.9440*	-1.0172*	-1.0067**	-1.0347**
	Job	-0.0984***	-0.1396***	-0.0708***	-0.0970***
	Nitaqat	0.0149#	0.0093#	0.0120#	0.0069#
	consumption	0.9267***	0.8249***	0.2011***	1.2517***
	cons	-0.3368#	-0.3368#	-0.3368#	-0.3368#
assumption 5					
explained	Total	-1.4757***	-1.2693***	-0.6450***	-2.0170***
	Worker	-0.0985***	-0.0388***	-0.0455***	-0.0176***
	Job	-0.0551***	-0.0247***	-0.0898***	-0.0635***
	Nitaqat	-0.0227***	-0.0184***	-0.0197***	-0.0146***
	consumption	-1.2993***	-1.1874***	-0.4899***	-1.9212***
Unexplained	Total	-0.3542***	-0.5605***	-1.1849***	0.1871***
	worker	-0.9537*	-1.0134*	-1.0067**	-1.0347**
	Job	-0.1054***	-0.1358***	-0.0708***	-0.0970#
	Nitaqat	0.0149#	0.0107#	0.0120#	0.0069#
	consumption	1.0268***	0.9148***	0.2174***	1.6487***
	cons	-0.3368#	-0.3368#	-0.3368#	-0.3368#
assumption 6					
explained	Total	-1.4819***	-1.2766***	-0.6464***	-2.0593***
	Worker	-0.0978***	-0.0391***	-0.0455***	-0.0176***
	Job	-0.0549***	-0.0254***	-0.0898***	-0.0635***
	Nitaqat	-0.0228***	-0.0186***	-0.0197***	-0.0146***
	consumption	-1.3064***	-1.1936***	-0.4913***	-1.9635***
Unexplained	Total	-0.3479***	-0.5533***	-1.1835#	0.2294***
	worker	-0.9544*	-1.0131*	-1.0067**	-1.0347**
	Job	-0.1056***	-0.1352***	-0.0708***	-0.0970#
	Nitaqat	0.0150#	0.0108#	0.0120#	0.0069#
	consumption	1.0339***	0.9210***	0.2188***	1.6910***
	cons	-0.3368#	-0.3368#	-0.3368#	-0.3368#
assumption 7					
explained	Total	-1.3826***	-1.1590***	-0.6270***	-1.5859***
	Worker	-0.1094***	-0.0346***	-0.0455***	-0.0176***
	Job	-0.0633***	-0.0208***	-0.0898***	-0.0635***
	Nitaqat	-0.0227***	-0.0169***	-0.0197***	-0.0146***
	consumption	-1.1873***	-1.0866***	-0.4719***	-1.4902***
Unexplained	Total	-0.4473***	-0.6709***	-1.2029***	-0.2439***
	worker	-0.9429*	-1.0176*	-1.0067**	-1.0347**
	Job	-0.0973***	-0.1398***	-0.0708***	-0.0970***
	Nitaqat	0.0149#	0.0092#	0.0120#	0.0069#
	consumption	0.9147***	0.8141***	0.1994***	1.2177***
	cons	-0.3368#	-0.3368#	-0.3368#	-0.3368#
assumption 8					
explained	Total	-1.4379***	-1.2249***	-0.6373***	-1.8087***
	Worker	-0.1028***	-0.0370***	-0.0455***	-0.0176***
	Job	-0.0575***	-0.0222***	-0.0898***	-0.0635***

	Nitaqat	-0.0225***	-0.0177***	-0.0197***	-0.0146***
	consumption	-1.2551***	-1.1480***	-0.4822***	-1.7129***
Unexplained	Total	-0.3920***	-0.6050***	-1.1926***	-0.0212#
	worker	-0.9494***	-1.0152*	-1.0067**	-1.0347**
	Job	-0.1031***	-0.1383#	-0.0708***	-0.0970***
	Nitaqat	0.0148#	0.0099#	0.0120#	0.0069#
	consumption	0.9825***	0.8754***	0.2097***	1.4404***
	cons	-0.3368#	-0.3368#	-0.3368#	-0.3368#
assumption 9					
explained	Total	-1.4828***	-1.2903***	-0.6568***	-0.6568***
	Worker	-0.0940***	-0.0359***	-0.0456***	-0.0456***
	Job	-0.0454***	-0.0146*	-0.0918***	-0.0918***
	Nitaqat	-0.0205***	-0.0162***	-0.0204***	-0.0204***
	consumption	-1.3229***	-1.2236***	-0.4990***	-0.4990***
Unexplained	Total	-0.3471***	-0.5396***	-1.1731***	-1.1731***
	worker	-0.9894*	-1.0476**	-1.0379**	-1.0379**
	Job	-0.1139***	-0.1447***	-0.0675***	-0.0675***
	Nitaqat	0.0086#	0.0044#	0.0085#	0.0085#
	consumption	1.0623***	0.9630***	0.2385***	0.2385***
	cons	-0.3147#	-0.3147#	-0.3147#	-0.3147#
assumption 10					
explained	Total	-1.4486***	-1.2418***	-0.6386***	-1.8400***
	Worker	-0.1011***	-0.0371***	-0.0464***	-0.0165***
	Job	-0.0524***	-0.0181***	-0.0937***	-0.0591***
	Nitaqat	-0.0218***	-0.0171***	-0.0207***	-0.0140***
	consumption	-1.2734***	-1.1696***	-0.4779***	-1.7506***
Unexplained	Total	-0.3813***	-0.5881***	-1.1913***	0.0102#
	worker	-0.9549*	-1.0189*	-1.0095*	-1.0395#
	Job	-0.1077***	-0.1419***	-0.0664***	-0.1010#
	Nitaqat	0.0108#	0.0061#	0.0097#	0.0030#
	consumption	1.0334***	0.9296***	0.2378***	1.5105#
	cons	-0.3630#	-0.3630#	-0.3630#	-0.3630#

9.4.5.2 2017

Table 10-32: the explained part for 2017.

	Consumption slop assumption	Omega	Pooled	W(0)	W(1)
1	4000>	85%	33%	21%	75%
2	Saudi .9 non-Saudi .7	85%	33%	22%	93%
3	Saudi .7 Non-Saudi 0.9	83%	30%	19%	56%
4	Saudi 0.77 Non-Saudi 0.85	84%	31%	20%	65%
5	Saudi 0.85 Non-Saudi 0.77	85%	33%	21%	80%
6	Saudi 0.9 Non-Saudi 0.8	85%	33%	21%	81%
7	Saudi 0.8	84%	31%	20%	64%

	Non-Saudi 0.9				
8	Both .9	84%	32%	21%	72%
9*	Saudi < 3000=0.9 non-Saudi < 3000=0.85 + 3000=.8 + 10000 =0.75	85%	35%	23%	79%

Table 10-33: coefficients of OB

Differences -1.584268***		omega	Pooled	W(0)	W(1)
assumption 2					
explained	Total	-1.3470***	-0.5227***	-0.3479***	-1.4755***
	Worker	-0.5349***	-0.0153***	0.0582***	-0.1420***
	Job	-0.1659***	-0.0718***	-0.0309***	-0.0757***
	Nitaqat	-0.0720***	-0.0201***	-0.0116***	-0.0119***
	consumption	-0.5742***	-0.4155***	-0.3636***	-1.2460***
Unexplained	Total	-0.2373***	-1.0616***	-1.2363***	-0.1088***
	worker	-0.0851***	-0.6047***	-0.6782***	-0.4780***
	Job	0.1832***	0.0891***	0.0482***	0.0930***
	Nitaqat	0.0583***	0.0065***	-0.0020***	-0.0017***
	consumption	0.3851***	0.2264***	0.1745***	1.0569***
	cons	-0.7789***	-0.7789***	-0.7789***	-0.7789***
assumption 3					
explained	Total	-1.3202***	-0.4747***	-0.3010***	-0.8859***
	Worker	-0.5786***	-0.0207***	0.0582***	-0.1420***
	Job	-0.1698***	-0.0624***	-0.0309***	-0.0757***
	Nitaqat	-0.0749***	-0.0170***	-0.0116***	-0.0119***
	consumption	-0.4969***	-0.3745***	-0.3167***	-0.6564***
Unexplained	Total	-0.2640***	-1.1096***	-1.2833***	-0.6984***
	worker	-0.0414***	-0.5993***	-0.6782***	-0.4780***
	Job	0.1871***	0.0797***	0.0482***	0.0930***
	Nitaqat	0.0613***	0.0034***	-0.0020***	-0.0017***
	consumption	0.3078***	0.1854***	0.1276***	0.4673***
	cons	-0.7789***	-0.7789***	-0.7789***	-0.7789***
assumption 4					
Explained	Total	-1.3312***	-0.4968***	-0.3178***	-1.0346***
	Worker	-0.5626***	-0.0206***	0.0582***	-0.1420***
	Job	-0.1672***	-0.0653***	-0.0309***	-0.0757***
	Nitaqat	-0.0734***	-0.0180***	-0.0116***	-0.0119***
	consumption	-0.5280***	-0.3928***	-0.3335***	-0.8051***
Unexplained	Total	-0.2531***	-1.0874***	-1.2665***	-0.5497***
	worker	-0.0574***	-0.5993***	-0.6782***	-0.4780***
	Job	0.1845***	0.0826***	0.0482***	0.0930***
	Nitaqat	0.0598***	0.0044***	-0.0020***	-0.0017***
	consumption	0.3389***	0.2037***	0.1444***	0.6160***
	cons	-0.7789***	-0.7789***	-0.7789***	-0.7789***
assumption 5					
explained	Total	-1.3414***	-0.5150***	-0.3361***	-1.2645***
	Worker	-0.5455***	-0.0182***	0.0582***	-0.1420***
	Job	-0.1659***	-0.0690***	-0.0309***	-0.0757***
	Nitaqat	-0.0723***	-0.0193***	-0.0116***	-0.0119***
	consumption	-0.5577***	-0.4084***	-0.3518***	-1.0349***

Unexplained	Total	-0.2428***	-1.0693***	-1.2482***	-0.3198***
	worker	-0.0744***	-0.6017***	-0.6782***	-0.4780***
	Job	0.1832***	0.0863***	0.0482***	0.0930***
	Nitaqat	0.0587***	0.0056***	-0.0020***	-0.0017***
	consumption	0.3686***	0.2193***	0.1627***	0.8458***
	cons	-0.7789***	-0.7789***	-0.7789***	-0.7789***
assumption 6					
explained	Total	-1.3422***	-0.5162***	-0.3377***	-1.2890***
	Worker	-0.5441***	-0.0179***	0.0582***	-0.1420***
	Job	-0.1658***	-0.0694***	-0.0309***	-0.0757***
	Nitaqat	-0.0723***	-0.0194***	-0.0116***	-0.0119***
	consumption	-0.5601***	-0.4095***	-0.3534***	-1.0594***
Unexplained	Total	-0.2420***	-1.0681***	-1.2466***	-0.2953***
	worker	-0.0759***	-0.6020***	-0.6782***	-0.4780***
	Job	0.1831***	0.0867***	0.0482***	0.0930***
	Nitaqat	0.0586***	0.0058***	-0.0020***	-0.0017***
	consumption	0.3710***	0.2204***	0.1643***	0.8703***
	cons	-0.7789***	-0.7789***	-0.7789***	-0.7789***
assumption 7					
explained	Total	-1.3300***	-0.4945***	-0.3158***	-1.0149***
	Worker	-0.5644***	-0.0207***	0.0582***	-0.1420***
	Job	-0.1674***	-0.0650***	-0.0309***	-0.0757***
	Nitaqat	-0.0735***	-0.0179***	-0.0116***	-0.0119***
	consumption	-0.5246***	-0.3909***	-0.3315***	-0.7853***
Unexplained	Total	-0.2543***	-1.0898***	-1.2684***	-0.5694***
	worker	-0.0555***	-0.5992***	-0.6782***	-0.4780***
	Job	0.1847***	0.0822***	0.0482***	0.0930***
	Nitaqat	0.0599***	0.0043***	-0.0020***	-0.0017***
	consumption	0.3355***	0.2018***	0.1424***	0.5962***
	cons	-0.7789***	-0.7789***	-0.7789***	-0.7789***
assumption 8					
explained	Total	-1.3368***	-0.5072***	-0.3274***	-1.1439***
	Worker	-0.5535***	-0.0197***	0.0582***	-0.1420***
	Job	-0.1663***	-0.0672***	-0.0309***	-0.0757***
	Nitaqat	-0.0727***	-0.0187***	-0.0116***	-0.0119***
	consumption	-0.5442***	-0.4016***	-0.3431***	-0.9143***
Unexplained	Total	-0.2475***	-1.0771***	-1.2569***	-0.4404***
	worker	-0.0664***	-0.6003***	-0.6782***	-0.4780***
	Job	0.1836***	0.0845***	0.0482***	0.0930***
	Nitaqat	0.0591***	0.0050***	-0.0020***	-0.0017***
	consumption	0.3551***	0.2125***	0.1540***	0.7252***
	cons	-0.7789***	-0.7789***	-0.7789***	-0.7789***
assumption 9					
explained	Total	-1.3512***	-0.5482***	-0.3587!	-1.2342#
	Worker	-0.524582***	-0.0190***	0.0606!	-0.1354***
	Job	-0.157575***	-0.0656***	-0.0328!	-0.0709***
	Nitaqat	-0.068641***	-0.0181***	-0.0121!	-0.0253#
	consumption	-0.600401***	-0.4455***	-0.3743!	-1.0025***
Unexplained	Total	-0.23307***	-1.0361***	-1.2256!	-0.3501#
	worker	-0.13977***	-0.6453***	-0.7250!	-0.5290***

	Job	0.17322***	0.0813***	0.0485!	0.0866***
	Nitaqat	0.0550366***	0.0045***	-0.0015!	0.0117#
	consumption	0.4010173***	0.2461***	0.1750!	0.8031***
	cons	-0.7225723!	-0.7226!	-0.7226!	-0.7226***
assumption 10					
Explained	Total	-1.3400***	-0.5159***	-0.3324!	-1.1851
	Worker	-0.5420***	-0.0156***	0.0634!	-0.1352***
	Job	-0.1651***	-0.0674***	-0.0339!	-0.0722***
	Nitaqat	-0.0718***	-0.0185***	-0.0125!	-0.0207
	consumption	-0.5612***	-0.4143***	-0.3495!	-0.9570***
Unexplained	Total	-0.2443***	-1.0684***	-1.2519!	-0.3991
	worker	-0.1297***	-0.6561***	-0.7351!	-0.5365***
	Job	0.1818***	0.0841***	0.0506!	0.0889***
	Nitaqat	0.0581***	0.0048***	-0.0013!	0.0070
	consumption	0.3871***	0.2403***	0.1754!	0.7830***
	cons	-0.7416!	-0.7416!	-0.7416!	-0.7416***