IS MORTALITY SPATIAL OR SOCIAL?

COLIN O'HARE[‡] AND YOUWEI LI[†]

ABSTRACT. Mortality modelling for the purposes of demographic forecasting and actuarial pricing is generally done at an aggregate level using national data. Modelling at this level fails to capture the variation in mortality within country and potentially leads to a mis-specification of mortality forecasts for a subset of the population. This can have detrimental effects for pricing and reserving in the actuarial context. In this paper we consider mortality rates at a regional level and analyse the variation in those rates. We consider whether variation in mortality rates within a country can be explained using local economic and social variables. Using Northern Ireland data on mortality and measures of deprivation we identify the variables explaining mortality variation. We create a population polarisation variable and find that this variable is significant in explaining some of the variation in mortality rates. Further, we consider whether spatial and non-spatial models have a part to play in explaining mortality differentials.

Keywords: Mortality rates; Frailty models; Deprivation measures.

Date: Latest version: July 21, 2014.

[‡] Department of Econometrics and Business Statistics, Monash University, 3800 Melbourne, Victoria, Australia, Tel: (+613) 9905 2327, and [†]School of Management, Queen's University of Belfast, BT9 5EE, Belfast, United Kingdom. Tel: (+44) 28 9097 4826. Fax: (+44) 28 9097 4201. Emails: colin.ohare@monash.edu and y.li@qub.ac.uk.

Acknowledgements: We are indebted to Professor Rob Hyndman of the Department of Econometrics and Business statistics, Monash University and to Professor Michael Sherris of the Australian Business School, University of New South Wales for their helpful comments. We benefited from detailed comments of two anonymous referees and Stephen Hall, the editor, on the earlier version of the paper. Financial support for Li from the Australian Research Council (ARC) under Discovery Grant (DP130103210) is gratefully acknowledged.

1. INTRODUCTION

How mortality rates are changing over time, and in particular the increase observed in life expectancy, has been a topic of considerable debate across the world over recent decades. Increases in the cost of providing for pensions, insurance and healthcare at older ages, driven by the rapid improvements in life expectancy, have led to life companies, pension schemes, individuals and governments giving more consideration to how these costs will be met in the future. Key to quantifying these future costs is the need for an accurate picture of how mortality rates vary over time and over populations, and as a result academics and practitioners have focused their efforts on accurately forecasting and quantifying expected future improvements in mortality rates. Aggregate mortality rates, that is mortality rates at a national level have shown an improving trend over many decades and it is variations in this trend that practitioners have provided advice on and that academics have wrestled with. It is the uncertainty in this trend that has been coined "longevity risk". In a financial context quantifying longevity risk has become a topic of great interest as the capital markets work to create ways to buy longevity risk as a diversification from their traditional financial risks (See Blake, Cairns,and Dowd, 2008, for example).

Within actuarial work, the pricing and reserving for life related products and pensions are based on the latest knowledge of mortality forecasts. Actuaries therefore also have a keen interest in accurately forecasting mortality rates. Implicit in the actuaries advice is the assumption of a homogeneous population, be that within a pension scheme work force or within a region of a country or in a whole country. In the past this has been sufficient and national mortality tables have been used to price products or to advise clients. Driven by the financial and actuarial sector less attention has been given in the actuarial space to considering the possibility that mortality rates within a population are not homogeneous. There is some implicit allowance for differing socioeconomic status within actuarial pricing or pensions since the pension amount is an indicator of socioeconomic status (those with larger pensions will on average have a higher socioeconomic status) however, this proxy is not very robust and indeed Richards (2008) shows that a mortality model using geographic classifications (a postcode mortality model) better fits United Kingdom annuitant mortality than a model using pension amounts. The actuarial profession is thus coming up with some way to recognising mortality variation within populations.

The approach used by some with disaggregating mortality experience using purely geographical location (postcode mortality) still does not help in the search for observable, meaningful variables that explain why mortality experience differs within populations. Developing a model that gives consideration to the variation in mortality rates within a population would be important for actuaries as the clients they advise do not usually have a mortality experience that matches that of the aggregated national data. For example, a pension scheme actuary advising the trustees of a scheme of London city workers will expect very different mortality experience than that of the actuary advising a company of coal miners. In addition many clients are international and therefore within one scheme there may be individuals or groups who have very different mortality experience. For these reasons consideration of mortality variation is valuable to actuaries and ignoring variations in mortality rates such as this can lead to mis pricing or mis reserving which ultimately may result in unsustainable or unanticipated liabilities. Combining the use of geographical information as proposed by Richards (2008) with observable variables is one of the aims of this paper.

We consider mortality data gathered from Northern Ireland and divided by geographical location. We use a suite of deprivation data, also divided by geographical location and gathered from the 2010 deprivation study[1] to analyse the variation in mortality rates as a function of various measures of deprivation. We further consider the use of frailty modelling techniques to explain the remaining variation after controlling for deprivation. Considering the deprivation measures separately we are able to identify which aspects of deprivation are more important in explaining the variation we see in mortality rates. We also draw some conclusions regarding the use of frailty modelling in the presence of good socioeconomic data.

The rest of the paper is laid out as follows. In Section 2 we outline the literature in mortality modelling, Section 3 describes the data we have used in this study, and in Section 4 we provide the empirical analysis. Section 5 concludes.

2. Literature review

Within the literature on mortality modelling there has been a large number of models developed over the last two decades. However, little consideration has been given to the issue of linking mortality rates to observable variables such as economic and social factors. This has

[1]The 2010 deprivation study reports deprivation measures across Northern Ireland compiled based on 2008 data, http://www.nisra.gov.uk/deprivation/archive/Updateof2005Measures/NIMDM_2010_Indicator_Summary.pdf

been a particular issue within the actuarial literature where the main focus has been to forecast and price longevity improvements whatever the cause of those improvements. The proliferation of models that have been developed in the last two decades have been helpfully categorised by Booth and Tickle (2008) into one of three types of model. (i) *Extrapolative models* - taking past data and extrapolating identified trends, (ii) *Explanatory models* - modelling mortality as a dependent variable explained by socioeconomic, biological and environmental factors, and (iii) *Expectations models* - taking advantage of the expert knowledge of actuaries and demographers and targeting future life expectancy at some expert held belief. Expectations models are the domain of practicing actuaries and rely heavily on their expert judgement to predict long term improvement rates in mortality, we do not consider them further here.

The academic and financial worlds have focused most of their attention on the method of extrapolation as a method to quickly identify and forecast patterns in mortality rates. With the desired outcome being the ability to accurately price future mortality linked financial and insurance products this has been a very successful endeavour. There are many papers in the area to testify to that success. See for example Lee and Carter (1992), Booth *et al.* (2002), Brouhns *et al.* (2002), Girosi and King (2005), Renshaw and Haberman (2006), Cairns *et al.* (2006), Currie *et al.* (2004), Currie (2006), Hári *et al.* (2008), Tuljapurkar (2008), Plat (2009), and O'Hare and Li (2012).

The explanatory approach has been primarily left to the medical and social science disciplines and there has been considerable literature linking mortality rates to observable factors. In particular there have been significant efforts to identify the relationship between wealth and mortality rates, see for example Acemoglu and Johnson (2007), Bhargava *et al.* (2001), Bloom, Canning, and Sevilla (2004), de la Croix and Licandro (1999), Lindahl (2005), and Preston (1975). The results of this analysis are mixed with strong evidence for a positive wealth effect and strong arguments to the contrary. There have also been many studies linking social and economic factors to mortality rates, see for example O'Hare and French (2013) who link the latent factor structure of mortality rates to observable economic and social variables from OECD data and identify factors such as GDP, fat intake and smoking habits to be significant in explaining mortality rates. Notably in this paper they identify a different number of observable variables for each of the countries they consider. Turrell *et al.* (2006) consider mortality rates in Tasmania

and associate the variance in mortality to measures of socioeconomic disadvantage, social capital and geographic remoteness. They conclude that socioeconomic disadvantage is associated with area variation in mortality rates but that social capital or remoteness does not explain the variation seen in mortality rates. Whilst the results were limited this paper does consider the question of mortality rates at a sub national level, dividing the mortality data into geographical regions along with the measures of socioeconomic disadvantage, social capital and remoteness also divided geographically.

In our paper we extend on this in several ways, firstly we consider a range of measures of socioeconomic disadvantage separately. These include measures of healthcare deprivation, income deprivation, employment deprivation and education deprivation. They also include measures of remoteness such as proximity to services. Secondly, in our paper we give consideration to the ability of spatial and non-spatial frailty modelling to explain mortality variation. Spatial models have been developed and applied to modelling house prices, crime levels and diseases amongst many other applications[2]. Rosen (1974) models house prices using spatial covariates including environmental attributes and geographical characteristics. Waller *et al.* (2007) models geographic variation in alcohol distribution and violent crime in Houston. Kazembe (2007) examined spatial clustering of malaria risks in northern Malawi. Geodemographic modelling, the spatial modelling of demographic data, is used in a range of applications. Richards (2008) uses geodemographic profiles based on postcodes to analyse life insurance and pension scheme mortality. Tuljapurkar and Boe (1998) outline mortality differentials by sex, education and socioeconomic variables. Richards and Jones (2004) discuss the impact of socioeconomic status on mortality rates in the UK. For Northern Ireland, there is limited formal modelling and analysis of mortality variation by geographical location using spatial models and limited analysis of variation of mortality according to socioeconomic risk factors.

The approach used in this paper can be broken down into three components. Firstly, as the mortality rates in each region will be heavily affected by the age distribution in the region the initial analysis will use a regression approach to age standardise the mortality rates. This will enable the remaining variability not due to age to be analysed. We secondly regress the socioeconomic data on these age standardised mortality rates carrying out a general to specific analysis to identify the significant deprivation measures. Finally, having done this we test the

[2]For a review of the various applications see Sherris and Tang (2010).

hypothesis that there may remain some unexplained variability or heterogeneity that cannot be explained by the socioeconomic variables that are used in this data set. We fit a non-spatial and spatial model to the residual data to test for this. We then repeat the spatial and non-spatial analysis in the case where we do not have socioeconomic data to assess the use of geographic structures in that case.

3. DATA

The analysis in this paper is based on mortality data from 2008 and deprivation data from 2008 provided by the Northern Ireland Statistical Research Agency NISRA and reported in the 2010 deprivation study (NIRSA, 2010). We use deprivation indices created by NISRA from raw government collected data. The data is divided geographically using Super Output Areas (SOA) of which there are 890 in Northern Ireland. SOA's are a set of geographies developed in the UK after the 2001 census[3]. For each super output area i the data used includes numbers of deaths, D_i, and mid year estimates of the exposure, E_i separated into male and female populations. We have also used a range of socioeconomic and demographic factors for each SOA including measures of employment levels, education, crime and violence, income, proximity to services, environment and health deprivation. No SOAs were omitted in the analysis as all had sufficiently large populations ranging from 997 to $3,667$[4]. The associated mortality rate for area i is labelled m_i and is calculated as

$$m_i = \frac{D_i}{E_i} \tag{1}$$

Death data and mid population estimates were available for each of the SOA's in 2008 for males and females individually from the NISRA website[5]. Non standardised mortality rates as calculated above were then standardised using a simple linear regression model with weighted age for each SOA as the covariate before analysing the data.

[3]In Northern Ireland SOA's typically contain 2000 lives but range from 1300 to 2800. They were designed to improve reporting of small area statistics and to ensure that each area was of similar size, unlike electoral wards which varied widely in size. The data is taken from the 2010 deprivation study. http://www.nisra.gov.uk/deprivation/nimdm_2010.htm

[4]A table of the SOA along with their names and SOA codes is available on request.

[5]see http://www.nisra.gov.uk/

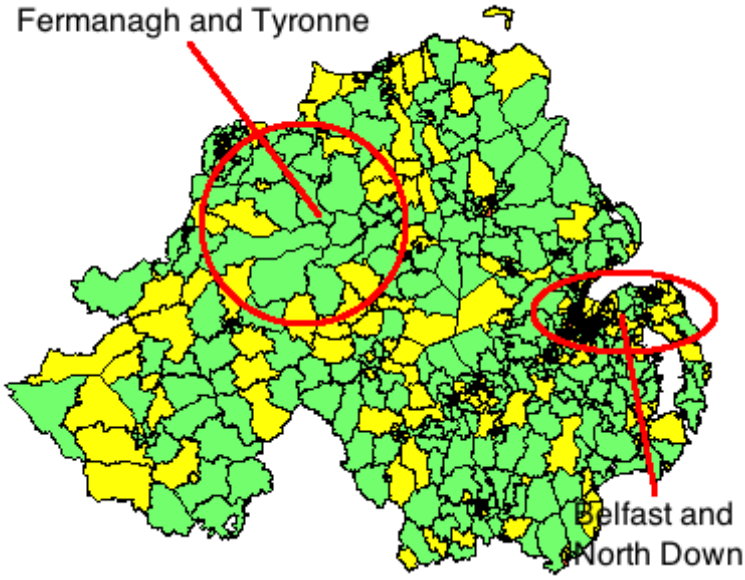

FIGURE 1. The normalised mortality rates for males in 2008 split by SOA.

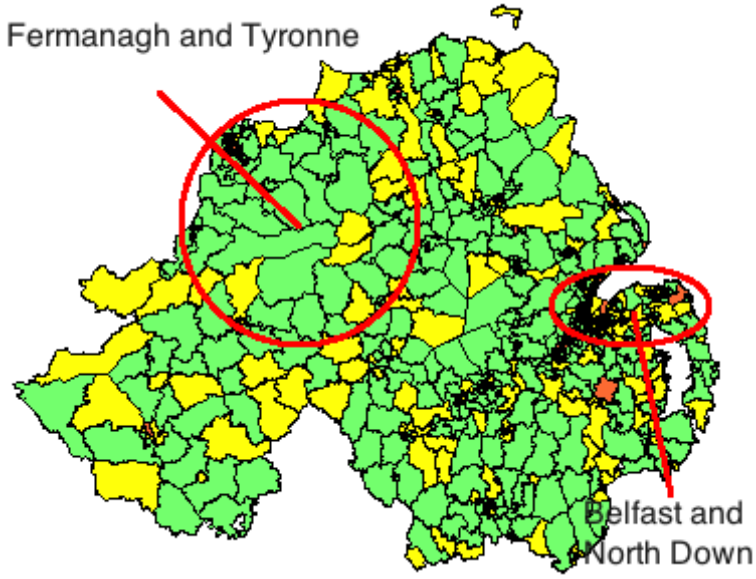

FIGURE 2. The normalised mortality rates for females in 2008 split by SOA.

The mortality rates used in the study are aggregated by age within each super output area. Given the refined nature of the super output areas across the Northern Ireland geography each region has a credible, but relatively small number of deaths upon which to base the mortality rate estimation. If we were to further subdivide the data (i.e. by separating out mortality by age) this would result in too few observations at many ages within each SOA. The resulting mortality data are a crude estimate for mortality rates and further research would consider how to rectify this.

Figs. 1 and 2 show the standardised mortality rates for males and females in 2008 split by SOA where low mortality is coloured yellow, high in green and higher in orange. Higher mortality rates occur in the more urban areas of Northern Ireland, Belfast, North Down, and Lisburn. Lower mortality rates occur in the more rural parts of Northern Ireland, for example Fermanagh and Tyrone. However, within urban areas the mortality can still be seen to experience a significant degree of variability not visible within the map of Northern Ireland. Similar maps of Northern Ireland for 1999 and 2003 show that the geographical variation in Northern Irish mortality rates has changed little over the period 1999-2008. Below in Table 1 some summary statistics for the age standardised mortality rates are set out. As can be seen the rates for females are slightly less than those for males on average but there is a wider spread of rates for females than for males.

TABLE 1. Age standardised mortality rates (expressed as a percentage)

	Males	Females
Minimum	0.35%	0.27%
Lower quartile	0.69%	0.65%
Mean	0.81%	0.80%
Upper quartile	0.96%	1.00%
Maximum	2.14%	2.84%

Deprivation data for each of the SOA's were selected to reflect the major factors expected to affect mortality. The deprivation measures were collated based on information available within the public sector. A detailed summary of the data upon which the deprivation factors were based can be found on the NIRSA website[6] and are also available by request.

In the Northern Ireland context population displacement has been significant leading to regions particularly dominated by catholics and regions particularly dominated by protestants. The mortality experience of these two types of population might be considered to be different owing to their history - the catholics predominantly coming from the Irish republic having suffered the famine etc., and the protestants coming predominantly from the Ulster Scots land owners etc. We test this in the paper where we have captured this Northern Ireland specific characteristic by defining a measure of "polarisation". We calculate this using the proportion of Catholics[7] in each region as the data item and converting this into a measure ranging from 0 to 1 reflecting the concentration of one particular background over another. The focus here

[6]http://www.nisra.gov.uk/deprivation/archive/Updateof2005Measures/NIMDM_2010_Indicator_Summary.pdf
[7]http://www.ninis.nisra.gov.uk/mapxtreme/viewdata/Census/CensusKS07b.xls

is on mortality rates in the two distinct types of area in Northern Ireland. "Polarised" areas are regions where there is a dominance of one particular background. The opposite to these sorts of areas are what are known as "mixed" areas where there is no particular concentration. If we denote α_i as the proportion of catholics in region i then we calculate the polarisation factor in an area as:

$$2\max(\alpha_i, 1-\alpha_i) - 1. \tag{2}$$

The aim of the factor is to capture areas of high polarisation. A similar factor may be calculated in other contexts or countries perhaps where there have been high levels of migration leading to pockets of populations with very different make up. Table 2 summarizes the NISRA covariates used for analysis.

TABLE 2. Deprivation covariates.

	Crime	Education	Employment	Environment	Healthcare	Income	Proximity	Political
Maximum	4.063	3.932	4.706	3.892	3.280	4.168	4.231	1.651
Upper quartile	0.439	0.435	0.464	0.514	0.658	0.502	0.496	0.960
Mean	-0.287	-0.310	-0.188	-0.285	-0.042	-0.206	-0.380	-0.192
Lower quartile	-0.762	-0.756	-0.678	-0.728	-0.668	-0.768	-0.709	-0.970
Minimum	-1.126	-1.185	-2.146	-1.417	-3.846	-1.621	-1.182	-1.235

Each deprivation measure is calculated as a combination of a range of government recorded statistics the details of which can be found in the deprivation study[8]. For example the income deprivation measure is compiled as a combination of records on housing benefit, income support, tax credits and state pension credits. Whereas the education deprivation measure combines information on key stages 2 and 3 teacher assessments, proportions of children in special needs education, proportion of children with 5 good GCSE's or above and the proportions of 18-21 year olds not attending further or higher education. The deprivation measures are relative to each other where a higher score on the deprivation measure means the super output area is more deprived on that measure. From the summary statistics we can see that the spread of healthcare deprivation in much wider (the interquartile range is 1.326) than for employment deprivation (the interquartile range is 1.142). We do not comment on the reasons for this in this paper but simply note that the variation in each of the deprivation measures is not similar across the super output areas.

[8]http://www.nisra.gov.uk/deprivation/archive/Updateof2005Measures/NIMDM_2010_Indicator_Summary.pdf

We plot the socioeconomic standardised factors from the 2010 deprivation study in Fig. 3 to show the distribution of socioeconomic characteristics in Northern Ireland. We also plot the same distributions but in a geographical fashion with thematic maps of the deprivation factors by SOA. These are available by request and also can be found on the NISRA website.

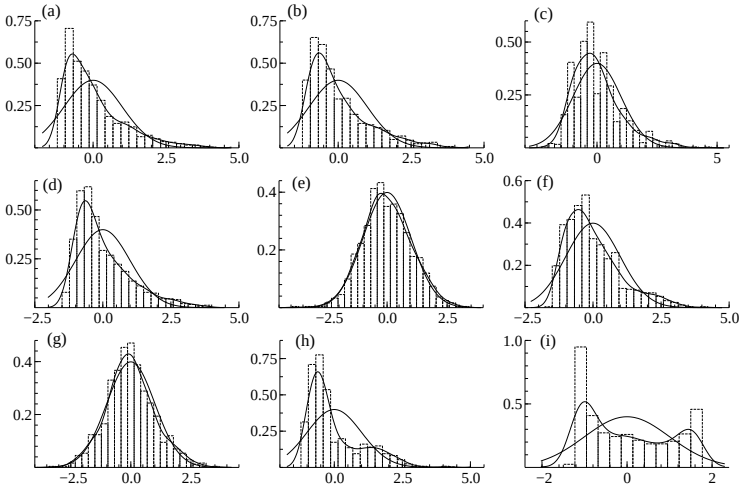

FIGURE 3. Distribution of deprivation measures and age where (a) Crime, (b) Education, (c) Employment, (d) Environment, (e) Healthcare, (f) Income, (g) Age, (h) Proximity to services, and (i) Polarisation.

Considering the distribution plots it can be seen that some deprivation measures are more unevenly distributed than others. For example Crime, Education, Income, Environment, and Proximity to services are very left skewed suggesting that there are many areas of high deprivation and only a few with low deprivation on these factors. However, healthcare and employment deprivation are more evenly distributed across the SOA. The geographical plots meanwhile show how all covariates (deprivation factors and age and polarisation) are spatially correlated, with similar measures of demographic and economic characteristics between nearby SOAs. High levels of Employment, Income, Education and Health deprivation exist in the North Western areas of Northern Ireland and in the South Down and Armagh areas of Northern Ireland. Deprivation of the living environment is high in the North Antrim and Fermanagh areas (these areas are significantly rural and have poor transportation links) and is also high in some inner city areas of Belfast and surrounding towns. Crime and Disorder is more significant in the cities but also shows high levels in some pockets of rural areas such as Tyrone, Armagh and Antrim. One significant factor to note for Northern Ireland is that nearly all areas (all areas except the major cities of Belfast, Lisburn, Newtownards, and Bangor) suffer from a lack of facilities.

We also analysed the covariance between deprivation measures. This is presented in Table 3 and shows that several of the deprivation measures are highly correlated, for example income and education which might be expected. Proximity is shown to be negatively correlated to almost all other deprivation measures suggesting that as individuals become wealthier or more educated as they move to more rural locations, i.e. locations with fewer facilities. Polarisation is not highly correlated with any of the other deprivation measures.

TABLE 3. Correlation between deprivation covariates.

	Employment	Income	Education	Environment	Healthcare	Proximity	Crime	Polarisation
Employment	1.000	0.946	0.809	0.533	0.915	-0.241	0.491	0.403
Income	0.946	1.000	0.866	0.590	0.919	-0.276	0.554	0.432
Education	0.809	0.866	1.000	0.647	0.802	-0.336	0.555	0.143
Environment	0.533	0.590	0.647	1.000	0.565	-0.386	0.688	0.055
Healthcare	0.915	0.919	0.802	0.565	1.000	-0.273	0.538	0.412
Proximity	-0.241	-0.276	-0.336	-0.386	-0.273	1.000	-0.531	0.107
Crime	0.491	0.554	0.555	0.688	0.538	-0.531	1.000	0.215
Polarisation	0.403	0.432	0.143	0.055	0.412	0.107	0.215	1.000

4. EMPIRICAL ANALYSIS

We first age standardise the mortality rates in each of the 890 super output areas. The results of this are shown in the Appendix A. Having done this we next show the results of fitting a generalised linear model to the age standardised mortality rates, still labelled m_i for simplicity, derived from the data in Northern Ireland. We then demonstrate how this model can be extended to allow for spatial frailties.

4.1. **Generalised linear model.** The modelling uses a hierarchical Bayes method. A prior distribution is assumed for each of the parameters which we then combine with the likelihood of the data given the parameters to give us the posterior distribution for the parameters given the data. Parameters are estimated using Markov Chain Monte Carlo methods.

The generalised linear model (GLM) of mortality rate m_i for region i has the log likelihood function:

$$L(m_i; \beta, \mathbf{x}_i) \propto \exp(\sum_{j=1}^{n} \beta_j x_{ij}), \tag{3}$$

where n is the number of covariates (socioeconomic or demographic factors), $\mathbf{x}_i$ is the vector of covariates respectively for each SOA i, $i = 1, \ldots, I$. The posterior distribution is:

$$p(\beta|\mathbf{x}_i, m_i) \propto \prod_{i=1}^{I} L(m_i; \beta, \mathbf{x}_i) p(\beta), \tag{4}$$

where the first term on the right represents the logistic likelihood, and the second is the prior distribution for the parameters. A vague uniform prior distribution is assumed with small mean and large variance because of a lack of prior knowledge about the parameters (Banerjee *et al.*, 2003). This prior for β is used in all the models.

The results of fitting the standardised mortality rates to a generalised linear regression model using all the standardised socioeconomic factors for Males 2008 and Females 2008 are shown in Table 4. We provide the parameter estimates and p-values of the covariates. In the case of males we find that environment, healthcare and polarisation in interaction with income deprivation and employment deprivation are significant at the 5% level. For females we find that healthcare and employment are significant at the 5% level with environment being significant at the 10% level. For females we find no significance for the polarisation variable either on its own or in interaction with income, education or employment.

To demonstrate the confidence we have for the parameter estimates for each covariate as Table 5 provides the 2.5%, 50% and 97.5% posterior percentiles for each of the predictors and interaction terms for the generalised linear model for both Males and Females. The quality of the fit using this model is measured using the DIC[9] measure which pits quality of fit against parsimony. Using this measure we have a fit given in Table 6. Given the correlations that exist between some of the covariates and the desire to simplify the model as best as we can, we next carry out a general to specific model selection procedure for both the Male and Female data to eliminate those variables that are not significant in explaining the variation we see in the age standardised mortality rates. We then extend the resulting models adding spatial and non-spatial frailties to explain any residuals.

4.2. **General to specific modelling of covariates.** In the general to specific methodology the specification of the general model from which reductions are made is crucial (Hendry, 2000)

[9]The DIC, an extension of the Akaike Information Criterion (AIC), is commonly used to compare the performance of different models (Spiegelhalter *et al.*, 2002). The DIC is based on the posterior distribution of the deviance statistic. It is defined analogously to the AIC as the expected deviance plus the effective number of parameters. See Appendix B for details.

TABLE 4. Results of generalised linear regression for Males and Females 2008.

MALES	Coefficient	Std.Error	t-value	p-value	Part R^2
Intercept	-4.9879	0.0286	-174.000	0.0000	0.9720
Crime	-0.0474	0.0377	-1.260	0.2090	0.0018
Education	-0.0937	0.0575	-1.630	0.1034	0.0030
Employment	-0.1008	0.0803	-1.250	0.2099	0.0018
Environment	0.1040	0.0358	2.910	0.0038	0.0095
Healthcare	0.3765	0.0649	5.800	0.0000	0.0369
Income	-0.0412	0.1045	-0.394	0.6937	0.0002
Proximity	0.0173	0.0288	0.601	0.5481	0.0004
Polarisation	-0.0050	0.0312	-0.162	0.8711	0.0000
Polarisation*Income	-0.2305	0.0830	-2.780	0.0056	0.0087
Polarisation*Education	0.0507	0.0508	0.998	0.3186	0.0011
Polarisation*Employment	0.1660	0.0671	2.480	0.0135	0.0069
FEMALES					
Intercept	-4.9284	0.0219	-225.000	0.0000	0.9829
Crime	-0.0008	0.0289	-0.027	0.9788	0.0000
Education	-0.0286	0.0440	-0.650	0.5157	0.0005
Employment	-0.1381	0.0615	-2.240	0.0251	0.0057
Environment	0.04781	0.0274	1.740	0.0814	0.0035
Healthcare	0.3369	0.0497	6.770	0.0000	0.0497
Income	0.0466	0.0801	0.582	0.5608	0.0004
Proximity	0.0287	0.0221	1.300	0.1944	0.0019
Polarisation	-0.0249	0.0239	-1.040	0.2975	0.0012
Polarisation*Income	-0.0878	0.0636	-1.380	0.1674	0.0022
Polarisation*Education	0.0018	0.0389	0.045	0.9641	0.0000
Polarisation*Employment	0.0790	0.0514	1.540	0.1243	0.0027

TABLE 5. Posterior percentiles for covariates and interaction terms for Male and Female data using the GLM model.

	Males			Females		
Coefficient	2.50%	median	97.50%	2.50%	median	97.50%
Intercept	-4.8680	-4.8440	-4.8190	-4.860	-4.834	-4.807
Employment	-0.2475	-0.1823	-0.0913	-0.146	-0.065	0.015
Income	-0.0569	0.0244	0.1107	-0.130	-0.033	0.067
Healthcare	0.2488	0.3190	0.3736	0.193	0.275	0.336
Education	-0.0966	-0.0428	0.0109	-0.042	0.030	0.079
Environment	-0.0138	0.0231	0.0595	-0.005	0.032	0.065
Crime	-0.0611	-0.0161	0.0221	-0.055	-0.010	0.027
Proximity	-0.0391	-0.0111	0.0182	-0.022	0.008	0.036
Polarisation	-0.0351	-0.0092	0.0183	-0.050	-0.023	0.005
Polarisation*Income	-0.1868	-0.1090	-0.0186	0.002	0.086	0.182
Polarisation*Education	-0.0265	0.0306	0.0836	-0.088	-0.038	0.025
Polarisation*Employment	0.0495	0.1266	0.1904	-0.089	-0.012	0.059

because a poorly specified general model stands little chance of leading to a good final specific model. To identify the appropriate simplified model we apply several mis-specification tests at

TABLE 6. Goodness of fit for Males and Females 2008 GLM.

	D	p_D	DIC
Males GLM	5266.00	12.43	5278.43
Females GLM	4493.71	12.83	4506.54

every stage of the reduction process. We follow the approach of Hendry and Krolzig (2001) and carry out the following tests on the general model: (1) Two F-tests for parameter constancy for breakpoints at the sample mid-point and 90th percentile; and (2) Doornik and Hansen (1994) χ^2 test for normality of the error terms.[10] The resulting reduced models following the general to specific analysis are shown in Tables 7 and 8.

TABLE 7. Results of linear regression for Males and Females 2008.

MALES	Coefficient	Std.Error	t-value	t-prob	Part R^2
Intercept	-4.9918	0.0243	-205.000	0.0000	0.9795
Health	0.2672	0.0395	6.770	0.0000	0.0493
Environment	0.0783	0.0300	2.610	0.0093	0.0076
Polarisation*Employment	0.1332	0.0639	2.080	0.0375	0.0049
Polarisation*Income	-0.1739	0.0637	-2.730	0.0065	0.0084
FEMALES					
Intercept	-4.9342	0.0173	-285.000	0.0000	0.9892
Employment	-0.1197	0.0430	-2.780	0.0055	0.0087
Health	0.3208	0.0441	7.270	0.0000	0.0563
Environment	0.0435	0.0210	2.070	0.0388	0.0048

TABLE 8. Posterior percentiles for covariates and interaction terms for Males and Females data using the GLM model.

Males Coefficient	2.50%	median	97.50%
Intercept	-4.852	-4.830	-4.807
Health	0.114	0.144	0.175
Environment	-0.018	0.009	0.038
Polarisation*Employment	0.024	0.098	0.163
Polarisation*Income	-0.148	-0.080	-0.006
Females Coefficient	2.50%	median	97.50%
Intercept	-4.847	-4.822	-4.797
Employment	-0.112	-0.053	0.003
Health	0.196	0.255	0.319
Environment	0.000	0.030	0.056

[10]The mis-specification test results are not presented here but are available on request.

We note from this analysis that polarisation does explain some of the variation we see in male data but not in female data. Potential reasons for this could relate to the specific problems that the Northern Ireland region has had over the past number of years and the greater impact that this would have had on the male population. We conclude from the general to specific analysis that the covariates to use in our simplified model for the standardised males mortality rates are: Healthcare, Environment and the interaction terms Polarisation by Income and Polarisation by employment. For the females we will use Healthcare, Employment and Environment.

4.3. **Introducing frailties.** Having identified the main socioeconomic factors that explain the variability of mortality rates across Northern Ireland, in this next section we extend this standard regression model to include a frailty aspect picking up the geographical location of each region.[11] We then look at the fitting quality again using the DIC measure. When we add the frailties the likelihood function becomes:

$$L(m_i;\beta,\mathbf{x}_i) \propto \prod_{i=1}^{I} \exp(\sum_{j=1}^{n} \beta_j x_{ij} + W_i), \tag{5}$$

where β, i, m_i and $\mathbf{x}_i$ are as previously defined and W_i is the frailty for SOA i, which captures any remaining effects not explained by the covariates. Under this non-spatial frailties setting, the frailties are assumed to be identical and independently distributed with the following distribution:

$$W_i \sim N(0,\sigma^2). \tag{6}$$

Eq. (6) assumes no spatial dependence since frailties in one SOA are independent of frailties in another. The hierarchical Bayes model is:

$$p(\beta, W_i, \sigma^2|\mathbf{x}_i) \propto L(m_i;\beta,\mathbf{x}_i)p(W_i|\sigma^2)p(\beta)p(\sigma^2), \tag{7}$$

where the likelihood is given by Eq. (5). As in Banerjee *et al.* (2003), a Gamma (0.001, 0.001) prior distribution is used for $\tau = 1/\sigma^2$ with a mean of 1 and a variance of 1000. A flat uniform prior was adopted for β.

[11] Alternative approaches to introduce frailties do exist; see for example Jonker *et al.* (2013) who use an alternative specification in the development of their healthy adjusted life expectancies model. We thank an anonymous referee for bringing this to our attention.

To consider the case of spatial clustering (i.e., adjacent SOA showing similar mortality characteristics), we allow for spatial correlations between nearby SOAs through a Conditional Autoregression specification $\mathbf{W}|\lambda \sim CAR(\lambda)$ as introduced in Besag *et al.* (1991). In this specification an adjacency matrix is defined to capture the geographical variations in mortality. The adjacency matrix is defined by assigning the ij^{th} entry a value of 1 if the SOA i is adjacent to SOA j and 0 otherwise. The hierarchical Bayes model becomes:

$$p(\beta, W_i, \lambda|\mathbf{x}_i) \propto L(m_i; \beta, \mathbf{x}_i)p(W_i|\lambda)p(\beta)p(\lambda), \tag{8}$$

where the prior $W_i|\lambda$ is given by:

$$\lambda_k^{1/2} \exp\Big[-\frac{\lambda_k}{2}\sum_{i \text{ adj } j}(W_{ik}-W_{jk})^2\Big] \propto \lambda_k^{1/2}\exp\Big[-\frac{\lambda_k}{2}\sum_{i=1}^{I} n_i W_{ik}(W_{ik}-\hat{W}_{ik})^2\Big], \tag{9}$$

where i adj j denotes that SOA i and SOA j are adjacent to each other, $\hat{W}_i$ is the average of the frailties W_i , adjacent to SOA i, and n_i represents the number of these adjacent regions (Bernardinelli and Montomoli, 1992). A consequence of the above prior is that:

$$W_i|W_{j\neq i} \propto N\big(\hat{W}_i, \frac{1}{\lambda n_i}\big). \tag{10}$$

Repeating the fitting analysis now using the independent frailties model and the spatial clustering model the parameter estimates along with confidence intervals are shown in Tables 9 and 10. Looking at the deviance measure now for the independent frailties and the spatial clustering models alongside the standard generalised model we have the results shown in Table 11. From this we can see that there is an improvement by allowing for non-spatial frailties. This can be observed in the lower DIC measure in the case of non-spatial frailties. The model fitting power is improved in the case of males and females with the allowance for non-spatial frailties. Under the spatial frailty model, the DIC measure is not an improvement over the non frailty model, but it apparently changes the statistical significance of some of the parameter estimates. This suggests that with sufficient social and economic data for each region we can adequately explain

any geographical structures visible in mortality rates. We also test the residuals[12] for normality for each of the three models above and note that the results are in line with other studies which have carried out similar tests on a range of stochastic mortality models. See for example Dowd *et al.* (2010).

TABLE 9. Posterior percentiles for covariates for Males and Females 2008 data with the independent frailties model.

Males			
Coefficient	2.50%	median	97.50%
Intercept	-4.936	-4.897	-4.821
Health	0.103	0.143	0.188
Environment	-0.019	0.019	0.060
Polarisation*Employment	0.001	0.104	0.193
Polarisation*Income	-0.186	-0.095	0.009
Females			
Coefficient	2.50%	median	97.50%
Intercept	-4.864	-4.835	-4.803
Employment	-0.128	-0.059	0.000
Health	0.195	0.259	0.334
Environment	0.000	0.031	0.062

TABLE 10. Posterior percentiles for covariates for Males and Females 2008 data with the spatial frailties model.

Males			
Coefficient	2.50%	median	97.50%
Intercept	-4.937	-4.896	-4.842
Health	0.132	0.196	0.265
Environment	-0.090	-0.032	0.025
Polarisation*Employment	-0.062	0.038	0.147
Polarisation*Income	-0.153	-0.043	0.055
Females			
Coefficient	2.50%	median	97.50%
Intercept	-4.859	-4.830	-4.800
Employment	-0.116	-0.045	0.025
Health	0.190	0.253	0.331
Environment	-0.026	0.013	0.047

[12]We follow the approach of Dowd *et al.* (2010) and test the age standardised mortality residuals for males and females using the three models proposed. The tests used aim to identify whether the mortality residuals are consistent with i.i.d. N(0,1) as assumed under the null hypothesis. The tests include: t-test of mean prediction, Variance ratio (VR) test (see Cochrane, 1988; and Lo and MacKinley, 1988 and 1989), and Jarque-Bera test. A statistically significant result for any of these tests - which we take to be any test which produces a p-value of less than 1% - indicates inconsistency with i.i.d. N(0,1). The results are available on request.

TABLE 11. Goodness of fit for Males and Females 2008 GLM with spatial and non-spatial frailties.

	D	p_D	DIC
Males			
GLM	5303.21	5.056	5308.27
Non-Spatial Frailties Model	4378.52	460.203	4838.72
Spatial Frailties Model	5590.56	1641.24	7231.8
Females			
GLM	4501.95	4.248	4506.2
Non-Spatial Frailties Model	4326.85	150.187	4477.04
Spatial Frailties Model	5492.28	1185.74	6678.02

4.4. **Can geographical information replace socioeconomic data?** We have demonstrated that in the case where we have significant amounts of deprivation data non-spatial frailties models can add some improvement in the fitting power of the models but that spatial frailty models do not. Next we further test this to see if there is a greater contribution from spatial and non-spatial frailty modelling when we have little other data upon which to fit our mortality rates. To test this we consider the situation where we do not have significant amounts of socioeconomic data available to us. We assume that the only covariate we have is an aggregate age parameter for each SOA and we fit a simple regression model and some extensions of it to the data for males and females in 2008 to the non-age standardised mortality rates. We chose this covariate to assess the impact of adding spatial structure as it is the most readily available piece of information we might have and the one which is likely to have the most significance when trying to explain raw mortality rates. Three models were fitted to the data, firstly a simple generalised linear model with no allowance for spatial dependence (the no frailties model), secondly a linear model allowing for independent, identically distributed frailties (non-spatial frailties model) and finally a linear spatial frailties model which allows for spatial dependence which employs a CAR specification to capture the spatial dependence between adjacent SOAs.

The models are again fitted using a Markov Chain Monte Carlo algorithm under a Bayesian hierarchical framework. A flat Uniform (-10,000,10,000) prior was adopted for the parameter and a Gamma (0.001,0.001) was chosen for the frailties W_i in the non-spatial frailties model. For the spatial frailties model the smoothness parameter λ of the CAR specification was given a Gamma (0.001,0.001) prior distribution. Table 12 compares the three models using the DIC, which is the sum of the expected deviance and the number of effective parameters p_D. Comparisons of DIC values show that the model with non-spatial frailties shows an improvement over

the no-frailties model, despite the increase in the number of effective parameters. However, due to the additional number of parameters the inclusion of spatial frailties does not improve the fit further. It seems that good quality social and economic data are able to describe a large portion of mortality variation by region. It is also noted that even with many socio-economic variables, there is still an unidentified structure to the variability of mortality rates across the region and it is this that is being picked up by the non-spatial frailty.

TABLE 12. Goodness of fit for Males and Females 2008 spatial and non-spatial modelling.

	Males		Females	
Model	p_D	DIC	p_D	DIC
Non-Spatial No frailties Model	1.94	5340.09	2.09	4772.49
Non-Spatial Frailties Model	452.19	4989.52	138.04	4743.69
Spatial Frailties Model	1698.49	7358.10	1259.65	6675.01

5. CONCLUSIONS

This paper has considered geographical variation of mortality and the effect of socioeconomic explanatory factors using Northern Ireland data. In particular we test the assertion that geographical location (or postcode) may provide a good proxy for the variation seen in mortality rates.

Regression models were fitted to age-standardised mortality rates using covariates extracted from the Northern Ireland Statistical Research Agency (NISRA). A general to specific analysis was carried out to simplify the models. In the case of males we identified that mortality rates were best explained by the variation in deprivation measures; healthcare, environment and polarisation (through income and education). This would suggest that the combined effect of living in a polarised area with poor education and income levels has an impact on mortality rates for males. For females we see that employment, healthcare and environment are important. In both cases of males and females the quality of the immediate environment (heating, double glazing etc.) came through as a significant factor in explaining mortality rate variation and would be a potential consideration for decision makers when looking to improve health outcomes.

Models including non-spatial and spatial frailties were also fitted to the data. Introducing frailties improved the fitting power measured using a DIC criteria despite the additional parameterisations. However the spatial frailty model did not improve the fitting power further due to additional number of parameters. We conclude that good quality social and economic data should be sufficient to describe mortality rates. Thus the best model was the non-spatial frailty followed by the GLM model and finally the spatial frailty model. Repeating the analysis with limited covariate information did not change the order of these results. The results of this paper show that there is a place in mortality modelling for allowing for underlying socioeconomic characteristics in the modelling process. In particular the result highlights the correlations between mortality and living environment. The inability of frailties to improve the fitting quality further could be due to some correlation between the socioeconomic variables included in this study and the frailties themselves. It could also highlight the fact that with good quality socioeconomic data we are able to explain the vast majority of the variation by region. The success of the frailty model suggests that even with many socio-economic variables, there is still some unidentified structure to the variability of mortality rates across the region and it is this that is being picked up by the non-spatial frailty. We acknowledge that there is more work to be done in this area. Having identified significant deprivation measures in the case of Northern Ireland, at least for the 2008 data, further research will investigate the issue of identifying time series for these deprivation measures and associating this with time varying mortality rates; in this way potentially leading to a model that might be used to better forecast mortality rates.

Appendix A: Age standardisation of mortality rates

The raw mortality rates we have are not standardised for age in each of the super output areas which will be the major explanatory variable for the variation in the true rates in each region. Since we are investigating the impact of socio-economic measures on the variation in mortality rates having already controlled for age we must first standardise the data for age. We do this by running a simple regression model on the natural logarithm of the mortality rate with only constant and age as the covariates. We are restricted somewhat in the age standardisation that we can carry out since we have only aggregated mortality data at the average age for each super output area. With better quality data we would be able to carry out full age standardisation along the lines of Breslow and Day (1987).

$$\ln\left(\frac{D_i}{E_i}\right) = \alpha + \beta \text{Age}_i + \epsilon_i \tag{11}$$

The results of this model for males and females are given in Table 13. From the results we can see that the constant explains a significant amount of the log mortality rates but this is to be expected since, whilst there is variation in the rates they all vary around a non-zero average value. Having identified the constant and age parameters we recalculate log transformed age-standardised mortality rates using the residuals specific to each area. The variation left in the resulting rates which is no longer due to age differentials can now be analysed in relation to the socio-economic variables we have in our data set. Figs. 4 and 5 show the age standardised mortality rates for males and females split by SOA.

TABLE 13. Results of Males and Females 2008 to control for age.

	Coefficient	Std.Error	t-value	t-prob	Part R^2
Males					
Intercept	-4.9342	0.0191	-258	0	0.9868
Age	0.3387	0.0191	17.7	0	0.2609
Females					
Intercept	-5.0132	0.0237	-211	0	0.9805
Age	0.4463	0.0237	18.8	0	0.285

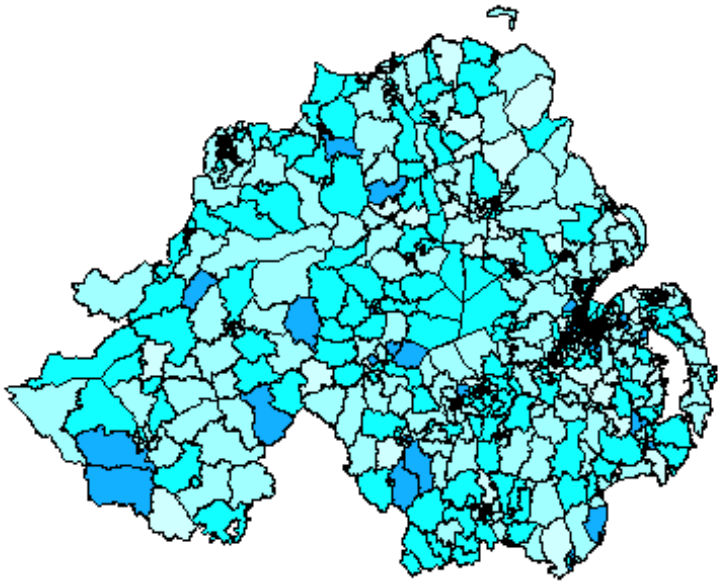

FIGURE 4. Male age standardised mortality rates; low rates are light blue, higher rates are darker blue.

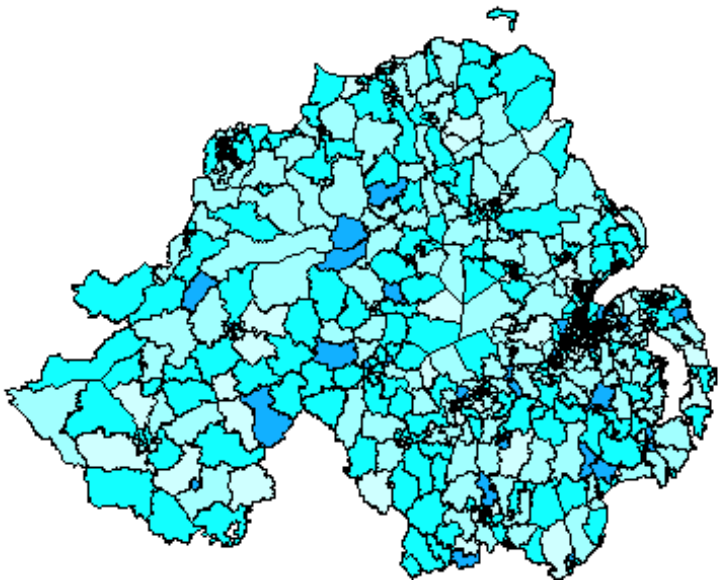

FIGURE 5. Female age standardised mortality rates; low rates are light blue, higher rates are darker blue.

Appendix B: Assessing Model Choice

When we fit various models we have to determine which model provides the best fit. We use an information measure, the Deviance Information Criterion (DIC), as we wish to ensure that when we add explanatory variables, spatial or socioeconomic, we are improving the model fit significantly. The DIC, an extension of the Akaike Information Criterion (AIC), is commonly used to compare the performance of different models (Spiegelhalter *et al.*, 2002). It is readily calculated using Markov Chain Monte Carlo (MCMC) methods (Banerjee and Carlin, 2003). The DIC is defined as:

$$\text{DIC} = \bar{D} + p_D, \tag{12}$$

with closeness of fit to the data measured by $\bar{D} = E_{\theta|y}[D]$ and the effective number of parameters measured by p_D, where y is the data vector and θ is the parameter vector. p_D is defined as,

$$p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \bar{D} - D(\bar{\theta}), \tag{13}$$

which is the deviance of the posterior mean subtracted from the posterior mean at the deviance. The deviance statistic is:

$$D(\theta) = -2\ln f(y|\theta) + 2\ln h(y), \tag{14}$$

where $f(y|\theta)$ is the likelihood, and $h(y)$ a standardising function of the data alone which thus has no impact on model selection. Small values of $\bar{D}$ represent a good fit and small values of p_D indicate a more parsimonious model. Smaller values of DICs are preferred. DICs are only used to compare models.

REFERENCES

[1] Acemoglu, D., and Johnson, S. (2007) Disease and Development: The Effect of Life Expectancy on Economic Growth. *Journal of Political Economy*, 115(6), s925-985.

[2] Banerjee, S., and Carlin, B. P. (2003) Semiparametric Spatio-Temporal Frailty Modeling. *Environmetrics*, 14, 523-535.

[3] Banerjee, S., Wall, M. M., and Carlin, B. P. (2003) Frailty Modeling for Spatially Correlated Survival Data, with Application to Infant Mortality in Minnesota. *Biostatistics*, 4, 123-142.

[4] Bernardinelli, L., and Montomoli, C. (1992) Empirical Bayes versus Fully Bayesian Analysis of Geographical Variation in Disease Risk. *Statistics in Medicine*, 11, 983-1007.

[5] Besag, J., York, J. C., and Mollie, A. (1991) Bayesian Image Restoration, with two Applications in Spatial Statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1-20.

[6] Bhargava, A., Jamison, D. T., Lau, L. J. and Murray, C. J. L. (2001) Modeling the effects of health on economic growth. *Journal of Health Economics*, 20, 42340.

[7] Blake, D., Cairns, A. and Dowd, K. (2008) The birth of the life market, *Asia-Pacific Journal of Risk and Insurance*, 3 (1) 6-36.

[8] Bloom, D.E., Canning, D., and Sevilla, J. (2004) The Effect of Health on Economic Growth: A Production Function Approach. *World Development*, 32(1), 1-13.

[9] Booth, H., Maindonald, J., Smith, L., (2002). Applying Lee-Carter under conditions of variable mortality decline.*Population Studies* 56, 325-336.

[10] Booth, H., and Tickle, L. (2008) Mortality Modelling and Forecasting: a Review of Methods , *Annals of Actuarial Science*, 3, 3-43.

[11] Breslow N, Day N. (1987). Statistical methods in cancer research, volume II. The design and analysis of cohort studies. Lyon: International agency for research on cancer. *WHO*

[12] Brouhns, N., Denuit, M., and Vermunt, J.K., (2002). A Poisson log-bilinear approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* 31(3), 373-393.

[13] Cairns, A.J.G., Blake, D., and Dowd, K., (2006), A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance* 73, 687-718.

[14] Cochrane, J. H. (1988) How big is the random walk in GNP? *Journal of Political Economy*, 96: 893-920.

[15] Currie, I.D., Durban, M., and Eilers, P.H.C., (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* 4, 279-298.

[16] Currie, I.D., (2006) Smoothing and forecasting mortality rates with P-splines. *Presentation to the Institute of Actuaries. Available at: http://www.ma.hw.ac.uk/ iain/research.talks.html.*

[17] de la Croix, D., and Licandro, O. (1999) Life expectancy and endogenous growth, *Economics Letters*, 65(2), 255-263.

[18] Doornik, J. A. and Hansen, H. (1994) An omnibus test for univariate and multivariate normality. Working Paper, Nuffield College, Oxford.

[19] Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M., (2010). Evaluating the Goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics*, 47, 255-265.

[20] Girosi, F., and King, G., (2005). *A reassessment of the Lee-Carter mortality forecasting method*, Working Paper, Harvard University.

[21] Hári, N., Waegenaere, A., Melenberg, B. and Nijman, T., (2008). Estimating the term structure of mortality. *Insurance: Mathematics and Economics* 42, 492-504.

[22] Hendry, D. F. (2000) Epilogue: the success of general-to-specific model selection, in D.F. Hendry, Econometrics: Alchemy or Science? (New Edition). Oxford: Oxford University Press.

[23] Hendry, D. F. and H.-M. Krolzig (2001). Automatic Econometric Model Selection using PcGets. London: Timberlake Consultants Press.

[24] Jonker, M.F., Congdon, P.D., Van Lenthe, F.J., Donkers, B., Burdorf, A., Mackenbach, J.P., (2013). Small-area health comparisons using health-adjusted life expectancies: A Bayesian random-effects approach, *Health and Place*, 23, 70-78,

[25] Kazembe, L. N. (2007) Spatial Modelling and Risk Factors of Malaria Incidence in Northern Malawi. *Acta Tropica*, 102, 126-137.

[26] Lee, R.D., and Carter, L.R. (1992), Modeling and Forecasting U. S. Mortality, *Journal of the American Statistical Association*, 87(419), 659-671.

[27] Lindahl, M. (2005) Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income, *Journal of Human Resources*, 40(1), 144-168.

[28] Lo, A. W., and A. C. MacKinley (1988) Stock prices do not follow random walks: Evidence based on a simple specification test. *Review of Financial Studies*, 1: 41-66.

[29] Lo, A. W., and A. C. MacKinley (1989) The size and power of the variance ratio test in finite samples: A monte carlo investigation. *Journal of Econometrics*, 40: 203-38.

[30] NIRSA (2010) Northern Ireland Multiple Deprivation Study. http://www.nisra.gov.uk/deprivation/archive/Updateof2005Measures/NIMDM_2010_Report.pdf

[31] O'Hare, C., and French, D. (2013) Forecasting death rates using exogenous determinants, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2259969

[32] O'Hare, C., and Li, Y. (2012) Explaining young mortality, *Insurance: Mathematics and Economics*, 50(1), 12-25.

[33] Plat, R., (2009), On stochastic mortality Modeling. *Insurance: Mathematics and Economics*, 45(3), 393-404.

[34] Preston, S. H. (1975) The Changing Relation Between Mortality and Level of Economic Development, *Population Studies*, 29(2), 231-248.

[35] Renshaw, A.E., Haberman, S., (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38, 556-70.

[36] Richards, S., and Jones, G. (2004) Financial Aspects of Longevity Risk. Staple Inn Actuarial Society, London.

[37] Richards, S. J. (2008) Applying Survival Models to Pensioner Mortality Data. *Institute of Actuaries Sessional Meeting Paper*.

[38] Rosen, S. (1974) Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *The Journal of Political Economy*, 82, 34-55.

[39] Sherris, M., and Tang, A. (2010) Spatial Variability in Mortality and Socioeconomic Factors for Australian Mortality, UNSW Australian School of Business Research Paper.

[40] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002) Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society*, 64, 583-639.

[41] Tuljapurkar, S., (2008), Mortality declines, Longevity risk and Aging. *Asia-Pacific Journal of Risk and Insurance*, 3(1), pp. 37-51.

[42] Tuljapurkar, S., and Boe, C. (1998) Mortality Change and Forecasting: How Much and How Little Do We Know? *North American Actuarial Journal*, 2, 13-47.

[43] Turrell, G., Kavanagh, A., and Subramanian, S. V. (2006) Area Variation in Mortality in Tasmania (Australia): the Contributions of Socioeconomic Disadvantage, Social Capital and Geographic Remoteness. *Health and Place*, 12, 291-305.

[44] Waller, L. A., Zhu, L., Gotway, C. A., Gorman, D. M., and Gruenewald, P. J. (2007) Quantifying Geographic Variations in Associations between Alcohol Distribution and Violence: A Comparison of Geographically Weighted Regression and Spatially Varying Coefficient Models. *Stochastic Environmental Research and Risk Assessment*, 21, 573-588.