AUSTRALIAN
SCHOOL OF BUSINESS™
THE UNIVERSITY OF NEW SOUTH WALES

www.business.unsw.edu.au

The University of New South Wales
Australian School of Business

Australian School of Business Research Paper No. 2010ACTL05

Spatial Variability in Mortality and Socioeconomic Factors for Australian Mortality

Michael Sherris & Andy Tang

23/04/2010
CRICOS Provider: 00098G

Spatial Variability in Mortality and Socioeconomic Factors for Australian Mortality

Andy Tang
AXA Asia Pacific
Hong Kong
asytang99@hotmail.com

Michael Sherris
Australian School of Business
University of New South Wales
m.sherris@unsw.edu.au

April 23, 2010

Abstract

Mortality rates are known to vary by geographical location and to depend on socio-economic factors. Demographic, ethnic and socio-economic mortality factors vary by geographical location. Regions that are in closer proximity are expected to have similar mortality because of similar socio-economic factors and demographic characteristics. In this paper the spatial variability of Australian mortality is assessed using a spatial model along with explanatory risk factors including age, income, labour force participation and unemployment rate. Geographical variation is based on statistical subdivisions, areas of similar social and economic backgrounds. Logistic regressions are estimated using an hierarchical Bayes model with Markov Chain Monte Carlo methods for mortality rates in 208 statistical subdivisions in Australia for census years 1996, 2001 and 2006. Spatial models explain mortality variation by geographical location better than non-spatial models when limited data is available for socio-economic factors. Explanatory factors, which also vary spatially, reduce the need for spatial models for mortality. The modeling has implications for pricing and risk management in life insurance companies. Geographical variation in risks can be quantified using spatial models especially if there is limited data for risk factors that generate mortality heterogeneity. Employment and workforce participation, ethnic background as well as income are found to be significant in explaining mortality variation by geographical location in Australia. Geographical location has been used recently in the UK based on postcode in pricing and risk management of mortality and longevity risk products. As demonstrated in this paper, spatial geodemographic models should be of significant interest to insurers in assessing mortality risk.

Keywords: Mortality, Logistic regression, Geodemographic, Spatial, Hierarchical Bayes

JEL Classifications: G22, C50

1 Introduction

Spatial variability in mortality rates as well as the effect of socio-economic factors has attracted increased attention recently. Life insurers issuing life and annuity products allow for significant risk factors known to affect mortality rates including age, gender, and smoking status. In life insurance adjustments are made to mortality rates used for premiums based on health status. With the increase in sales of annuity products insurers are offering impaired lives annuity rates. Postcode underwriting reflecting geographical varication in mortality is increasingly used by insurers in the UK. Mortality risk has traditionally been quantified using a life table of age-based survivor probabilities. These life tables differ by risk factors including gender and, for life insurance purposes, smoking status. There are other risk factors related to socio-economic status that also impact mortality.

Spatial models have been developed and applied to modeling house prices, crime levels and diseases amongst many others. Rosen (1974) [21] models house prices using spatial covariates including environmental attributes and geographical characteristics. Waller *et al.* (2007) [28] models geographic variation in alcohol distribution and violent crime in Houston. Kandala and Madise (2004) [18] use spatial analysis to study geographical variations in the impact of diarrhea and fever among infants in Africa. Kazembe (2007) [16] examined spatial clustering of malaria risks in northern Malawi.

Geodemographic modeling, the spatial modeling of demographic data, is used in a range of applications. Commercial applications include customer profiling for product marketing and development. Grubesic, 2004 [14] applies geodemographic models to assess broadband access. Richards (2008) [20] uses geodemographic profiles based on postcodes to analyse life insurance and pension scheme mortality. Tuljapurkar and Boe (1998) [24] outline mortality differentials by sex, education and socio-economic variables. Richards and Jones (2004) [19] discuss the impact of socio-economic status on mortality rates in the UK. For Australia, there is limited formal modeling and analysis of mortality variation by geographical location using spatial models and limited analysis of variation of mortality according to socio-economic risk factors. Jain, 1994 [2], Wilkinson *et al.*, 2000 [30] and Turrell *et al.*, 2004 [25] provide only a descriptive analysis of Australian mortality from a spatial perspective.

Mortality is known to vary spatially and geographic regions in close proximity have similar mortality because of similar socio-economic and demographic characteristics. Significant geographical variation in mortality occurs in many countries including Australia. In the United Kingdom, socio-economic factors were implicitly allowed in insurance based on policy or annuity amounts. Individuals with larger annuity amounts have lower mortality rates. Richards (2008) [20] shows that a mortality model using geodemographic classifications better fits United Kingdom annuitant mortality than a model using pension amount.

This paper assesses the geographical variation in mortality rates in Australia. Socio-economic and demographic factors including unemployment rate, age, indigenous proportion, income, occupation, birthplace and labour force participation rate. Spatial models and covariate models are used to explain the geographical variation. The analysis covers the period 1993-2007 using Statistical Subdivisions. Statistical Subdivisions (SSDs) are defined by the Australian Bureau of Statistics (ABS, 2008 [4]) to be "socially and economically homogeneous regions characterized by identifiable links between the inhabitants" covering the whole of Australia. Logistic regression along with spatial frailties and covariates are used. The methodology can be readily applied to studying spatial variations of mortality in other countries and for more refined geographical subdivisions.

Spatial models are found to explain mortality variation by geographical location more effectively than non-spatial models when limited data is available. Spatial models capture geographical variation in mortality. This is especially the case if there is limited data for socio-economic factors that generate mortality heterogeneity. Explanatory factors, which also vary spatially, reduce the need for spatial models and provide information about causes of geographical variation in mortality.

Life insurance companies need to assess mortality heterogeneity for pricing and risk management of insurance policies. Modeling geographical variation in mortality has the potential to reduce the need to take socio-economic factors into account in both pricing and risk management. At the same time identifying and quantifying the significant socio-economic and demographic risk factors that impact mortality provides valuable information that spatial modeling does not.

The next sections outline models used for covariates (risk factors) and spatial modeling for mortality rates. The data used in the models is then summarized. Following that the results are presented and discussed. Finally conclusions are summarized.

2 Modeling Geographical Variation in Mortality

Mortality can be modeled using survival (time to death) data or aggregate death rate data. If data on individual characteristics including death dates are available then hazard rates can be estimated using proportional hazards models to quantify the effect of covariates. For aggregate data on deaths and exposures, the effect of covariates on death rates can be estimated using logistic regressions. Both approaches to modeling can be modified to include spatial variation. Frailty models are used for heterogeneity in mortality rates to account for unobserved covariates (Vaupel *et al.* (1979) [26]). Frailty models can also be used to capture spatial variation and unobserved heterogeneity. Banerjee and Carlin (2003) [6] provide details of spatial frailty models in the hierarchical Bayes model.

2.1 Proportional Hazards Models

A proportional hazards model including covariates for different geographic regions can be written:

$$h(t_{ij};\mathbf{x_{ij}}) = h_0(t_{ij})\, exp(\beta^T \mathbf{x_{ij}}), \tag{1}$$

where h_0 is the *baseline hazard*, t_{ij} is the time to death or censoring and $\mathbf{x_{ij}}$ is the individual-specific covariates for the jth individual in region i respectively, where $i = 1, ..., I$ and $j = 1, ..., n_i$. The model assumes that the shape of the hazard functions are similar for different regions. Region-specific frailties are incorporated by introducing $W_i = log\, w_i$, the frailty for region i, to obtain:

$$
\begin{aligned}
h(t_{ij};\mathbf{x_{ij}}) &= h_0(t_{ij})\, w_i\, exp(\beta^T\mathbf{x_{ij}}) \\
&= h_0(t_{ij})\, exp(\beta^T\mathbf{x_{ij}} + W_i),
\end{aligned} \tag{2}
$$

Simple *i.i.d.* specifications including the gamma distribution, the log-normal distribution and the normal distribution, have been assumed for W_i (McGilchrist and Aisbett, 1991 [17] and Wienke, 2003 [29]).

Using the indicator variable $\gamma_{ij} = \begin{cases} 1 & \text{if dead} \\ 0 & \text{if alive} \end{cases}$, the likelihood of the model is:

$$
L(\beta, \mathbf{W}, \mathbf{t}, \mathbf{x}, \gamma) \propto \prod_{i=1}^{I}\prod_{j=1}^{n_i} \{h_0(t_{ij};\mathbf{x_{ij}})\}^{\gamma_{ij}}\, exp\{-H_0(t_{ij})\, exp(\beta^T\mathbf{x_{ij}} + W_i)\}, \tag{3}
$$

where

$$
H_0(t) = \int_0^t h_0(u)du,
$$

is the integrated baseline hazard function (Banerjee and Carlin (2003) [6]).

2.2 Spatial Logistic Modelling

Proportional hazards models require data at the individual level including the time until death. If data is only available at the aggregate level for death rates then the logistic regression model can be used. The event time data t_{ij} is replaced with an indicator $Y_{ij} = \begin{cases} 1 & \text{survived} \\ 0 & \text{otherwise.} \end{cases}$ and $\rho_{ij} = Pr(Y_{ij} = 1)$, has a logistic frailties model:

$$
logit(\rho_{ij}) = \beta^{\mathbf{T}}\mathbf{x_{ij}} + W_i, \tag{4}
$$

where W_i is the frailty term for region i, β are the parameters and $\mathbf{x_{ij}}$ are the individual-specific covariates for the jth subject in region i.

Ingram and Kleinman (1989) [15] and Doksum and Gasko (1990) [11] show that the results for the β parameters can be quite similar in the two different models when the probability of death is small or where there is no censoring. However, since the proportional hazards model is based on more information than the logistic regression model, Banerjee *et al.* (2003) [7] note that the proportional hazards model should be more powerful in detecting significant covariate effects.

2.3 Geostatistical Modelling

Because of the assumption of independent frailties, the models do not account for spatial correlations. Spatial effects can be modeled using continuous *geostatistical* models or discrete *lattice* models.

The geostatistical approach uses the exact geographical location of a region (Cressie (1993) [10]). Frailties $\mathbf{W}$ are indexed continuously throughout a geographical region D.

A prior distribution given observations W_i for known locations i, $i = 1, ..., I$, is used for the unobserved frailty values at other target locations $t \in D$ by assuming:

$$\mathbf{W} \mid \theta \sim N_I(0, H(\theta)), \tag{5}$$

where $H(\theta)_{ij}$ denotes the covariance structure between region i and region j. This is usually assumed to be *isotropic* where the spatial correlation between regions depend only on the Euclidean distance d_{ij} between region i and region j. Cressie (1993) [10] and Stein (1999) [23], amongst others, discuss other isotropic specifications for H, including the powered exponential given by:

$$H(\theta)_{ij} = \sigma^2 exp(-\phi\, d_{ij}^{\kappa}), \quad \sigma^2 > 0,\ \phi > 0, \kappa \in (0, 2]. \tag{6}$$

Zimmerman (1993) [31] discusses how the strength of spatial dependence may also depend on the direction, referred to as *anisotropy.* Ecker and Gelfand (1999) [13], use a combination of Bayesian and semivariogram estimation techniques for geostatistical anistropic models.

2.4 Lattice Modelling

When $\mathbf{W}$ is defined only for discrete regions such that the regions form a partition of the geographical study space D, then this is the lattice model. Banerjee and Carlin (2002) [5] use the CAR *conditionally autoregressive* model for the prior distribution:

$$\mathbf{W} \mid \lambda \sim CAR(\lambda), \tag{7}$$

introduced in Besag *et al.* (1991) [9]. Bernardinelli and Montomoli (1992) [8] refer to the most common form of this prior having the following joint distribution:

$$\mathbf{W} \,|\, \lambda \,\propto\, \lambda^{I/2}\, exp \left[\frac{\lambda}{2} \sum_{i=1}^{I} \sum_{j=1}^{n_i} \alpha_{ij}\, W_i \,(W_i - \frac{1}{\sum_{j=1}^{n_i} \alpha_{ij}} \sum_{j=1}^{I} \alpha_{ij}\, W_j) \right], \tag{8}$$

where α_{ij} represents the weights between region i and region j. With spatial correlation, higher weights should be assigned to regions in closer proximity to each other. For example $\alpha_{ij} = \begin{cases} 1 & i \text{ adj } j, \\ 0 & \text{otherwise.} \end{cases}$, in which case Equation (8) simplifies to:

$$\mathbf{W} \,|\, \lambda \,\propto\, \lambda^{I/2} exp \left[-\frac{\lambda}{2} \sum_{i=1}^{I} m_i W_i (W_i - \bar{W}_j) \right], \tag{9}$$

where $\bar{W}_i$ is the average of the frailties $W_{j \neq i}$ that are adjacent to region i, and m_i is the number of these adjacencies. This then gives

$$W_i \,|\, W_{j \neq i} \;\sim\; N(\bar{W}_i, \frac{1}{\lambda m_i}). \tag{10}$$

The CAR model displaces each individual region-effect estimates towards the local 'mean effect' $\bar{W}_i$. Banerjee *et al.* (2003) [7] note that the lattice model is computationally simpler compared to the geostatistical approach.

2.5 Spatio-Temporal Modelling

Mortality rates evolve through time (Jain, 1994 [2]; Wilkinson *et al.*, 2000 [30] and ABS, 2008 [1]) so that a model with both dependence in space and through time would be ideal if sufficient data were available. Such a model is referred to as a spatio-temporal model. Banerjee and Carlin (2003) [6] discuss models that incorporate temporal dependence in an analysis of the survival of women diagnosed with breast cancer in Iowa. If t_{ijk} denotes the time to death for the jth subject in region i in the year k with $i = 1, 2, ..., I$, $k = 1, 2, ..., K$ and $j = 1, 2, ..., n_{ik}$, $\mathbf{x_{ijk}}$ denotes the vector of covariates, W_{ik} the spatio-temporal frailties corresponding to the ith region in the kth year, then, Equation (2) becomes:

$$h(t_{ijk}; \mathbf{x_{ijk}}) = h_0(t_{ijk})\, exp(\beta^T \mathbf{x_{ijk}} + W_{ik}). \tag{11}$$

Assuming a lattice structure and a $CAR(\lambda_k)$ model (see Besag *et al.*, 1991 [9]) the prior distribution is:

$$\lambda_k^{I/2}\, exp\left[-\frac{\lambda_k}{2}\sum_{i\, adj\, j}(W_{ik} - W_{jk})^2\right] \propto \lambda_k^{I/2}\, exp\left[-\frac{\lambda_k}{2}\sum_{i=1}^{I} m_i\, W_{ik}\,(W_{ik} - \bar{W}_{jk})\right] \tag{12}$$

where i adj j denotes that region i and region j are adjacent to each other, and $\bar{W}_{ik}$ is the average of the frailties W_{jk} adjacent to region i for the kth year and m_i represents the number of these adjacencies. The conditional distribution of W_{ik} becomes:

$$W_{ik}\,|\,W_{(j\neq i)k} \sim N(\bar{W}_{ik}, \frac{1}{\lambda_k\, m_i}), \tag{13}$$

2.6 Non-Spatial Frailties

Waller *et al.* (1997) [27] include non-spatial frailties V_k to capture unexplained residuals. Eberly and Carlin (2000) [12] and Banerjee and Carlin (2003) [6] proposed that $V_{ik} \sim N(0, \frac{1}{\tau_k})$ and noted that the choice of priors for τ_k must be carefully defined. The joint likelihood function becomes:

$$L(\beta, \mathbf{W}, \mathbf{V}; \mathbf{t}, \mathbf{x}, \gamma) \propto \prod_{k=1}^{K}\prod_{i=1}^{I}\prod_{j=1}^{n_{ik}}\{h_0(t_{ijk}; \mathbf{x_{ijk}})\}^{\gamma_{ijk}} \cdot$$
$$exp\{-H_0(t_{ijk})exp(\beta^T \mathbf{x_{ijk}} + W_{ik} + V_{ik})\}, \tag{14}$$

where γ_{ijk} is an indicator (0 for survive, 1 for dead) for the jth subject in region i in the kth year.

3 Logistic Regression Models for Australian Geographical Regions

The ideal data for analysis of mortality risk is socio-economic, demographic and deaths data at the individual level through time. This would allow the estimation of survival models with spatial data and proportional hazards models for risk factors. This data is usually only available from longitudinal studies or possibly from census and deaths data at the individual level. Only aggregate data was available for this study. As a result a logistic regression (GLM) model is used along with allowance for spatial frailties. Ingram and Kleinman (1989) [15], Doksum and Gasko (1990) [11] and Banerjee *et al.* (2003) [7] compare proportional hazards and logistic models and find only marginal differences in the efficiency of the covariate parameters of the two different models.

Models used are hierarchical Bayes. A prior distribution is assumed for the parameters, which along with the likelihood of the data given the parameters, is the posterior distribution for the parameters given the data. Parameters were estimated using Markov Chain Monte Carlo. The Conditionally Autoregressive (CAR) model for spatial variation is used.

3.0.1 Logistic Regression

The generalized linear model with a logit link function has the following likelihood:

$$L(\beta;\mathbf{x}) \propto \prod_{i=1}^{I}\{\frac{e^{\sum_{j=1}^{m}\beta_j x_{ij}}}{1+e^{\sum_{j=1}^{m}\beta_j x_{ij}}}\}, \tag{15}$$

where m is the number of covariates, $\mathbf{x_i}$ the vector of covariates respectively for subdivision i, $i = 1, 2, ..., I$. The posterior distribution is:

$$p(\beta|\mathbf{x}) \propto L(\beta;\mathbf{x})\,p(\beta), \tag{16}$$

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 adopted with small mean and large variance because of a lack of knowledge about the parameters (Banerjee *et al.* (2003) [7]. This prior for β is used in all the models.

3.0.2 Non-Spatial Frailties Model

The model is extended to include frailties with likelihood:

$$L(\beta,\mathbf{W};\mathbf{x}) \propto \prod_{i=1}^{I}\{\frac{e^{\sum_{j=1}^{m}\beta_j x_{ij}+W_i}}{1+e^{\sum_{j=1}^{m}\beta_j x_{ij}+W_i}}\}, \tag{17}$$

where W_i is the frailty for subdivision 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:

$$W_i \sim N(0,\sigma^2). \tag{18}$$

Equation (18) assumes no spatial dependence since frailties in one SSD are independent of frailties in another. The hierarchical Bayes model is:

$$p(\beta,\mathbf{W},\sigma^{\mathbf{2}}|\mathbf{x}) \propto L(\beta,\mathbf{W};\mathbf{x})p(\mathbf{W}|\sigma^{\mathbf{2}})p(\beta)p(\sigma^{\mathbf{2}}), \tag{19}$$

where the likelihood is given by Equation (17). As in Banerjee *et al.* (2003) [7], a $Gamma(0.001, 0.001)$ prior distribution is used for $\tau = \frac{1}{\sigma^2}$ with mean 1 and variance 1000. A flat Uniform$(-10000, 10000)$ prior was adopted for β.

3.0.3 Spatial Frailties Model

Adjacent subdivisions are expected to share similar characteristics. To allow for spatial correlations between nearby SSDs, a CAR specification is assumed with an adjacency matrix to capture the geographical variations. In this adjacency matrix, the ijth entry is assigned a value of 1 if subdivision i is adjacent to subdivision j and 0 otherwise. The hierarchical Bayes model is:

$$p(\beta, \mathbf{W}, \lambda|\mathbf{x}) \propto L(\beta, \mathbf{W}; \mathbf{x})\, p(\mathbf{W}|\lambda)\, p(\beta)\, p(\lambda), \tag{20}$$

where the prior $\mathbf{W} \,|\, \lambda$ is given by:

$$\lambda^{I/2}\, exp\left[\frac{\lambda}{2}\sum_{i\, adj\, j}(W_i - W_j)^2\right] \propto \lambda^{I/2}\, exp\left[\frac{\lambda}{2}\sum_{i=1}^{I} m_i\, W_i\,(W_i - \bar{W}_j)\right], \tag{21}$$

where i adj j denotes that subdivision i and subdivision j are adjacent to each other, $\bar{W}_i$ is the average of the frailties W_j, adjacent to subdivision i and m_i represents the number of these adjacencies (Bernardinelli and Montomoli (1992) [8]). A consequence of the above prior is that:

$$W_i \,|\, W_{j\neq i} \sim N(\bar{W}_i, \frac{1}{\lambda\, m_i}). \tag{22}$$

A higher value of λ indicates lower deviance from the mean of the frailties in adjacent SSDs. The higher the value of λ, the higher the clustering between adjacent subdivisions. The previous priors are used along with a prior distribution for the smoothness parameter λ in the CAR specifications for the frailties W_i. To reflect a lack of prior knowledge of a suitable value for λ, a vague $Gamma(0.001, 0.001)$ specification is chosen with mean 1 and variance 1000.

3.0.4 Assessing Model Choice

The Deviance Information Criterion (DIC), an extension of the Akaike Information Criterion (AIC), is commonly used to compare the performance of different models (Spiegelhalter *et al.* (2002) [22]). It is readily calculated using MCMC methods (Banerjee and Carlin (2003) [6]). The DIC is defined as:

$$DIC = \bar{D} + p_D, \tag{23}$$

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. p_D is defined as

$$p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \bar{D} - D(\bar{\theta}), \tag{24}$$

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

$$D(\theta) = -2\, log\, f(\mathbf{y}|\theta) + 2\, log\, h(\mathbf{y}), \tag{25}$$

where $f(\mathbf{y}|\theta)$ is the likelihood, $\mathbf{y}$ the data vector, θ the parameter vector, and $h(\mathbf{y})$ a standardising function of the data alone. It does not have any 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.

4 Data

4.1 Australian Standard Geographical Classification

Data used for the analysis consists of deaths data, population data and covariate data from the Australian Bureau of Statistics (ABS) by geographical area using the Australian Standard Geographical Classifications (ASGC) 2006 [4]. Under this classification structure, Australia is split up into different hierarchical regions. The following hierarchical spatial units are used: 9 States and Territories (S/T), 69 Statistical Divisions (SD), 217 Statistical Subdivisions (SSD), 1,426 Statistical Local Areas (SLA) and 38,704 Census Collection Districts (CD), where the last spatial unit is used only during Population Census years.

Figure 1: Statistical Subdivisions Classification by ABS, 2006

The data used was for Statistical Subdivisions (SSDs). Individuals within each SSD exhibit similar characteristics and therefore represent homogeneous risk profiles. Figure 1 shows the classification of SSDs for Australia. There are 8 SSDs within the Australian Capital Territory, whereas only a slightly larger number of 12 SSDs exist in the vast landmass of the Northern Territory. Over 200,000 males live in the SSD of Newcastle

(SSD Code: 11005), whereas the SSD of Australian Capital Territory - Bal (SSD Code: 81005) has a total population of only 160 males in 2006.

Statistical Subdivisions (SSDs) are used because the number of 69 SDs is not refined enough for insurance applications. Using the 1,426 Statistical Local Areas (SLAs) or 38,704 Census Collection Districts (CDs) involved confidentiality issues and provides less data for each spatial area. Of the 217 SSDs in Australia, only 208 SSDs were included because of data availability and confidentiality. Census data and deaths data was collected for the 208 SSDs. The SSDs that were omitted in the analysis represent regions which are off-shore with population sizes well under 1,000. The 9 SSDs that were omitted were not critical to the analysis and were: Off-Shore Areas & Migratory (NSW) - SSD Code: 18501; Off-Shore Areas & Migratory (VIC) - SSD Code: 28501; Off-Shore Areas & Migratory (QLD) - SSD Code: 38501; Off-Shore Areas & Migratory (SA) - SSD Code: 48501; Off-Shore Areas & Migratory (WA) - SSD Code: 58501; Off-Shore Areas & Migratory (TAS) - SSD Code: 68501; Off-Shore Areas & Migratory (NT) - SSD Code: 78501; Other Territories (OT) - SSD Code: 95001; and Off-Shore Areas & Migratory (OT) - SSD Code: 98501.

4.2 Deaths Data

The number of deaths for five year periods for both males and females in each SSD was obtained from the ABS. The 5-year periods used were: 1993-1997, 1998-2002 and 2003-2007 to match the classifications of ASGC (2006) [4] and the data for census years 1996, 2001 and 2006. Mortality rates were determined by dividing the total number of deaths in the five year periods by five times the corresponding census population for each SSD. They represent a 5 year level of mortality and the study considers a 15 year period in total. The years 1996, 2001 and 2006 refer to mortality rates for the periods 1993-1997, 1998-2002 and 2003-2007 respectively.

Figure 2 shows the male mortality rates for 2006. Higher mortality rates occur in the south eastern region of Australia and lower mortality rates occur in Western Australia, South Australia and the capital cities. There are lower mortality rates in capital cities. Figure 3 maps the female mortality rates for 2006. A similar geographical pattern is evident in the female mortality rates. Lighter shades in the female map reflect the lower overall magnitude of female mortality rates compared to male mortality rates in Australia. Similar maps for 1996 and 2001 show that the geographical variation in Australian mortality rates changed little over the period 1993-2006.

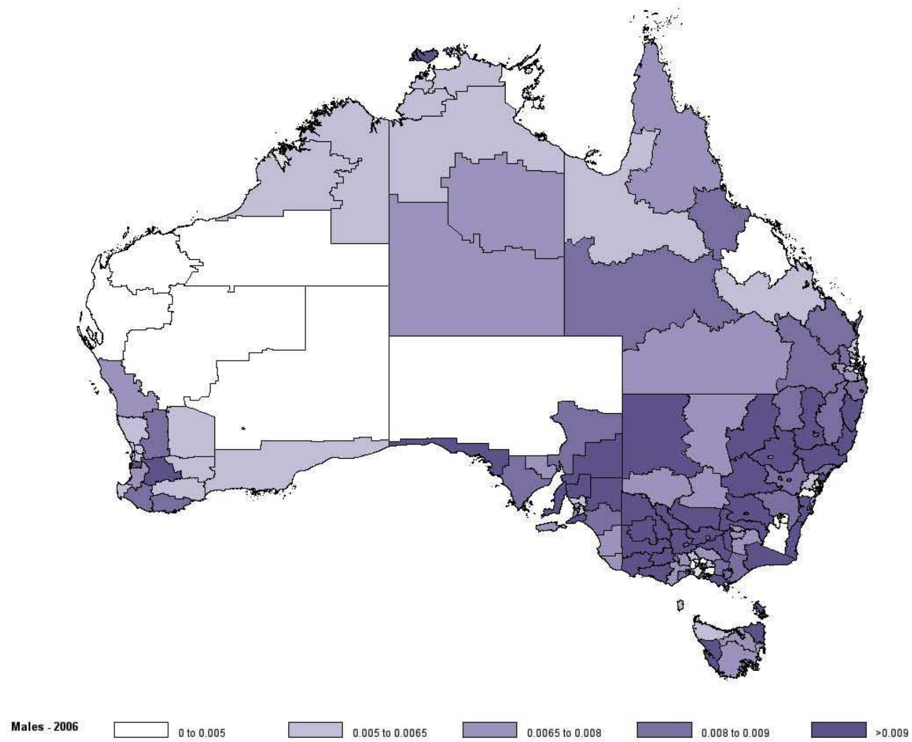

Figure 2: Males 2006 Mortality Rates

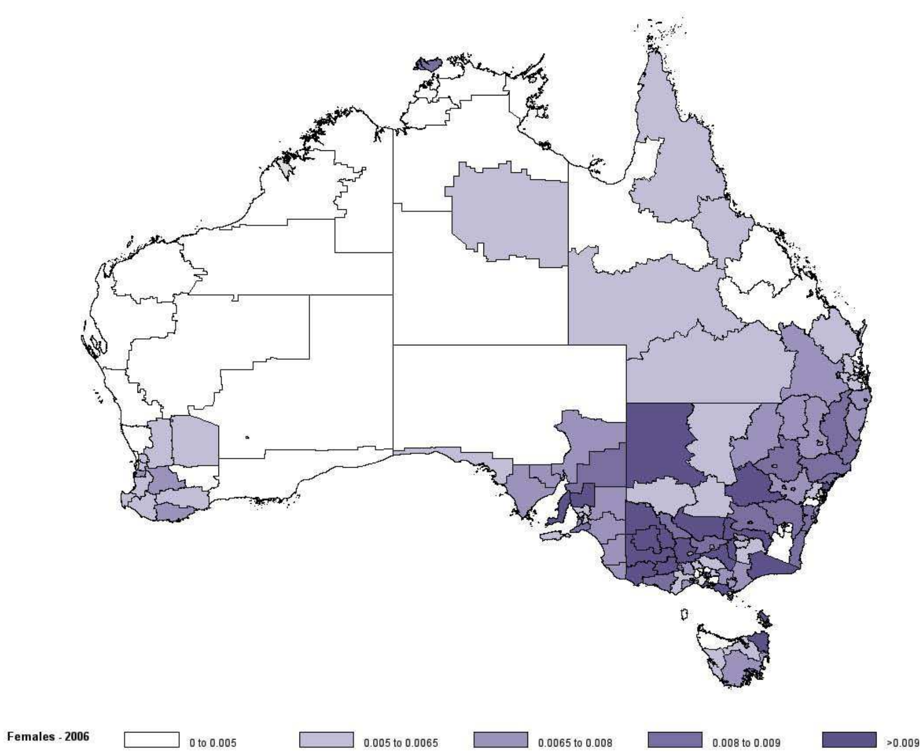

Figure 3: Females 2006 Mortality Rates

4.3 Census Data

Census data for covariates reflecting socio-economic factors were obtained from the ABS for males and females for 1996, 2001 and 2006. Census data dates back earlier, but were not available for the current ASGC (2006) [4] classifications. The data series obtained were selected to reflect the major factors expected to affect mortality. They were:

- Unemployment rate
- Age structure of the population
- Median individual income levels
- Place of birth
- Occupation levels (according to the Australian and New Zealand Standard Classification of Occupations, 2006 [3])
- Number of people with indigenous origins
- Labour force participation rate
- Education levels

Some of the data series had a significant amount of data which was "inadequately described" or "not stated" in the census. Only variables with less than 10% of inadequate data were selected for use to ensure a high level of accuracy. Unfortunately education levels had approximately 25% of the data as inadequate and was excluded. There are other factors included that are strongly correlated with education such as median income and occupation.

Table 1 summarizes the notation and definitions of the covariates used for analysis. Each covariate is standardized before being used in the models to provide clearer interpretation of parameters in terms of variation of the factors.

Table 1: *Notation and Definition of Covariates*

Characteristics	**Covariates**
Unemployment	Unemployment Rate of SSD in %
Age	Proportion of Population 65 or Above in %
Income	Median Individual Income in AUD
Overseas	Proportion of Population Born Overseas in %
Occupation	Proportion of Low-Skilled Workers in % (Machinery Operators, Drivers and Labourers) [3]
Indigenous	Proportion of Population with Indigenous Origins in %
Labour	Labour Force Participation Rate of SSD in %

In Appendix A Figures 23 - 36 plots of the standardized covariates by gender and census years are provided. Only the Males 2006 and Females 2006 standardised covariate plots are shown for space reasons. In the plots, red (lighter) indicates positive standardised data values and blue (darker) indicates negative standardised data values. These plots show how all covariates are spatially correlated, with similar measures of demographic and economic characteristics between nearby SSDs. There are similar geographical variations between standardised covariates for males and females. Only relatively small changes occur in the pattern of standardised data for the covariates from 1996 to 2006.

High Unemployment, high Overseas and low Occupation (unskilled workers proportion) variables concentrate in the urban and capital cities. Rural regions are often associated with low Unemployment, low Overseas and high Occupation variables. A lower than average Labour (workforce participation rate) variable is observed in the coastal areas. Higher Income variables occur in capital cities around Australia. There is also higher than national average Income variables in rural and less populated regions in Western Australia and South Australia. Higher proportions of population with indigenous origins occur in the Northern Territory. Victoria shows the lowest level of indigenous proportion in Australia.

5 Results and Discussion

A generalised linear logistic regression model with the selected standardised covariates was fitted to the mortality rates for Males 2006 and Females 2006. Table 2 provides the parameter estimates and p-values of the covariates. All covariates are significant at the 5% level. Higher levels of Unemployment, Age, Indigenous and Labour variables have higher mortality rates and higher levels of Income, Overseas and Occupation variables have lower mortality rates.

Table 2: *Results of Simple GLM for Males 2006 and Females 2006*

Covariates	**Males 2006**		**Females 2006**	
	β	**p-value**	β	**p-value**
Intercept	-4.941072	<2e-16	-5.079094	<2e-16
Unemployment	0.033225	<2e-16	0.021947	5.62e-09
Age	0.203857	<2e-16	0.331761	<2e-16
Income	-0.097017	<2e-16	-0.041879	<2e-16
Overseas	-0.044966	<2e-16	-0.011762	1.51e-06
Occupation	-0.063694	<2e-16	-0.023508	1.48e-09
Indigenous	0.034164	3.45e-08	0.031776	3.72e-05
Labour	0.017509	0.00138	0.091902	<2e-16

Table 3 shows the fitted model including interaction terms. The effect of income is seen to depend on Overseas, Occupation and Labour variables and the Overseas effect depends on the Occupation and Labour variables.

Table 3: *GLM with Interaction Terms for Males 2006 and Females 2006*

Covariates	**Males 2006**		**Females 2006**	
	β	**p-value**	β	**p-value**
Intercept	-4.923961	<2e-16	-5.067186	<2e-16
Unemployment	0.057167	<2e-16	0.019644	4.65e-06
Age	0.260605	<2e-16	0.309064	<2e-16
Income	-0.122080	<2e-16	-0.127572	<2e-16
Overseas	-0.030401	<2e-16	-0.027687	<2e-16
Occupation	-0.052651	<2e-16	-0.059814	<2e-16
Indigenous	0.085992	<2e-16	0.039646	1.73e-07
Labour	0.093241	<2e-16	0.096293	<2e-16
Income*Overseas	0.041217	<2e-16	0.045536	5.59e-15
Income*Occupation	-0.020306	1.08e-13	-0.070073	<2e-16
Income*Labour	-0.060051	<2e-16	-0.091559	<2e-16
Overseas*Occupation	0.016799	2.44e-09	0.014250	2.26e-08
Overseas*Labour	-0.036501	<2e-16	-0.036109	<2e-16

These models do not include spatial frailties.

5.1 Model with Covariate Age and Spatial Frailties

In order to compare the use of spatial models with additional covariates, only the Age covariate was used in the model for Australian mortality rates for Males 2006 and Females 2006. The non-spatial frailties model and the spatial frailties model were then fitted and compared. While the former assumes that frailties in each SSD are independent and identically distributed, the latter employs a CAR specification to capture the spatial dependence between adjacent SSDs.

Table 4 shows the Deviance Information Criterion (DIC) for the three models.

Table 4: *Goodness of Fit for Models with Covariate Age Only*

Model	$\mathbf{p_D}$	**DIC**
Non-Spatial No-Frailties Model	1.998	8221.78
Non-Spatial Frailties Model	197.53	2224.03
Spatial Frailties Model	191.816	2222.94

Comparisons of DIC values show that the models with geo-spatial frailties are a significant improvement over the no-frailties model, despite the increase in the number of effective parameters. Although the spatial frailties model and the non-spatial frailties model have similar DICs, the spatial frailties model has a smaller p_D value demonstrating the importance of including spatial dependence.

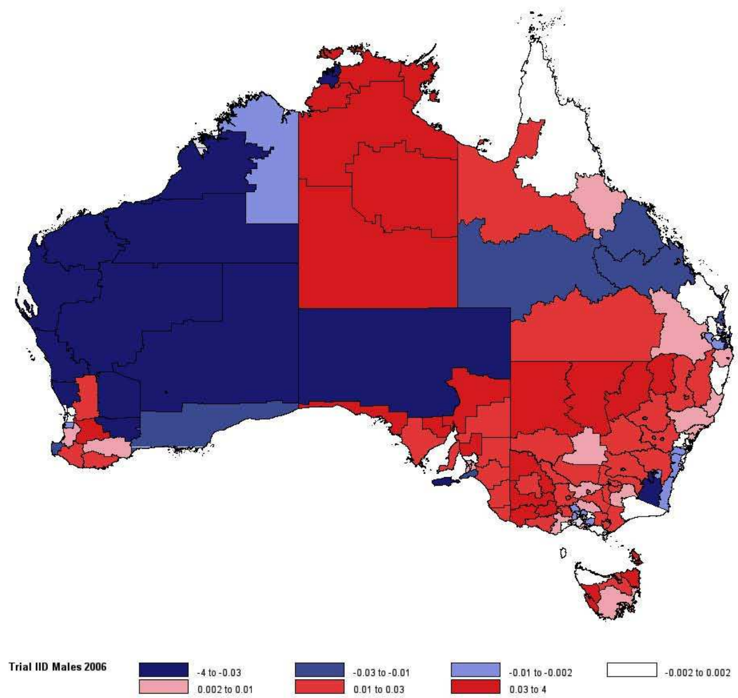

Figure 4: Standardized Error Map for Non-Spatial Model with Covariate Age (Males 2006)

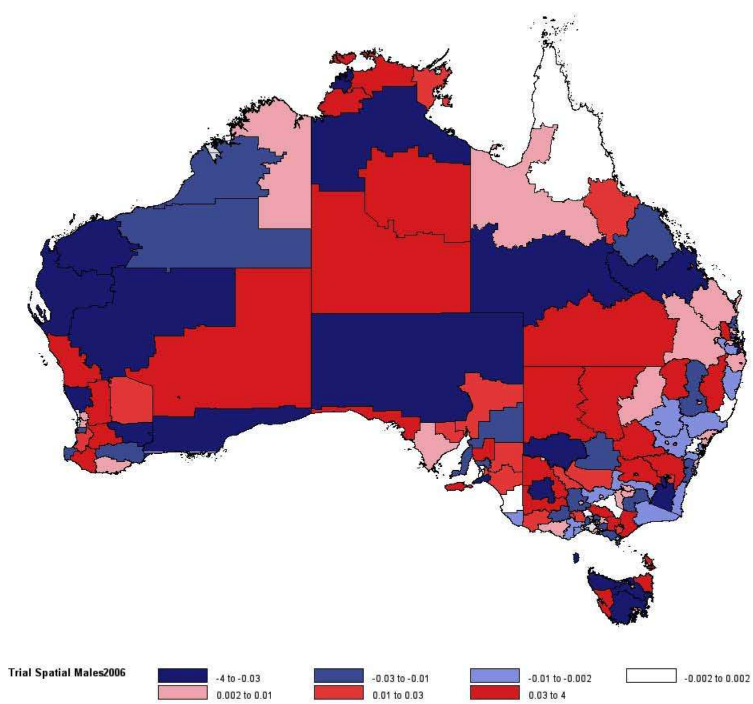

Figure 5: Standardized Error Map for Spatial Model with Covariate Age (Males 2006)

Figure 4 and Figure 5 map the standardized error plots for Males for 2006 under

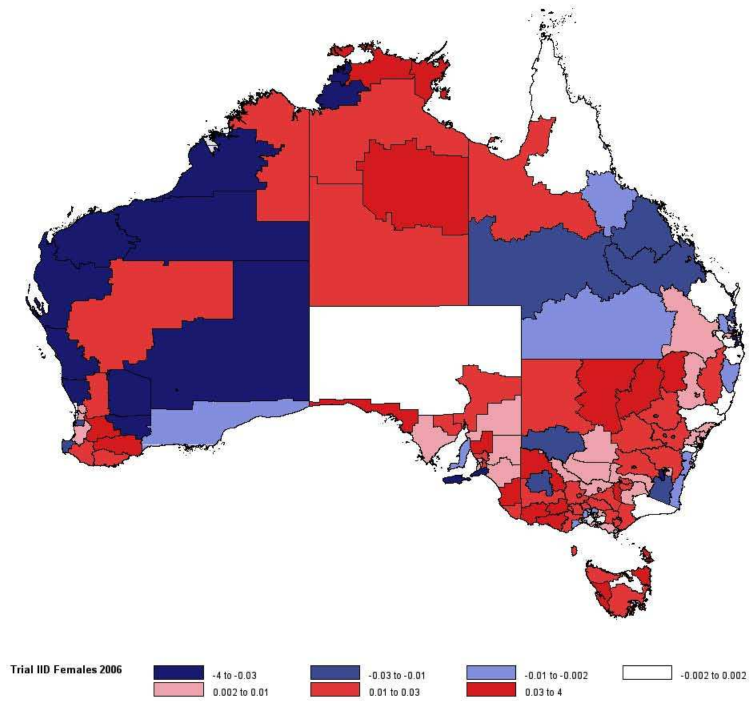

Figure 6: Standardized Error Map for Non-Spatial Model with Covariate Age (Females 2006)

the non-spatial model and the spatial model respectively. Figure 6 and Figure 7 are the standardized error plots for Females for 2006. While the errors of the spatial-frailties model appear random, a degree of spatial clustering is evident for the errors of the non-spatial frailties model. This is seen for the Males 2006 error map where overestimation occurs in Western Australia and South Australia, while death rates are underestimated in the Northern Territory and the south eastern parts of Australia. Although less apparent, similar observations can be made for the error plots for Females 2006.

Using only the Age covariate, the models show that spatial dependence is beneficial in explaining geographical variation in Australian mortality rates. This spatial variation has been captured in the spatial frailty terms. The non-spatial frailties model, although accounting for the heterogeneity from the missing risk factors, has spatial dependence in the model errors.

5.2 Models including all Covariates

All the covariates Unemployment, Age, Income, Overseas, Occupation, Indigenous, Labour and five interaction terms including Income*Overseas, Income*Occupation, Income*Labour, Overseas*Occupation and Overseas*Labour were fitted for both males and females for the census years 1996, 2001 and 2006. Tables 5 - 8 provide the 2.5%, 50% and 97.5% posterior percentiles for each of the predictors and interaction terms for both non-spatial and spatial models for Males 2006 and Females 2006. **Bold text** indicates statistical significance.

Both the spatial and non-spatial models have the same significant covariates. Higher

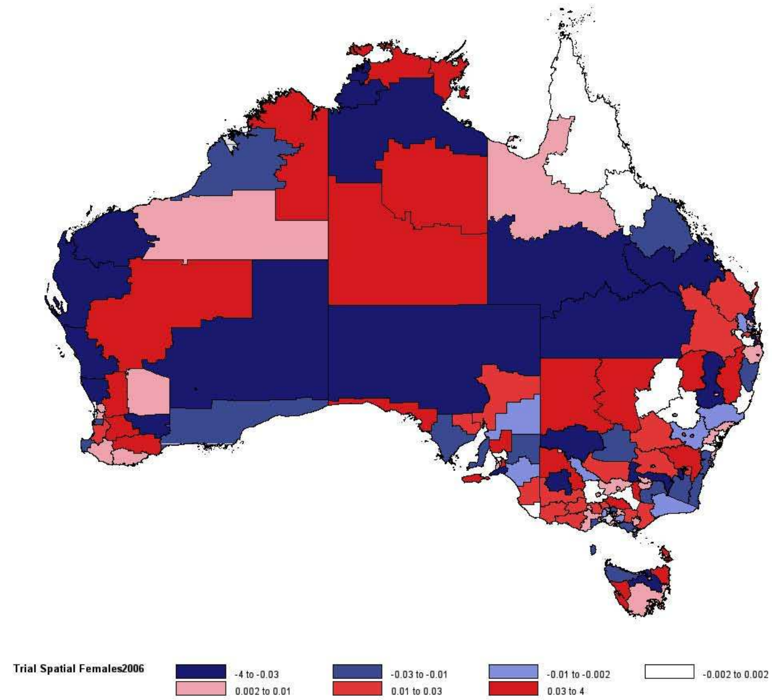

Figure 7: Standardized Error Map for Spatial Model with Covariate Age (Females 2006)

death rates in a SSD are associated with higher levels of Unemployment, Age, Indigenous and Labour covariates. Lower Overseas and Income covariates correspond to higher mortality rates. The lower mortality rates for higher values of the Overseas covariate is assumed to reflect selection of migrants to Australia as younger, healthier and wealthier than the average population.

Not all the covariates and interaction terms are significant at the 5% level. For the non-spatial frailties model for Males 2006, the covariates Overseas and Occupation are not significant, while for the spatial frailties model, the covariates Income and Occupation are not significant. The interaction terms of Income*Overseas, Income*Labour and Overseas*Labour are significant in the non-spatial frailties model. The interaction term Income and Labour is the only significant interaction term for the spatial frailties model.

For the Females 2006, Income, Overseas and Occupation covariates are not significant at the 5% level for the non-spatial frailties model. Overseas, Occupation and Labour are not significant in the spatial frailties model. The interactions between Income*Overseas and Income*Labour are significant for the non-spatial model while Income*Occupation and Income*Labour are significant for the spatial frailties model. The negative parameter value for the interaction term Income*Labour shows that a SSD with a higher proportion of people participating in the labour force reduces the effect of changes in Income.

A summary of the significant parameters for the models for all the years including the non-spatial (IID) and spatial frailties model for males and females is shown in Table 9. Age and Indigenous covariates and the interaction terms for Income*Overseas and Income*Labour are significant across these census years. Occupation and other interac-

Table 5: *Posterior Summaries for Non-Spatial Frailties Model (Males 2006) bold font indicates significant co-variate*

Covariates	**2.5%**	**50%**	**97.5%**
Intercept	-4.9740	-4.9480	-4.9190
Unemployment	0.0451	0.0741	0.1019
Age	0.2143	0.2556	0.2998
Income	-0.1475	-0.1031	-0.0576
Overseas	-0.0552	-0.0274	0.0016
Occupation	-0.0503	-0.0157	0.0177
Indigenous	0.0421	0.0822	0.1230
Labour	0.0479	0.0898	0.1359
Income * Overseas	0.0238	0.0697	0.1123
Income * Occupation	-0.0209	-0.0050	0.0112
Income * Labour	-0.0628	-0.0379	-0.0139
Overseas * Occupation	-0.0286	0.0045	0.0317
Overseas * Labour	-0.0719	-0.0381	-0.0013
τ	41.81	52.76	66.11

Table 6: *Posterior Summaries for Spatial Frailties Model (Males 2006)*

Covariates	**2.5%**	**50%**	**97.5%**
Intercept	-4.9440	-4.9250	-4.9050
Unemployment	0.0396	0.0748	0.1077
Age	0.2557	0.2961	0.3412
Income	-0.0796	-0.0205	0.0313
Overseas	-0.0853	-0.0490	-0.0100
Occupation	-0.0403	-0.0081	0.0219
Indigenous	0.1049	0.1551	0.2042
Labour	0.0128	0.0587	0.1043
Income * Overseas	-0.0016	0.0402	0.0885
Income * Occupation	-0.0140	0.0009	0.0163
Income * Labour	-0.0710	-0.0483	-0.0253
Overseas * Occupation	-0.0136	0.0168	0.0448
Overseas * Labour	-0.0671	-0.0333	0.0001
λ	15.48	19.80	25.04

tion terms including Income*Occupation, Overseas*Occupation and Overseas*Labour are not significant for the majority of the models. Interestingly, Income, Overseas and Labour covariates are not significant in earlier years but are more significant in recent years. The impact of Unemployment on mortality rates has increased over this period.

Table 7: *Posterior Summaries for Non-Spatial Frailties Model (Females 2006)*

Covariates	**2.5%**	**50%**	**97.5%**
Intercept	-5.1050	-5.0730	-5.0370
Unemployment	0.0540	0.0895	0.1276
Age	0.3551	0.4019	0.4529
Income	-0.0912	-0.0389	0.0146
Overseas	-0.0243	0.0090	0.0423
Occupation	-0.0243	0.0119	0.0488
Indigenous	0.0891	0.1331	0.1782
Labour	0.1001	0.1581	0.2123
Income * Overseas	0.0175	0.0847	0.1361
Income * Occupation	-0.0457	-0.0221	0.0013
Income * Labour	-0.1123	-0.0803	-0.0479
Overseas * Occupation	-0.0188	0.0075	0.0344
Overseas * Labour	-0.0692	-0.0244	0.0196
τ	34.88	44.37	56.00

Table 8: *Posterior Summaries for Spatial Frailties Model (Females 2006)*

Covariates	**2.5%**	**50%**	**97.5%**
Intercept	-5.0680	-5.0460	-5.0240
Unemployment	0.0358	0.0727	0.1084
Age	0.3286	0.3777	0.4288
Income	0.0030	0.0523	0.0961
Overseas	-0.0727	-0.0355	0.0045
Occupation	-0.0482	-0.0143	0.0181
Indigenous	0.1414	0.1945	0.2463
Labour	-0.0066	0.0542	0.1156
Income * Overseas	-0.0163	0.0305	0.0792
Income * Occupation	-0.0480	-0.0286	-0.0080
Income * Labour	-0.1283	-0.0998	-0.0717
Overseas * Occupation	-0.0069	0.0175	0.0402
Overseas * Labour	-0.0463	-0.0045	0.0316
λ	15.25	19.52	24.88

Table 9: *Significance of Covariates for Different Models*

Covariates	Males 2006 (Spatial)	Males 2006 (IID)	Females 2006 (Spatial)	Females 2006 (IID)	Males 2001 (Spatial)	Males 2001 (IID)	Females 2001 (Spatial)	Females 2001 (IID)	Males 1996 (Spatial)	Males 1996 (IID)	Females 1996 (Spatial)	Females 1996 (IID)
Unemployment	✓	✓	✓	✓	✓				✓			
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Income		✓	✓			✓				✓		
Overseas	✓				✓	✓						
Occupation												
Indigenous	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓
Labour	✓	✓		✓								
Income * Overseas		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
Income * Occupation			✓		✓						✓	
Income * Labour	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Overseas * Occupation								✓			✓	
Overseas * Labour		✓								✓		

5.2.1 Unemployment

Figure 8 plots the confidence intervals of the estimated Unemployment parameters for the spatial and non-spatial models. While the covariate is not significant in earlier years, a positive relationship between the unemployment and mortality rates is observed in 2006 for both males and females. SSDs with higher unemployment rates exhibit higher mortality levels.

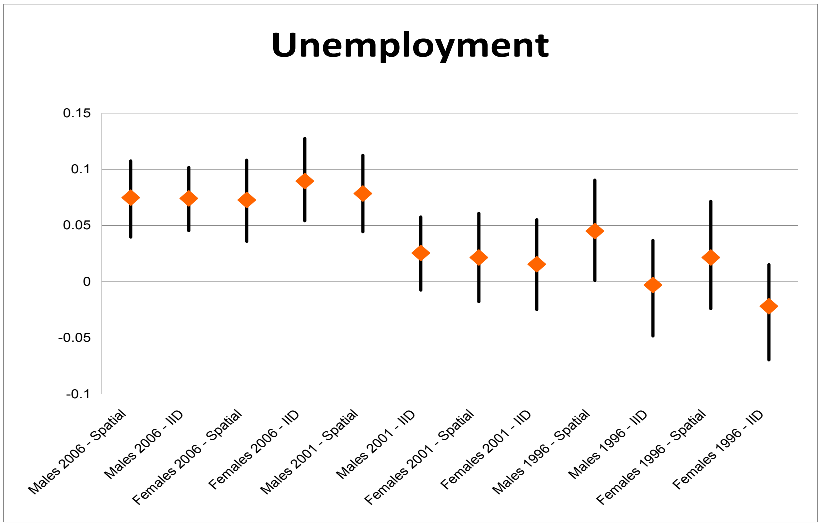

Figure 8: Boxplot of Unemployment Covariate

5.2.2 Age

The boxplots in Figure 9 show that the higher the proportion of people aged 65 or above, the higher the mortality rates within a SSD. The proportion of females 65+ has a greater impact on mortality levels than the proportion of males 65+.

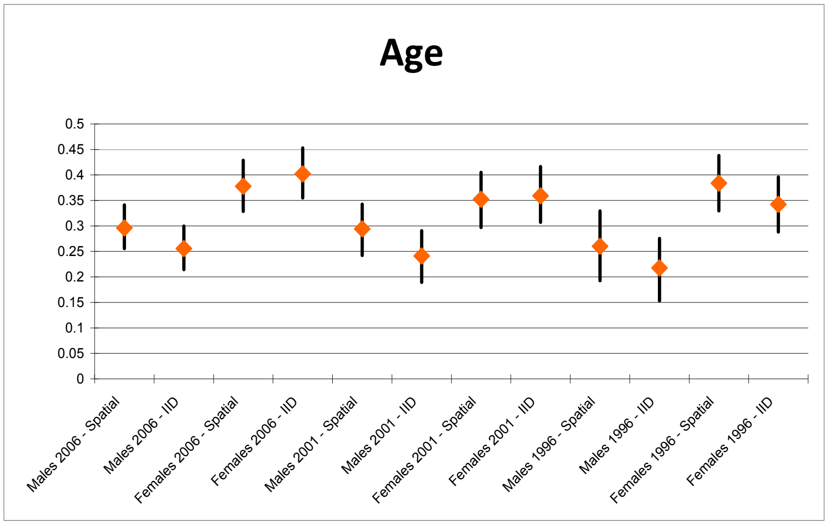

Figure 9: Boxplot of Age Covariate

5.2.3 Income

Income is not significant for many of the models. Figure 10 shows a statistically significant inverse relationship between Income and mortality rates for most of the non-spatial IID models. Spatial frailties models reduce the effect of income on mortality within a SSD.

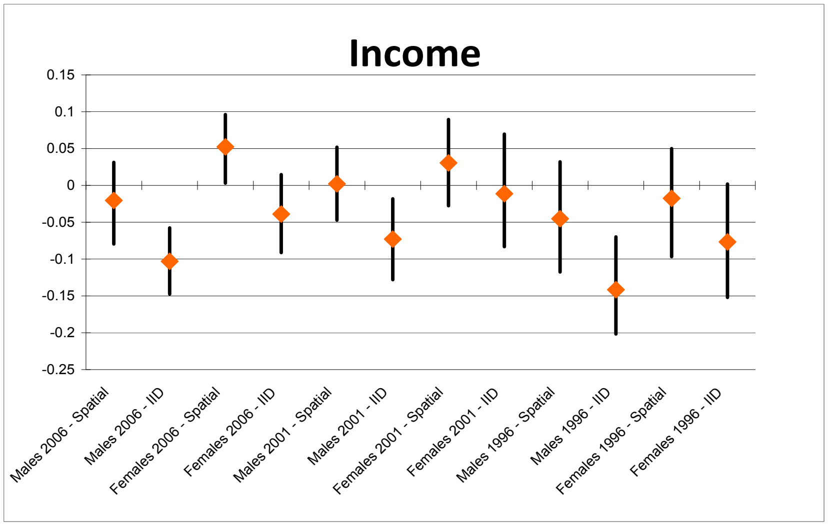

Figure 10: Boxplot of Income Covariate

Non-spatial (IID) frailties models show a more negative relationship between mortality rates and Income. Differences in parameter values for the spatial and non-spatial models are statistically significant for most of the models. The impact of Income on mortality rates are less significant when the frailties are assumed to be correlated with adjacent SSDs. Not including spatial dependence affects the significance and magnitude of estimated parameters.

5.2.4 Indigenous Population

Figure 11 shows that the proportion of individuals with indigenous origins is an important factor in explaining geographical variation in mortality. A positive coefficient shows that the higher the indigenous proportion in a SSD, the higher the mortality rate, consistent with previous observations of Wilkinson *et al.* (2000) [30]. Also most indigenous populations live away from urban centres and health facilities. Spatial models generally produce higher positive parameter values, so that when spatial dependence is allowed in the frailties specifications, the effect of the Indigenous covariate is greater. The impact of the Indigenous covariate has been increasing, indicating that differentials between indigenous and non-indigenous population in Australia are widening.

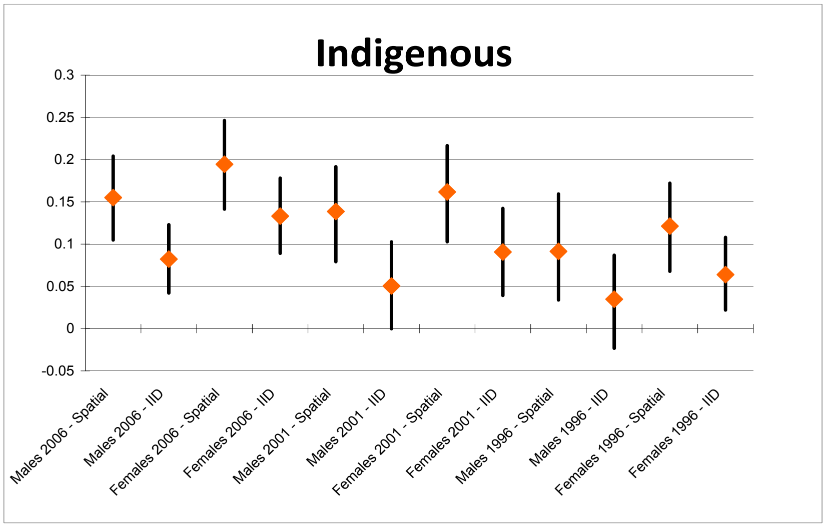

Figure 11: Boxplot of Indigenous Covariate

5.2.5 Interaction Terms

For males and females in 1996, 2001 and 2006, two interaction terms were found to be significant for the majority of the models: Income*Overseas and Income*Labour. Figure 12 reveals a positive coefficient for the interaction term Income*Overseas, so that the effect of Income on Australia mortality is higher for SSDs with a high proportion of people born overseas. Examination of the geographical variations in the proportion of people born overseas in Australia shows that SSDs with a high overseas proportion are generally located in major urban centres around Australia. The effect of income is greater in these major cities than in more rural regions. A downward trend shows this effect is diminishing over the past 10 years.

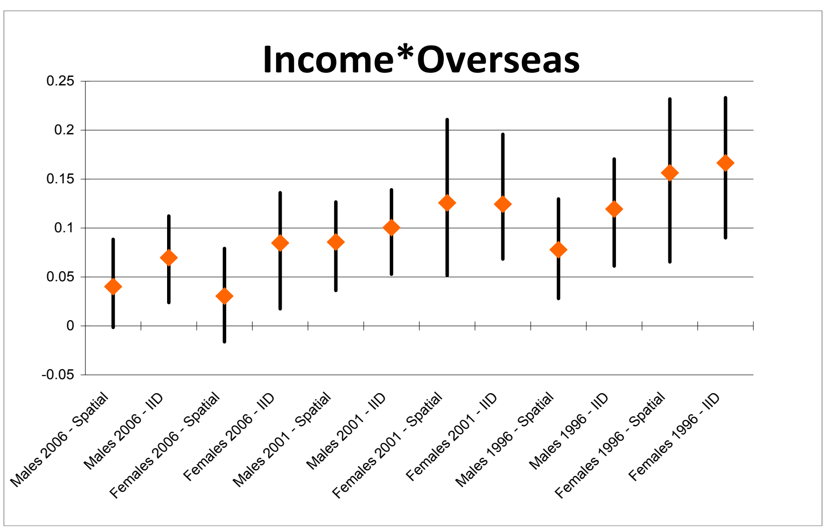

Figure 12: Boxplot of Income*Overseas Interaction Term

The interaction between income levels and labour force participation rate is also significant, as shown in Figure 13. Income has an effect that depends on labour force participation rate. Higher labour force participation rates reduce the effect of income.

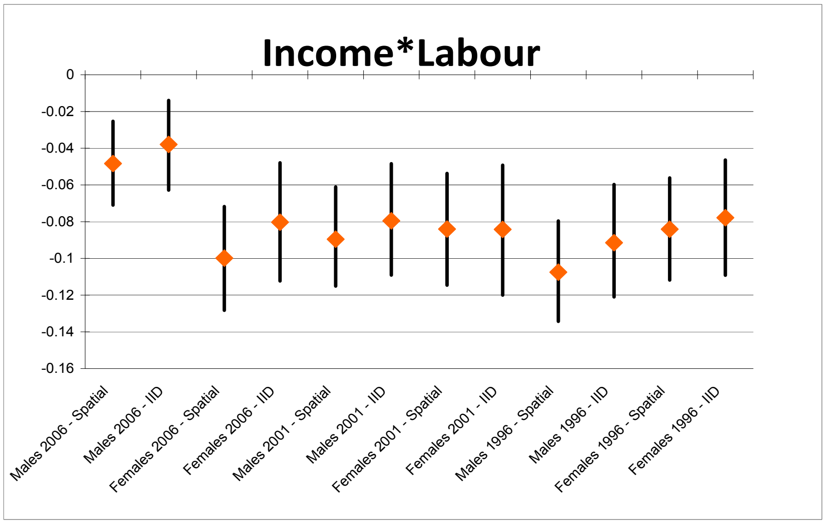

Figure 13: Boxplot of Income*Labour Interaction Term

5.2.6 Other Covariates

The covariate Overseas has had an increased impact on mortality rates of Australia in more recent years. While it was largely insignificant in 1996, there is a statistically significant inverse relationship between the Overseas covariate and mortality rates in recent years. This is assumed to be from the selection process where those who migrate into Australia have to meet age, health and wealth requirements.

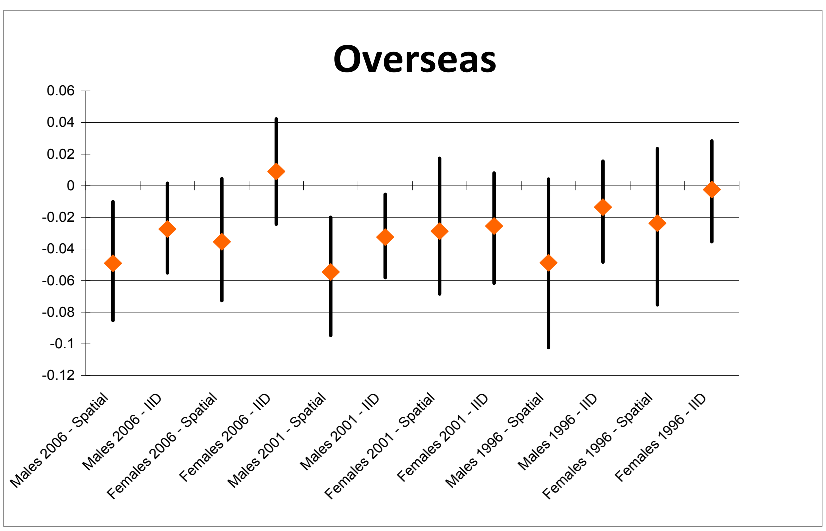

Figure 14: Boxplot of Overseas Covariate

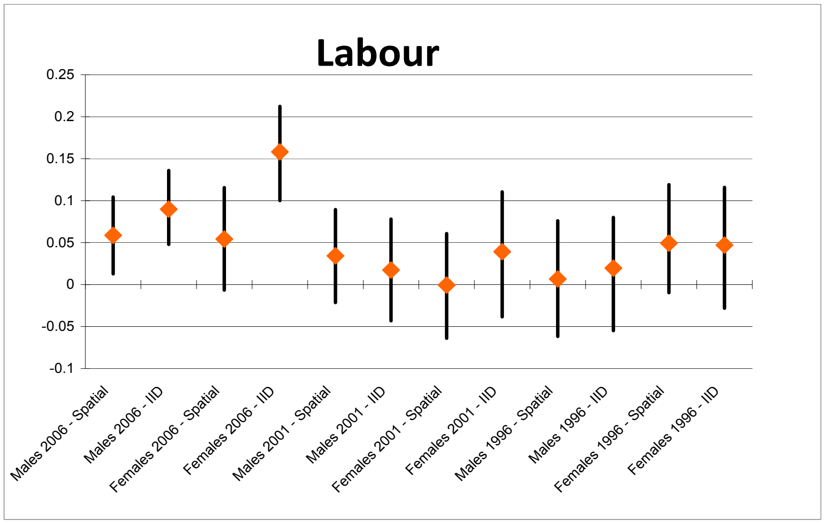

Figure 15: Boxplot of Labour Covariate

Although insignificant in the early years, the Labour covariate shows a positive relationship with mortality rates in 2006. A SSD with a higher labour force participation rate has a higher mortality rate than a SSD with a lower participation rate.

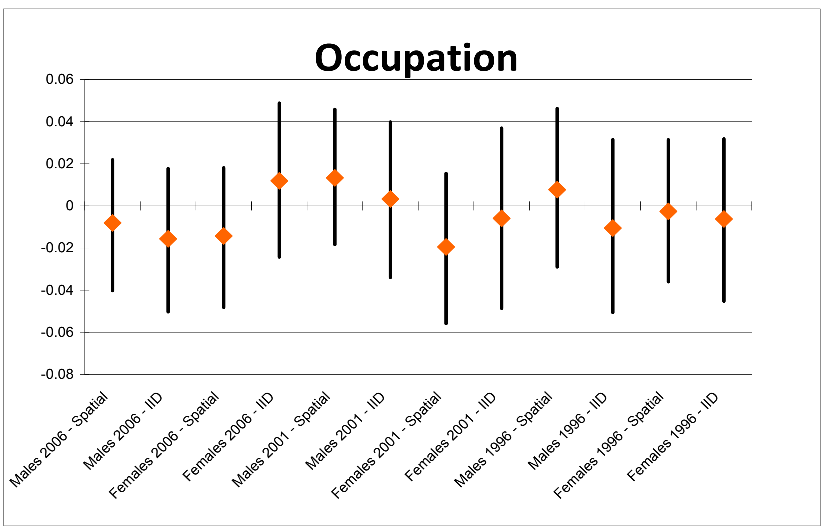

Figure 16: Boxplot of Occupation Covariate

Figure 16 shows that occupation is not significant when frailties are included in the models.

5.2.7 Prior Distribution Parameters: λ and τ

Under the spatial frailties model, a CAR specification was chosen to describe the dependence between frailties in adjacent SSDs. A high value of λ indicates smaller deviance from the average adjacent frailties $\bar{W}_i$. This means a higher degree of clustering and spatial correlation between frailties of adjacent SSDs. The boxplot in Figure 17 shows a slight increase in the estimated value of λ. There is increased spatial clustering evident in more recent years.

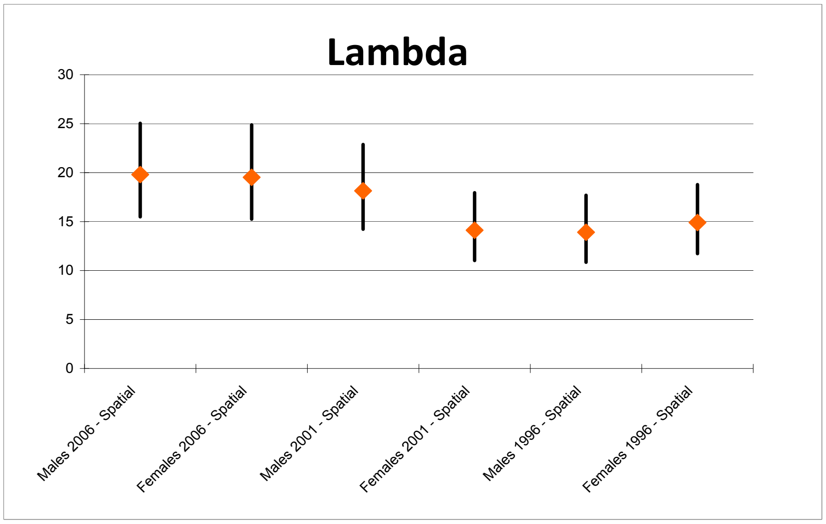

Figure 17: Boxplot of Parameter λ

Under the non-spatial model, the frailties are assumed to be independent and identically distributed with $W_i \sim N(0, \sigma^2)$ and $\tau = \frac{1}{\sigma^2}$. An increasing trend for the value of τ is shown in Figure 18. This indicates that there is less need for frailty terms in more recent years.

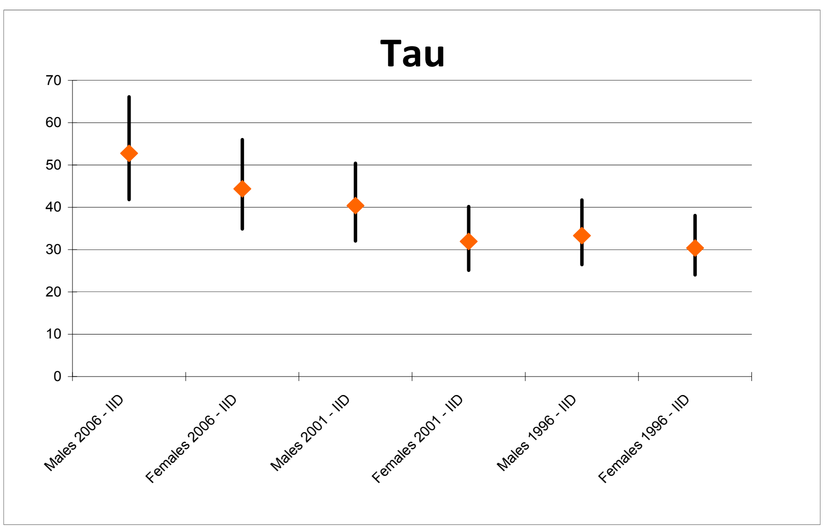

Figure 18: Boxplot of Parameter τ

5.2.8 Goodness of Fit

Table 10 shows that the frailties models have very similar DIC values. All the p_D values for the spatial frailties model are smaller than the corresponding values in the non-spatial frailties model showing that spatial frailties model reduce the effective model size p_D and thus the complexity of the model by including spatial correlation between adjacent SSDs.

Table 10: *Goodness of Fit of Models*

Model	**Males 2006**		**Males 2001**		**Males 1996**	
	$\mathbf{p_D}$	**DIC**	$\mathbf{p_D}$	**DIC**	$\mathbf{p_D}$	**DIC**
Spatial Model	189.499	2217.37	190.974	2213.25	194.708	2226.74
IID Model	191.586	2214.71	194.314	2211.22	196.264	2217.17

Model	**Females 2006**		**Females 2001**		**Females 1996**	
	$\mathbf{p_D}$	**DIC**	$\mathbf{p_D}$	**DIC**	$\mathbf{p_D}$	**DIC**
Spatial Model	185.355	2178.8	190.47	2184.97	188.897	2160.16
IID Model	190.285	2183.58	193.478	2177.61	193.743	2161.49

The spatial frailties and the non-spatial frailties model are similar when spatially correlated covariates are included. In the circumstances of limited data, spatial models provide substantial advantages over non-spatial models. There is less need for spatial models as additional factors are incorporated into the modeling. Geographical variations in mortality rates are captured by the covariates since they are also spatially correlated and capture the heterogeneity in mortality rates.

5.2.9 Model Errors

Figures 19-22 show the model error terms for the state of New South Wales to more clearly illustrate the effect of the spatial models. Although differences are small between the error terms of the spatial and non-spatial models, they show more randomness in the spatial models and are more clustered in the non-spatial models.

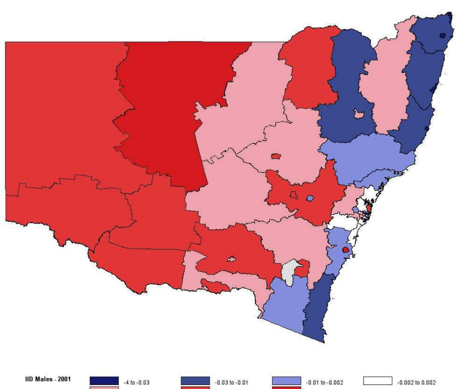

Figure 19: NSW Errors Plot Males 2001 (IID)

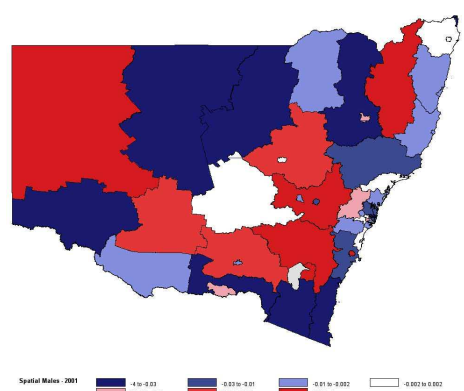

Figure 20: Errors Plot Males 2001 (Spatial)

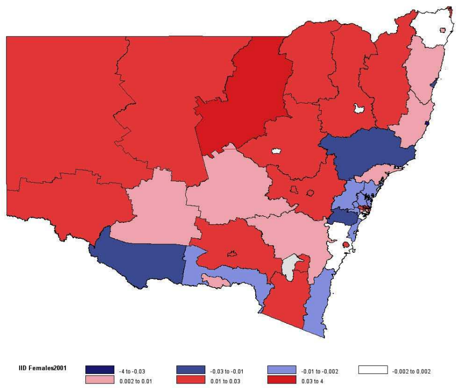

Figure 21: NSW Errors Plot Females 2001 (IID)

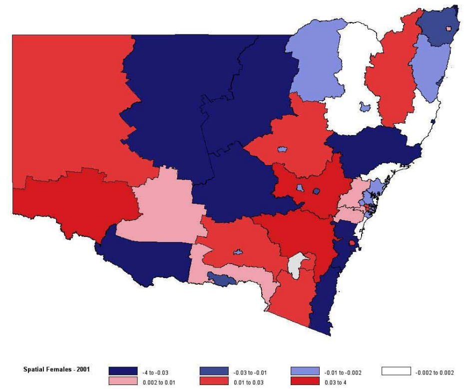

Figure 22: Errors Plot Females 2001 (Spatial)

Figures 19 and 21 show the error terms for a non-spatial specification for males and females mortality data in 2001. Clustering of underestimation of mortality rates in parts of the State are evident, while overestimation of mortality rates are concentrated along the coastal line of New South Wales. The error plots for the spatial models show randomness, indicating that spatially correlated covariates relevant to mortality rates in Australia have been omitted. For Australia there is evidence of the clustering of underestimation of mortality rates in New South Wales and to some extent, inland Australia. Clustering of overestimation in mortality rates occurs in the coastal regions of Queensland and Western Australia.

6 Conclusions

This paper has considered geographical variation of mortality and the effect of explanatory factors using Australian data. Logistic regressions in an hierarchical Bayes model were used to estimate the impact of covariates extracted from ABS deaths and census data on mortality rates. Models included spatial frailties to account for geographical variation.

Socio-economic variables are known to affect mortality rates. Despite this, in insurance and pensions the main risk factors taken into account are age, smoking status and certain diseases with known impact on mortality. Increasingly geo-spatial data is being considered for insurance risks and postcode underwriting is being used in the UK for life annuities. Socio-economic factors are known to vary spatially, as does observed mortality rates. Modeling a wider range of risk factors as well as spatial variation for mortality rates should be of significant practical interest to the life insurance industry.

In existing insurer pricing and risk management, where spatially-correlated risk factors are often overlooked, spatial models allow a more accurate assessment of pricing and reserving strategies in practice. The implementation of spatial frailties models provides improved mortality risk estimation. Incorporating additional spatially-correlated risk factors reduces the need for spatial models since these additional factors also vary geographically. Even so, including spatial dependence improves the models even with covariates.

More detailed data at an individual level would allow better assessment of mortality risk including the application of proportional hazards models. The models can be applied in wider insurance risk applications and to assess more detailed data. Although the focus has been on Australian mortality, the methodologies and framework covered are readily applied in analyzing spatial variations of mortality and insurance risk factors in other countries.

7 Acknowledgements

Andy Tang acknowledges the support of the Australian School of Business. Sherris acknowledges the support of ARC Linkage Grant Project LP0883398 Managing Risk with Insurance and Superannuation as Individuals Age with industry partners PwC and APRA and financial support from the Institute of Actuaries of Australia UNSW Actuarial Foundation.

8 Appendix A - Spatial Distribution of Covariates

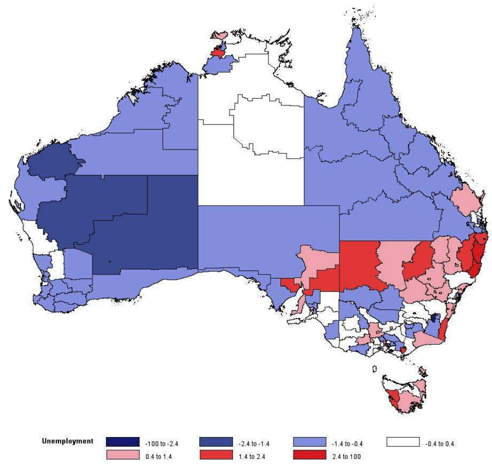

Figure 23: Males 2006 Unemployment

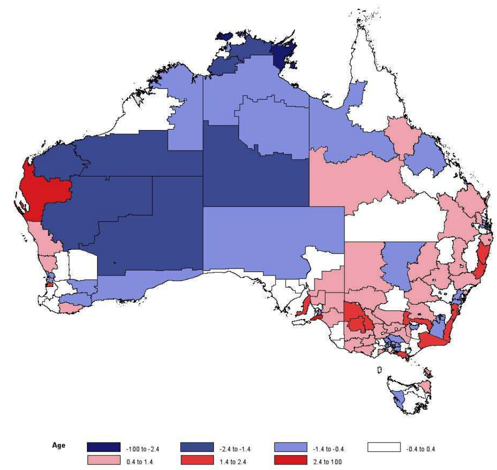

Figure 24: Males 2006 Age

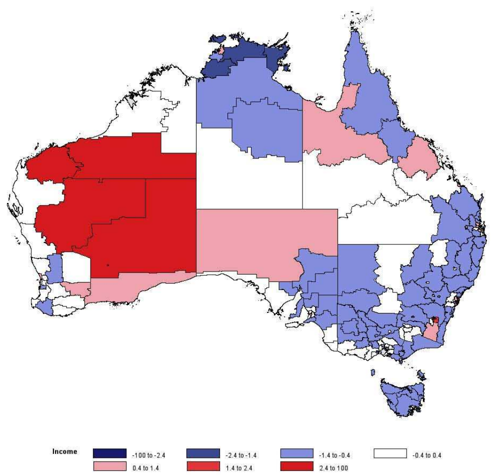

Figure 25: Males 2006 Income

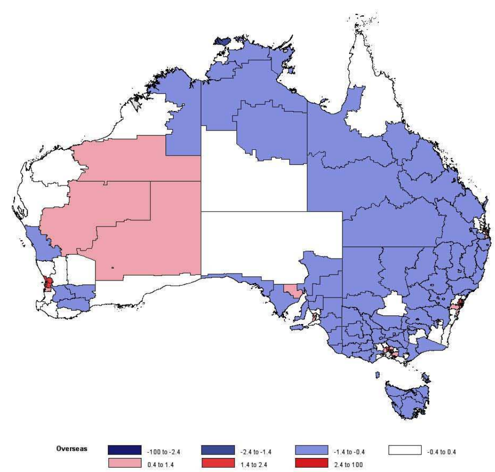

Figure 26: Males 2006 Overseas

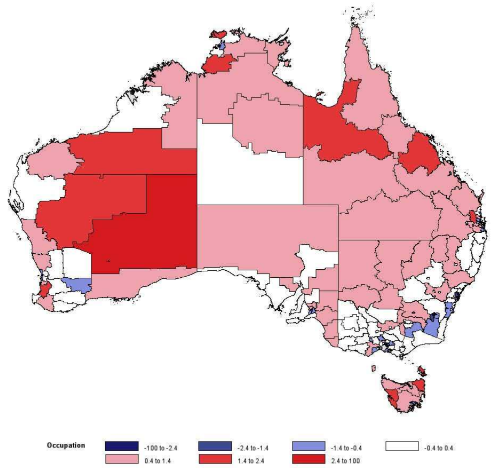

Figure 27: Males 2006 Occupation

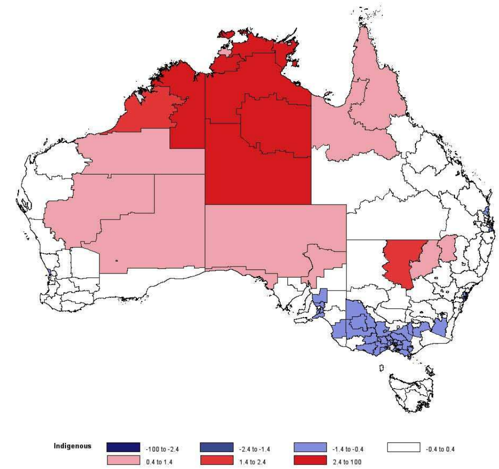

Figure 28: Males 2006 Indigenous

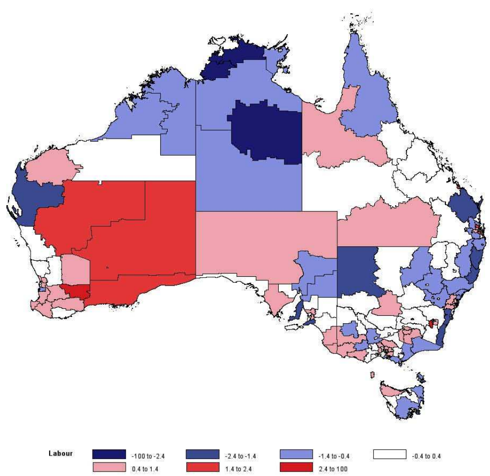

Figure 29: Males 2006 Labour

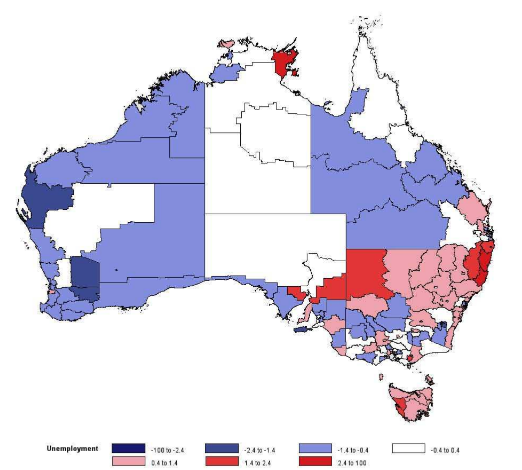

Figure 30: Females 2006 Unemployment

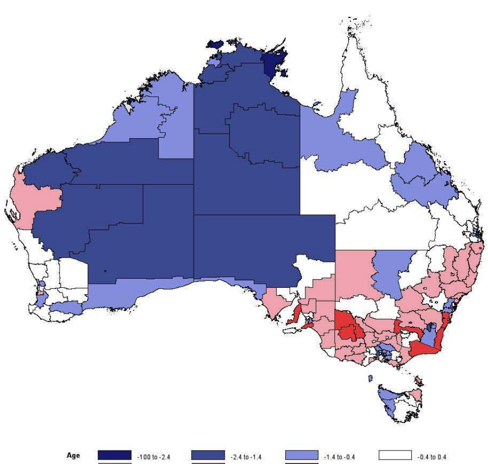

Figure 31: Females 2006 Age

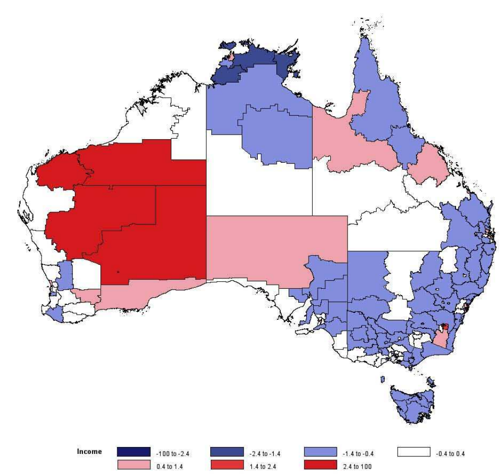

Figure 32: Females 2006 Income

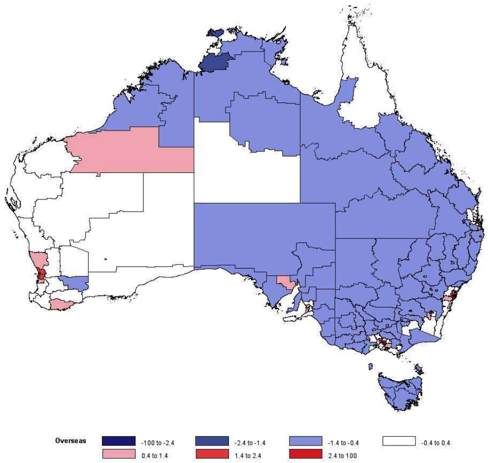

Figure 33: Females 2006 Overseas

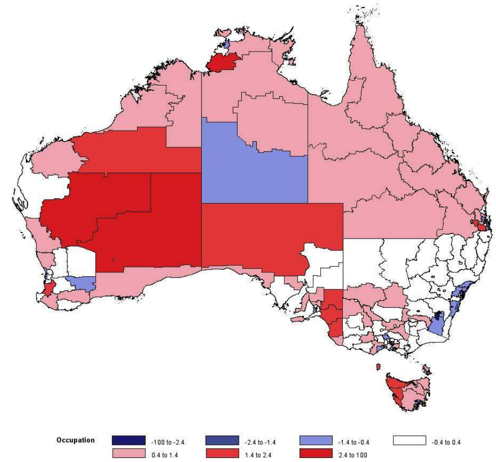

Figure 34: Females 2006 Occupation

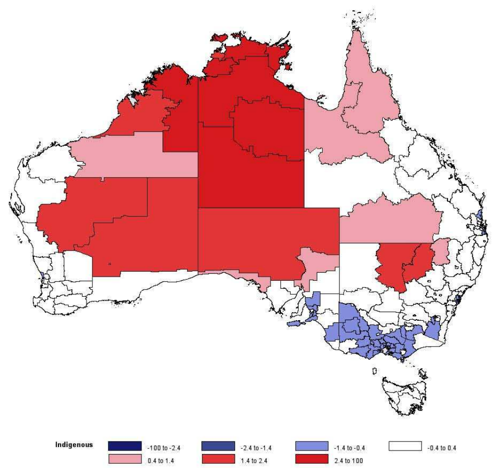

Figure 35: Females 2006 Indigenous

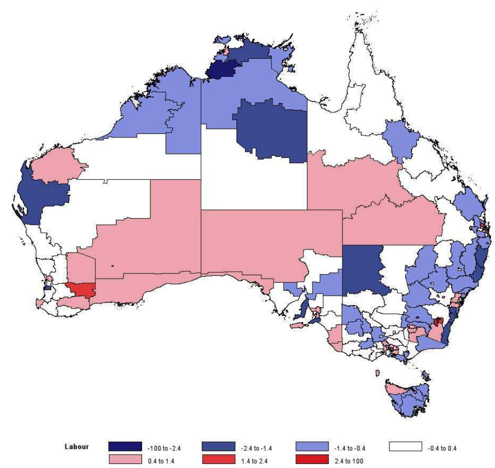

Figure 36: Females 2006 Labour

References

[1] Australian Bureau of Statistics. Causes of Death, Australia. Cat. No. 3303.0, Australian Government Publishing Service, Canberra.

[2] Australian Bureau of Statistics. Trends in Mortality - By Causes of Death in Australia, the States and Territories during 1971-92, and in Statistical Divisions and Sub-divisions during 1991-92. Cat. No. 3313.0, Australian Government Publishing Service, Canberra, 1994.

[3] Australian Bureau of Statistics. Australian and New Zealand Standard Classification of Occupations. Cat. No. 1220.0, Australian Government Publishing Service, Canberra, 2006.

[4] Australian Bureau of Statistics. Australian Standard Geographical Classification. Cat. No. 1216.0, Australian Government Publishing Service, Canberra, 2006.

[5] Banerjee, S., and Carlin, B. P. Spatial Semiparametric Proportional Hazards Models for Analyzing Infant Mortality Rates in Minnesota Counties. Appear in Case Studies in Bayesian Statistics, Volume VI, New York: Springer, 2002.

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

[7] Banerjee, S., Wall, M. M., and Carlin, B. P. Frailty Modeling for Spatially Correlated Survival Data, with Application to Infant Mortality in Minnesota. *Biostatistics 4*, 1 (2003), 123–142.

[8] Bernardinelli, L., and Montomoli, C. Empirical Bayes versus Fully Bayesian Analysis of Geographical Variation in Disease Risk. *Statistics in Medicine 11*, 8 (1992), 983–1007.

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

[10] Cressie, N. A. *Statistics for Spatial Data*. New York: Wiley, 1993.

[11] Doksum, K. A., and Gasko, M. On a Correspondence Between Models in Binary Regression Analysis and in Survival Analysis. *International Statistical Review 58* (1990), 243–252.

[12] Eberly, L. E., and Carlin, B. P. Identifiability and Convergence Issues for Markov Chain Monte Carlo Fitting of Spatial Models. *Statistics in Medicine 19* (2000), 2279–2294.

[13] Ecker, M. D., and Gelfand, A. E. Bayesian Modeling and Inference for Geometrically Anisotropic Spatial Data. *Mathematical Geology 31*, 1 (1999), 67–83.

[14] Grubesic, T. H. The Geodemographic Correlates of Broadband Access and Availability in the United States. *Telematics and Informatics 21* (2004), 335–358.

[15] Ingram, D. D., and Kleinman, J. C. Empirical Comparisons of Proportional Hazards and Logistic Regression Models. *Statistics in Medicine 8* (1989), 525–538.

[16] Kazembe, L. N. Spatial Modelling and Risk Factors of Malaria Incidence in Northern Malawi. *Acta Tropica 102*, 2 (2007), 126–137.

[17] McGilchrist, C. A., and Aisbett, C. W. Regression with Frailty in Survival Analysis. *Biometrics 47* (1991), 461–466.

[18] Ngianga-Bakwin, K., and Madise, N. The Spatial Epidemiology of Childhood Diseases in Malawi and Zambia. *African Population Stuides 19* (2004), 191–218.

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

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

[21] Rosen, S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *The Journal of Political Economy 82*, 1 (1974), 34–55.

[22] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society 64*, 4 (2002), 583–639.

[23] Stein, M. L. *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer, 1999.

[24] Tuljapurkar, S., and Boe, C. Mortality Change and Forecasting: How Much and How Little Do We Know? *North American Actuarial Journal 2*, 4 (1998), 13–47.

[25] Turrell, G., Kavanagh, A., and Subramanian, S. V. Area Variation in Mortality in Tasmania (Australia): the Contributions of Socioeconomic Disadvantage, Social Capital and Geographic Remoteness. *Health and Place 12*, 3 (2006), 291–305.

[26] Vaupel, J. W., Manton, K. G., and Stallard, E. The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography 16*, 3 (1979), 439–454.

[27] Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. Hierarchical Spatio-Temporal Mapping of Disease Rates. *Journal of the American Statistical Association 92* (1997), 607–617.

[28] Waller, L. A., Zhu, L., Gotway, C. A., Gorman, D. M., and Gruenewald, P. J. 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*, 5 (2007), 573–588.

[29] Wienke, A. Frailty Models. Working Paper, Max Planck Institute for Demographic Research, Germany, 2003.

[30] Wilkinson, D., Hiller, J., Moss, J., Ryan, P., and Worsley, T. Mortality Variation Across Australia: Descriptive Data for States and Territories, and Statistical Divisions. *Australian and New Zealand Journal of Publich Health 24*, 3 (2000), 226–233.

[31] Zimmerman, D. Another Look at Anisotropy in Geostatistics. *Mathematical Geology 25*, 4 (1993), 453–470.