ESRC Centre for Population Change Working Paper
Number 13

Does Specification Matter? Experiments with Simple Multiregional Probabilistic Population Projections

James Raymer
Guy J Abel
Andrei Rogers

July 2011

ISSN2042-4116

ABSTRACT

Population projection models that introduce uncertainty are a growing subset of projection models in general. In this paper, we focus on the importance of decisions made with regard to the model specifications adopted. We compare the forecasts and prediction intervals associated with four simple regional population projection models: a total growth rate model, a component model with net migration, a component model with in-migration and out-migration rates, and a multiregional model with destination-specific out-migration rates. Vector autoregressive models are used to forecast future rates of growth, birth, death, net migration, in-migration and out-migration, and destination-specific out-migration for the North, Midlands and South regions in England (additional specification decisions once again come into play). They are also used to forecast different international migration measures. The base data represent a time series of annual data provided by the Office for National Statistics from 1976 to 2008. The results illustrate how both the forecasted subpopulation totals and the corresponding prediction intervals differ for the multiregional model in comparison to other simpler models, as well as for different assumptions about international migration. The paper ends end with a discussion of our results and possible directions for future research.

KEYWORDS

Population forecasting, multiregional projections, time series models, forecast uncertainty, England

EDITORIAL NOTE

Dr. James Raymer and Dr. Guy J. Abel are members of the ESRC Research Centre for Population Change, University of Southampton. Dr. Guy Abel is also at the *Wittgenstein Centre* for Demography and Global Human Capital, Vienna Institute of Demography. Professor Andrei Rogers is affiliated with the Population Program, Institute of Behavioral Science, University of Colorado, Boulder.

Corresponding author: J.Raymer@soton.ac.uk

ACKNOWLEDGEMENTS

This research has been supported by ESRC Grant number RES-625-28-0001, as well as an ESRC-SSRC Collaborative Visiting Fellowship. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and should not be attributed in any manner to the Office for National Statistics (ONS) or the National Records of Scotland (NRS). The authors would like to thank Jakub Bijak, Peter W.F. Smith, Jonathan J. Forster and Arkadiusz Wiśniowski for their advice and suggestions concerning the modelling and direction of this research.

The ESRC Centre for Population Change Working Paper Series is edited by Teresa McGowan

ESRC Centre for Population Change

The ESRC Centre for Population Change (CPC) is a joint initiative between the Universities of Southampton, St Andrews, Edinburgh, Stirling, Strathclyde, in partnership with the Office for National Statistics (ONS) and the National Records of Scotland (NRS). The Centre is funded by the Economic and Social Research Council (ESRC) grant number RES-625-28-0001.

Website | Email | Twitter | Facebook | Mendeley

DOES SPECIFICATION MATTER? EXPERIMENTS WITH SIMPLE MULTIREGIONAL PROBABILISTIC POPULATION PROJECTIONS

TABLE OF CONTENTS

1. INTRODUCTION

Since the 1990s, there has been an increasing need to move away from deterministic and variant-style projections to probabilistic projections. Probabilistic projections have the advantage over variant style projections in that they specify the likelihood that a particular future population value will occur (Ahlburg and Land 1992; Lee and Tuljapurkar 1994; Lutz 1996; Bongaarts and Bulatao 2000). With variant projections, on the other hand, the user has no idea how likely they are, only that they are plausible scenarios representing the "most likely" and the "extreme" high and low possibilities. Despite the advantages of probabilistic projections, they have yet to be widely adopted by statistical agencies for several reasons (Lutz and Goldstein 2004). First, there are many types of uncertainties to consider, and including them in projections is not always straightforward, and it can be misleading to include them incorrectly. Second, national statistical offices do not always have the necessary expertise to develop probabilistic models or to extend their current models to include probabilities. Finally, while much has been done, there is still a lot of work needed to produce probabilistic models that are usable at a detailed demographic level, and that are capable of incorporating expert knowledge of demographic experts.

In this paper, we focus on only a small part of the picture, that is, to identify the consequences of choosing a particular projection model in terms its forecasted populations and measures of uncertainty. We develop a probabilistic time series framework for multiregional projection models (Rogers 1995), extending some of the ideas in Gullickson (2001), Sweeney and Konty (2002) and Wilson and Bell (2007). The overall aim is to learn about issues arising from simple probabilistic multiregional projection modelling.

Deterministic models are first used to illustrate why specification matters. Second, we show how adding probabilities, obtained in a time series framework, gives rise to additional issues of specification. The illustrations are carried out with a three-region multiregional model of England using demographic data collected from 1976 to 2008. Multivariate time series methods are used to forecast various future crude rates of subnational demographic change (i.e., births, deaths, internal migration and international migration). These models account for the strong correlations over time and across regions. The forecasted demographic rates are then used to produce four different sets of future regional populations in England for the purpose of comparing the consequences of different 'closed' (to international migration) projection models. Finally, we introduce three different specifications for including international migration, and then assess the consequences for the resulting forecasts.

2. ALTERNATIVE SPECIFICATIONS OF A THREE REGION POPULATION SYSTEM

Because a large number of different subnational projections are possible within an individual country, it is not feasible to consider here all of the combinations that might be relevant. Consequently, we focus on a particular example of multiregional mathematical demography: a three region population model of England. However, it should be clear that the methods described here are also applicable to more than three regions, indeed even to regions that are not regions in the geographical sense but that are states of existence, such as the states of being married or divorced, healthy or sick, employed or unemployed (Land and Rogers 1982; Schoen 1988).

2.1 DATA

This study requires data on populations, births, deaths, interregional migration and international migration. It also requires us to produce future values of these components to be inserted into the projection models described below. To keep things simple, we decided to focus on just three regions in England: the North, Midlands and South, which can be aggregated from England's nine Government Office Regions.[1] The data were obtained from the Office for National Statistics for the years 1976 to 2008. Finally, as we are primarily interested in the consequences of projection model specification, we utilise 'crude' rates of demographic change which exclude the effects of age and sex.

The mid-year population estimates for the North, Midlands and South regions in England from 1976 to 2008 are presented in the top panel of Figure 1. These plots illustrate the different growth regimes in England with the North population remaining at pretty much the same level over time (around 14.6 million), the Midlands population rising slightly from 9.0 million in 1976 to 9.8 million in 2008 and the South population rising more rapidly from 23.0 million in 1976 to 26.9 million in 2008. The bottom panel of Figure 1 contains the corresponding annual rates of growth. With the exception of the late 1980s, early 1990s and after 2000 periods, the North exhibited negative growth rates, whereas those for the Midlands and South regions were positive (with the exception of the Midlands in 1981). Since 2000, the growth rates increased considerably for all three regions.

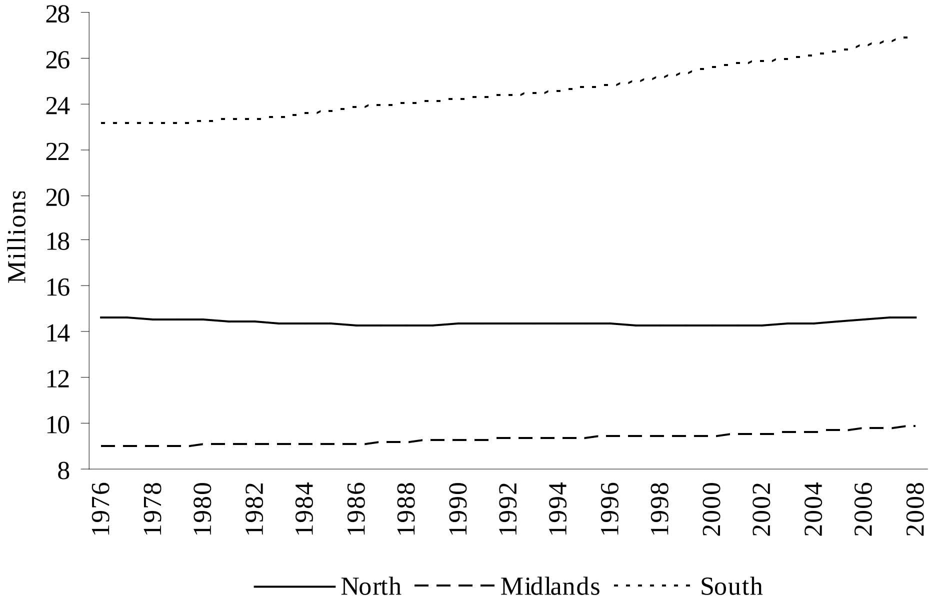
The crude rates of birth and death are presented for the three regions in Figure 2. Over time, the regional birth rates fluctuated (in parallel) between 0.010 and 0.014. The regional death rates, on the other hand, steadily declined with the North consistently exhibiting the

---

[1] North = North East, North West and Yorkshire and the Humber; Midlands = East Midlands and West Midlands; South = East, London, South East and South West.

highest rates, and the South (after the mid-1980s) the lowest. Finally, the crude rates of internal migration and international migration are presented for the three regions in Figures 3 and 4, respectively. From 1976 to 2008, net internal migration rates for the North tended to be negative, while those for the Midlands and South regions tended to be positive. Net international migration, on the other hand, increased steadily for all regions, with the South region exhibiting the highest rates. In terms of flows, the Midlands exhibited relatively high rates of internal migration whereas, for international migration, the South did.

A. Population size (millions)



B. Annual growth rate

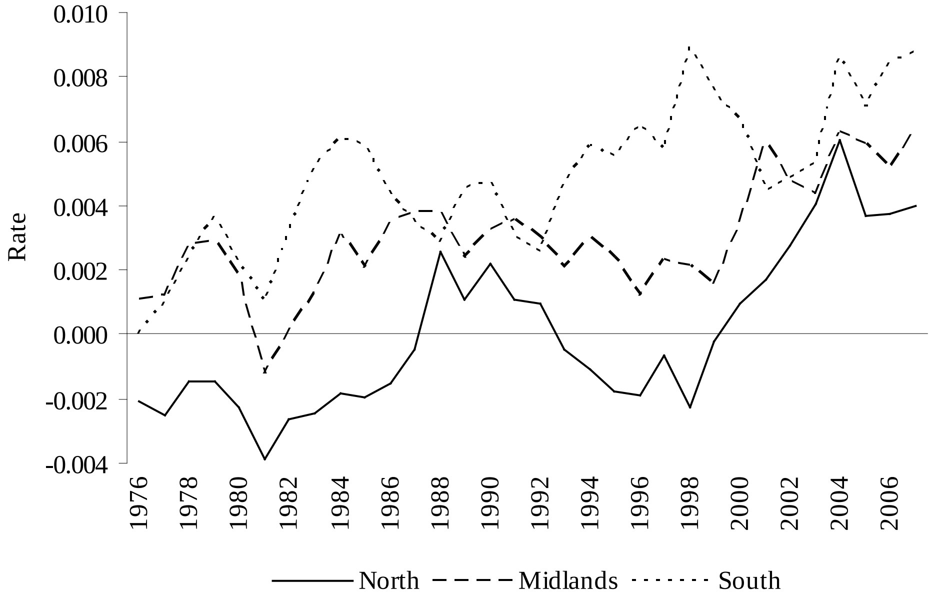
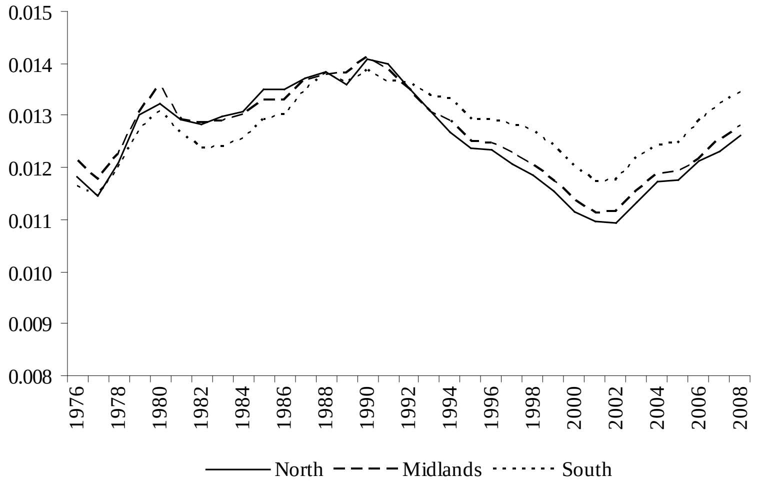


Figure 1 Population sizes and annual rates of growth for the North, Midlands and South regions of England, 1976-2008

A. Birth



B. Death

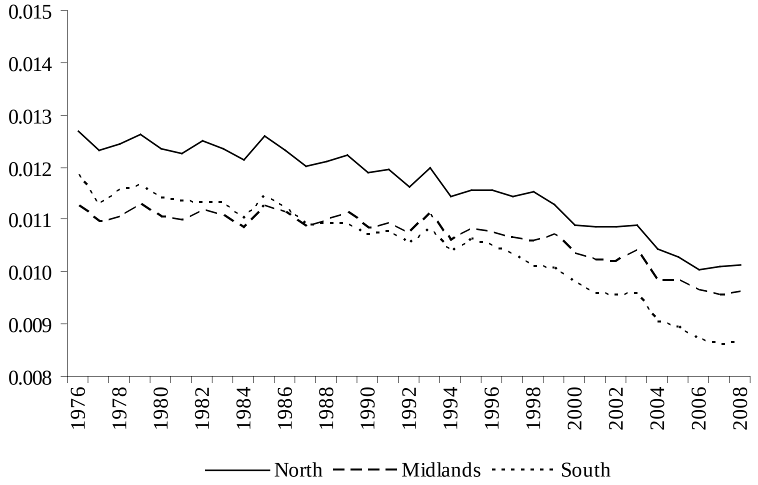
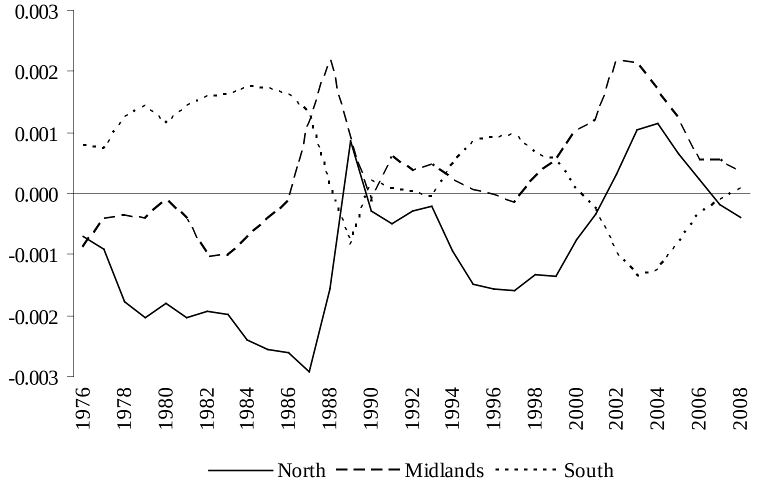
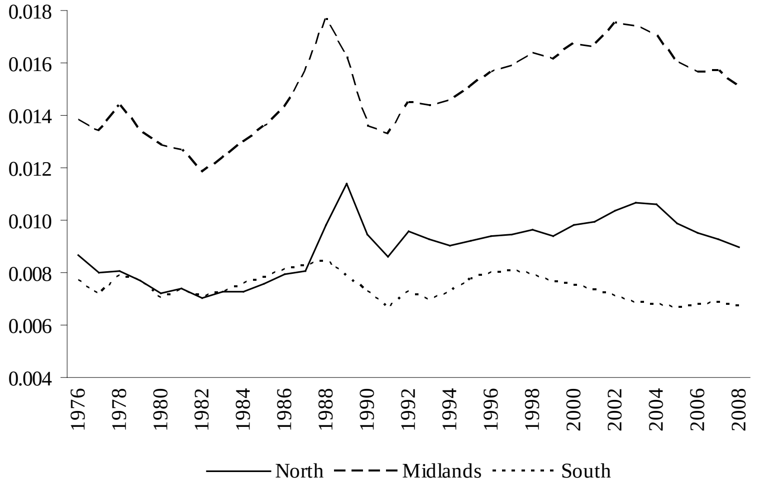


Figure 2 Crude rates of birth and death for the North, Midlands and South regions of England, 1976-2008

A. Net internal migration



B. In-migration



C. Out-migration

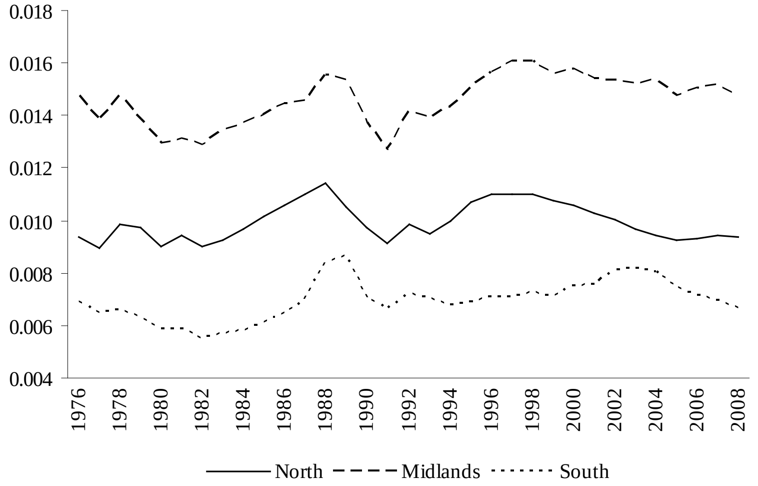
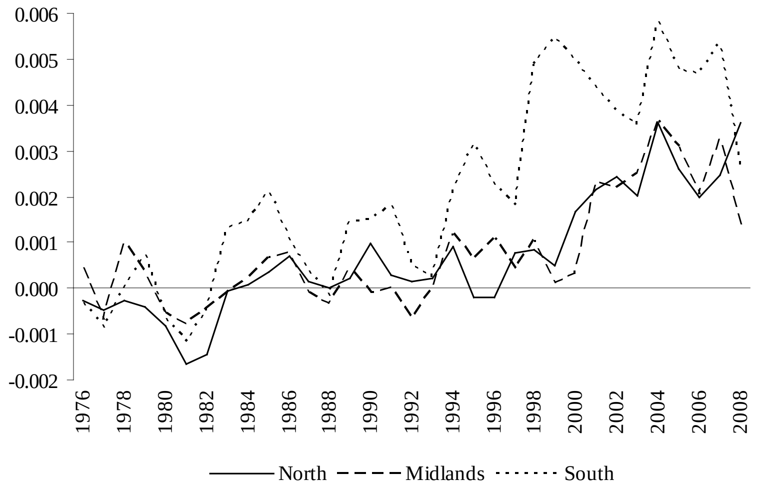
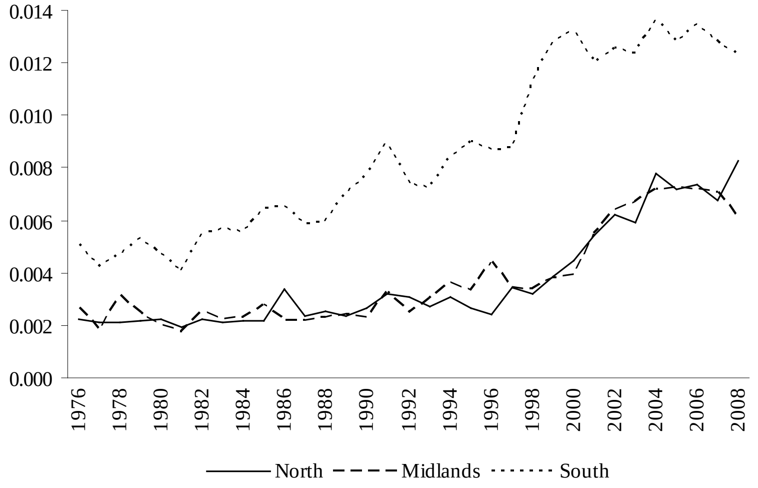


Figure 3 Crude rates of net internal migration, in migration and out-migration for the North, Midlands and South regions of England, 1976-2008

A. Net international migration



B. Immigration



C. Emigration

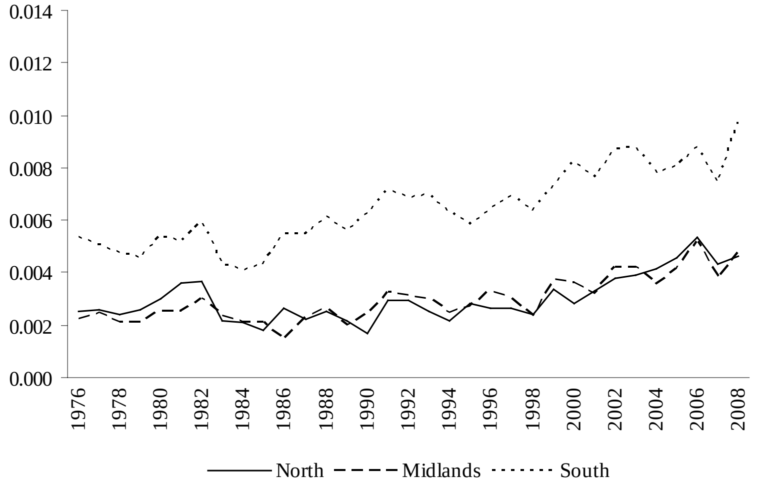


Figure 4 Crude rates of net international migration, immigration and emigration for the North, Midlands and South regions of England, 1976-2008

2.2 CLOSED UNIREGIONAL AND MULTIREGIONAL PROJECTION MODELS

The simplest ‘closed’ model is the global projection model, which for a one-year projection, is specified for three regions as:

$$P_{t+1}^{N} = P_{t}^{N}(1 + r_{t,t+1}^{N}),$$

$$P_{t+1}^{M} = P_{t}^{M}(1 + r_{t,t+1}^{M}),$$

$$P_{t+1}^{S} = P_{t}^{S}(1 + r_{t,t+1}^{S}), \qquad (1)$$

where P denotes population, t denotes year and r denotes the annual growth rate. The superscripts N, M and S denote the North, Midlands and South regions, respectively. For this model, the emphasis is on forecasting the inputs, i.e., the annual rates of growth, to the year 2021.

The global model provides a useful benchmark but does not contain the demographic rates underlying the annual growth rate, i.e., a disaggregation of fertility, mortality and migration often considered necessary for more accurate projections. The incorporation of fertility and mortality rates is relatively straightforward. The inclusion of migration rates, however, is more complicated because two populations are involved simultaneously (i.e., an out-migrant from one place is an in-migrant to another). In practice, there are at least three principal ways of incorporating internal migration into ‘closed’ subnational population projections. The first focuses on net migration, the other two on gross migration. These models are specified below.

The second projection model considered in this paper is the component projection model with net migration rates, which for the three regions, are specified as:

$$P_{t+1}^{N} = P_{t}^{N}(1 + b_{t,t+1}^{N} - d_{t,t+1}^{N} + nm_{t,t+1}^{N}),$$

$$P_{t+1}^{M} = P_{t}^{M}(1 + b_{t,t+1}^{M} - d_{t,t+1}^{M} + nm_{t,t+1}^{M}),$$

$$P_{t+1}^{S} = P_{t}^{S}(1 + b_{t,t+1}^{S} - d_{t,t+1}^{S} + nm_{t,t+1}^{S}), \qquad (2)$$

where b is a crude birth rate, d is a crude death rate and nm is a crude net migration rate. Note, $r = b - d + nm$. Net migration rates are problematic because they only describe the difference in movements, that is, they are difficult to model behaviourally because there is no such individual as a net migrant, and they generally introduce a bias into the projection process because both the numerators and the denominators of the net migration rates are changing (Rogers 1990).

Gross migration may be entered into the projection process either by considering only inflows and outflows irrespective of other regions in the system (a uniregional perspective) or by keeping track of the various origins and destinations (a multiregional perspective). In both cases, one obtains a considerable increase in useful information over the net migration projection. Thus, the third model is the component projection model with crude rates of in-migration (*i*) and out-migration (*o*) instead of the net migration rates (i.e., $nm = i - o$) used in the previous model. This model, specified for the three regions, is:

$$P_{t+1}^{N} = P_{t}^{N}(1 + b_{t,t+1}^{N} - d_{t,t+1}^{N} + i_{t,t+1}^{N} - o_{t,t+1}^{N}),$$

$$P_{t+1}^{M} = P_{t}^{M}(1 + b_{t,t+1}^{M} - d_{t,t+1}^{M} + i_{t,t+1}^{M} - o_{t,t+1}^{M}),$$

$$P_{t+1}^{S} = P_{t}^{S}(1 + b_{t,t+1}^{S} - d_{t,t+1}^{S} + i_{t,t+1}^{S} - o_{t,t+1}^{S}). \qquad (3)$$

The final model is the multiregional model with destination-specific out-migration rates,

$$P_{t+1}^{N} = P_{t}^{N}(1 + b_{t,t+1}^{N} - d_{t,t+1}^{N} - o_{t,t+1}^{N-M} - o_{t,t+1}^{N-S}) + P_{t}^{M} o_{t,t+1}^{M-N} + P_{t}^{S} o_{t,t+1}^{S-N},$$

$$P_{t+1}^{M} = P_{t}^{M}(1 + b_{t,t+1}^{M} - d_{t,t+1}^{M} - o_{t,t+1}^{M-N} - o_{t,t+1}^{M-S}) + P_{t}^{N} o_{t,t+1}^{N-M} + P_{t}^{S} o_{t,t+1}^{S-M},$$

$$P_{t+1}^{S} = P_{t}^{S}(1 + b_{t,t+1}^{S} - d_{t,t+1}^{S} - o_{t,t+1}^{S-N} - o_{t,t+1}^{S-M}) + P_{t}^{N} o_{t,t+1}^{N-S} + P_{t}^{M} o_{t,t+1}^{M-S}, \qquad (4)$$

where, for example, o^{M-N} represents the destination-specific out-migration rate between the Midland and North regions. In this model, there are no rates of in-migrations, only rates of out-migration applied to the correct populations “at risk”. The multiregional model can also be expressed in matrix form, i.e.,

$$\begin{bmatrix} P_{t+1}^{N} \\ P_{t+1}^{M} \\ P_{t+1}^{S} \end{bmatrix} = \begin{bmatrix} 1 + b_{t,t+1}^{N} - d_{t,t+1}^{N} - o_{t,t+1}^{N} & o_{t,t+1}^{M-N} & o_{t,t+1}^{S-N} \\ o_{t,t+1}^{N-M} & 1 + b_{t,t+1}^{M} - d_{t,t+1}^{M} - o_{t,t+1}^{M} & o_{t,t+1}^{S-M} \\ o_{t,t+1}^{N-S} & o_{t,t+1}^{M-S} & 1 + b_{t,t+1}^{S} - d_{t,t+1}^{S} - o_{t,t+1}^{S} \end{bmatrix} \begin{bmatrix} P_{t}^{N} \\ P_{t}^{M} \\ P_{t}^{S} \end{bmatrix}$$

(5)

or, more simply, as

$$\mathbf{P}_{t+1} = \mathbf{G}_{t,t+1}\mathbf{P}_{t}, \qquad (6)$$

where **G** is the growth matrix and **P** is a vector of subnational populations. In the next subsection, this model is extended to include flows of international migration.

To summarise, there are several ways to specify subnational population projections. Uniregional perspectives of population growth and change (i.e., Equations 1-3) are simpler to construct but are problematic because they can easily introduce biases and inconsistencies into regional population projections (e.g., the overall national net migration total may not come out to be zero). The problems arise because both net migration totals and in-migration flows are assessed only with respect to the population in the region of destination. Thus, changes in the size of the destination population, arising out of changes in the patterns of, say, natural increase for a given year, will produce a higher (or lower) net migration or in-migration total in the following year as a result. For example, one could imagine the origin population being ultimately reduced to zero, but a fixed and positive in-migration rate in the destination region will nevertheless continue to generate a flow of migrants from other regions in the population. A multiregional perspective removes these biases. Furthermore, a projected multiregional population system must yield a zero net internal migration total for the nation, but net internal migration-based models never do that.

To illustrate the differences that can arise between a uniregional projection and a multiregional projection, consider a simple case where the rates of demographic change from the most recent period (i.e., 2008) are kept fixed for 13 years to project the regional populations for the year 2021. In the uniregional case (Equations 1-3), the projection model for 2009, expressed in matrix form, is equal to:

$$\begin{bmatrix} 14695 \\ 9879 \\ 27070 \end{bmatrix} = \begin{bmatrix} 1.00208 & 0 & 0 \\ 0 & 1.00351 & 0 \\ 0 & 0 & 1.00490 \end{bmatrix} \begin{bmatrix} 14664 \\ 9844 \\ 26938 \end{bmatrix},$$

where the regional populations at time *t* and *t*+1 are in thousands. Note, since the rates are fixed, the results for the models in Equations 1-3 are the same. The corresponding multiregional projection model (Equations 4-6) is:

$$\begin{bmatrix} 14695 \\ 9879 \\ 27070 \end{bmatrix} = \begin{bmatrix} 0.99311 & 0.00504 & 0.00304 \\ 0.00350 & 0.98845 & 0.00360 \\ 0.00586 & 0.00969 & 0.99817 \end{bmatrix} \begin{bmatrix} 14664 \\ 9844 \\ 26938 \end{bmatrix}$$

In the first projection, the estimated values from both models are the same as they both use rates calculated on the basis of the previous year. The differences become more apparent over time, as shown in Figure 5. By 2021, the uniregional model projects a North population that is 23 thousand less than the multiregional projection. The same is true for the Midlands population but with the difference being much less at around five thousand. For the South region, the opposite occurs: the uniregional projection results in a population that is 29 thousand more than produced by the multiregional projection.

2.3 OPENING THE MULTIREGIONAL MODEL TO INTERNATIONAL MIGRATION

The previous subsection specified different models for a closed population system. In this subsection, the impacts of different assumptions regarding the inclusion of international migration are assessed. The base model from which to make comparisons is the multiregional projection model described above in Equations 4-6. There are several options for adding international migration to this model. The first and simplest option is to include net international migration rates within the diagonal elements of Equation 6. The second option is to model immigration and emigration rates separately and then include them in the diagonal elements of Equation 6. The third option includes immigration counts as an additional vector and crude rates of emigration are placed in the diagonal elements of the growth matrix,

$$\mathbf{P}_{t+1} = \mathbf{G}_{t,t+1}\mathbf{P}_t + \mathbf{I}_t \,. \qquad (7)$$

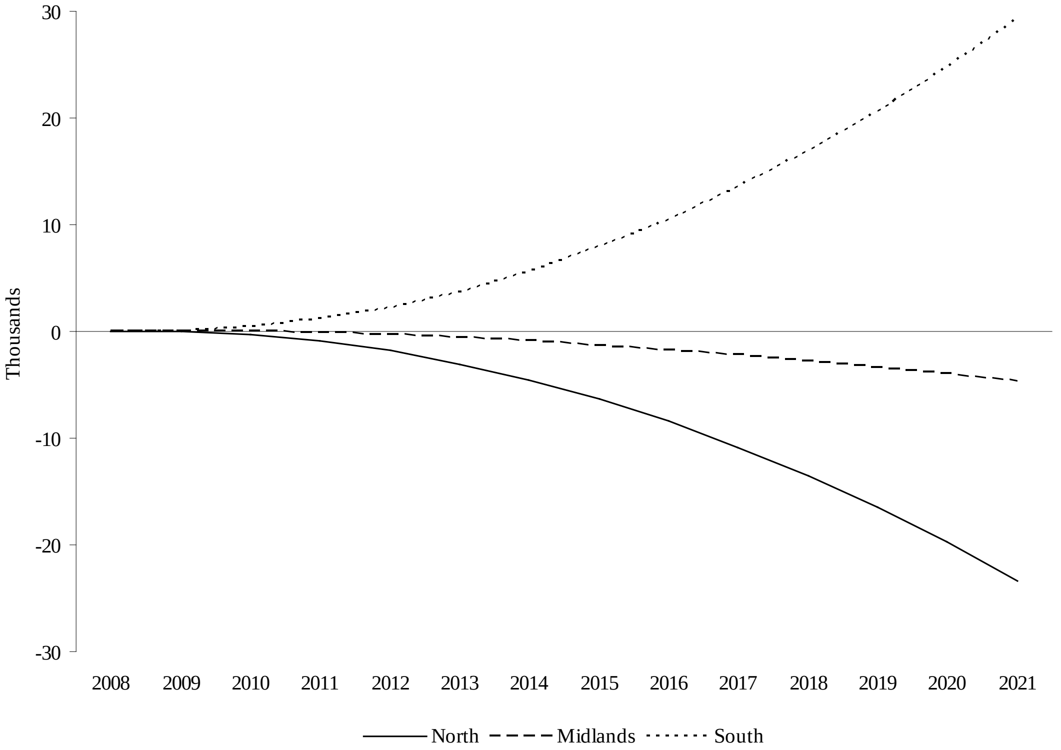


Figure 5 Differences between uniregional and multiregional projections: 2008 fixed rates example

To illustrate the differences created by the different assumptions of international migration, we continue our example of fixed rates from the previous subsection. Here, the multiregional model is used as the basis for projection. The multiregional projection model that incorporates fixed rates of international migration in the diagonal (i.e., net immigration or immigration and emigration rates) is:

$$\begin{bmatrix} 14748 \\ 9892 \\ 27139 \end{bmatrix} = \begin{bmatrix} 0.99672 & 0.00504 & 0.00304 \\ 0.00350 & 0.98977 & 0.00360 \\ 0.00586 & 0.00969 & 1.00073 \end{bmatrix} \begin{bmatrix} 14664 \\ 9844 \\ 26938 \end{bmatrix}.$$

The multiregional model that uses emigration rates in the diagonal and adds immigration numbers (Equation 7) is specified as:

$$\begin{bmatrix} 14748 \\ 9892 \\ 27139 \end{bmatrix} = \begin{bmatrix} 0.98847 & 0.00504 & 0.00304 \\ 0.00350 & 0.98368 & 0.00360 \\ 0.00586 & 0.00969 & 0.98841 \end{bmatrix} \begin{bmatrix} 14664 \\ 9844 \\ 26938 \end{bmatrix} + \begin{bmatrix} 121 \\ 60 \\ 332 \end{bmatrix}.$$

Again, the first projection (2009) results in the same regional populations and the differences only become apparent over time. The differences between the two projection models with international migration and the closed model are presented in Figure 6. Here, not surprisingly, we see that opening the models to international migration greatly increases the projection population totals by 2021: 652-707 thousand for the North, 171-196 thousand for the Midlands and 772-963 thousand for the South. Applying immigration in the diagonal elements of the projection model results in higher projected populations, especially for the South region, where the difference is 191 thousand for 2021.

3. MULTIVARIATE TIME SERIES FORECASTING MODELS

The previous section showed how various subnational projection models may be specified. In this section, we are interested in obtaining forecasts of the demographic components of change with measures of uncertainty. To do this, we rely on multivariate time series models, for which additional specification decisions need to be made. We rely on models that are able to capture the correlations both over time and amongst regions. Simulations of the results from the models fitted to the crude rates are then used to quantify the future uncertainty in the forecasts based on the historical patterns in the demographic components. As before, the projection models are initially closed to international migration to simplify the comparison. Afterwards, we add international migration to the multiregional specification. The incorporation of age cohorts and sex will be carried in future research.

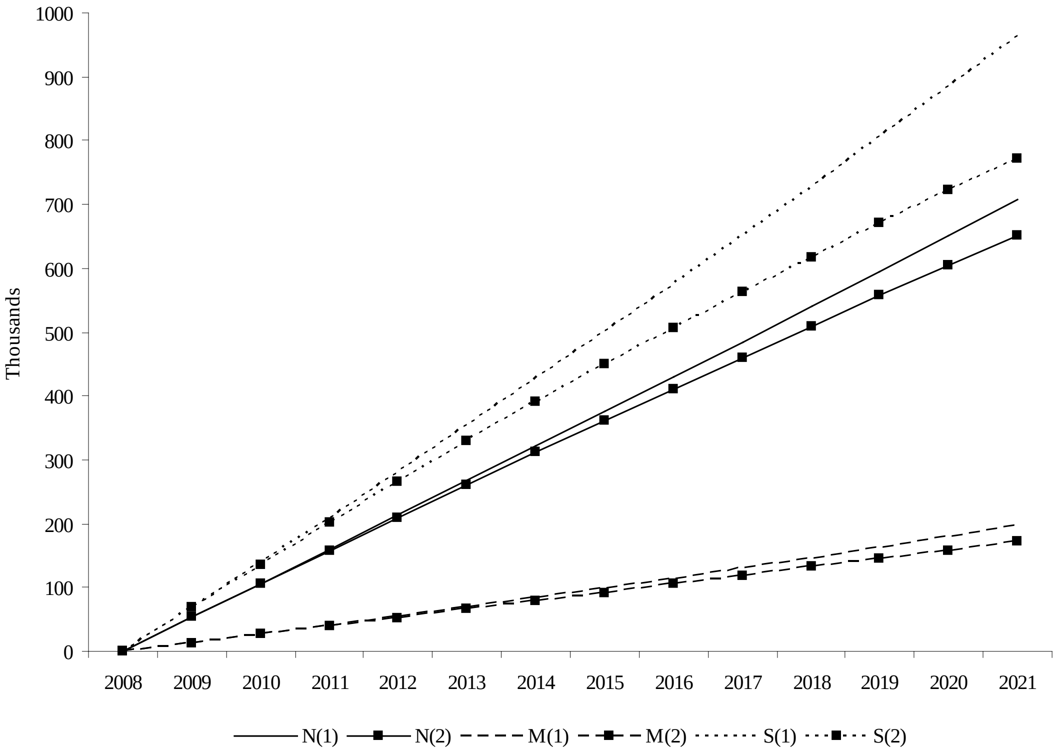


Figure 6 The differences between two open multiregional projections and the closed projection: 2008 fixed rates example

Notes: N(1), M(1) and S(1) = North, Midlands and South projections, respectively, with immigration and emigration rates in the diagonal; N(2), M(2) and S(2) = North, Midlands and South projections, respectively, with emigration rates in the diagonal and immigration numbers added as a vector.

3.1 CORRELATIONS AMONG REGIONAL DEMOGRAPHIC COMPONENTS OVER TIME

In our data, there exist strong correlations in the demographic components of change both over time and amongst regions. When considering probabilistic subnational projections, one needs to account for these correlations to obtain accurate prediction intervals. We first describe the correlations amongst the different demographic inputs for the four closed projection models, specified in Equations 1-4 and presented in Figures 1-3, followed by the correlations for the time series of international migration data presented in Figure 4.

To start, consider the global projection model specified in Equation 1. The inputs for this model are presented in Figure 1 and the corresponding correlations are presented in Table 1A. Here, we see that the correlation between the North and Midlands is strong (0.76), while the correlations between the other two series are relatively weak (Midlands and South, 0.36)) and not significant (North and South, 0.22) at the 5% significance level, based on the standard t-test for the null hypothesis of no correlation between two variables.

The correlations amongst crude regional birth, death and net migration rates, used as inputs in the projection model specified in Equation 2, are set out in Table 1B. Here, we find that the patterns of regional births and deaths, considered separately, are all highly and positively correlated. The correlations amongst regional net migration rates are also significant but positive between the North and Midlands (0.59) and negative between North and South (-0.95) and Midlands and South (-0.81). As far as the correlations amongst the different demographic components (i.e., births, deaths and net migration) are concerned, the patterns are less clear and not very strong. For example, the South's birth rates are not correlated with regional death rates or net-migration rates, however, those in the North and Midlands are (with the exception of net-migration in the Midlands).

The correlations amongst regional in-migration and out-migration rates are presented in Table 1C, along with the correlations between birth rates and death rates. Interestingly, the only significant correlation amongst the three in-migration rates is between North and Midlands. The correlations amongst regional out-migration rates, on the other hand are all significant and positive, as are the correlations between in-migration and out-migration rates (with the exception of South in-migration and Midlands out-migration). The correlations amongst the destination-specific out-migration rates are presented in Table 1D. All of the correlations amongst these rates were positive; however, not all were significant.

Finally, the correlations amongst net international migration rates, immigration and emigration rates and immigration totals and emigration rates, used as inputs for the projection models specified in Section 2.3, are presented in Table 2. Here, all correlations are positive and highly significant.

In summary, for the modelling of demographic components, we decided that there was plenty of evidence to include the correlations amongst the regional rates of each demographic component, as well as between the separate components of migration (e.g., in-migration and out-migration and immigration and emigration). The correlations between other demographic components, e.g., birth rates and death rates or death rates and in-migration rates, however, are not included as they are not as strong and do not exhibit clear patterns.

3.2 VECTOR AUTOREGRESSIVE MODELS

Uncertainty in population projections come from four main sources: the projection model(s), parameter estimates, expert judgments and historical data (Alho and Spencer 2005, pp. 238-240). Uncertainty can also be based on the results of past projections (Keilman 2001, 2008). As Gullickson (2001, p. 2) points out, there are two important issues that must be addressed when producing multiregional population forecasts with uncertainty:

> "First, one must consider the spatial correlation between component rates across regions. Second, one must develop a parsimonious method of modeling and forecasting a larger number of migration rates."

We focus the first aspect. Uncertainty measures are derived from historical time series by using multivariate time series models (described below) for forecasting crude rates of regional growth, birth, death, net migration, in-migration, out-migration and destination-specific out-migration. More specifically, we apply vector autoregressive (VAR) time series models to account for correlations both over time and across regions.

A. Overall growth rates

	N	M
M	0.76	
S	*0.22*	0.36

B. Birth (B), death (D) and net migration (NM) rates

		B			D			NDM	
		N	M	S	N	M	S	N	M
B	M	0.99							
	S	0.82	0.83						
D	N	0.52	0.52	*0.03*					
	M	0.50	0.49	*0.06*	0.98				
	S	0.48	0.48	*-0.01*	0.99	0.97			
NM	N	-0.39	-0.35	*-0.06*	-0.65	-0.60	-0.66		
	M	*-0.28*	*-0.28*	*0.08*	-0.58	-0.50	-0.60	0.59	
	S	0.38	0.36	*-0.01*	0.70	0.64	0.71	-0.95	-0.81

C. Birth (B), death (D), in-migration (I) and out-migration (O) rates

		B			D			I			O	
		N	M	S	N	M	S	N	M	S	N	M
B	M	0.99										
	S	0.82	0.83									
D	N	0.52	0.52	*0.03*								
	M	0.50	0.49	*0.06*	0.98							
	S	0.48	0.48	*-0.01*	0.99	0.97						
I	N	-0.37	-0.35	*0.07*	-0.61	-0.51	-0.63					
	M	-0.48	-0.48	*-0.05*	-0.64	-0.55	-0.65	0.86				
	S	*0.24*	*0.23*	*0.11*	0.52	0.55	0.52	*-0.12*	*0.12*			
O	N	*0.01*	*0.00*	*0.20*	*0.03*	*0.13*	*0.02*	0.35	0.57	0.81		
	M	-0.56	-0.55	*-0.17*	-0.54	-0.46	-0.54	0.74	0.89	*0.32*	0.68	
	S	*-0.29*	*-0.27*	*0.07*	-0.47	-0.38	-0.49	0.95	0.90	*0.06*	0.44	0.73

D. Birth (B), death (D) and destination-specific out-migration (O) rates

		B			D			O				
		N	M	S	N	M	S	N-M	N-S	M-N	M-S	S-N
B	M	0.99										
	S	0.82	0.83									
D	N	0.52	0.52	*0.03*								
	M	0.50	0.49	*0.06*	0.98							
	S	0.48	0.48	*-0.01*	0.99	0.97						
O	N-M	-0.53	-0.53	*-0.08*	-0.57	-0.46	-0.56					
	N-S	*0.33*	*0.31*	*0.32*	0.37	0.44	0.35	0.37				
	M-N	-0.56	-0.53	*-0.11*	-0.61	-0.51	-0.61	0.82	*0.09*			
	M-S	-0.47	-0.47	*-0.18*	-0.40	-0.35	-0.39	0.86	0.50	0.64		
	S-N	*-0.12*	*-0.09*	*0.16*	*-0.26*	*-0.16*	*-0.28*	0.45	*0.14*	0.79	*0.25*	
	S-M	-0.40	-0.39	*-0.01*	-0.59	-0.52	-0.61	0.77	*0.21*	0.83	0.66	0.73

Table 1 Correlations amongst crude regional demographic rates, 1976-2008
Note: *Italics* = not significant at 0.05 level; N = North, M = Midlands and S = South.

A. Net international migration rates

	N	M
M	0.83	
S	0.78	0.76

B. Immigration (IM) and emigration (EM) rates

		IM			EM	
		N	M	S	N	M
IM	M	0.94				
	S	0.87	0.88			
EM	N	0.84	0.80	0.64		
	M	0.83	0.84	0.80	0.85	
	S	0.86	0.81	0.87	0.74	0.90

C. Immigration (IM) totals and emigration (EM) rates

		IM			EM	
		N	M	S	N	M
IM	M	0.94				
	S	0.88	0.90			
EM	N	0.84	0.80	0.67		
	M	0.84	0.84	0.81	0.85	
	S	0.86	0.82	0.88	0.74	0.90

Table 2 Correlations amongst crude regional rates of international migration over time, 1976-2008
Note: N = North, M = Midlands and S = South.

First consider autoregessive (AR) models, which have a long history of being used to forecast populations (see, e.g., Saboia 1974; Ahlburg 1987; Pflaumer 1992; Alho and Spencer 2005). An AR model of order 1, denoted AR(1), is defined as

$$y_t = \mu + \alpha y_{t-1} + u_t \qquad (8)$$

where y denotes a particular demographic rate, the subscript t denotes time period, μ represents the mean level of the process, α is the autoregressive coefficient representing the correlation between observations y_t and y_{t-1} and u_t is assumed to be independently normally distributed with zero mean and constant variance, σ^2. Predictions from this model can be obtained as

$$y_{T+1|T} = \mu + \alpha y_T \, , \qquad (9)$$

where T is the last observation of y_t. The 95% prediction intervals for this value are

$$y^U_{T+1|T} = y_{T+1|T} + 1.96 \frac{\sigma}{\sqrt{T}}$$

$$y_{T+1|T}^{L} = y_{T+1|T} - 1.96\frac{\sigma}{\sqrt{T}} \tag{10}$$

Once fitted, AR models can be used to forecast future values of the time series process.

When observations are taken simultaneously on two or more time series, a multivariate model to describe the interrelationships amongst several series of data can be developed (for an introduction, refer, e.g., to Chatfield 2004 or Lütkepohl 2005). In other words, VAR models are the multivariate equivalent of the AR model outlined above. A VAR model describes the evolution of *m* variables as a linear function of their past observed values. The variables can be arranged into a set of *m* × 1 vectors y'_t = (y_{it}, …, y_{mt}). A standard VAR(1) model, when for example m=3, is specified as:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} \tag{11}$$

This is can be expressed in matrix notation as:

$$\mathbf{y}_t = \mathbf{C} + \mathbf{A}\mathbf{y}_{t-1} + \mathbf{u}_t \tag{12}$$

where **C** is a *m* × 1 vector of constants, **A** is a *m* × *m* matrix and $\mathbf{u}_t$ is a *m* × 1 vector of error terms. The matrix **A** captures the correlations over time and amongst regions. In this paper, we also make us of a simple extension of Equation 12 for the inclusion of trend terms:

$$\mathbf{y}_t = \mathbf{C} + \mathbf{D}_t + \mathbf{A}\mathbf{y}_{t-1} + \mathbf{u}_t \tag{13}$$

where $\mathbf{D}_t$ is a 1 × *m* vector of additional parameters that represent time dependent trend in $\mathbf{y}_t$. As the regional data are highly correlated, we apply VAR models to predict all of the crude rates used in the various projection models. These include the crude rates of growth, birth, death, net migration, in-migration and out-migration, destination-specific out-migration and immigration and emigration.

For simplicity, we only consider VAR(1) models in this paper. Most of the patterns are explained by the first lag, although we admit that alternative specifications with longer lags may be used (e.g., Abel et al. 2010b). However, given the relatively short time series it is difficult to test what the best model may be. We also do not restrict the structure of the VAR models, and allow some parameters, that might not be significant, to be included in the projection model. One major advantage of this approach is that the forecasts of the demographic inputs are predicted, not only based on past trends, but also by trends exhibited simultaneously in other regions. For example, we know that there is strong symmetry in origin-destination migration flow tables that persist over time, and that subnational patterns of fertility often follow the same pattern as each other, albeit at different levels. Our models take these factors into account.

4. CLOSED UNIREGIONAL AND MULTIREGIONAL FORECASTS

In this section, we first present the results from the VAR models applied to forecast the crude rates of growth, birth, death, net migration, in-migration and out-migration, and destination-specific out-migration. Second, we present and compare the forecasted populations according to the four projection models described in Section 2.2.

4.1 PROBABILISTIC TIME SERIES FORECASTS OF DEMOGRAPHIC COMPONENTS

Six VAR(1) models were applied to forecast crude rates of growth, birth, death, net migration, in-migration and out-migration, and destination-specific out-migration.These were fitted using the *vars* package (Pfaff 2008a, 2008b) in R (R Development Core Team 2010). The models for rates of growth, birth, death and net migration consisted of three time series each, whereas the models for in-migration and out-migration and destination-specific out-migration consisted of six series each.

In Table 3, the estimated VAR(1) coefficients and goodness-of-fit for the model applied to the regional crude rates of growth are presented. The model did well in capturing the patterns in the data with adjusted R^2 values of 0.76 for the North, 0.60 for the Midlands and 0.66 for the South. However, only three parameters were significant: the autocorrelation parameters for the North and South and the trend parameter for the South. The reason why the autocorrelation parameter for the Midlands was not significant is simply that there is no autocorrelation when the model controls for the other variables in the model. Here, we find that there is correlation between the North and Midlands growth rates over time, whereas the South's patterns are not correlated at all with the North's or the Midlands'. In other words, the Midlands patterns can be predicted by the patterns in the North but not in the South.

Simulations for the predicted values from the model of regional growth rates are set out in Figure 7. These were produced by simulating 10,000 values from multivariate normal distributions. The time varying mean vectors in this distribution were based on the estimated coefficients illustrated in Table 3. The variance-covariance matrix in these distributions was estimated from the VAR model (not shown in Table 3). A generic function was written in R to provide these simulated values for any size VAR model. This function was depended on the `rmvnorm` routine from the *mvtnorm* R package (Genz & Bretz, 2009 and Genz et al. (2010). From Figure 7, we see that the model predicts increases in the future growth rates for all three regions.

		North	Midlands	South	Constant	Trend	Adj. R^2
Coefficient	North	0.644	0.099	-0.103	-0.001	0.000	0.764
	Midlands	0.351	0.239	0.182	0.002	0.000	0.603
	South	0.081	-0.349	0.442	0.002	0.000	0.656
Std. Error	North	0.203	0.251	0.166	0.001	0.000	
	Midlands	0.187	0.231	0.152	0.001	0.000	
	South	0.213	0.264	0.174	0.001	0.000	
Pr(>|t|)	North	0.004	0.695	0.540	0.234	0.110	
	Midlands	0.071	0.309	0.242	0.084	0.982	
	South	0.708	0.198	0.017	0.079	0.033	

Table 3 Vector autoregressive model coefficients and goodness-of-fit for regional growth rates, 1976-2008

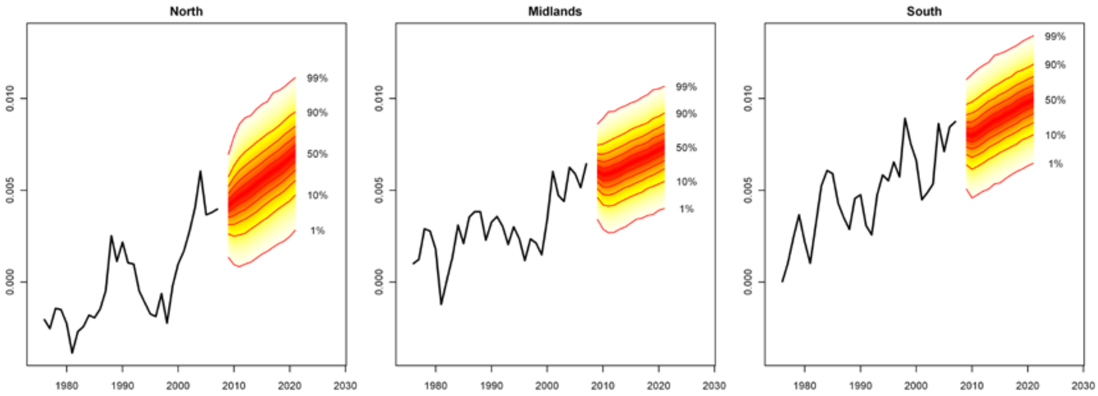


Figure 7 Vector autoregressive forecasts of total growth rates for the North, Midlands and South regions, 2009-2021

The coefficients for the VAR(1) models were applied to crude rates of birth and death are set out in Table 4A and Table 4B, respectively. For these models, the adjusted R^2 values are considerably higher than for the growth rates, particularly for mortality. For the model applied to births, six parameters are significant: the autocorrelation parameters for the North and South, the South-North parameter and the three constant terms (trend is not included in the model). Nearly all parameters in the mortality model are significant. The exceptions are the North-South parameter and the trend parameters for the North and South. The predicted crude rates of birth and death and corresponding predicted intervals are set out in Figure 8.

Both the crude birth and death rates are predicted to decline for the forecasted period, albeit with considerably less certainty in the regional fertility forecasts.

		North	Midlands	South	Constant	Trend	Adj. R^2
A. Birth rates							
Coefficient	North	1.089	0.038	-0.349	-0.971		0.869
	Midlands	0.741	0.261	-0.256	-1.101		0.854
	South	0.660	-0.608	0.726	-0.958		0.790
Std. Error	North	0.504	0.573	0.167	0.465		
	Midlands	0.478	0.544	0.158	0.441		
	South	0.435	0.494	0.144	0.401		
Pr(>|t|)	North	0.040	0.948	0.045	0.046		
	Midlands	0.133	0.635	0.118	0.019		
	South	0.140	0.229	0.000	0.024		
B. Death rates							
Coefficient	North	1.046	-1.893	1.214	-2.909	0.003	0.946
	Midlands	0.970	-1.762	1.201	-2.842	0.005	0.884
	South	0.831	-1.881	1.591	-2.196	0.003	0.957
Std. Error	North	0.425	0.582	0.393	0.659	0.002	
	Midlands	0.429	0.588	0.397	0.666	0.002	
	South	0.494	0.677	0.458	0.767	0.002	
Pr(>|t|)	North	0.020	0.003	0.005	0.000	0.143	
	Midlands	0.032	0.006	0.005	0.000	0.009	
	South	0.104	0.010	0.002	0.008	0.244	

Table 4 Vector autoregressive model coefficients and goodness-of-fit for regional birth rates and death rates, 1976-2008

The parameters for the VAR(1) model applied to the regional rates of net migration are presented in Table 5. For this model, the adjusted R^2 values are 0.88 for the North, 0.66 for the Midlands and 0.90 for the South. The only significant parameters were the constant terms for the North and Midlands. The predicted rates and corresponding predicted intervals are set out in Figure 9. The regional net migration rates are forecasted to remain relatively flat, with negative rates for the North, positive rates for the Midlands and near zero rates for the South.

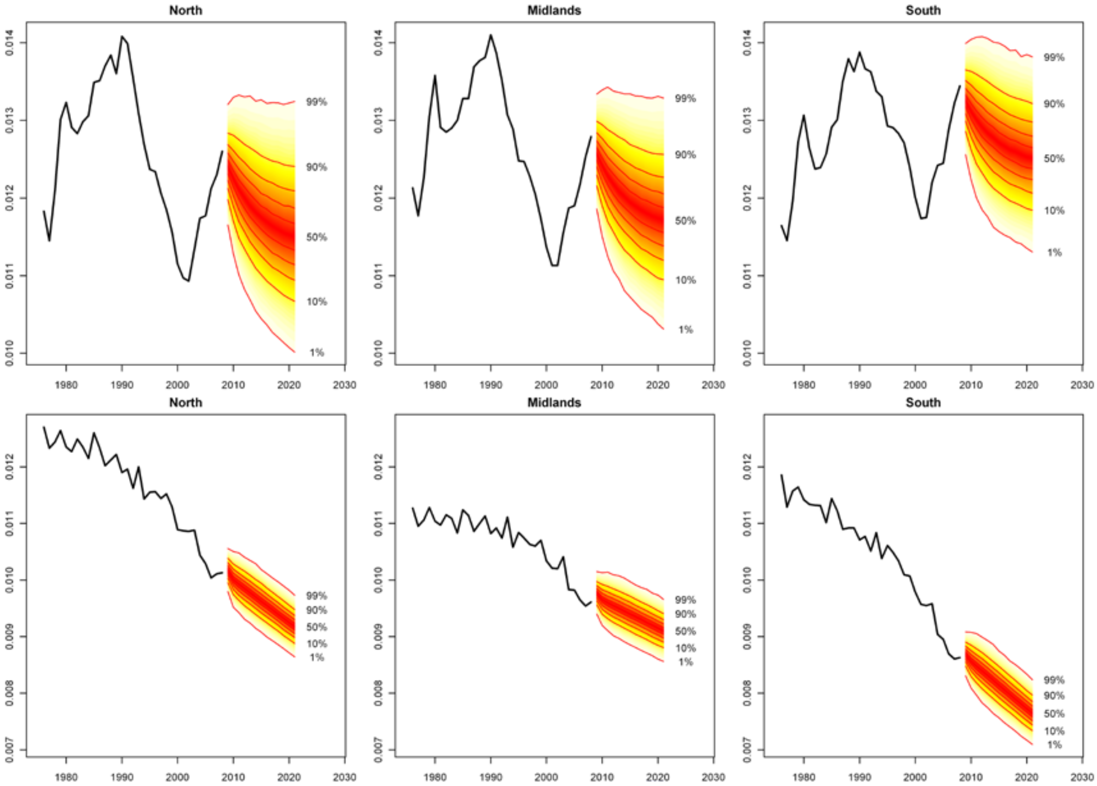


Figure 8 Vector autoregressive forecasts of crude birth (top) and death (bottom) rates for the North, Midlands and South regions, 2009-2021

		North	Midlands	South	Constant	Adj. R^2
Coefficient	North	-0.295	0.175	-1.344	-0.001	0.880
	Midlands	-2.039	-0.332	-3.171	0.000	0.655
	South	1.499	0.375	2.927	0.000	0.896
Std. Error	North	1.261	0.832	2.130	0.000	
	Midlands	1.666	1.100	2.815	0.000	
	South	0.955	0.630	1.612	0.000	
Pr(>|t|)	North	0.817	0.835	0.533	0.000	
	Midlands	0.231	0.765	0.269	0.941	
	South	0.128	0.557	0.080	0.000	

Table 5 Vector autoregressive model parameters for regional net migration rates, 1976-2008

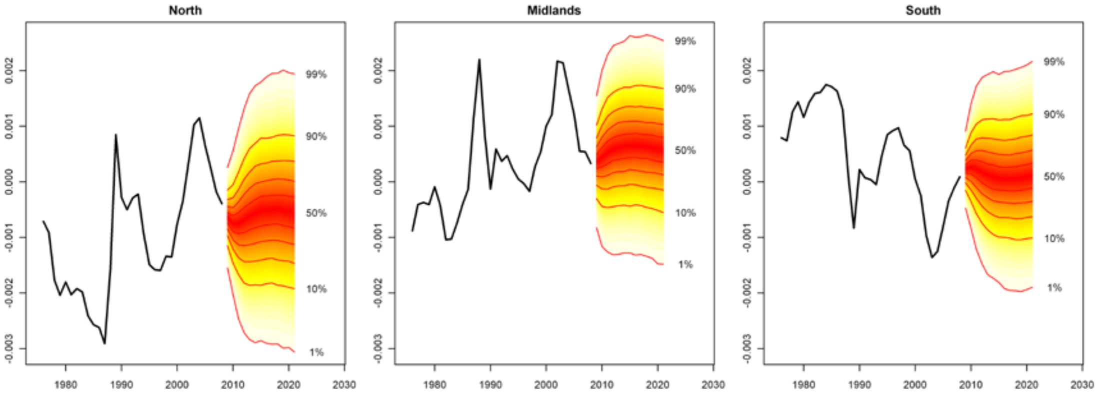


Figure 9 Vector autoregressive forecasts of crude net internal migration rates for the North, Midlands and South regions, 2009-2021

		In			Out					
		North	Mid-lands	South	North	Mid-lands	South	Con-stant	Trend	Adj. R2
Coefficient	In North	-7.713	-3.948	-13.070	7.660	4.150	13.010	0.007	0.000	0.931
	In Midlands	-9.380	-3.995	-15.100	8.790	5.156	14.050	0.010	0.000	0.818
	In South	-2.967	-1.576	-4.404	2.956	1.703	4.422	0.006	0.000	0.675
	Out North	-5.030	-2.762	-8.410	5.610	2.918	7.677	0.007	0.000	0.726
	Out Midlands	-7.563	-3.850	-12.980	7.435	4.934	11.430	0.011	0.000	0.767
	Out South	-5.691	-2.632	-8.629	5.093	2.790	9.345	0.005	0.000	0.878
Std. Error	In North	1.940	1.133	3.072	1.690	1.183	3.189	0.002	0.000	
	In Midlands	4.466	2.609	7.072	3.892	2.724	7.341	0.004	0.000	
	In South	1.853	1.082	2.935	1.615	1.130	3.046	0.001	0.000	
	Out North	2.381	1.391	3.771	2.075	1.452	3.914	0.002	0.000	
	Out Midlands	3.008	1.757	4.765	2.622	1.835	4.946	0.002	0.000	
	Out South	1.781	1.041	2.821	1.553	1.087	2.929	0.001	0.000	
Pr(>|t|)	In North	0.001	0.002	0.000	0.000	0.002	0.000	0.000	0.710	
	In Midlands	0.046	0.139	0.043	0.033	0.071	0.068	0.007	0.900	
	In South	0.122	0.158	0.147	0.080	0.145	0.160	0.001	0.472	
	Out North	0.045	0.059	0.035	0.012	0.056	0.062	0.001	0.724	
	Out Midlands	0.019	0.038	0.012	0.009	0.013	0.030	0.000	0.564	
	Out South	0.004	0.018	0.005	0.003	0.017	0.004	0.002	0.978	

Table 6 Vector autoregressive model parameters and correlation matrix of residuals for regional in-migration and out-migration rates, 1976-2008

Finally, the parameters for the VAR(1) models applied to the crude rates of in-migration and out-migration and destination-specific migration rates are set out in Table 6 and Table 7, respectively. The adjusted R^2 values are all relatively large, ranging from 0.63 (Midlands to South in Table 7) to 0.93 (in-migration to North in Table 6). For the in-migration and out-migration rate model, most parameters were significant, except those relating to in-migration to the South (only the constant was significant) and the trend parameters. The predicted rates and corresponding predicted intervals are presented in Figure 10. Note that the rates for the North and Midlands are considerably higher than for the South.

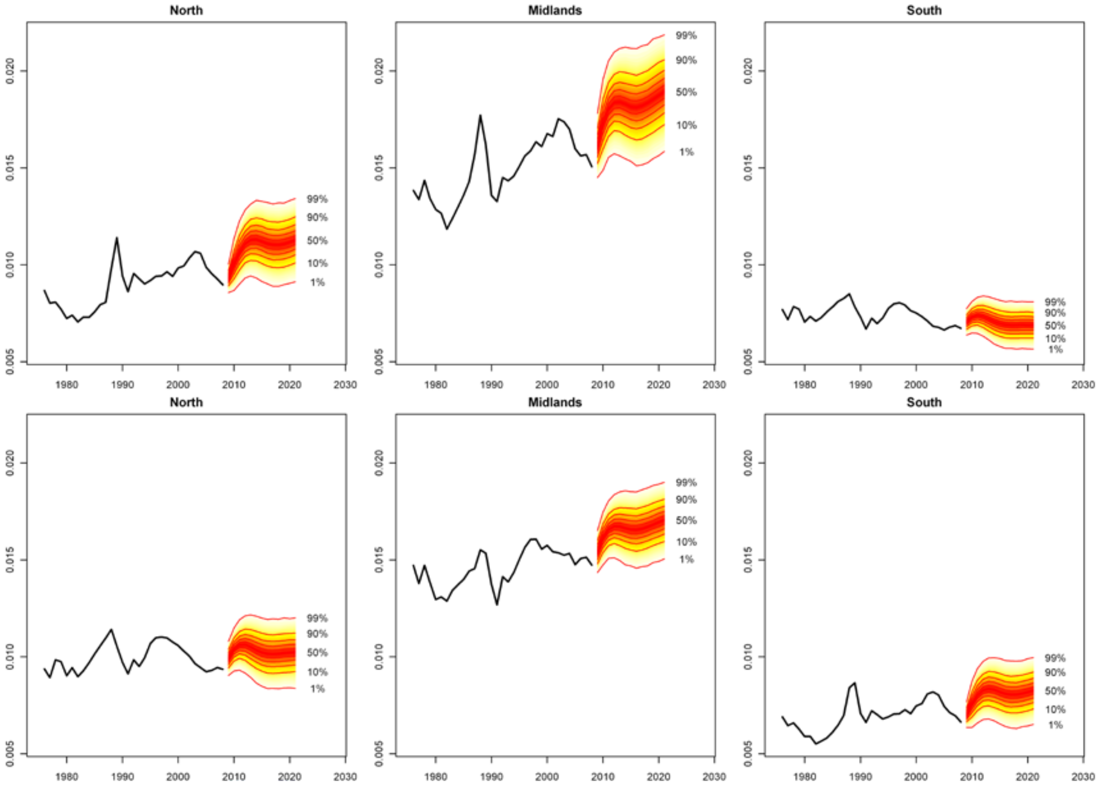


Figure 10 Vector autoregressive forecasts of crude in-migration (top) and out-migration rates (bottom) for the North, Midlands and South regions, 2009-2021

		N-M	N-S	M-N	M-S	S-N	S-M	Constant	Trend	Adj. R2
Coefficient	N-M	1.264	-0.010	-0.325	-0.060	0.068	-0.161	0.002	0.000	0.752
	N-S	1.098	0.697	-0.671	-0.064	0.040	-0.224	0.003	0.000	0.730
	M-N	1.483	-0.101	-0.404	-0.156	0.225	0.399	0.002	0.000	0.820
	M-S	1.294	-0.045	-0.387	0.404	-0.334	-0.195	0.005	0.000	0.629
	S-N	0.920	0.077	-0.858	-0.214	0.726	0.680	0.001	0.000	0.843
	S-M	0.625	-0.010	-0.752	0.087	0.070	0.854	0.001	0.000	0.808
Std. Error	N-M	0.305	0.099	0.276	0.135	0.244	0.160	0.001	0.000	
	N-S	0.542	0.177	0.491	0.240	0.434	0.285	0.001	0.000	
	M-N	0.379	0.124	0.344	0.168	0.304	0.199	0.001	0.000	
	M-S	0.729	0.238	0.661	0.323	0.584	0.383	0.001	0.000	
	S-N	0.308	0.100	0.279	0.136	0.247	0.162	0.001	0.000	
	S-M	0.388	0.127	0.352	0.172	0.311	0.204	0.001	0.000	
Pr(>|t|)	N-M	0.000	0.920	0.251	0.662	0.784	0.325	0.011	0.135	
	N-S	0.054	0.001	0.184	0.792	0.927	0.439	0.010	0.861	
	M-N	0.001	0.423	0.252	0.362	0.467	0.057	0.015	0.999	
	M-S	0.089	0.851	0.563	0.222	0.572	0.616	0.001	0.328	
	S-N	0.006	0.449	0.005	0.129	0.007	0.000	0.051	0.568	
	S-M	0.121	0.939	0.043	0.615	0.824	0.000	0.198	0.183	

Table 7 Vector autoregressive model parameters and correlation matrix of residuals for destination-specific out-migration rates, 1976-2008
Note: N = North, M = Midlands and S = South.

For the destination-specific out-migration rate model (Table 7), most of the parameters were not significant, except for the constant terms, the autocorrelation terms for North to Midlands, North to South, South to North and South to Midlands, and some of the cross-flow lag terms. The predicted rates and corresponding predicted intervals are set out in Figure 11. Note that the rates for the North to South and Midlands to South are the highest.

4.2 RESULTS

The forecasted rates of the demographic components described above were used as inputs into four separate regional populations for the purpose of identifying the key differences in both the forecasts and prediction intervals. The results are presented in Figure 12. As expected, the component and multiregional projection models resulted in narrower prediction intervals than the global and net migration models, as they include more information. Likewise, the widest intervals were consistently produced by the global projection model, which contained the least amount of information.

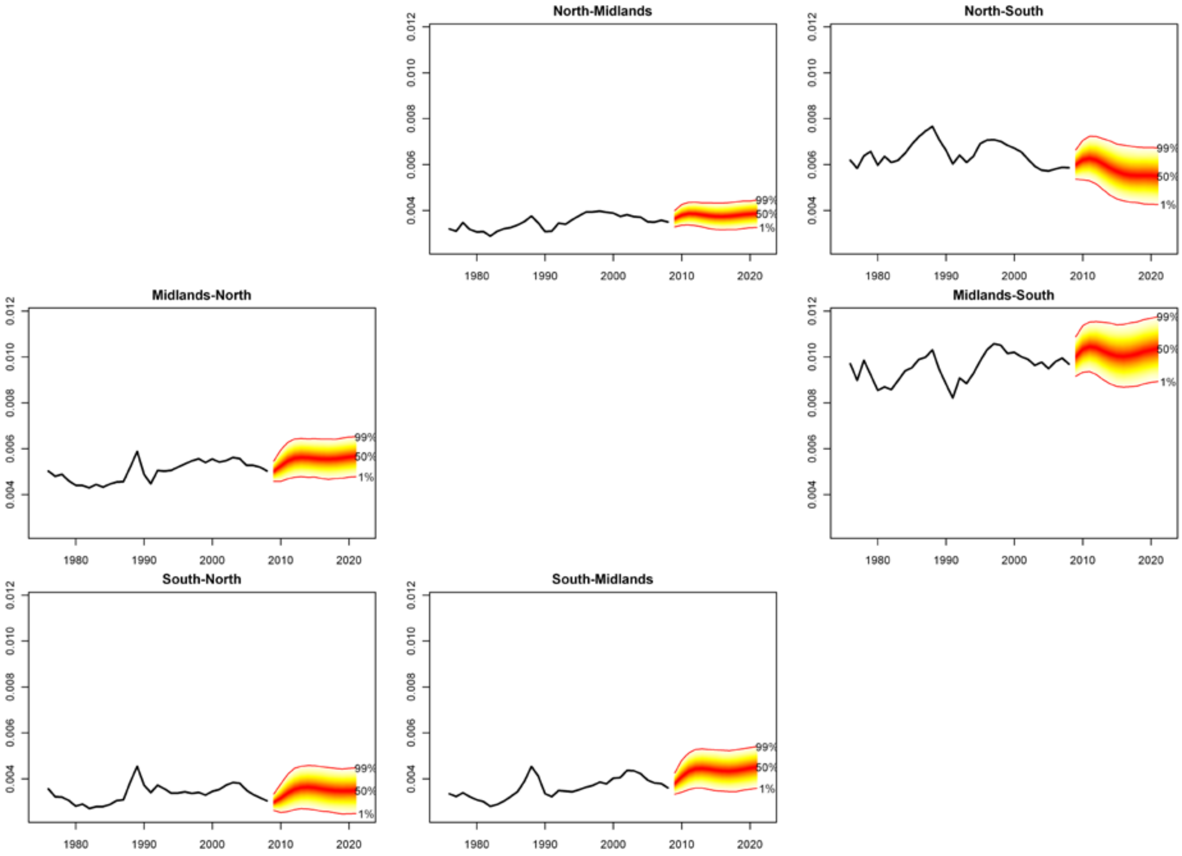


Figure 11 Vector autoregressive forecasts of crude destination-specific out-migration rates from the North, Midlands and South regions, 2009-2021

In terms of median forecasts for the year 2021, the North ranged from 14.94 million (net migration model) to 15.21 million (global model); the Midlands from 10.23 million (net migration model) to 10.37 million (global model); and the South from 28.14 million (component model) to 29.15 million (global model). These represent differences of 260 thousand, 140 thousand and one million, respectively, over a 13-year forecast period. The differences between the medians of the multiregional forecasts and the component forecasts were much smaller at -9 thousand for the North, 18 thousand for the Midlands and 23 thousand for the South. The multiregional prediction intervals were slightly narrower (+/- 3.9 percent) than the component prediction intervals (+/- 4.1 percent).

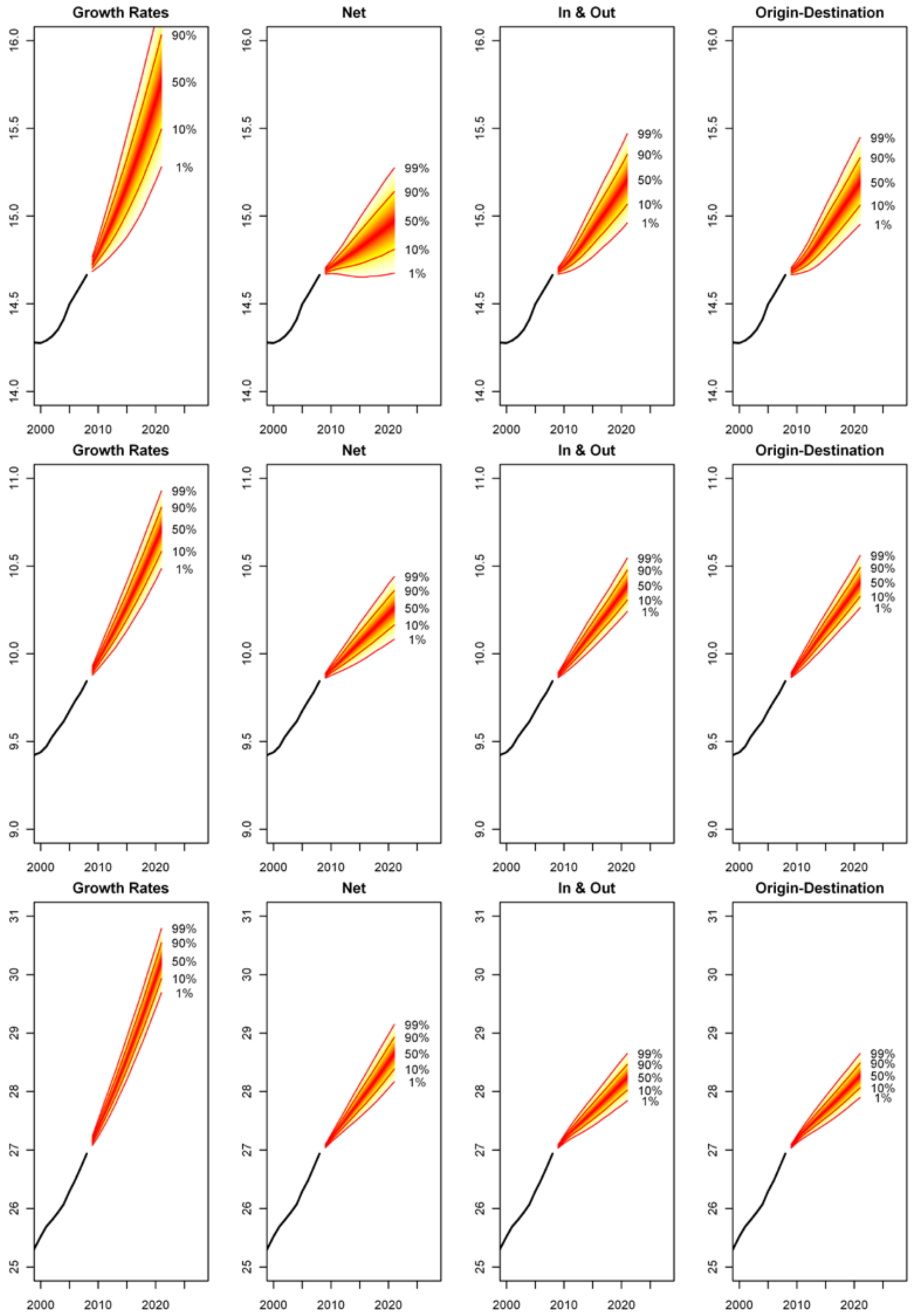


Figure 12 Four closed regional population forecasts (in thousands) for the North (top), Midlands (middle) and South (bottom) regions, 2009-2021

To summarise, the projection model specification clearly makes a difference in the results, even with a simple and relatively stable example, such as ours. We would expect the differences to be even larger if more regions were considered. For instance, if one were to model the population dynamics in the nine Government Office Regions in England, the multiregional model would contain 72 interregional migration flows to be modelled, whereas the in-migration and out-migration rate model would only contain 18 flows to be modelled. We prefer the more complex multiregional model with destination-specific out-migration rates because they are multiplied to the correct populations at risk of migrating. However, the VAR models utilised in this paper are not designed to handle large matrices of time series flows. To overcome this obstacle, one could disaggregate the flows into multiplicative components consisting of main effects and interaction terms (Sweeney and Konty 2002) and model just the time-varying components. This would make the number of series to model similar to the component projection model.

5. OPEN MULTIREGIONAL FORECASTS

In this section, we extend the multiregional population forecasts presented in the previous section to include international migration. As discussed in Section 2.3, there are three ways to include forecasts of international migration: (1) net international migration rates, (2) immigration and emigration rates and (3) immigration counts and emigration rates. We show the differences and consequences arising from these different assumptions of including international migration.

VAR(1) forecasts of the rates of net international migration, immigration and emigration are presented in Figure 13. We find that that the prediction intervals are wider for forecasts of net international migration than they are for immigration and emigration. In all three cases, the forecasts result in increased migration during the forecast period. The forecasts of immigration counts, used for the projection model specified in Equation 7, are presented in Figure 14. Here, we see that the levels of immigration are expected to increase substantially in the South region. Note, these results were obtained from a VAR(1) forecast of regional immigration counts and emigration rates.

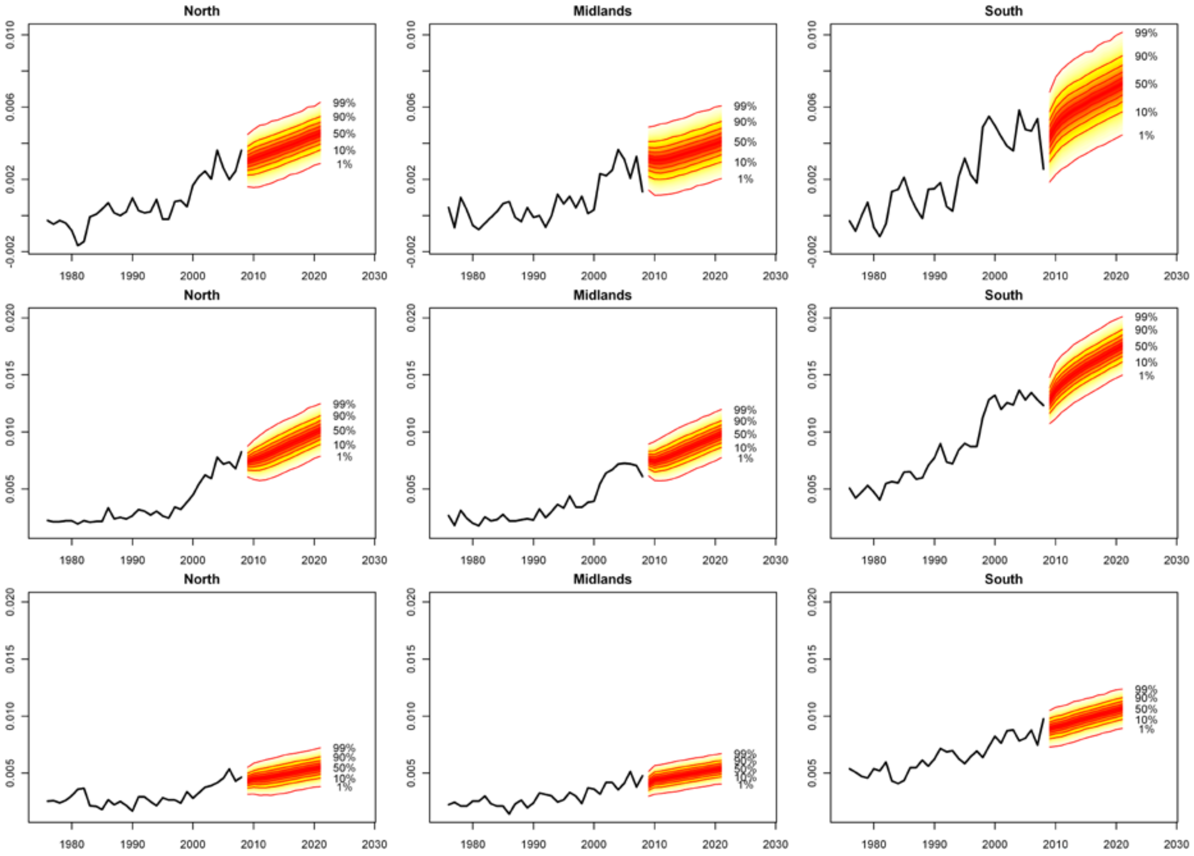


Figure 13 Vector autoregressive forecasts of crude rates of net international migration (top), immigration (middle) and emigration (bottom) for the North, Midlands and South regions, 2009-2021

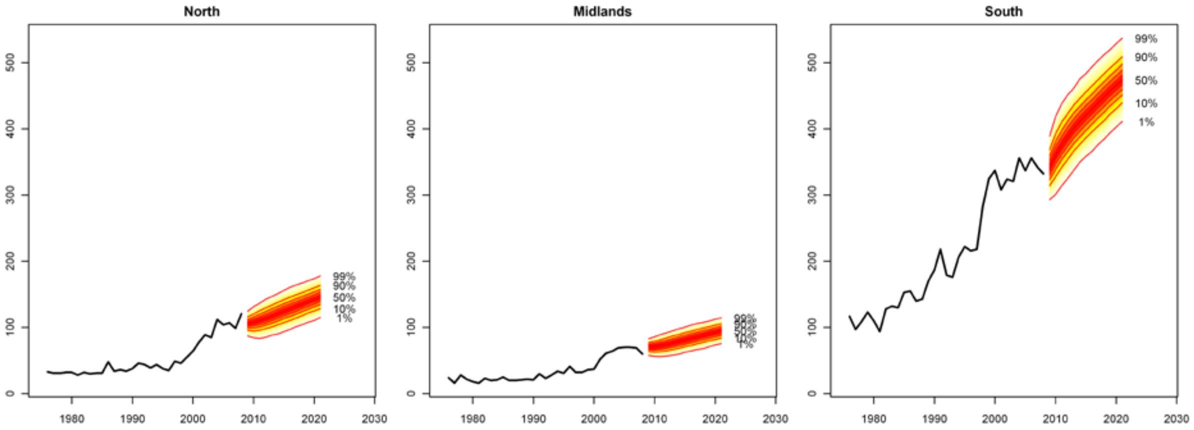


Figure 14 Vector autoregressive forecasts of immigration counts for the North, Midlands and South regions, 2009-2021

The results from integrating the three types of international migration components with the multiregional model are presented in Table 8. In comparison to the closed model, all regional populations are expected to be larger from the inclusion of international migration.

However, the model that includes forecasts of the counts of immigration resulted in the greatest increases in the median forecasts: 1.55 million for the North, 1.05 million for the Midlands and 4.95 million for the South. The differences between the net international migration model and the immigration and emigration rate model were relatively small, especially for the North and Midlands regions.

In terms of prediction intervals, the inclusion of international migration increases the uncertainty in comparison to the closed projection model, although there's not a large difference between the three models in relative terms. The model with immigration counts and emigration rates resulted in the widest relative prediction intervals for the North and Midlands, whereas the model with immigration rates and emigration rates resulted in the widest intervals for the South.

Region	Percentile	Closed	NIM Rates	IM and EM Rates	IM Counts and EM Rates
North	25	15.08	15.76	15.76	16.60
	50	15.15	15.84	15.85	16.70
	75	15.21	15.93	15.93	16.79
Midlands	25	10.32	10.76	10.78	11.36
	50	10.36	10.81	10.83	11.41
	75	10.40	10.87	10.88	11.47
South	25	28.07	30.07	29.96	32.97
	50	28.17	30.23	30.12	33.12
	75	28.27	30.40	30.28	33.28
Total	25	53.47	56.60	56.50	60.93
	50	53.67	56.89	56.80	61.23
	75	53.88	57.19	57.09	61.54

Table 8 Closed and open multiregional population forecasts (in thousands) for the North (top), Midlands (middle) and South (bottom) regions, 2009-2021

Note: NIM = Net international migration rate, IM = immigration and EM = emigration.

6. ADDITIONAL DISAGGREGATION, FUTURE RESEARCH AND CONCLUSIONS

In this paper, we have presented a number of population forecasts for regional populations in England. Its clear that the assumptions included in population projections matter, even for simple ones such as the ones presented in this paper. The ideas included in this paper can be extended to include age and sex in the projection framework. Here, one would need to consider the correlations or regularities in age patterns of demographic events, was well as across regions and over time. We hope to pursue this in future work.

Extending the approach used in this paper to include more regions, such as the nine Government Office Regions in England, let alone the nearly fifty counties, would require a different approach. The VAR models, as used in this paper, are not designed to handle so many different series. One idea would be to include some structure in the VAR models. Another would be to focus on modelling just the time-dependent structures in the migration flow tables, as Sweeney and Konty (2002) did for regions in California. By reducing the dimensionality of the migration flow tables, the modelling of the migration flow tables were greatly simplified. For example, a multiregional region with nine subpopulations requires 72 origin-destination-specific flows. If one were to just focus on the time-dependent parameters, then one would model just nine overall levels of migration, nine origin distributions and nine destination distributions (assuming that the three way interactions between origin, destination and time are mostly insignificant). This means that the time series modelling would focus on 27 time series instead of 72 time series.

The major contribution of this paper is its analysis of the influence of specification, particularly in regional population modelling exercises that include uncertainty. Aside from Gullickson (2001), Sweeney and Konty (2002) and Wilson and Bell (2007), very little work has been done in this area. We utilised multivariate time series models, VAR(1), to capture correlations over time and amongst regions in England. We then used these forecasts as inputs into various subnational projection models with the uncertainty coming directly from the time series of regional demographic components for England. In doing so, we reinforced the notion that specification is important for subnational projections (Rogers 1990), and showed that it also matters for measures of uncertainty. More work, however, needs to be done for specifying uncertainty for a greater number of regions and model selection. While we have demonstrated the existence of strong correlations amongst regional demographic components over time, capturing them for a larger number of regions is likely to be cumbersome. If age and sex disaggregations are introduced, then the use of “shrinking” and of parameterised model schedules can further reduce the dimensionality of the data (Rogers 1976, 1986; McNown and Rogers 1989; Knudsen et al. 1993; McNown et al. 1995). Further work could also be undertaken to consider a wider range of multivariate time series models and to incorporate model uncertainty (e.g., Abel et al. 2010a).

The future of producing population estimates will require more emphasis on specifying uncertainty so that more informed decisions can be made by population planners and policy makers. We hope our research contributes to this endeavour by illustrating the importance of choosing the projection model itself.

REFERENCES

Abel, G.J., Bijak, J., Forster, J.J., Raymer J. and Smith, P.W.F (2010a) What do Bayesian methods offer population forecasters? Centre for Population Change Working Paper 6/2010, ESRC Research Centre for Population Change, University of Southampton.

Abel, G.J., Bijak, J. and Raymer, J. (2010b) A comparison of official population projections with Bayesian time series forecasts for England and Wales. *Population Trends* 141(Autumn):95-114.

Ahlburg, D.A. (1987) Population Forecasts for South Pacific Nations using Autoregressive Models 1985 - 2000. *Journal of Population Research* 4(2):157-167.

Ahlburg, D.A. and Land, K.C eds. (1992) Population forecasting: Guest Editors' introduction. *International Journal of Forecasting* 8:289-299.

Alho, J. and Spencer, B. (2005) *Statistical demography and forecasting*. Springer, New York.

Bongaarts, J. and Bulatao, R.A. eds. (2000) *Beyond six billion: Forecasting the world's population*. National Academy Press, Washington DC.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting* 22(3):547–581.

Chatfield, C. (2004) *The analysis of time series: An introduction*, 6th edition. Chapman & Hall / CRC, Boca Raton.

Genz, A., Bretz, F., Miwa, T., Mi, X,. Leisch, F., Scheipl, F., Bornkamp, B. and Hothorn, T. (2010) mvtnorm: Multivariate Normal and t Distributions. R package, Version 0.9-94. URL: http://CRAN.R-project.org/package=mvtnorm

Genz, A. and Bretz, F. (2009) *Computation of multivariate normal and t probabilities*. Lecture Notes in Statistics Vol. 195, Springer-Verlage, Heidelberg.

Gullickson, A. (2001) Multiregional probabilistic forecasting. Unpublished Paper, Department of Demography, University of California, Berkeley. Available at: www.demog.berkeley.edu/~aarong/PAPERS/gullick_iiasa_stochmig.pdf.

Keilman, N. (2001) Data quality and accuracy of United Nations population projections, 1950–95. *Population Studies* 55:149-164.

Keilman, N. (2008) European demographic forecasts have not become more accurate over the past 25 years. *Population and Development Review* 34(1):137-153.

Knudsen, C., McNown, R. and Rogers, A. (1993) Forecasting fertility: An application of time series methods to parameterized model schedules. *Social Science Research* 22:1-23.

Land, K.C. and Rogers, A. eds. (1982) *Multidimensional mathematical demography*. Academic Press, New York.

Lee, R.D. and Tuljapurkar, S. (1994) Stochastic population projections for the United States: Beyond high, medium and low. *Journal of the American Statistical Association* 89(428):1175-1189.

Lütkepohl, H. (2005) *New introduction to multiple time series analysis*. Springer, Berlin.

Lutz, W. ed. (1996) *The future population of the world: What can we assume today?* Earthscan, London.

Lutz, W. and Goldstein, J.R. (2004) Introduction: How to deal with uncertainty in population forecasting? *International Statistical Review* 72(1):1-4.

McNown, R. and Rogers, A. (1989) Forecasting mortality: A parameterized time series approach. *Demography* 26(4):645-660.

McNown, R., Rogers, A. Little, J. (1995) Simplicity and complexity in extrapolative population forecasting models. *Mathematical Population Studies* 5(3):235-257.

Pfaff, B. (2008a) VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software* 27(4):1-32.

Pfaff, B. (2008b) *Analysis of integrated and cointegrated time series with R,* Second Edition. Springer, New York.

Pflaumer (1992) Forecasting US population totals with the Box-Jenkins approach. *International Journal of Forecasting* 8(3):329-338.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rogers, A. (1976) Shrinking large-scale population projection models by aggregation and decomposition. *Environment and Planning A* 8(6):515-541.

Rogers, A. (1986) Parameterized multistate population dynamics and projections. *Journal of the American Statistical Association* 81(393):48-61.

Rogers, A. (1990) Requiem for the net migrant. *Geographical Analysis* 22:283-300.

Rogers, A. (1995) *Multiregional demography: Principles, methods and extensions*. Wiley, Chichester.

Saboia, J.L.M. (1974) Modeling and Forecasting Populations by Time Series: The Swedish Case. *Demography* 11(3):483-492.

Schoen, R. (1988) *Modeling multigroup populations*. Plenum Press, New York.

Sweeney, S.H. and Konty, K.J. (2002) Population forecasting with nonstationary multiregional growth matrices. *Geographical Analysis* 34(4):289-309.

Wilson, T. and Bell, M (2007) Probabilistic regional population forecasts: The example of Queensland, Australia. *Geographical Analysis* 39:1-25.

ESRC Centre for Population Change
Building 58, Room 2001
Faculty of Social and Human Sciences
University of Southampton
SO17 1BJ

T: +44 (0)2380 592579
E: cpc@soton.ac.uk
www.cpc.ac.uk

To subscribe to the CPC newsletter and keep up-to-date with research activity, news and events, please register online: www.cpc.ac.uk/newsletter

You can also follow CPC on Twitter, Facebook and Mendeley for our latest research and updates:

www.facebook.com/CPCpopulation

www.twitter.com/CPCpopulation

www.mendeley.com/groups/3241781/centre-for-population-change

The ESRC Centre for Population Change (CPC) is a joint initiative between the University of Southampton and a consortium of Scottish universities including St Andrews, Edinburgh, Stirling and Strathclyde, in partnership with the Office for National Statistics and National Records of Scotland.