POPULATION, SPACE AND PLACE
*Popul. Space Place* **21**, 335–353 (2015)
Published online 24 March 2014 in Wiley Online Library
(wileyonlinelibrary.com) **DOI**: 10.1002/psp.1847

# New Evaluations of Simple Models for Small Area Population Forecasts

Tom Wilson*
*Queensland Centre for Population Research, School of Geography, Planning and Environmental Management, University of Queensland, Brisbane, Queensland, Australia*

## ABSTRACT

**At the small area scale simple methods for forecasting total populations are often employed because of a lack of data for cohort-component models, concerns about the reliability of these models for forecasting small population totals, and resource constraints. To date, a select number of authors have assessed the forecast accuracy of several individual, averaged, and composite models. This paper extends this stream of work by evaluating a large number of models on new datasets. The aims of the paper are to examine the performance of (a) 10 individual forecasting models (some of which are well known; others less so); (b) averages of every combination of 2, 3, 4, and 5 of the individual models (627 in total); and (c) composite models based on population size and growth rates (200,000 in total). Do averaged and composite models outperform individual models? Using new small area population datasets, forecasts from 2001 to 2031 were produced for three case study countries, Australia, New Zealand, and England & Wales. Both forecast accuracy and credibility (avoidance of negatives; degree of constraining to state populations) were assessed in 2011; for 2031, just credibility was evaluated. Of the individual models, constant share of growth (positive shares only) and constant share of population performed the best. A small proportion of averaged and composite models outperformed the best individual models in forecast accuracy. Several recommendations for the practice of small area population forecasting are made. Copyright © 2014 John Wiley & Sons, Ltd.**

*Correspondence to: Tom Wilson, Queensland Centre for Population Research, School of Geography, Planning and Environmental Management, University of Queensland, Brisbane, Queensland, Australia.
E-mail: tom.wilson@uq.edu.au

*Accepted 4 February 2014*

**Keywords:** small area; forecasts; averaging; composite; credibility; forecast accuracy

## INTRODUCTION

Population forecasts for small areas form valuable inputs to a wide variety of policy development and planning activities, including councils' local area plans, future power and water infrastructure assessments, educational enrolment and facility forecasting, health services planning, traffic forecasting, electoral redistributions, business site decisions, and market profitability assessments (Bell, 1997; Swanson *et al.*, 2010). In addition to their usefulness at a fine spatial scale, small area forecasts can be aggregated up to organisation-specific regional geographies such as government departments' service areas and commercial organisations' sales regions.

There is no universally agreed precise definition of a 'small area' in demographic studies (Smith & Morrison, 2005). Generally, however, small areas are considered to be spatial units at the detailed end of any geographical classification for which limited data exist. Typical examples include SA2 areas in Australia, area units in New Zealand, census tracts in the US, and census super output areas and wards in the UK, all of which usually have resident populations under 20,000. 'Small area' is strictly something of a misnomer as 'small' refers to population size and not geographical extent (which can be considerable in remote Australia).

Despite the extensive use and importance of small area forecasts, there exists little consensus on which is the 'best' small area model (or models). This situation contrasts with population forecasting at the national, State, and large regional scales where the cohort-component model, and especially its multiregional extension, is widely regarded as the 'gold standard' (Wilson & Rees, 2005). But at the small area scale the multiregional, or even single region, cohort-component model is less easily applied. The amount of data required to implement these models for large numbers of small areas is substantial and simply may not be available. Even in cases where data are accessible, data quality issues or high costs may prevent their use. And when reliable data *are* readily available, they will often prove too sparse to yield robust age schedules of rates. Although indirect estimation methods have been developed to tackle some of these data challenges (e.g. Assuncao *et al*., 2005; Raymer *et al*., 2013), the resource and time frame constraints faced by many practitioners mean they often cannot be implemented. Furthermore, several studies have shown that for forecasting small area population totals, cohort-component models do not appear to possess a clear advantage over simpler methods in terms of forecast accuracy (e.g. Smith & Sincich, 1992; Smith, 1997).

There is one small area forecasting model, however, which is commonly used under certain circumstances. For fast growing urban areas, especially those experiencing major residential development projects, the housing unit method is the obvious choice (e.g. Foss, 2002). But it does depend on reliable dwelling forecasts being available for the entire forecast horizon. In many parts of the world such data simply do not exist or are of questionable quality. So apart from situations where the housing unit method is appropriate and its application feasible, agreement on the 'best' model for small area population forecasting is lacking, and there is relatively little guidance from the literature on model choice.

Why is this? Part of the explanation is probably due to considerable variation in the types, coverage and quality of population data, which limits the choice of models for any particular data situation. The heterogeneity of small area population dynamics probably also contributes so that different models work well under different demographic environments. A further part the explanation almost certainly lies in the limited amount of research dedicated to small area demographic forecasting (summarised in the following section).

This paper makes an attempt to expand the evidence base for selecting small area population forecasting models. It does so through both retrospective and prospective empirical tests. The retrospective tests consisted of 'forecasting' from the past up to a recent date and assessing the success (or otherwise) of those forecasts in predicting population estimates for that date. Although forecasting success in the past does not, of course, guarantee similar success in the future, it nonetheless provides a useful guide. Previous research has observed a reasonable degree of temporal stability in population forecast errors (e.g. Smith & Sincich, 1988; Smith & Rayer, 2011; Wilson & Rowe, 2011). In addition, a demonstration of past forecasting success is appreciated by projection model clients as it proves that, at the very least, selected models work under some circumstances, rather than under no circumstances! The prospective testing involved continuing the forecasts beyond the present to ascertain their characteristics in the long-run forecast horizons.

This study makes use of new small area population datasets for Australia (ABS, 2013), England & Wales (Norman, 2013), and New Zealand (Statistics New Zealand, 2012), countries for which small area population models have rarely been tested. Most previous evaluations have been undertaken with US data, and there is no guarantee those findings will be relevant for other countries. All three datasets were carefully compiled to ensure that all population estimates were based on temporally fixed sets of geographical boundaries (Wilson & Rees, 1999).

Forecasts were evaluated in terms of both forecast accuracy and credibility, with credible defined as the avoidance of (i) negative projected populations and (ii) significant constraining to sum to independent State or national populations. Population forecasts can also be evaluated on many other criteria, such as the amount of output detail, conceptual adequacy, production costs, ease of validation, ease of scenario creation, political acceptability, and timeliness, all of which are important (Smith *et al*., 2001). Nonetheless, forecast accuracy and credibility are especially important and are the focus of this contribution.

A particular emphasis of the paper is a test of the theory that combining forecasts can improve accuracy (Ahlburg, 1995; Armstrong, 2001). Goodwin (2009 p 34) argues that 'when several methods are combined, there is a likelihood that biases in different directions will counteract each other, thereby improving accuracy'. The types of



**DOI**: 10.1002/psp

combined forecasts evaluated here consist of *averaged models*, calculated as simple averages of two or more individual models, and *composite models*, which comprise different models for different types of areas (Isserman, 1977). Specifically, the paper addresses the following questions:

(1) Which individual models provide the most accurate and credible forecasts of total population?
(2) Do averaged models improve performance beyond that of the best individual model, and if so, which are the best?
(3) Do composite models improve performance beyond that of the best individual model, and if so, which are the best?
(4) What recommendations can be made about model choice for small area population forecasting?

The paper continues as follows. The next section provides a brief overview of previous work on the evaluation of small area forecasting models. The population estimate datasets, projection models, and assessment measures are described in the following 'Data and Methods' section. Results of the forecast error and credibility evaluations are presented in the subsequent section, followed by a discussion. The conclusions include recommendations for small area population forecasting.

## PREVIOUS RESEARCH

The existing literature on the evaluation of models for forecasting small area populations is quite limited. Key papers include Isserman (1977), Openshaw and van der Knapp (1983), Smith (1987), Smith and Shahidullah (1995), Rayer (2008), Chi and Voss (2011), and Rayer and Smith (2010). Additional studies have assessed forecasting models at higher levels of geography (e.g. White, 1954; Smith & Sincich, 1992). Primarily, analyses have focused on the errors of simple models for total populations by using retrospective forecasts. Models evaluated often include extrapolative types such as linear and exponential extrapolation, and share models that distribute an independent forecast of population, or population growth, for a larger 'parent' region to small areas. These retrospective projections start at some point in the past and 'forecast' out to the present where the 'forecasts' can be compared with actual population estimates. Although the performance of models in the past is no guarantee of their performance in the future, a number of studies have found a fair degree of temporal stability in forecast accuracy (e.g. Rayer, 2008).

Which models have proved most accurate according to these papers? In a retrospective test of 10 models which forecast subcounty populations out a decade, Isserman (1977) discovered that several gave very similar levels of error. Smith and Shahidullah (1995) projected census tract populations out 10 years using four models, finding the lowest errors were generated by a share of growth model. Applying 66 individual models to Dutch municipalities, Openshaw and van der Knapp (1983) concluded that 'the best forecasts are likely to be provided by either a Holt-Winters model or a ratio correction model [share of growth] or a low order exponential smoothing model' (p. 301). The forecast accuracy of five models applied to county populations by Rayer (2008) revealed linear extrapolation to be the most accurate. Analysing subcounty areas in Florida, Rayer and Smith (2010) also discovered linear extrapolation was the most accurate, followed by constant size and share of growth models.

The share of growth model, found to produce relatively accurate projections by several researchers, unfortunately possesses an undesirable characteristic (Smith *et al.*, 2001; Hachadoorian *et al.*, 2011). It is revealed when a local area experiences population change in the base period which is of a different sign to that of the State, and the State is then forecast to experience a change of trend. For example, a local area declines in the base period whilst the State grows modestly (giving a negative share); the State is then forecast to grow rapidly in population. In such circumstances the local area's population would be projected to decline even more than during the base period, which seems unlikely. Negative projections can result. Because of this problem, White (1954) adapted the model so that shares of growth are calculated only for areas which increased in population over the base period, with areas experiencing negative growth allocated a share of zero.

In addition to individual models, a few authors have examined the performance of averaged and composite models. In most cases the findings are that averaged and composite models are about



**DOI**: 10.1002/psp

as accurate, or marginally more accurate, than the best individual model (Isserman, 1977; Rayer, 2008; Rayer & Smith, 2010). For example, Rayer (2008) found that average or trimmed average models (average excluding the highest and lowest) were about as accurate as the best individual model. He also determined that two composite models performed slightly better than individual models, specifically (i) an Exponential model (EXP) for rapidly declining areas, a Linear model (LIN) for moderately changing areas, and a Constant Share of Population (CSP) for rapidly growing areas; and (ii) an EXP model for rapidly declining areas and a LIN model for moderately declining or growing areas. Similarly positive findings on averaged and composite models have been found in other fields, such as marketing (Armstrong, 2001), the forecasting of economic indicators (e.g. Kapetanios *et al*., 2008), and election forecasting (e.g. Graefe *et al*., 2014). Despite these encouraging findings, averaged and composite models have rarely been used in small area population modelling, with the Florida Bureau of Economic and Business Research's trimmed average approach being one of the few applications (Smith & Rayer, 2012).

There are a number of directions in which this small body of research could be extended. First, the number of averaged and composite models tested to date has been quite limited, so there is scope to evaluate many more combinations of individual models. Second, the share-of-growth model has provided relatively good results in a number of studies, but the shortcoming mentioned previously is unfortunate. Are there other variations of the share-of-growth model which would work as well but avoid the problems? Third, few studies have evaluated the performance of small area forecasting models outside the US. Fourth, existing studies have measured the performance of models almost exclusively in terms of the average forecast accuracy across all areas. For many users, average forecast accuracy is an important criterion on which to select a model, but other criteria are also significant. It would also be useful to assess the credibility of projection models in terms of characteristics such as the frequency of generating negative populations and runaway growth. For those situations where small area projections have to be constrained to regional, state, or national projections, the greater the constraining, the less the constrained small area projections resemble the original values. This paper attempts to contribute to these four areas.

## DATA AND METHODS

### Population Estimates

The study datasets comprised of mid-year population estimate totals for small areas in Australia, England & Wales, and New Zealand in 1991, 2001, and 2011 on temporally consistent sets of boundaries. The Australian small areas are SA2 areas for which the Australian Bureau of Statistics has recently released total population estimates on 2011 boundaries back to 1991 (ABS, 2013). Population data for New Zealand area units on a consistent set of boundaries were supplied directly by Statistics New Zealand (Statistics New Zealand, 2012). Population estimates for Census Area Statistics wards in England & Wales on 2001 boundaries were created by Paul Norman of Leeds University (Norman, 2013) based on the methods described in Norman *et al*. (2003; 2008). Table 1 summarises some of the characteristics of these small areas.

### Individual Models

Only simple total population models which could be easily fitted in an Excel spreadsheet were selected for assessment. Models requiring fitting to lengthy annual time series and/or considerable additional socio-economic data were not considered because such data are often unavailable. Examples of these include autoregressive integrated moving average (e.g. Ahlburg, 1987), regression (e.g. Chi & Voss, 2011), land-use/housing-unit (e.g. Bell *et al*., 2000), exponential smoothing (e.g. Walters & Cai, 2008), all-age component (e.g. Rees 1990), employment-led (e.g. BEA, 1996), and microsimulation models (e.g. Vidyattama & Tanton, 2010). Ten individual models were selected for analysis on the basis of common usage, desirable characteristics, or success in previous forecasting assessment studies. Many of these models are common to the literature though, to the best of the author's knowledge, the Variable Share of Growth model (VSG) is presented for the first time. The first four models produce forecasts which are a function solely of each local area's past population trends, whereas the following six are linked to an independent projection for



**DOI**: 10.1002/psp

Table 1. Summary statistics on the small areas.

| | Australia | England & Wales | New Zealand |
|---|---|---|---|
| Type of area | SA2 area | CAS ward | Area unit |
| No. of areas | 2072 | 8839 | 1725 |
| Median population, 2001 | 7704 | 4842 | 2110 |
| *No. of areas by 2001 total population* | | | |
| 0–999 | 17 | 45 | 470 |
| 1,000–1999 | 34 | 994 | 348 |
| 2,000–4999 | 528 | 3555 | 828 |
| 5,000–9999 | 745 | 2729 | 78 |
| 10,000–14,999 | 378 | 1263 | 1 |
| 15,000+ | 370 | 253 | 0 |
| Median annual average 1991–2001 growth rate (%) | 0.59 | 0.24 | 0.58 |
| *No. of areas by annual average 1991–2001 growth rate* | | | |
| <−0.5% | 371 | 1336 | 332 |
| −0.5% to 0.5% | 636 | 4232 | 494 |
| 0.5–1.5% | 393 | 2880 | 394 |
| 1.5–2.5% | 208 | 591 | 238 |
| >2.5% | 464 | 400 | 267 |

*Source:* calculated from Australian Bureau of Statistics, Statistics New Zealand and Office for National Statistics data
CAS, Census Area Statistics.

the State, country, or some other encompassing 'parent' region. The models are summarised in Table 2 and described briefly as follows.

(1) The simple LIN model can be expressed as

$$P_i(t+1) = P_i(t) + G_i$$

where $P_i(t)$ is the jump-off year population of small area $i$, $G_i$ is the annual average population growth over the base period, and $P_i(t+1)$ is the forecast population 1 year ahead. For growing areas the LIN projects a gradually declining growth rate, preventing runaway growth. However, for rapidly declining areas it can potentially result in projections of negative populations.

(2) The well-known EXP model forecasts population as

$$P_i(t+1) = P_i(t)\, e^{r_i}$$

where $r_i$ is the annual average population growth rate of small area $i$ over the base period. For declining areas the EXP model possesses the helpful property of gradually reducing projected decline. For rapidly growing areas, however, there exists the risk of projecting runaway growth.

(3) To avoid the extreme forecasts of the LIN and EXP models it is possible to create what is effectively a basic composite model: linear extrapolation is used if the base period population change is positive, whereas the EXP model is applied if it is negative. This is described here as the *Linear/Exponential* model (LIN/EXP). It is similar to one of the composite models found to be quite accurate by Rayer (2008).

(4) Alternatively, the EXP model can be modified so that as population increases the growth rate is dampened, similar to a logistic function. A *Modified Exponential* model (MEX) can be formulated (Baker *et al.*, 2008):

$$P_i(t+1) = P_i(t)\, e^{[r_i(1-P_i(t)/K_i)]}$$

where $K_i$ is the upper limit to population for small area $i$. Ideally, this limit would be based on a maximum population density based on historical experience and/or land use zoning regulations. This information was not readily available for all the case study countries so $K_i$ was set to five times the jump-off year population. Although to some extent arbitrary, this was selected as a compromise between the prevention of runaway growth and excessive dampening of growth for fast-growing areas. For declining areas a modification is necessary to ensure that



**DOI**: 10.1002/psp

Table 2. Summary of the 10 individual projection models.

| Label | Name | Equation |
|---|---|---|
| *Models based solely on local area population trends* | | |
| LIN | Linear | $P_i(t+1)=P_i(t)+G_i$ |
| EXP | Exponential | $P_i(t+1)=P_i(t)\,e^{r_i}$ |
| LIN/EXP | Linear/Exponential<br>上一个period + annual growth<br>上一个period * exp(annual growth rate) | Linear model if base period growth positive:<br>$P_i(t+1)=P_i(t)+G_i$<br>Exponential model if base period growth negative:<br>$P_i(t+1)=P_i(t)\,e^{r_i}$ |
| MEX | Modified Exponential | If base period growth positive: base period rate (ri) 在CA列<br>$P_i(t+1)=P_i(t)\,e^{[r_i(1-P_i(t)/K_i)]}$ C-AF列是上一年<br>If base period growth negative:<br>$P_i(t+1)=P_i(t)\,e^{[r_i(1-K_i/P_i(t))]}$ CL列是Ki |
| *Models linked to an independent projection for the State (or other higher geography)* | | |
| CGD | Constant Growth Rate Difference | $P_i(t+1)=P_i(t)\,e^{(r_{State}(t,t+1)+grd_i)}$ |
| CSP | Constant Share of Population | $P_i(t+1)=P_{State}(t+1)\ SHAREPOP_i(t)$ 上一年 * share of pop |
| FSP | Forecast Share of Population | $P_i(t+1)=P_{State}(t+1)\ SHAREPOP_i(t+1)$ |
| CSG | Constant Share of Growth | $P_i(t+1)=P_i(t)+SHAREGROWTH_i\ G_{State}(t,t+1)$ |
| CSG+ | Constant Share of Growth with + shares only | $P_i(t+1)=P_i(t)+SHAREGROWTH_i\ G_{State}(t,t+1)$ |
| VSG | Variable Share of Growth<br>Cell6222开始是output<br>Cell6216是pos/neg factor<br>Cell6开始是上一年数据 | If base period growth positive: $P_i(t+1)=P_i(t)+G_i(t,t+1)\ POSFACTOR_i(t,t+1)$<br>If base period growth negative: $P_i(t+1)=P_i(t)+G_i(t,t+1)\ NEGFACTOR_i(t,t+1)$.<br>整体计算出现在Cell4148 |

Notation:
$P_i(t)$ jump-off year population of small area $i$
$P_i(t+1)$ projected population of small area $i$ at time $t+1$
$G_i$ annual average population growth over the base period
$r_i$ annual average population growth rate of small area $i$ over the base period
$K_i$ limit of population for small area $i$
$grd_i$ base period growth rate difference, defined as the annual average growth rate of the small area over the base period minus that of the State: $grd_i=r_i(t-10,t)-r_{State}(t-10,t)$
$SHAREPOP_i(t)$ share of State population in small area $i$ at jump-off year $t$
$SHAREGROWTH_i$ small area's share of State population growth in the base period
$G_{State}$ forecast State population growth
$POSFACTOR_i(t,t+1)$ plus–minus adjustment factor for positive growth
$NEGFACTOR_i(t,t+1)$ plus–minus adjustment factor for negative growth

rates of population change become less negative as a lower boundary is approached. For these areas $K_i$ denotes a lower limit, and the projection equation is

$$P_i(t+1)=P_i(t)\,e^{[r_i(1-K_i/P_i(t))]}.$$

In this paper the lower limit is set to one-fifth of the jump-off population. Clearly the need to determine $K_i$ exogenously is a limitation of the MEX model.

(5) The *Constant Growth Rate Difference* model (CGD) assumes that the base period difference between the population growth rate of the small area and the State (or national population) remains constant into the future. A conceptually attractive feature is the combination of State-level factors through the State's forecast growth rate and local factors in the form of growth rate differences. The growth rate difference model shown below originates from Davis (1995) but is rewritten in order to bring out the connection to the EXP. Populations are forecast as

$$P_i(t+1)=P_i(t)\,e^{(r_{State}(t,t+1)+grd_i)}$$

where $r_{State}(t,t+1)$ refers to the forecast growth rate of the State in the period $t$ to $t+1$, and $grd_i$ is the base period growth rate difference, defined as the annual average growth rate of the small area over the base period minus that of the State:

$$grd_i = r_i(t-10,t) - r_{State}(t-10,t).$$

(6) The *CSP* model assumes each small area maintains its share of the State's population. It is calculated as

$$P_i(t+1) = P_{State}(t+1)\ \ SHAREPOP_i(t)$$

where $SHAREPOP_i(t)$ is the share of State population in small area *i* at jump-off year *t*. Implicit in this model is the assumption that all small areas experience the same future growth rates as the State. Because the State's population is being shared between small areas, the CSP model by definition creates small area forecasts which sum to that of the State.

(7) The *Forecast Share of Population* model (FSP) has the same equation as the CSP model, but the share of population is forecast to change over time. In this application the shares were forecast by linear extrapolation. A weakness of this model is its ability to forecast negative shares and therefore, negative populations, in some circumstances. Whether this occurs in typical applications is demonstrated in the results section.

(8) The *Constant Share of Growth* model (CSG) distributes forecast State population growth to small areas on the basis of their share of growth in the base period. It is defined as

$$P_i(t+1) = P_i(t) + SHAREGROWTH_i\ \ G_{State}(t,t+1)$$

where $SHAREGROWTH_i$ is the small area's share of State population growth in the base period and $G_{State}$ is forecast State population growth. Like the CSP model, this model requires an independent forecast for the State and generates small area forecasts, which automatically sum to that of the State.

(9) As noted earlier, the CSG model can generate implausible forecasts when a small area experiences population change in the base period, which is of a different sign to that of the State. Therefore White's (1954) adaptation of the CSG model is used. It employs the same equation as the CSG model previously but the base period shares of growth are calculated only for areas which increased in population over the base period. $SHAREGROWTH_i$ is thus the small area's base period growth divided by the total growth of only those areas experiencing growth in the base period. Areas that experienced base period decline are allocated a share of growth of zero in the projections. This is described as the *CSG with Positive Shares only* model (CSG+). A limitation of this model is its unsuitability in situations of widespread population decline.

(10) A second approach to overcoming the limitations of the CSG model is termed here the *VSG* model. Unlike CSG+, this model permits areas to decline in population even if the State is projected to grow. It produces initial forecasts of growth, $G_i(t,t+1)$, using the LIN/EXP model, which are then modified with the plus–minus method (Shryock & Siegel, 1973) so that they sum to the forecast growth of the State. If base period population growth is positive then forecasts are calculated as

$$P_i(t+1) = P_i(t) + G_i(t,t+1)\ POSFACTOR_i(t,t+1)$$

whereas if it is negative, they are calculated as

$$P_i(t+1) = P_i(t) + G_i(t,t+1)\ NEGFACTOR_i(t,t+1).$$

*POSFACTOR* and *NEGFACTOR* are plus–minus method variables defined as follows:

$$POSFACTOR_i(t,t+1) = \frac{\sum_i |G_i(t,t+1)| + (G_{State}(t,t+1) - \sum_i G_i(t,t+1))}{\sum_i |G_i(t,t+1)|}$$

$$NEGFACTOR_i(t,t+1) = \frac{\sum_i |G_i(t,t+1)| - (G_{State}(t,t+1) - \sum_i G_i(t,t+1))}{\sum_i |G_i(t,t+1)|}.$$

All model parameters were calculated from population change between 1991 and 2001. Then two sets of forecasts were created. The *forecast-constrained* set adjusted all small area forecasts,



**DOI**: 10.1002/psp

so that they summed to the official principal series State or national population forecasts available at around the time of the jump-off year (2001). Constraining in this manner is common to many sets of State and local government population forecasts. Note that the last five models listed in Table 2 are by definition automatically constrained to State or national forecasts, so constraining applies only to LIN, EXP, LIN/EXP, MEX, and CGD models. The *estimate-constrained* set removed the effect of incorrect official population forecasts by constraining to 2011 population estimates, thus revealing the degree of distributional error. Constraining was applied by multiplying each small area's forecast by the same scaling factor, calculated as the official State (Australia) or national (England & Wales; New Zealand) forecast divided by the unconstrained forecasts summed across all small areas in the State/country.

### Averaged and Composite Forecasting Models

An inductive approach was taken to the assessment of averaged and composite models. Averaged models were calculated as the mean of 2, 3, 4, and 5 of every combination of the 10 individual models, giving 627 averaged models in total.[1] Forecasts were averaged first, and then constrained to State (Australia) or national (New Zealand; England & Wales) totals. Both estimate-constrained and forecast-constrained sets of forecasts were generated.

Composite forecasting models were created by using different models for different categories of base period growth and jump-off year population size. Five categories were distinguished for both growth rate and population size; these are shown in Tables 4 and 5. The combination of 10 individual models and 5 categories resulted in 100,000 different composite models being created for both growth rate and population size composites.[2] As before, forecasts were constrained to State or national totals, with both estimate-constrained and forecast-constrained sets being created.

### Forecast Error and Credibility Assessment

Error was assessed by the Median Absolute Percentage Error (MedAPE), the median value of the distribution of Absolute Percentage Error (APE),

$$APE^i = \frac{|F^i - A^i|}{A^i} 100$$

where $F$ denotes the forecast population and $A$ the actual value. MedAPE was chosen in preference to Mean Absolute Percentage Error because of the tendency for APE distributions to include outliers, which inflate its value (Tayman and Swanson 1999). Forecast error was assessed for all small areas, and then by category of base period growth rate and jump-off year population size.

The ability of a model to give low average error values is undoubtedly welcome, but if it generates implausible or, worse, non-sensical forecasts for a proportion of small areas then its value is severely compromised. Whilst manual intervention or rules automated in a spreadsheet or projection code can be implemented to fix up undesirable forecasts, it is preferable to use a model that has no need for such interventions. The forecast-constrained set was therefore extended 20 years beyond 2011 to examine the longer term credibility of the models being evaluated. (It would have been useful to assess forecast accuracy for longer horizons as well but unfortunately the necessary time series of small area population data on a consistent geography was unavailable for the three case study countries.) Credibility was assessed by the following measures:

(1) the proportion of small areas with negative populations; and
(2) the ratio of the sum of the unconstrained small area projections to the national projection.

Note that ratios can only deviate from 1.0 for LIN, EXP, LIN/EXP, MEX, and CGD; the remaining models, CSP, FSP, CSG, CSG+, and VSG, are by definition automatically constrained to State or national population forecasts.

## RESULTS

### Individual Models

Table 3 presents a summary of the individual models' performance at forecast horizons of 10 and 30 years. The most accurate forecasts after 10 years, when constrained to official population forecasts, were given by CSG+ for Australia, CSP for England & Wales, and CSG+ for New Zealand. For Australia, the CSP model also



**DOI**: 10.1002/psp

Table 3. Performance of individual models at 10-year and 30-year forecast horizons.

| | 10-year horizon | | | | | | 30-year horizon | |
|---|---|---|---|---|---|---|---|---|
| | Forecast-constrained | | | Estimate-constrained | | | Forecast-constrained | |
| Model | MedAPE | %-ve | Ratio | MedAPE | %-ve | Ratio | %-ve | Ratio |
| *Australia* | | | | | | | | |
| LIN | 8.5 | 0.00 | 0.99 | 7.9 | 0.00 | 0.95 | 0.39 | 1.02 |
| EXP | 21.1 | 0.00 | 1.40 | 18.4 | 0.00 | 1.34 | 0.00 | 1424.64 |
| LIN/EXP | 8.3 | 0.00 | 1.00 | 7.7 | 0.00 | 0.95 | 0.00 | 1.03 |
| MEX | 10.5 | 0.00 | 1.05 | 8.5 | 0.00 | 1.00 | 0.00 | 1.12 |
| CGD | 22.1 | 0.00 | 1.41 | 19.2 | 0.00 | 1.41 | 0.00 | 1273.51 |
| CSP | 7.9 | 0.00 | 1.00 | 10.0 | 0.00 | 1.00 | 0.00 | 1.00 |
| FSP | 10.5 | 0.05 | 1.00 | 9.9 | 0.05 | 1.00 | 3.57 | 1.00 |
| CSG | 8.7 | 0.00 | 1.00 | 12.2 | 0.34 | 1.00 | 0.29 | 1.00 |
| CSG+ | 7.1 | 0.00 | 1.00 | 7.6 | 0.00 | 1.00 | 0.00 | 1.00 |
| VSG | 8.3 | 0.00 | 1.00 | 8.5 | 0.00 | 1.00 | 0.00 | 1.00 |
| *England & Wales* | | | | | | | | |
| LIN | 6.7 | 0.01 | 1.00 | 6.7 | 0.01 | 0.96 | 0.19 | 1.00 |
| EXP | 7.5 | 0.00 | 1.02 | 6.7 | 0.00 | 0.98 | 0.00 | 2.19 |
| LIN/EXP | 6.6 | 0.00 | 1.00 | 6.6 | 0.00 | 0.96 | 0.00 | 1.01 |
| MEX | 6.3 | 0.00 | 1.00 | 6.0 | 0.00 | 0.97 | 0.00 | 1.02 |
| CGD | 7.5 | 0.00 | 1.02 | 6.7 | 0.00 | 1.02 | 0.00 | 2.19 |
| CSP | 4.6 | 0.00 | 1.00 | 5.5 | 0.00 | 1.00 | 0.00 | 1.00 |
| FSP | 6.9 | 0.01 | 1.00 | 6.9 | 0.01 | 1.00 | 0.31 | 1.00 |
| CSG | 6.9 | 0.01 | 1.00 | 12.0 | 0.09 | 1.00 | 0.23 | 1.00 |
| CSG+ | 5.0 | 0.00 | 1.00 | 6.0 | 0.00 | 1.00 | 0.00 | 1.00 |
| VSG | 6.6 | 0.00 | 1.00 | 6.8 | 0.00 | 1.00 | 0.00 | 1.00 |
| *New Zealand* | | | | | | | | |
| LIN | 8.5 | 0.06 | 1.00 | 8.6 | 0.06 | 0.97 | 0.87 | 1.07 |
| EXP | 13.9 | 0.00 | 1.13 | 11.7 | 0.00 | 1.09 | 0.00 | 16.84 |
| LIN/EXP | 8.3 | 0.00 | 1.01 | 8.4 | 0.00 | 0.97 | 0.00 | 1.08 |
| MEX | 9.0 | 0.00 | 1.03 | 8.0 | 0.00 | 1.00 | 0.00 | 1.14 |
| CGD | 13.9 | 0.00 | 1.11 | 11.7 | 0.00 | 1.11 | 0.00 | 14.95 |
| CSP | 9.3 | 0.00 | 1.00 | 10.9 | 0.00 | 1.00 | 0.00 | 1.00 |
| FSP | 9.7 | 0.12 | 1.00 | 9.4 | 0.12 | 1.00 | 0.00 | 1.00 |
| CSG | 8.2 | 0.06 | 1.00 | 9.8 | 0.17 | 1.00 | 0.58 | 1.00 |
| CSG+ | 7.4 | 0.00 | 1.00 | 7.7 | 0.00 | 1.00 | 0.00 | 1.00 |
| VSG | 8.2 | 0.00 | 1.00 | 8.5 | 0.00 | 1.00 | 0.00 | 1.00 |

*Source:* Author's calculations.
MedAPE is the Median Absolute Percentage Error; %-ve is the percentage of areas with negative projected populations; ratio refers to the ratio of the sum of unconstrained projected populations for all areas to the forecast or estimate for the country as a whole to which the initial projections are constrained.

performed well whilst for England & Wales CSG + was also quite accurate. All these small area forecasts were, of course, affected by error in the constraining state/national forecasts. All three national populations happened to be under forecast, with the percentage errors after 10 years being −4.3% for Australia, −3.7% for England & Wales, and −3.6% for New Zealand.

When the effect of incorrect official forecasts was removed by constraining to *actual* population estimates for 2011, the most accurate models remained unchanged, though their error increased a little. For EXP, MEX, and CGD errors were reduced. In Australia, CSG+ just remained the most accurate, but LIN/EXP was nearly as good. For England & Wales, CSG+ and MEX were not much less accurate than CSP whilst for New Zealand, MEX proved nearly as accurate as CSG+.

In terms of credibility, LIN, FSP, and CSG generated negative projections, although even after a 30-year horizon relatively few areas suffered from this problem. However, the constraining ratios were much more variable. Even after 10 years, preliminary forecasts from the EXP and CGD models required significant constraining to sum to independent forecasts and estimates in Australia and New Zealand. After 30 years, they were a serious problem in all three countries, indicating that initial forecasts were being completed altered through constraining. The problem is due to runaway growth in areas with high base period growth rates.

Previous research on the US has found certain models perform better for specific types of areas categorised by base period growth rate and jump-off year population size. Is this also the case for our three case study countries? Table 4 presents the most accurate models by base period growth rate according to their MedAPEs. Two or more models are listed if they all achieved the same MedAPE rounded to one decimal place. In all six sets of results the CSG+ model proved the most accurate for areas experiencing annual average population change below −0.5% in the base period. At the other end of the growth scale, those areas increasing the most during the base period are best forecast with the CSP model (in five out of six sets of forecasts), although errors are higher than in other growth rate categories. For the middle three growth rate categories various models did well, but the best were often best by the slimmest of margins. For example, in the New Zealand forecast-constrained set for the 0.5–1.5% growth rate category the CSP model achieved a MedAPE of 6.0% CSG 6.1%, LIN/EXP, FSP, CSG+, and VSG gave 6.2%; and LIN 6.3%.

Table 5 lists the models achieving the lowest MedAPE by jump-off year population size category. Confirming previous research, the very smallest populations in Australia and New Zealand proved the hardest to forecast, though curiously not in England & Wales. The results do not appear to show any clear patterns except that for many population size categories the best model was also the best model overall (Table 2).

## Averaged Models

How accurate were the averaged models? Table 6 reveals what percentage of averaged models proved more accurate than the best individual

Table 4. Models with the lowest Median Absolute Percentage Error (MedAPE) after 10 years, by base period growth rate.

| | Forecast-constrained | | Estimate-constrained | |
|---|---|---|---|---|
| Growth rate* | Model | MedAPE | Model | MedAPE |
| *Australia* | | | | |
| <−0.5% | CSG+ | 5.6 | CSG+ | 5.6 |
| −0.5–0.5% | CSP | 5.5 | LIN | 4.8 |
| 0.5–1.5% | LIN/EXP | 5.5 | MEX | 6.1 |
| 1.5–2.5% | MEX | 7.6 | MEX; CSP | 8.4 |
| >2.5% | CSP | 12.7 | CSP | 12.5 |
| *England & Wales* | | | | |
| <−0.5% | CSG+ | 4.9 | CSG+ | 4.9 |
| −0.5–0.5% | CSP | 4.1 | CSG+ | 4.3 |
| 0.5–1.5% | CSP | 4.6 | CSP | 5.2 |
| 1.5–2.5% | CSP | 5.7 | CSP | 5.8 |
| >2.5% | CSP | 7.7 | CSP | 7.4 |
| *New Zealand* | | | | |
| <−0.5% | CSG+ | 7.4 | CSG+ | 7.4 |
| −0.5–0.5% | CSG+ | 6.4 | LIN; LIN/EXP | 6.2 |
| 0.5–1.5% | CSP | 6.0 | MEX | 6.2 |
| 1.5–2.5% | CSG+ | 7.3 | CSP | 7.3 |
| >2.5% | CSG+ | 12.0 | CSP | 14.1 |

*Source:* Author’s calculations.
*Annual average base period growth rate.



**DOI**: 10.1002/psp

Table 5. Models with the lowest Median Absolute Percentage Error (MedAPE) after 10 years, by jump-off year population.

| | Forecast-constrained | | Estimate-constrained | |
|---|---|---|---|---|
| Population* | Model | MedAPE | Model | MedAPE |
| *Australia* | | | | |
| 0–1999 | CSG+ | 21.9 | VSG | 25.1 |
| 2000–4999 | CSG+ | 8.0 | CSG+ | 8.4 |
| 5000–9999 | CSG+ | 7.7 | CSG+; LIN/EXP | 8.3 |
| 10,000–14,999 | CSP | 6.3 | LIN/EXP | 6.6 |
| 15,000+ | CSP | 4.8 | LIN/EXP | 5.6 |
| *England & Wales* | | | | |
| 0–1999 | CSP | 4.6 | CSP | 5.7 |
| 2000–3999 | CSP | 4.2 | CSP | 5.6 |
| 4000–5999 | CSP | 4.4 | CSP | 5.4 |
| 6000–9999 | CSP | 4.8 | MEX | 5.7 |
| 10,000+ | CSP | 5.1 | CSP | 5.2 |
| *New Zealand* | | | | |
| 0–999 | CSG+; MEX | 12.4 | MEX | 12.1 |
| 1000–1999 | CSG+ | 7.8 | CSG+ | 7.8 |
| 2000–3999 | CSG+ | 5.9 | CSG+ | 6.2 |
| 4000–5999 | CSG+ | 5.5 | MEX | 5.6 |
| 6000+ | FSP | 4.1 | CSG+; LIN/EXP | 4.1 |

*Source:* Author's calculations.
*At the jump-off year.

Table 6. Percentage of averaged models more accurate than the best individual model at a 10-year forecast horizon.

| | Forecast-constrained | Estimate-constrained |
|---|---|---|
| | | % |
| Australia | 7.0 | 13.7 |
| England & Wales | 0.2 | 2.1 |
| New Zealand | 0.6 | 1.3 |

*Source:* Author's calculations.

model at a forecast horizon of 10 years. In the forecast-constrained set, 7.0% of the averaged models for Australia proved more accurate, whilst for England & Wales it was just 0.2%, and for New Zealand 0.6%. In the estimate-constrained set, these percentages increased to 13.7% for Australia, 2.1% for England & Wales, and 1.3% for New Zealand. In other words, when the state or national populations were 'forecast' accurately, the averaged models performed better.

Tables 7–9 list the averaged models which gave the lowest MedAPEs at a forecast horizon of 10 years for the three countries. For reasons of space only the top five models for each set of forecasts are shown. Small gains in accuracy were achieved by the best averaged models for Australia and New Zealand, but there was almost no gain for England & Wales.

The best averaged model for Australia in the forecast-constrained set, CSP VSG, achieved about a 10% lower MedAPE than the best individual model (Table 7). Although not one of the top five models in the estimate-constrained set, it also achieved a 4% improvement over the best individual model. The best averaged model of the estimate-constrained set, LIN/EXP CSP CSG+ VSG, also achieved a MedAPE about 10% lower than the best individual model and was also slightly more accurate than the best individual model in the forecast-constrained set. Although the top five models differ between the two sets of forecasts, there was considerable commonality in the averaged models which outperform the best individual model. All but two of the models in the estimate-constrained set with a lower MedAPE than the best individual model also outperformed the best individual model in the forecast-constrained set.



**DOI**: 10.1002/psp

Table 7. Lowest Median Absolute Percentage Error (MedAPE) averaged and individual models: Australia.

| | | 10-year horizon | | | 30-year horizon | |
|---|---|---|---|---|---|---|
| | Average of: | MedAPE | %-ve | Ratio | %-ve | Ratio |
| *Forecast-constrained* | | | | | | |
| Best individual model | CSG+ | 7.1 | 0.00 | 1.00 | 0.00 | 1.00 |
| Best averaged models | CSP, VSG | 6.4 | 0.00 | 1.00 | 0.00 | 1.00 |
| | CSP, CSG | 6.4 | 0.00 | 1.00 | 0.00 | 1.00 |
| | LIN/EXP, CSP | 6.4 | 0.00 | 1.00 | 0.00 | 1.01 |
| | CSP, CSG+ | 6.5 | 0.00 | 1.00 | 0.00 | 1.00 |
| | LIN, CSP | 6.5 | 0.00 | 1.00 | 0.00 | 1.01 |
| *Estimate-constrained* | | | | | | |
| Best individual model | CSG+ | 7.6 | 0.00 | 1.00 | *0.00* | *1.00* |
| Best averaged models | LIN/EXP, CSP, CSG+, VSG | 6.8 | 0.00 | 0.99 | *0.00* | *1.01* |
| | CSP, CSG+, VSG | 6.8 | 0.00 | 1.00 | *0.00* | *1.00* |
| | LIN, MEX, CSP, CSG+ | 6.8 | 0.00 | 0.99 | *0.00* | *1.03* |
| | MEX, CSP, CSG+ | 6.8 | 0.00 | 1.00 | *0.00* | *1.04* |
| | MEX, LIN/EXP, CSP, CSG+ | 6.8 | 0.00 | 0.99 | *0.00* | *1.04* |

*Source:* Author's calculations.
Numbers in italics for the estimate-constrained set are credibility measures from the forecast-constrained set, shown in order to reveal the longer term characteristics of these averaged models. %-ve is the percentage of areas with negative projected populations; ratio refers to the ratio of the sum of unconstrained forecast populations for all areas to the forecast or estimate for the country as a whole to which the initial projections are constrained.

Table 8. Lowest Median Absolute Percentage Error (MedAPE) averaged and individual models: England & Wales.

| | | 10-year horizon | | | 30-year horizon | |
|---|---|---|---|---|---|---|
| | Average of: | MedAPE | %-ve | Ratio | %-ve | Ratio |
| *Forecast-constrained* | | | | | | |
| Best individual model | CSP | 4.6 | 0.00 | 1.00 | 0.00 | 1.00 |
| Best averaged models | CSP, CSG+ | 4.6 | 0.00 | 1.00 | 0.00 | 1.00 |
| | MEX, CSP, CSG+ | 4.9 | 0.00 | 1.00 | 0.00 | 1.01 |
| | MEX, CSP | 4.9 | 0.00 | 1.00 | 0.00 | 1.01 |
| | LIN/EXP, CSP, CSG+ | 5.0 | 0.00 | 1.00 | 0.00 | 1.00 |
| | CSP, CSG+, VSG | 5.0 | 0.00 | 1.00 | 0.00 | 1.00 |
| *Estimate-constrained* | | | | | | |
| Best individual model | CSP | 5.5 | 0.00 | 1.00 | *0.00* | *1.00* |
| Best averaged models | EXP, CSP | 5.4 | 0.00 | 0.99 | *0.00* | *1.60* |
| | CGD, CSP | 5.4 | 0.00 | 1.01 | *0.00* | *1.59* |
| | CSP, CSG+ | 5.4 | 0.00 | 1.00 | *0.00* | *1.00* |
| | MEX, CSP, CSG+ | 5.4 | 0.00 | 0.99 | *0.00* | *1.01* |
| | MEX, CSP | 5.4 | 0.00 | 0.98 | *0.00* | *1.01* |

*Source:* Author's calculations.
Numbers in italics for the estimate-constrained set are credibility measures from the forecast-constrained set, shown in order to reveal the longer term characteristics of these averaged models. %-ve is the percentage of areas with negative projected populations; ratio refers to the ratio of the sum of unconstrained forecast populations for all areas to the forecast or estimate for the country as a whole to which the initial projections are constrained.

For England & Wales few averaged models outperformed CSP, the best individual model (Table 8). In the forecast-constrained set only CSP CSG+ gave a lower MedAPE, whilst in the estimate-constrained set a small number of models outperformed CSP by the slimmest of margins. However, some of these models, such as EXP CSP and CGD CSP, proved problematic for long-term forecasts because of high constraining ratios. These ratios are shown in the table in italics

Table 9. Lowest Median Absolute Percentage Error (MedAPE) averaged and individual models: New Zealand.

| | | 10-year horizon | | | 30-year horizon | |
|---|---|---|---|---|---|---|
| | Average of: | MedAPE | %-ve | ratio | %-ve | ratio |
| *Forecast-constrained* | | | | | | |
| Best individual model | CSG+ | 7.4 | 0.00 | 1.00 | 0.00 | 1.00 |
| Best averaged models | CSP, CSG+, VSG | 7.3 | 0.00 | 1.00 | 0.00 | 1.00 |
| | LIN/EXP, CSP, CSG+ | 7.3 | 0.00 | 1.00 | 0.00 | 1.03 |
| | CSP, CSG, CSG+ | 7.3 | 0.00 | 1.00 | 0.06 | 1.00 |
| | LIN/EXP, CSP | 7.3 | 0.00 | 1.00 | 0.00 | 1.04 |
| | CSP, FSP, CSG+ | 7.4 | 0.00 | 1.00 | 0.00 | 1.00 |
| *Estimate-constrained* | | | | | | |
| Best individual model | CSG+ | 7.7 | 0.00 | 1.00 | *0.00* | *1.00* |
| Best averaged models | EXP, LIN/EXP, CSP, CSG+ | 7.3 | 0.00 | 1.02 | *0.00* | *4.98* |
| | LIN, CGD, CSP, CSG+ | 7.3 | 0.00 | 1.02 | *0.06* | *4.50* |
| | EXP, MEX, CSP, CSG+ | 7.3 | 0.00 | 1.02 | *0.00* | *5.00* |
| | EXP, LIN/EXP, CSP | 7.3 | 0.00 | 1.02 | *0.00* | *6.31* |
| | LIN/EXP, CGD, CSP | 7.3 | 0.00 | 1.03 | *0.00* | *5.68* |
| Best averaged models | CSP, CSG+, VSG | 7.6 | 0.00 | 1.00 | *0.00* | *1.00* |
| with 2031 ratios | CSP, FSP, CSG+, VSG | 7.7 | 0.00 | 1.00 | *0.00* | *1.00* |
| within 0.95–1.05 | CSP, FSP, CSG+ | 7.7 | 0.00 | 1.00 | *0.00* | *1.00* |
| and no negatives | CSP, FSP, VSG | 7.7 | 0.00 | 1.00 | *0.00* | *1.00* |
| | CSP, VSG | 8.0 | 0.00 | 1.00 | *0.00* | *1.00* |

*Source:* Author's calculations.
Numbers in italics for the estimate-constrained set are credibility measures from the forecast-constrained set, shown in order to reveal the longer term characteristics of these averaged models. %-ve is the percentage of areas with negative projected populations; ratio refers to the ratio of the sum of unconstrained forecast populations for all areas to the forecast or estimate for the country as a whole to which the initial projections are constrained.

because they relate to 2031 and have to come from the forecast-constrained set. The best model which avoided excessive constraining was CSP CSG+.

For New Zealand, a small minority of averaged models in the forecast-constrained set were more accurate than the best individual model (Table 9), and they were only more accurate by tiny margins. CSP CSG+ VSG was the most accurate model. In the estimate-constrained set a larger proportion of averaged models outperformed the best individual model, and the error reduction was greater. Unfortunately, the top five models proved suitable only for short-term forecasts because the 30-year constraining ratios were very high. The lower panel of Table 9 therefore shows the top five averaged models without credibility problems. Of these, CSP CSG+ VSG was the most accurate.

## Composite Models

Table 10 reveals what percentage of composite models proved more accurate than the best individual models. For forecast-constrained growth rate category composites only 1.5% of the composite models for Australia proved more accurate than the best individual model, whilst for England & Wales and New Zealand it was 0.0%. In the estimate-constrained forecasts these percentages improved to 2.1% for Australia, 3.2% for England & Wales, and 1.3% for New Zealand. For the population size category

Table 10. Percentage of composite models more accurate than the best individual model at a 10-year forecast horizon.

| Model type<br>Country | Forecast-constrained | Estimate-constrained |
|---|---|---|
| *Growth rate composite models* | | % |
| Australia | 1.5 | 2.1 |
| England & Wales | 0.0 | 3.2 |
| New Zealand | 0.0 | 1.3 |
| *Population size composite models* | | |
| Australia | 0.0 | 0.1 |
| England & Wales | 0.0 | 0.0 |
| New Zealand | 0.0 | 0.0 |

*Source:* Author's calculations.



**DOI**: 10.1002/psp

composite models, almost none proved more accurate than the best individual model. They are not considered further.

Tables 11–13 list the top five growth rate composite models for the three countries. Generally, the best composite models achieved a small improvement in accuracy over the best individual models. For Australia, the best forecast-constrained composite was CSG+ (<−0.5% annual average growth in the base period), CSP (−0.5% to 0.5%), CSP (0.5–1.5%), CSG+ (1.5–2.5%), and CSG+ (2.5%) (Table 11). In this model previously declining areas are forecast to experience no population change, low and medium base period growth areas are forecast with State growth rates, whilst previously fast growing areas are projected with gradually declining growth rates. In the estimate-constrained set, the best model acts similarly, with MEX and LIN gradually reducing growth rates.

For England & Wales the best forecast-constrained model, CSG+ for the lower base period growth categories and CSP for previously higher growth areas, was only just more accurate than the best individual model (Table 12). However, had the national forecast been correct (as shown in the estimate-constrained set), the gains in accuracy would have been more significant. The lowest forecast error models, with CGD and EXP in the 2.5%+ category, unfortunately gave high constraining ratios at 30-year horizons. However, the third best model, CSP CSP CSP CSP CSG, avoided this problem without sacrificing much in the way of accuracy. In this composite, the CSG component for previously fast growing areas projects gradually decelerating growth rates.

For New Zealand, the best forecast-constrained composite was the same as the best individual model (Table 13). In the estimate-constrained set the best composite was CSG+ CGD MEX CSP LIN, which reduces rates of decline and growth in the <−0.5% and 2.5% growth rate categories, respectively. CGD for those areas that changed little in the base period (−0.5% to 0.5%) maintains low growth rates but MEX and CSP for categories 0.5–1.5% and 1.5–2.5% generally lower growth rates. However, a large number of composite models gave very similar levels of error.

Table 11. Lowest Median Absolute Percentage Error (MedAPE) growth rate composite and individual models: Australia.

| | Model (by base period growth rate) | | | | | 10-year horizon | | | 30-year horizon | |
|---|---|---|---|---|---|---|---|---|---|---|
| | <−0.5% | −0.5–0.5% | 0.5–1.5% | 1.5–2.5% | 2.5%+ | MedAPE | %-ve | Ratio | %-ve | Ratio |
| *Forecast-constrained* | | | | | | | | | | |
| Best individual model | CSG+ | | | | | 7.1 | 0.00 | 1.00 | 0.00 | 1.00 |
| Best composite models | CSG+ | CSP | CSP | CSG+ | CSG+ | 6.6 | 0.00 | 1.03 | 0.00 | 1.08 |
| | CSG+ | CSP | CSP | MEX | CSG+ | 6.7 | 0.00 | 1.04 | 0.00 | 1.09 |
| | CSG+ | CSP | CSP | LIN | CSG+ | 6.7 | 0.00 | 1.04 | 0.00 | 1.09 |
| | CSG+ | CSP | CSP | LIN/EXP | CSG+ | 6.7 | 0.00 | 1.04 | 0.00 | 1.09 |
| | CSG+ | CSP | CSP | CSP | CSG+ | 6.7 | 0.00 | 1.03 | 0.00 | 1.07 |
| *Estimate-constrained* | | | | | | | | | | |
| Best individual model | CSG+ | | | | | 7.6 | 0.00 | 1.00 | *0.00* | *1.00* |
| Best composite models | CSG+ | CSG+ | MEX | MEX | LIN | 6.9 | 0.00 | 0.97 | *0.00* | *1.06* |
| | CSG+ | CSG+ | MEX | MEX | LIN/EXP | 6.9 | 0.00 | 0.97 | *0.00* | *1.06* |
| | CSG+ | CSG+ | MEX | LIN | LIN | 6.9 | 0.00 | 0.97 | *0.00* | *1.06* |
| | CSG+ | CSG+ | MEX | LIN | LIN/EXP | 6.9 | 0.00 | 0.97 | *0.00* | *1.06* |
| | CSG+ | CSG+ | MEX | LIN/EXP | LIN | 6.9 | 0.00 | 0.97 | *0.00* | *1.06* |

*Source:* Author's calculations.
Numbers in italics for the estimate-constrained set are credibility measures from the forecast-constrained set, shown in order to reveal the longer term characteristics of these averaged models. %-ve is the percentage of areas with negative projected populations; ratio refers to the ratio of the sum of unconstrained projected populations for all areas to the projection or estimate for the country as a whole to which the initial projections are constrained.



**DOI**: 10.1002/psp

Table 12. Lowest Median Absolute Percentage Error (MedAPE) growth rate composite and individual models: England & Wales.

| | Model (by base period growth rate) | | | | | 10-year horizon | | | 30-year horizon | |
|---|---|---|---|---|---|---|---|---|---|---|
| | <−0.5% | −0.5–0.5% | 0.5–1.5% | 1.5–2.5% | 2.5% + | MedAPE | %-ve | Ratio | %-ve | Ratio |
| *Forecast-constrained* | | | | | | | | | | |
| Best individual model | CSP | | | | | 4.6 | 0.00 | 1.00 | 0.00 | 1.00 |
| Best composite models | CSG+ | CSG+ | CSP | CSP | CSP | 4.5 | 0.00 | 0.98 | 0.00 | 0.95 |
| | CSG+ | CSP | CSP | CSP | CSP | 4.6 | 0.00 | 1.00 | 0.00 | 0.99 |
| | CSP | CSP | CSG+ | CSP | CSP | 4.6 | 0.00 | 1.01 | 0.00 | 1.01 |
| | CSP | CSG+ | CSP | CSP | CSP | 4.6 | 0.00 | 0.99 | 0.00 | 0.97 |
| | CSP | CSP | CSP | CSP | CSP | 4.6 | 0.00 | 1.00 | 0.00 | 1.00 |
| *Estimate-constrained* | | | | | | | | | | |
| Best individual model | CSP | | | | | 5.5 | 0.00 | 1.00 | *0.00* | *1.00* |
| Best composite models | CSP | CSP | CSP | CSP | CGD | 4.8 | 0.00 | 1.03 | *0.00* | *2.19* |
| | CSP | CSP | CSP | CSP | EXP | 4.8 | 0.00 | 1.03 | *0.00* | *2.19* |
| | CSP | CSP | CSP | CSP | CSG | 4.8 | 0.00 | 1.03 | *0.00* | *1.04* |
| | CSP | CSP | CSP | MEX | EXP | 4.9 | 0.00 | 1.03 | *0.00* | *2.22* |
| | CSP | CSP | CSP | LIN | CGD | 4.9 | 0.00 | 1.04 | *0.00* | *2.22* |

*Source:* Author's calculations.
Numbers in italics for the estimate-constrained set are credibility measures from the forecast-constrained set, shown in order to reveal the longer term characteristics of these averaged models. %-ve is the percentage of areas with negative projected populations; ratio refers to the ratio of the sum of unconstrained projected populations for all areas to the projection or estimate for the country as a whole to which the initial projections are constrained.

## DISCUSSION

The results presented above have demonstrated the importance of examining the credibility of forecasting models alongside error. It has been shown that some simple models in common usage – whilst often relatively accurate overall – are susceptible to generating implausible forecasts. LIN, FSP, and CSG are able to produce negative populations and in practice, do so, whereas EXP and CGD permit runaway growth, which necessitates severe constraining to independent state or national forecasts. As a general rule these five models are best avoided.

The strong performance of CSP for England & Wales but not the other two countries may be related to the relatively narrow distribution of growth rates across its small areas. The interquartile range of 2001–2011 annual average growth rates was 0.95% in England & Wales compared with 1.67% in Australia and 1.72% in New Zealand. If growth rates do not deviate too much from the average then an assumption of national projected growth rates (CSP) is probably reasonable. The difference in growth rate distributions may well be related to slower national population growth in the UK, the relative maturity of its settlement system, and different residential planning controls and processes. The important point is that small area forecasting models behave differently under different demographic regimes and that findings from small area forecasting model assessments are not necessarily generalisable across countries.

The lack of clear patterns in the results by jump-off year population category may be partly due to the small range of population sizes in the case study datasets. The results by base period growth rate are much clearer and supportive of earlier research. CSG+ proved the best model for areas experiencing population decline in the base period. Recall that this model projects these areas' populations as a fixed size. Thus, the best approach to forecasting previously declining areas in the case study countries is to assume decline abates, and the population then remains constant into the future. The best model for areas with the fastest base period growth was CSP. This model assumes all small areas grow at the projected State growth rate, suggesting that the best assumption for previously fast-growing areas is one of more modest growth in the future. To a



**DOI**: 10.1002/psp

Table 13. Lowest Median Absolute Percentage Error (MedAPE) growth rate composite and individual models: New Zealand.

| | Model (by base period growth rate) | | | | | 10-year horizon | | | 30-year horizon | |
|---|---|---|---|---|---|---|---|---|---|---|
| | <−0.5% | −0.5–0.5% | 0.5–1.5% | 1.5–2.5% | 2.5%+ | MedAPE | %-ve | Ratio | %-ve | Ratio |
| *Forecast-constrained* | | | | | | | | | | |
| Best individual model | CSG+ | | | | | 7.4 | 0.00 | 1.00 | 0.00 | 1.00 |
| Best composite models | CSG+ | CSG+ | CSG+ | CSG+ | CSG+ | 7.4 | 0.00 | 1.00 | 0.00 | 1.00 |
| | CSG+ | CSG+ | CSG | CSP | CSG+ | 7.4 | 0.00 | 1.00 | 0.00 | 0.99 |
| | CSG+ | CSG+ | EXP | CSG | CSG+ | 7.4 | 0.00 | 1.01 | 0.00 | 1.05 |
| | CSG+ | CSG+ | EXP | VSG | CSG+ | 7.4 | 0.00 | 1.01 | 0.00 | 1.05 |
| | CSG+ | CSG+ | VSG | CSP | CSG+ | 7.4 | 0.00 | 1.00 | 0.00 | 0.99 |
| *Estimate-constrained* | | | | | | | | | | |
| Best individual model | CSG+ | | | | | 7.7 | 0.00 | 1.00 | *0.00* | *1.00* |
| Best composite models | CSG+ | CGD | MEX | CSP | LIN | 7.4 | 0.00 | 0.98 | *0.00* | *1.03* |
| | CSG+ | CGD | MEX | CSP | LIN/EXP | 7.4 | 0.00 | 0.98 | *0.00* | *1.03* |
| | CSG+ | CSG+ | LIN | MEX | LIN | 7.4 | 0.00 | 0.98 | *0.00* | *1.11* |
| | CSG+ | CSG+ | LIN | MEX | LIN/EXP | 7.4 | 0.00 | 0.98 | *0.00* | *1.11* |
| | CSG+ | CSG+ | LIN/EXP | MEX | LIN | 7.4 | 0.00 | 0.98 | *0.00* | *1.11* |

*Source:* Author's calculations.
Numbers in italics for the estimate-constrained set are credibility measures from the forecast-constrained set, shown in order to reveal the longer term characteristics of these averaged models. %-ve is the percentage of areas with negative projected populations; ratio refers to the ratio of the sum of unconstrained projected populations for all areas to the projection or estimate for the country as a whole to which the initial projections are constrained.

large extent, these findings can be summarised as regression towards the mean, a statistical phenomenon in which very large or small values are followed by ones which are closer to the mean (Stigler 1997; Smith, 1987). In the context of small area population forecasts, areas experiencing large growth rates of either sign in the base period are likely to undergo more moderate rates of change in the future.

With regard to averaged models, a small proportion was shown to give slightly lower errors than the best individual models. The best averaged models generally contained CSP and either or both of CSG+ and VSG, indicating that an average of the projected state or national growth rate (CSP) and a local component (CSG+ or VSG) is a good combination. Such a finding is consistent with the view of Ahlburg and Lutz (1999 p. 5) who argue that 'combining forecasts is most likely to improve accuracy when the forecasts combined use different methods and approaches that capture different information sets or different specifications'. These best-performing averaged models also incorporate regression towards the mean: they increase the growth of previously declining areas, maintain average growth in moderately growing areas, and reduce the growth of fast growing areas. For England & Wales averaged models proved less useful, possibly because of the more limited distribution of growth rates.

A small proportion of growth rate composite models produced slightly lower errors than the best individual models, provided that the State or national constraining forecasts were accurate. The lack of accuracy gains with the population size composites may, again, be due to the small range of population sizes in the case study datasets, or perhaps a limited role of population size in affecting accuracy at this spatial scale. Overall, the best composite models were generally ones that reduced both decline and fast growth relative to the base period, which aligns with the above findings.

What should be made of the differences in results between the estimate-constrained and forecast-constrained sets? The estimate-constrained sets revealed what the small area models' errors would be in the absence of error in the constraining state or national population forecasts, whereas the



**DOI**: 10.1002/psp

forecast-constrained results demonstrated the outcome of constraining to typical official population forecasts. Differences between the best performing models do add to the challenge of recommending a model, but the differences should not be overstated. For example, results from the averaged models in both sets of forecasts have correlation coefficients of 0.93 for Australia, 0.83 for New Zealand, and 0.60 for England & Wales. Many of the most accurate models were found to be common to both estimate-constrained and forecast-constrained sets.

## CONCLUSIONS

This paper has reported on an evaluation of a large number of individual, averaged, and composite models for forecasting the total populations of small areas. It assessed both forecast error and credibility (absence of negative populations and large constraining ratios) for three case study countries. The individual models CSG+, CSP, VSG, LIN/EXP, and MEX were shown to avoid credibility problems and, in many cases, give relatively low forecast errors. A small proportion of averaged and growth rate composite models generated slightly more accurate forecasts than the best individual models. From one perspective, combining forecasting models is only *just* worthwhile. The results are not overwhelmingly favourable: many averaged and composite models gave large forecast errors, with almost all composite models based on population size failing to improve on the best individual models. Nonetheless, there is an additional reason to consider combining, or at least averaged models. It reduces the risk of selecting a single model which turns out to be very inaccurate. Averaged forecasts are less likely to be highly inaccurate (Hibon & Evgeniou, 2005).

It is important to bear in mind the limitations of this study when interpreting results. First, all forecasts were based on, and compared against, population estimates. Estimates will always contain a degree of error themselves, both from census processes and the conversion of data to a consistent geography. Second, a limited selection of individual models was considered, and there are others that could be evaluated in further work. Third, forecast error was calculated for 10-year horizons only due to data limitations. Although a decade is a common forecast horizon, both shorter and longer horizons are worth exploring. Fourth, models generating the smallest forecast errors for specific periods in the past may not always do so in the future. Whilst it is hoped those models that performed best in the test period 2001–2011 will give the most accurate forecasts in the future, this cannot be guaranteed.

Further research could address some of these limitations. Through geographical data conversion, it may be possible to extend the time series of population data for the three case study countries to assess the temporal stability of the results. Other countries could be added to the evaluation, and in particular, it would be interesting to include some with widespread population decline. Other simple models could be included in the analysis, as could a comparison with cohort-component models. In addition, it may be worth devising and testing an 'individual' model which acts like the best averaged or composite models but is quicker to implement. Furthermore, it would be valuable to investigate whether information on the distribution of base period growth rates could be usefully incorporated into the model selection process.

Finally, given the results reported in the paper, what can be recommended for the production of small area population forecasts if the decision is to make use of simple models? Averaged models which incorporate both State/national growth and local growth are both appealing conceptually and empirically. From a conceptual standpoint they imply that population growth at the local level is due to a combination of local factors and wider regional or national forces. From an empirical perspective, results from this study demonstrated such models to be relatively successful in terms of both accuracy and credibility. Therefore,

(1) For Australia and New Zealand, the averaged model CSP CSG+ VSG is a good choice. If you just want to use one model, the best is CSG+.
(2) For England & Wales, consider the averaged model CSP CSG+. If you just want to use a single model then CSP is the best option. Alternatively, consider the composite model with CSP for areas growing under 2.5% per annum on average over the previous decade and CSG for those areas exceeding this rate.



**DOI**: 10.1002/psp

(3) Generally avoid models prone to generating implausible forecasts, LIN, FSP, CSG, EXP, and CGD, except for very short forecast horizons.

However, simple models, whether individual, averaged, or composite, will not provide satisfactory population forecasts for all types of local area. Therefore,

(3) Treat the numbers from the previous models as a 'base layer' of forecasts, which can be improved in certain places. Areas with significant residential development anticipated are ideally forecast with a housing-unit model, as mentioned in the Introduction. Certain other areas, such as those with relatively large populations in communal accommodation, may also need special treatment.
(3) Finally, if age–sex forecasts are required, consider applying a cohort-component model (preferably a bi-regional rather than a net migration version) constrained to the total population forecasts created from the above methods (e.g. Wilson & Cooper, 2013).

ACKNOWLEDGEMENTS

The author is grateful to Dr Paul Norman of Leeds University for supplying ward population estimates for England & Wales on a consistent geography and to Kim Dunstan of Statistics New Zealand for supplying area unit population data for New Zealand. Thanks are also due to the anonymous reviewers whose comments were very helpful in improving the original draft of this paper.

ENDNOTES

[1] The number of averaged models created when averaging every combination of 5, 4, 3 and 2 of the ten individual models is 627, i.e.

$$\frac{10}{5!(10-5)!}+\frac{10}{4!(10-4)!}+\frac{10}{3!(10-3)!}+\frac{10}{2!(10-2)!}=627.$$

[2] The number of composite models obtained with 5 categories of growth rate or population size and ten individual models is 100,000, i.e. $10^5$=100,000.

REFERENCES

ABS (Australian Bureau of Statistics). 2013. Regional Population Growth, Australia, 2012. [online dataset] Catalogue No. 3218.0. ABS: Canberra.

Ahlburg DA. 1987. Population forecasts for South Pacific nations using autoregressive models, 1985–2000. *Journal of the Australian Population Association* **4**: 157–167.

Ahlburg DA. 1995. Simple versus complex models: evaluation, accuracy and combining. *Mathematical Population Studies* **5**: 281–290.

Ahlburg D, Lutz W. 1999. Introduction: the need to rethink approaches to population forecasts. In Frontiers of Population Forecasting, Lutz W, Vaupel J, Ahlburg D (eds). Population Council: New York; 1–14.

Armstrong JS. 2001. Combining forecasts. In Principles of Forecasting: A Handbook for Researchers and Practitioners, Armstrong JS (ed.). Springer: New York; 417–439.

Assuncao RM, Schmertmann C P, Potter JE, Cavenaghi SM. 2005. Empirical Bayes estimation of demographic schedules for small areas. *Demography* **423**: 537–558.

Baker J, Ruan X, Alcantara A, Jones T, Watkins K, McDaniel M, Frey M, Crouse N, Rajbhandari R, Morehouse J, Sanchez J, Inglis M, Baros S, Penman S, Morrison S, Budge T Stallcup W. 2008. Density-dependence in urban housing unit growth: an evaluation of the Pearl-Reed model for predicting housing unit stock at the census tract level. *Journal of Social and Economic Measurement* **33**: 155–163.

BEA (Bureau of Economic Analysis). 1996. Metropolitan area and BEA economic area projections of economic activity and population to the year 2005. *Survey of Current Business*, June: 56-72. http://www.bea.gov/scb/pdf/regional/proj/1996/0696rea.pdf

Bell M. 1997. Small Area Forecasting for Infrastructure Planning: Towards a Better Practice. Department of Industry, Science and Tourism: Canberra.

Bell M, Dean C, Blake M. 2000. Forecasting the pattern of urban growth with PUP: a web-based model interfaced with GIS and 3D animation. *Computers, Environment and Urban Systems* **24**: 559–581.

Chi G, Voss P. 2011. Small-area population forecasting: borrowing strength across space and time. *Population, Space and Place* **17**: 505–520.

Davis HC. 1995. Demographic Projection Techniques for Regions and Smaller Areas. University of British Columbia Press: Vancouver.

Foss W. 2002. Small area population forecasting. *The Appraisal Journal* **70**: 163–172.

Goodwin P. 2009. New evidence on the value of combining forecasts. *Foresight* **12**: 33–35.

Graefe A, Armstrong JS, Jones RJ, Cuzan AG. 2014. Combining forecasts: an application to elections. *International Journal of Forecasting* **30**: 43–54. 10.1016/j.ijforecast.2013.02.005

Achadoorian L, Gaffin SR, Engelman R. 2011. Projecting a gridded population of the world using ratio methods of trend extrapolation. In Human Population: Its Influences on Biological Diversity, Cincotta RP, Gorenflo LJ (eds). Springer-Verlag: Berlin; 13–25.

Hibon M, Evgeniou T. 2005. To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting* **12**: 15–24.

Isserman AM. 1977. The accuracy of population projections for subcounty areas. *Journal of the American Institute of Planners* **43**: 247–259.

Kapetanios G, Labhard V, Price S. 2008. Forecast combination and the Bank of England's suite of statistical forecasting models. *Economic Modelling* **25**: 772–792.

Norman P. 2013. Population estimates for CAS wards in England & Wales, 1981, 1991, 2001 and 2011 [dataset]. School of Geography, University of Leeds.

Norman P, Rees P, Boyle P. 2003. Achieving data compatibility over space and time: creating consistent geographical zones. *International Journal of Population Geography* **9**: 365–386.

Norman P, Simpson L, Sabater A. 2008. 'Estimating with Confidence' and hindsight: new UK small area population estimates for 1991. *Population, Space and Place* **14**: 449–472.

Openshaw S, van der Knapp G. 1983. Small area population forecasting: some experience with British models. *Tijdschrift voor economische en sociale geografie* **74**: 291–304.

Rayer S. 2008. Population forecast errors: a primer for planners. *Journal of Planning Education and Research* **27**: 417–430.

Rayer S, Smith SK. 2010. Factors affecting the accuracy of subcounty population forecasts. *Journal of Planning Education and Research* **30**: 147–161.

Raymer J, Wiśniowski A, Forster JJ, Smith PWF, Jakub B. 2013. Integrated modeling of European migration. *Journal of the American Statistical Association*. DOI: 10.1080/01621459.2013.789435

Rees P. 1990. Manual on subnational population projections. Unpublished manuscript, School of Geography, University of Leeds.

Shryock H, Siegel J. 1973. The Methods and Materials of Demography. US Government Printing Office: Washington DC.

Smith SK. 1987. Tests of forecast accuracy and bias for county population projections. *Journal of the American Statistical Association* **82**: 991–1003.

Smith S. 1997. Further thoughts on simplicity and complexity in population projection models. *International Journal of Forecasting* **13**: 557–565.

Smith SK, Morrison PA. 2005. Small area and business demography. In Handbook of Population, Dudley Poston D, Micklin M (eds). Springer Publishers: New York; 761–785.

Smith S, Rayer S. 2011. An evaluation of population forecast errors for Florida and its counties, 1980–2010. Special Population Reports Number 9. Bureau of Economic and Business Research, University of Florida.

Smith SK, Rayer S. 2012. Projections of Florida population by county, 2011–2040. Bulletin 162 (Revised), March 2012. Bureau of Economic and Business Research, University of Florida, Gainesville.

Smith S, Shahidullah M. 1995. An evaluation of population projection errors for census tracts. *Journal of the American Statistical Association* **90**: 64–71.

Smith S, Sincich T. 1988. Stability over time in the distribution of population forecast errors. *Demography* **25**: 461–474.

Smith S, Sincich T. 1992. Evaluating the forecast accuracy and bias of alternative population projections for states. *International Journal of Forecasting* **8**: 495–508.

Smith S, Tayman J, Swanson D. 2001. State and Local Population Projections: Methodology and Analysis. Springer: New York.

Statistics New Zealand. 2012. Population estimates and census counts for area units, 1991, 1996, 2001, 2006 and 2011 [dataset]. Statistics New Zealand: Christchurch.

Stigler S M. 1997. Regression towards the mean, historically considered. *Statistical Methods in Medical Research* **6**: 103–114.

Swanson D A, Schlottmann A, Schmidt B. 2010. Forecasting the population of census tracts by age and sex: an example of the Hamilton–Perry method in action. *Population Research and Policy Review* **29**: 47–63.

Tayman J, Swanson D A. 1999. On the validity of MAPE as a measure of population forecasting accuracy. *Population Research and Policy Review* **18**: 299–322.

Vidyattama Y, Tanton R. 2010. Projecting small area statistics with Australian Spatial Microsimulation Model (SPATIALMSM). *Australasian Journal of Regional Studies* **16**: 99–126.

Walters A, Cai Q. 2008. Investigating the use of Holt-Winters time series model for forecasting population at the State and sub-State levels. Paper prepared for the Population Association of American Annual Meeting, 17-19 April 2008, New Orleans, USA.

White HR. 1954. Empirical study of the accuracy of selected methods of projecting state populations. *Journal of the American Statistical Association* **29**: 480–498.

Wilson T, Cooper J. 2013. Overview of the new Queensland demographic projection systems. Research note. Queensland Centre for Population Research, School of Geography, Planning and Environmental Management, The University of Queensland, Brisbane.

Wilson T, Rees P. 1999. Linking 1991 population statistics to the 1998 local government geography of Great Britain. *Population Trends* **97**: 37–45.

Wilson T, Rees P. 2005. Recent developments in population projection methodology: a review. *Population, Space and Place* **11**: 337–360.

Wilson T, Rowe F. 2011. The forecast accuracy of local government area population projections: a case study of Queensland. *Australasian Journal of Regional Studies* **17**: 204–243.




**DOI**: 10.1002/psp