Insights from the evaluation of past local area population forecasts

Tom Wilson*, Huw Brokensha‡, Francisco Rowe§, Ludi Simpson†

* ‡ Northern Institute, Charles Darwin University, Darwin, Australia
§ Department of Geography and Planning, University of Liverpool, UK
† School of Social Sciences University of Manchester, UK

* Email: tom.wilson@cdu.edu.au

Paper prepared for the Population Association of New Zealand conference, 24th - 25th July 2017, Christchurch, New Zealand

Abstract
Local area population forecasts have a wide variety of uses in the public and private sectors. But not enough is known about the errors of such forecasts, particularly over the longer-term (20 years or more). Understanding past errors is valuable for both forecast producers and users. This paper (i) evaluates the forecast accuracy of past local area population forecasts published by Australian State and Territory Governments over the last 30 years, and (ii) illustrates ways in which past error distributions can be employed to quantify the uncertainty of current forecasts. Population forecasts from the past 30 years were sourced from State and Territory Governments. Estimated Resident Populations to which the projections were compared were created for the geographical regions of the past projections. The key features of past forecast error patterns are described. Forecast errors mostly confirm earlier findings with regards to the relationship between error and length of projection horizon and population size. The paper then introduces the concept of a forecast 'shelf life', which indicates how far into the future a forecast is likely to remain reliable. It also illustrates how past error distributions can be used to create empirical prediction intervals for current forecasts. These two complementary measures provide a simple way of communicating the likely magnitude of error that can be expected with current local area population forecasts.

Key words
Population forecasts; local area; Australia; forecast error; shelf life; empirical prediction intervals

1. Introduction

Local area population forecasts have many uses across the public and private sectors. They assist governments with planning local areas' future needs for housing, education, health care, transport, water, power, waste removal, and other infrastructure and services. For the business sector, they can inform decisions about retail and office locations and help assess future markets for goods and services. For researchers and analysts, they provide inputs to many other types of demand and budget forecasting models. Huge financial commitments are made, or not made, as a result of decisions informed by local area population forecasts.

Many users presume that local area population forecasts are highly accurate. Others may be more aware of the inherent error in forecasts but their organisations' decision-making processes and models are unable to accommodate anything other than a single forecast. Unfortunately large errors, such as 10% or more after just 10 years into the forecast, are common for local and small area populations, as shown by earlier research on forecast error and accuracy, e.g. Isserman (1977), Rayer (2008), Smith and Shahidullah (1995), Tayman, Schafer, and Carter (1998), and Wilson and Rowe (2011).

Studies on the accuracy of past local area population forecasts such as these make several contributions. First, they may shed light on problems with methods or data which were not previously apparent, and which can be rectified in future sets of forecasts. Second, descriptions of error distributions can assist those producing local area forecasts with managing user expectations about the size of errors likely to eventuate in new forecasts. Third, past errors might help population forecasters to decide what aspects of their forecasts to hold back from publication, such as forecasts beyond a certain time in the future or those for areas below a particular population size. Fourth, evaluations of past forecasts can assist clients seeking forecasts to select the forecaster whose previous efforts have proved the most accurate. Fifth, the expected error can help planners to design contingency plans that respond to the costs of different errors. Finally, past errors can be used to quantify the likely error in the latest set of population forecasts. Past error distributions can be easily adapted to create empirical prediction intervals for total populations (e.g. Keyfitz 1981; Rayer, Smith, and Tayman 2009; Rayer and Wang 2015; Tayman 2011). This is the main focus of our paper.

Studies of past subnational population forecast errors are less common than those focusing on national populations. Examples from the last decade include the evaluation of UK local authority district population forecasts by the Office for National Statistics (2008, 2015), an analysis by Statistics New Zealand (2008) of forecasts at various spatial scales, and an assessment of Japanese subnational forecasts by Yamauchi, Koike, and Kamata (2017). Examples from the US include Rayer (2008), Rayer and Smith (2010, 2014), Smith and Rayer (2011), and Tayman, Smith, and Rayer (2011) who all assessed forecasts for counties and/or various sub-county local areas. Some of these studies report the error patterns from experiments which applied competing forecasting models retrospectively, while others were assessments of previously published forecasts. In Australia evaluations have been undertaken of local government area population forecasts in Queensland (Wilson and Rowe 2011), State and Territory forecasts produced by the Australian Bureau of Statistics (Wilson 2012), and selected local area forecasts over five year periods (Wilson 2015). In general these studies have found that errors are negatively associated with population size, and

positively associated with forecast horizon length and the volatility of net migration; average forecast error increases roughly linearly over time, but bias is unpredictable; and mining and Indigenous areas tend to be forecast with greater error. Overall, much remains to be learned about subnational population forecast error generally, and especially for countries other than the US. Of the handful of studies which have examined previously published local area forecasts, many are limited in their spatial and temporal coverage.

The aims of this paper are to (i) present an evaluation of Australian local area population forecast errors involving a much greater number of areas than has been possible to date, and (ii) illustrate ways in which past errors can be employed to quantify the uncertainty of current sets of forecasts. Our study assessed 33 sets of local area population forecasts published over the last 30 years by State and Territory governments; it includes a total of 3,118 areas for which forecasts were assessed over horizons of up to 20 years.

The large-scale nature of our study gave us the confidence to use past error distributions to quantify uncertainty in current forecasts. We take two approaches. First, the distribution of past errors permits the calculation of 'shelf life' estimates for current forecasts which provide a simple way for users to determine how far into the future forecasts are likely to remain usable. Second, we illustrate the application of empirical 80% prediction intervals to current forecasts. We encourage users to explicitly incorporate these prediction intervals in their use of the forecasts, entering the upper and lower 80% bounds into their decision-making processes and spreadsheets in order to assess whether a different decision would be made. The promotion of empirical prediction intervals to illustrate uncertainty is especially significant given the reluctance of statistical offices to adopt complex and data-hungry subnational probabilistic methods.

A brief note on terminology is useful at this point. In common with other forecast error studies, we refer to population *forecasts* even though some producers emphasise that they only produce population *projections*. The former are often defined as predictions; the latter are simply the outcome of calculations based on specified input data assumptions (Smith, Tayman, and Swanson 2013), and may be likely, implausible, or illustrative of extreme scenarios. Given that many users interpret projections as forecasts, and that the purpose of this paper is to assess how well the projections / forecasts fared as predictions of population, we choose the term 'forecasts'.

Following this introduction, section 2 describes the forecasts and population estimates data, and methods used in this study. In section 3 the main results are presented and compared with other studies. The focus of section 4 is on making use of past error distributions to inform users of the likely error in current sets of local area population forecasts. Concluding remarks and recommendations comprise the final section.

2. Data and evaluation methods

2.1. Population forecast data

Thirty-three sets of local area population forecasts produced by State and Territory governments, covering a total of 3,118 local areas, were assessed in this study. The evaluation was undertaken for total populations only and at forecast horizons of 5, 10, 15 and 20 years as these comprised the common data values across all sets of forecasts. Medium series projections were selected as forecasts

for analysis where there was a choice of series with the exception of a couple of cases where other series were designated the "preferred" or "recommended" series at the time of publication. Forecasting methods included housing-unit methods, cohort-component models, and simple extrapolative approaches (Wilson 2011); unfortunately for some sets of forecasts it was not possible to ascertain what methods were used. The majority of local areas consisted of local government areas, while those for the two territories, the Northern Territory and Australian Capital Territory, were sub-territory regions for which those two jurisdictions had prepared forecasts. The earliest sets of forecasts were 1986-based and the most recent were 2011-based. All forecasts which had been prepared 20 or more years ago were evaluated over 20 year forecast horizons while more recent forecasts were assessed over shorter periods up to 2016.

2.2. Population estimates data

All past forecasts were compared to Estimated Resident Populations (ERPs), the official resident population estimates published by the ABS. These were obtained for the period 1986-2016. However, because of changes to local area boundaries over time it was necessary to convert the current time series of ERPs to the geographical areas of past projections. This proved a substantial task which was largely completed in ArcGIS. The general approach was to create ERPs for the spatial units of the projections by aggregating up ERPs of very small areas which fell within the boundaries of the old local areas. ERPs for SA1 areas (with populations mostly between 200 and 800) were available from 2001 onwards, whilst ERPs for Census Collection Districts (containing similar sized populations) were available for earlier years. ERPs for both these types of area often had to be disaggregated into smaller units using census data for mesh blocks (the smallest spatial unit in the ABS geographic hierarchy).

2.3. Evaluation measures

Percentage Error (PE) was used to measure the forecast error of individual local area forecasts:

$$PE_t = \frac{F_t - ERP_t}{ERP_t}\ 100\%$$

where F denotes forecast, t a year in the future, and ERP Estimated Resident Population. Positive values of PE indicate forecasts which were too high; negative values are obtained for forecasts which proved too low.

One of the challenges faced in the project was dealing with ABS's decision to "recast" (adjust) ERPs for 1991-2011 following the results of the 2011 Census (ABS 2013). This resulted in discrepancies between jump-off populations in the forecast datasets and the recast ERPs, even for those forecasts launched from 'finalised' ERPs for census years. We therefore decided to remove jump-off discrepancies for all forecasts. We calculated an alternative version of PE (Keilman 1999) to remove the effect of the initial discrepancy:

$$PE_t = \frac{F_t - ERP_t - \left(F_0 - ERP_0\right)}{ERP_t}\ 100\%$$

where 0 refers to the jump-off year. In much of the paper, Absolute Percentage Error (APE), the unsigned value of PE, is used.

Average absolute errors are reported with Median Absolute Percentage Error (MedAPE), the middle value of a set of ranked errors, and Mean Absolute Percentage Error (MAPE). Average signed errors are measured with Median Percentage Error (MedPE).

3. Forecast error patterns

3.1. Average error

Median Absolute Percentage Errors (MedAPEs) for local area population forecasts are presented in Figure 1. For areas of all population sizes, illustrated by the set of bars at the right of the graph, MedAPE was 2.8% after 5 years, 5.4% after 10, 8.0% after 15, and 11.7% after 20 years. This represents a near-linear increase in average error over time.

Not surprisingly, average errors were highest for the smallest populations, echoing previous findings in the literature (e.g. Tayman, Schafer, and Carter 1998). For areas with a jump-off population under 1,000 the MedAPE was already 7.9% after 5 years, reaching 25.1% after 20 years. Average errors were notably reduced in the next population size category of 1,000-1,999 persons, with a MedAPE of 4.6% after 5 years and 14.1% after 20 years. Smaller, but nonetheless important, reductions in error can be seen with increasing population size categories up to jump-off populations of about 10,000. Increases in population beyond 10,000 reveal a general trend of small reductions in error with increasing population size, but the pattern is uneven, especially at forecast horizons of 20 years (for which the number of data points is smaller).

[Figure 1 about here]

The average errors shown in Figure 1 cover all forecasts evaluated in this study. But does the magnitude of error vary much over time? For example, are more recent forecasts more accurate than those produced in the 1980s? Figure 2 shows MedAPEs for groups of jump-off years by forecast horizon. Some temporal variation is evident: MedAPEs after 5 years varied between 2 and 4%, and between 5 and 6% after 10 years, while greater variation occurred after 15 and 20 years. Forecasts produced with 1996-2000 jump-off years have slightly more error than the other forecasts. But on the basis of the 5 and 10 year MedAPEs it is not possible to say that recent forecasts have been more accurate than earlier forecasts, or that there is any clear change in error magnitudes over the time period covered by Figure 2.

[Figure 2 about here]

How do these errors for Australian local areas compare to those found in other studies? Unfortunately comparisons are difficult because other studies often deal with different population sizes, use different error measures, and report much of their findings graphically. However, it is possible to compare errors from population forecasts produced for Florida's counties with jump-off years 1980-2005 using the evaluation by Smith and Rayer (2011) and by calculating Mean Absolute Percentage Errors (MAPEs) for the Australian local area forecasts. MAPEs for Florida counties by forecast horizon are shown on the left-hand side of Table 1; these are very close to those for Australian local areas. Greater variation is evident for MAPEs by broad population size in Australia than in Florida after 10 years (right-hand side of the table).

[Table 1 about here]

3.2. Bias

Figure 3 shows the extent to which local area forecasts were subject to bias – whether the forecasts were too high or too low overall. For the forecasts as whole (shown by the set of bars at the right of the graph) there was very little bias. Median Percentage Error was 0.4% after both 5 and 10 years, -1.3% after 15, and 0.6% after 20 years.

Forecasts by population size category revealed the smaller population sizes were generally over-forecast (positive bias) while the larger populations were under-forecast (negative bias). The explanation is possibly due to pressure on State and Territory government demographers to be optimistic with the smallest local area populations, especially those with population decline. For the more populous areas, there has probably been some conservatism in forecasting migration in urban local area populations which has resulted in under-forecasting of these populations overall.

[Figure 3 about here]

Is bias stable over time? Figure 4 presents Median Percentage Errors for forecasts in five year jump-off year categories for forecast horizons up to 20 years. The graph suggests that bias is variable and unpredictable over time, a finding which echoes previous US research (e.g. Rayer, Smith, and Tayman 2009; Tayman 2011).

[Figure 4 about here]

3.3. Error distributions

Average error measures perform a useful function in any evaluation of forecast error, but it is also important to examine error distributions. The same average errors may be accompanied by narrow or wide distributions. Wide absolute error distributions with a very long tail of high errors may point to serious problems with the forecasts of specific areas. Narrower error distributions with fairly high average errors may indicate a general problem with forecast assumptions across all areas; narrow error distributions with low average errors represent a good outcome.

Figure 5 displays boxplots of Absolute Percentage Errors by population size category for forecast horizons of 5, 10, 15, and 20 years. Outliers, those values more than three times the inter-quartile range away from the first or third quartiles, are not shown in order to maintain the clarity of the graphs. For all local areas as a whole (shown at the bottom of each graph), the third quartiles were 5.3% for 5 year forecast horizons, 9.7% after 10 years, and 19.6% after 20 years. These errors exhibit near-linear trends over time.

[Figure 5 about here]

The positioning of the third quartiles and right-hand whiskers in Figure 5 reveals skewed error distributions with longer tails of errors above the median, particularly for the smaller population size categories and particularly at horizons of 5 and 10 years. But skewed distributions are also evident for most of the other population size categories at most forecast horizons. This is not surprising and matches findings from earlier research (e.g. Rayer, Smith, and Tayman 2009). Mirroring the trend in MedAPEs, the third quartiles and maximum values were very high for the smallest population size category (up to 999), but considerably lower in the next size category (1,000-1,999), and lower still

for slightly larger populations (2,000-2,999), certainly for forecast horizons of up to 10 years. There is greater randomness in the patterns at 15 and 20 year horizons, but it is still possible to discern a general trend of substantial narrowing of the error distribution with increasing population size over the smallest 5 or 6 population size categories.

The local areas experiencing highly erroneous forecasts – those with APEs above the 95th percentile in each population size category – were found to be quite diverse in character. Due to the lack of data for local area geographies at the time the projections were produced it was not possible to model the correlates of very high errors. A qualitative assessment had to be made instead. Some local areas had tiny populations where the smallest numerical population change can yield large forecast errors (e.g. Sandstone, Western Australia, with a population of under 200). Others were predominantly Indigenous areas where ERPs are known to be less reliable (e.g. Ngaanyatjarraku, Western Australia). High forecast errors occurred in many mining areas where employment, and thus population, has fluctuated considerably over time in response to global commodity prices and demand (e.g. Mount Morgan shire, Queensland). Some bad forecasts occurred in metropolitan fringe areas (e.g. Camden, New South Wales), perhaps due to the anticipated timing of residential development changing, and there were also poor forecasts in some central urban areas (e.g. City of Adelaide, South Australia). The variety of places with very high population forecast errors suggests a range of factors are likely to have been at work. However, the challenge of unpicking the causes of these erroneous forecasts must be left to future detailed case studies.

3.4. Summary

The analysis of past error patterns has revealed several key features. Most obviously, errors in forecasting local area populations can be substantial. They regularly exceed the 10% error found to be acceptable to users by Tye (1994). Forecasts for local areas with very small populations are likely to be highly erroneous, even just 5 years into the forecast. This is especially the case for those with under 2,000 people. For many such areas it is not possible to state with any confidence at the outset whether the population is likely to increase or decrease in the future. Population forecasters would be best to merge such areas with others to obtain larger populations; forecast users should avoid making important decisions on the basis of forecasts for individual populations of this size. It is also worth emphasising that although error tends to reduce as population size increases, areas with larger populations are not completely immune from the occasional large forecast error, as Figure 5 demonstrates.

As a general rule, the further into the future forecasts go, the greater the error. This study shows that after 20 years, most local area forecasts were subject to large errors. Even the largest local areas with 150,000 people or more at jump-off experienced a MedAPE of 7.6% after 20 years and a third quartile APE of 10.2%. The lesson for both forecast producers and users is that few local area population forecasts are reliable 20 years into the future, and those for the smallest populations are reliable for much shorter periods.

Absolute error exhibits some variation over time, as shown by Figure 2, but without a discernible pattern. It is therefore reasonable to use past error distributions as a basis for estimating the likely errors of current forecasts. It is not really possible to predict bias – whether sets of population forecasts will be under- or over-forecast overall – as revealed by Figure 4. It would therefore be better to focus on estimating only future absolute error.

4. Warning users about likely forecast error

The renowned demographer the late Professor Nathan Keyfitz argued for the provision of information about forecast uncertainty very nicely when he wrote,

> "Demographers can no more be held responsible for inaccuracy in forecasting population 20 years ahead than geologists, meteorologists, or economists when they fail to announce earthquakes, cold winters, or depressions 20 years ahead. What we can be held responsible for is warning one another and our public what the error of our estimates is likely to be". Keyfitz (1981 p. 579)

Datasets of past errors provide valuable information which can be used to warn users about the likely error in current sets of local area population forecasts. In this section we discuss two complementary measures. First, we introduce the concept of forecast 'shelf lives' which indicate how far into the future a forecast is likely to remain within a certain error margin, and then second, we demonstrate how empirical prediction intervals can be applied.

4.1. Forecast shelf lives

If the past errors illustrated in this paper prove a reasonable guide to the magnitude of errors in the future, then approximate 'shelf life' estimates for current local area forecasts can be assumed. For example, the shelf life might be defined as the number of years ahead 80% of forecasts for local areas remain within 10% APE. Taking 80% as the cut-off covers the majority of forecasts but excludes the more volatile tail-end of the error distribution (Lutz et al. 2004).

Figure 6 shows the number of years ahead 80% of areas of a particular population size can be expected to lie within 10% APE. Beyond this shelf life less than 80% of areas will have errors within 10%. The shelf life estimates were created by taking 80th percentile APEs for 28 population size categories at 5, 10, 15 and 20 years ahead, smoothing the trends over population size, and then interpolating over time to find the number of years ahead the 80th percentile APE reached 10%. Reading off the smoothed trend line, we can estimate that for areas of about 10,000 people 80% of forecasts can be expected to be within 10% APE for about 9½ years ahead. For areas of 50,000 people the equivalent shelf life is around 12½ years, whilst for areas with jump-off populations of 100,000 it is about 13½ years.

[Figure 6 about here]

Of course, these shelf life estimates contain several limitations. They are only approximate indicators of the longevity of forecasts. They apply error distributions from large numbers of past forecasts to individual areas based on smooth curves fitted to fairly noisy original data, and they do so simply on the basis of jump-off population size (and not any other characteristics). The shelf life estimates also assume zero jump-off error, which is only valid for forecasts launched from finalised census year ERPs. Nonetheless they provide a simple indication of how far into the future a forecast is likely to remain usable and when it has reached its 'use by' date.

Forecasts beyond the recommended shelf life should be disregarded by users in the majority of situations. Forecast producers could consult shelf lives to decide how much of their forecasts to report. If they applied the shelf life concept as a cut-off for publication it would result in much

shorter forecast horizons being published than the 20 to 30 years commonly found today. As Figure 6 shows, the smallest populations have very short shelf lives. However, it is important to stress that it would not be sensible to consider the shelf life as the interval necessary between rounds of forecasts. While Figure 6 shows that 80% of forecasts of nearly ten years ahead for areas of population 10,000 can be expected to have errors of less than 10%, this does not mean that no new forecasts are needed for ten years. The forecast of 2025 made from 2015 becomes closer and less useful for some purposes as each year passes; furthermore, a forecast updated with the latest population estimates in say 2020 would reduce the expected errors for 2025.

4.2. Empirical prediction intervals

Empirical prediction intervals can be created from past error distributions, and then applied to current local area forecasts to quantify the likely range of error. The idea is not new. Keyfitz (1981) examined past errors of several sets of United Nations population forecasts for all countries with more than one million people. On the basis of the error patterns he estimated a two-thirds prediction interval for a population forecast for the USA. The official forecast at the time was for a population 260 million by the year 2000. Keyfitz (1981 p.587) wrote,

> "If the forecasts now being made are as good as those of the past and no better, and if the unpredictable twists and turns of the components of the population are as great as those of the past but no greater, then we can take it that the error now being made is drawn from the same distribution as past error, so that the chance that the 2000 population will fall between 245 and 275 million is two thirds ..."

A few other researchers have also created prediction intervals for current population forecasts from past error distributions, but such contributions remain rare. At the national scale they include Stoto (1983) and the US National Research Council's Panel on Population Projections (Bongaarts and Bulatao 2000, chapter 7). At the subnational scale, empirical prediction intervals have been created for US states by Smith and Sincich (1988), for US counties by Rayer, Smith, and Tayman (2009), for small areas of San Diego by Tayman, Schafer, and Carter (1988), and for Australian states and territories by Wilson (2012). Interestingly, the official county population projections for Florida include low and high values which represent the upper and lower bounds of a 75% prediction interval based on past error distributions (Rayer and Wang 2015). Many demographers have employed probabilistic methods to quantify forecast uncertainty (Bijak et al. 2015) but subnational applications of these methods are rare. The few subnational examples which do exist are for large subnational regions and are complex, data-hungry, and time-consuming to produce (e.g. Lee, Miller, and Edwards 2003; Wilson 2013). Empirical prediction intervals are much simpler and easier to handle at the local scale. They can also simply be 'added on' to existing local area forecasts without requiring any change to forecasting methods and processes.

We calculated empirical prediction intervals for total local area populations using Absolute Percentage Error distributions. There were three steps. First, the 80th percentiles of APE distributions were calculated for forecasts at 5, 10, 15 and 20 year forecast horizons with jump-off populations in 28 size categories (e.g. 0-999, 1,000-1,999, etc.). Second, the 80th percentiles were plotted against the mid-point values of the population size categories at 5, 10, 15, and 20 year horizons. Third, fractional polynomial curves were fitted to the original data to eliminate noise. Figure 7 shows the smooth curves and original data for the 80th percentile APEs. The equations for the curves enable the 80th percentile of APE to be estimated for any jump-off population size from

500 to 150,000 at 5, 10, 15 and 20 years ahead. The equations, and the R code used to estimate them, are given in the Appendix.

[Figure 7 about here]

The 80th percentile of past APE distributions can be interpreted as an 80% prediction interval for a current population forecast. It is based on the assumption that (i) future errors will be the same as those of the past, and (ii) individual local areas will experience the same errors as all those in their population size category. The upper and lower bounds of the 80% prediction interval for any local area population forecast may be calculated as:

Upper bound = Forecast + 80th percentile APE / 100 × Forecast

Lower bound = Forecast – 80th percentile APE /100 × Forecast.

where the 80th percentile APE refers to the smoothed curves shown in Figure 7.

Prediction intervals calculated in this way are, of course, approximate only. They are symmetrical because they are based on *absolute* PE applied either side of the forecast. More refined intervals could incorporate asymmetry, which in many cases would allow for more uncertainty above, rather than below, the main forecast. And like the shelf life estimates, these prediction intervals can only be applied to forecasts based on finalised census year ERPs. But they at least enable 'ballpark' guides to the uncertainty inherent in local area population forecasts to be calculated and visualised with minimal data inputs.

In describing prediction intervals calculated in this way to users, forecast producers should note their strengths and weaknesses, and stress that they constitute conditional measures of uncertainty. They represent the possible range of error in a current set of population forecasts for the majority of forecasts providing that the error distributions of the past, and thus implicitly the demographic regime of the past, apply in the future. Extensive changes in demographic trends or local planning schemes, the closure of an area's main employer, as well as major disasters, will invalidate this assumption. Even without these types of shocks the 80% interval is not a maximum-minimum range: an estimated 20% of population trends will turn out to lie outside the 80% interval. And some local areas may well experience very large errors due to specific local factors.

To illustrate the use of our empirical prediction intervals they were calculated for selected New South Wales local government area population forecasts (New South Wales Department of Planning & Environment 2016). Figure 8 shows the forecasts and estimated 80% prediction intervals for the local government areas of Brewarrina, Forbes, Georges River, and Western Plains Regional. Brewarrina is the smallest of the four example populations and has very wide prediction intervals. After 20 years they span 30% either side of the main forecast, and make it impossible to say whether there will be long-run population growth or decline. The larger population of Forbes has narrower prediction intervals, but it is nonetheless subject to considerable forecast uncertainty. The two forecasts shown at the bottom of Figure 8 have much narrower intervals, though the prediction interval for Georges River is only marginally narrower than that for Western Plains Regional even though its jump-off population is about three times the size. This reflects the modest declines in error with increasing population size beyond about 10,000 (Figures 1 and 7).

[Figure 8 about here]

Prediction intervals and their lower and upper bounds could be helpful to users who input population forecasts into their planning or budgeting models, or decision-making processes. In addition to the main population forecast, they could also use the population numbers at the upper and lower bounds of the prediction interval to determine whether decisions, plans, budgets or policies would remain the same if those lower or higher populations were to eventuate.

5. Conclusions

In this paper we have presented the results of a large-scale evaluation of local area population forecast errors in Australia. In line with many other studies it was found that:

- local area population forecasts suffer from quite large errors relative to national and state populations, and relative to the amount of error acceptable to most users,
- forecast error declines in a nonlinear manner with increasing population size,
- error increases approximately linearly with increasing forecast horizon,
- there is some degree of variability in absolute error over the decades but without clear or interpretable pattern, and
- bias is variable over time and largely unpredictable.

We made use of the past error distributions to provide information for users about the magnitude of error likely in current local area population forecasts. We discussed the concept of a forecast shelf life, defined as the number of years into the future a specified proportion of forecasts (e.g. 80%) of a certain population size is likely to remain within a particular error margin (e.g. 10%). We went on to calculate 80% prediction intervals for local area forecasts based on the distribution of past errors, and illustrated them on some recently published local government area forecasts.

Clearly, there are limitations to our evaluation of past forecasts and our methods of using past error distributions to illustrate likely future in current forecasts. Due to data availability we were only able to assess total (and not age-specific) populations, and we were also unable to include jump-off error due to the recasting of 1991-2011 ERPs by the ABS. The shelf life estimates and empirical prediction intervals assume that future error will resemble that of the past, and that error is dependent only on jump-off population size and forecast horizon. Obviously these are simplifications.

Nevertheless, we believe that shelf lives and empirical prediction intervals will be useful for many users and producers of local area population forecasts. The methods suggested are simple, quick and easy to implement, and do not require specialist expertise or lots of staff time. It is better to include approximate indications of uncertainty for users than none at all, which is often the current practice. And these methods offer a simpler alternative to highly complex, data-hungry and time-consuming probabilistic methods, which in any case have yet to be fully developed for local areas. They also provide an evidence-based approach to uncertainty, which traditional high and low projection variants do not. Many studies have shown how such variants give misleading and inconsistent indications of forecast uncertainty (e.g. Lee 1999; Keilman, Pham, and Hetland 2002; Bell, Wilson, and Charles-Edwards 2011).

Further research could focus on creating more refined 80th percentile APEs which take into account the variability of past population change, growth rates, and other factors which have been shown to affect error, such as the proportion of mining employment, the variability of past growth, and the

proportion of the population identifying as Indigenous (Wilson and Rowe 2011; Wilson 2012). It would also be sensible to validate the prediction intervals based on past errors by applying them to sets of old local area population forecasts not included in the present study. The key question would be 'Did the 80% prediction intervals contain 80% of actual populations?'. It would also be interesting to see if the error distributions calculated for Australian local areas are similar to those in other countries.

In addition to research on communicating forecast uncertainty, efforts are needed to try to reduce the magnitude of errors in local area population forecasts. This is more challenging. It may also be worth investigating totally new methods of forecasting, moving beyond the limitations of current models to experiment with new tools, such as automatic forecasting software (Hyndman and Khandakar 2008) and machine learning (Mullainathan and Spiess 2017).

Acknowledgements
This study was supported by the Australian Research Council (Discovery Project DP150103343). The authors would like to extend their thanks to the state and territory government demographers who generously supplied past projections and advice for this study.

References

ABS (2013). *Australian Demographic Statistics*. December quarter 2012. Catalogue No. 3101.0. Canberra: ABS.

Bell, M., Wilson, T., and Charles-Edwards, E. (2011). Australia's population future: probabilistic forecasts incorporating expert judgement. *Geographical Research,* 49(3), 261-275.

Bijak, J., Alberts, I., Alho, J., Bryant, J., Buettner, T., Falkingham, J., Forster, J. J., Gerland, P., King, T., Onorante, L., Keilman, N., O'Hagan, A., Owens, D., Raftery, A., Ševčíková, H., and Smith, P. W. F. (2015). Probabilistic Population Forecasts for Informed Decision Making. *Journal of Official Statistics,* 31(4), 537–544.

Bongaarts, J., and Bulatao, R. A. (Eds.) (2000). *Beyond Six Billion: Forecasting the World's Population*. Washington DC: National Academy Press.

Hyndman, R., and Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software,* 27(3), 1-22.

Isserman, A. M. (1977). The accuracy of population projections for subcounty areas. *Journal of the American Institute of Planners,* 43, 247-259.

Keilman, N. (1999). How Accurate are the United Nations World Population Projections? In Lutz, W., Vaupel, J. W., and Ahlburg, D. A. (Eds.) *Frontiers of Population Forecasting* (pp. 15-41). New York: Population Council.

Keilman, N., Pham, D.Q., and Hetland, A. (2002). Why population forecasts should be probabilistic – illustrated by the case of Norway. *Demographic Research,* 6, 409-454.

Keyfitz, N. (1981). The Limits of Population Forecasting. *Population and Development Review,* 7(4), 579-593.

Lee, R. (1999). Probabilistic approaches to population forecasting. In Lutz, W., Vaupel, J. W., and Ahlburg, D.A. (Eds.) *Frontiers of Population Forecasting* (pp. 156-190). New York: Population Council.

Lee, R., Miller, T., Edwards, R. D. (2003). The Growth and Aging of California's Population: Demographic and Fiscal Projections, Characteristics and Service Needs. UC Berkeley: Center for the Economics and Demography of Aging.

Lutz, W., Sanderson, W. C., and Scherbov, S. (2004). The end of world population growth. In Lutz, W., Sanderson, W. C., and Scherbov, S. (Eds.) *The End of World Population growth in the 21st Century* (pp. 17-83). London: Earthscan.

Mullainathan, S., and Spiess. J, (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31(2), 87-106.

New South Wales Department of Planning & Environment (2016). 2016 NSW population and household projections. Online at http://www.planning.nsw.gov.au/projections. Sydney: Department of Planning & Environment.

Office for National Statistics (2008). Subnational Population Projections Accuracy Report. Office for National Statistics: London.

Office for National Statistics (2015) Subnational Population Projections accuracy report. Office for National Statistics: Titchfield.

Rayer, S., and Wang, Y. (2015). Projections of Florida Population by County, 2015–2040, with Estimates for 2014. Bureau of Economic and Business Research, University of Florida.

Rayer, S., Smith, S. K., and Tayman, J. (2009). Empirical Prediction Intervals for County Population Forecasts. *Population Research and Policy Review*, 28(6), 773-793.

Rayer, S., Smith, S. K. (2010). Factors affecting the accuracy of subcounty population forecasts. *Journal of Planning Education and Research*, 30, 147-161.

Rayer, S. (2008). Population Forecast Errors: a Primer for Planners. *Journal of Planning Education and Research*, 27, 417-430.

Rayer, S., and Smith, S.K. (2014). Population Projections by Age for Florida and its Counties: Assessing Accuracy and the Impact of Adjustments. *Population Research and Policy Review*, 33(5), 747-770.

Smith, S. K., Tayman, J., & Swanson, D. A, (2013). *A Practitioner's Guide to State and Local Population Projections*. Dordrecht: Springer.

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

Smith, S., and Sincich, T. (1988). Stability Over Time in the Distribution of Population Forecast Errors. *Demography*, 25, 461-474.

Smith, S.K., and Rayer, S. (2011). An Evaluation of Population Forecast Errors for Florida and Its Counties, 1980–2010. Bureau of Economic and Business Research, University of Florida.

Statistics New Zealand (2008). How Accurate are Population Projections? An Evaluation of Statistics New Zealand Population Projections, 1991-2006. Statistics New Zealand: Wellington

Stoto, M. A. (1983). The Accuracy of Population Projections. *Journal of the American Statistical Association*, 78(381), 13-20.

Tayman, J., Smith, S. K., and Rayer, S. (2011). Evaluating Population Forecast Accuracy: A Regression Approach Using County Data. *Population Research and Policy Review*, 30(2), 235-262.

Tayman, J. (2011). Assessing Uncertainty in Small Area Forecasts: State of the Practice and Implementation Strategy. *Population Research and Policy Review*, 30, 781-800.

Tayman, J., Schafer, E., and Carter, L. (1998). The Role of Population Size in the Determination and Prediction of Population Forecast Errors: An Evaluation Using Confidence Intervals for Subcounty Areas. *Population Research and Policy Review*, 17, 1–20.

Tye, R. (1994). Local Authority Estimates and Projections: How are They Used and How Accurate Do They Need to Be? Working Paper 9, Estimating with Confidence project. Department of Social Statistics, University of Southampton, UK.

Wilson, T. (2011). A Review of Sub-Regional Population Projection Methods. Report to the Office of Economic and Statistical Research. Queensland Centre for Population Research, School of Geography, Planning and Environmental Management, The University of Queensland, Brisbane.

Wilson, T. (2015). Short-term forecast error of Australian local government area population projections. *Australasian Journal of Regional Studies*, 21(2), 253-275.

Wilson, T. (2012). Forecast Accuracy and Uncertainty of Australian Bureau of Statistics state and territory population projections. *International Journal of Population Research*, Volume 2012

Wilson, T. (2013). Quantifying the uncertainty of regional demographic forecasts. *Applied Geography*, 42, 108-115.

Wilson, T., and Rowe, F. (2011). The Forecast Accuracy of Local Government Area Population Projections: a Case Study of Queensland. *Australasian Journal of Regional Studies*, 17(2), 204-243.

Yamauchi, M., Koike, S., and Kamata, K. (2017) How Accurate Are Japan's Official Subnational Projections? Comparative Analysis of Projections in Japan, English-Speaking Countries and the EU. In Swanson, D. A. (ed.) *The Frontiers of Applied Demography* (pp. 305-328). Dordrecht: Springer.

Appendix

80th percentile Absolute Percentage Errors for the calculation of empirical prediction intervals can be estimated from the following equations. For 5 years after the jump-off year:

$$\text{80th percentile APE} = 2.310501 + 10.97458\ (\text{jump-off population}/1000)^{-0.5}$$

and for 10 years ahead:

$$\text{80th percentile APE} = 6.023633 + 13.12815\ (\text{jump-off population}/1000)^{-0.5}.$$

For 15 years ahead:

$$\text{80th percentile APE} = 9.316877 + 17.74707\ (\text{jump-off population}/1000)^{-0.5}$$

and for 20 years ahead:

$$\text{80th percentile APE} = 14.11185 + 21.22259\ (\text{jump-off population}/1000)^{-0.5}.$$

These assume zero jump-off error and are applicable to populations between 500 and 150,000 at jump-off.

The R code used to estimate the equations was as follows.

```
#Clean memory
rm(list = ls())
#Load packages
library(foreign)
#Read data
apedata <- read.dta("apedata.dta")
head(apedata)
summary(apedata)
#Scale population by 1000
avpop_sc <- apedata$avpop/1000
#Fractional polynomial terms used
pop_weight05 <- avpop_sc^(-.5)
#5 years
m5y <- lm(pct80_5y ~ pop_weight05, data=apedata)
summary(m5y) # show results
#10 years
m10y <- lm(pct80_10y ~ pop_weight05, data=apedata)
summary(m10y) # show results
#15 years
m15y <- lm(pct80_15y ~ pop_weight05, data=apedata)
summary(m15y) # show results
#20 years
m20y <- lm(pct80_20y ~ pop_weight05, data=apedata)
summary(m20y) # show results
```

The input data file apedata.dta read in by the above code was:

avpop	pct80_5y	pct80_10y	pct80_15y	pct80_20y
500	18.3	23.5	35.19	47.05
1500	11.37	18.94	21.16	25.16
2750	8.67	13.53	19.57	26.43
4250	6.96	12.27	19.4	26.04
6000	6.85	12.84	18.86	22.41
8000	5.37	8.67	13.39	19.33
10000	5.83	12.03	18.64	24.07
12000	4.93	10.04	13.81	19.28
14000	5.23	10.25	14.45	21.64

17000	4.82	8.47	12.75	18.55
20500	3.85	7.28	10.92	13.51
24000	4.31	7.62	13.24	25.67
28000	4.43	8.09	10.06	15.07
32500	4.49	7.56	12.18	18.71
37500	4.38	8.43	14.12	18.35
42500	4.31	9.7	15.03	25.57
47500	4.26	6.88	10.88	14.54
52500	4.81	7.63	11.07	14.51
57500	4.31	8.59	10.34	11.15
65000	4.55	7.05	12.06	20.75
75000	3.7	9.57	15.81	25.51
85000	3.34	9.13	10.96	15.54
95000	2.91	6.01	10.21	12.14
105000	2.24	7.37	11.61	16.07
115000	3.71	6.25	7.21	10.01
125000	3.34	6.76	12.36	18.27
135000	3.63	6.49	9.07	13.97
145000	3.38	7.7	11.69	17.47

Table 1: Mean Absolute Percentage Errors (%): Australian local areas and Florida counties

Forecast horizon	Australian local areas	Florida counties	Jump-off population	Australian local areas	Florida counties
5 years	4.5	4.9		*10 year forecast horizon*	
10 years	8.2	7.8	0-14,999	10.2	8.8
15 years	11.0	10.9	15,000-49,999	5.5	8.9
20 years	15.4	14.7	50,000-199,999	4.3	8.0

Sources: Authors' calculations; Smith and Rayer 2011

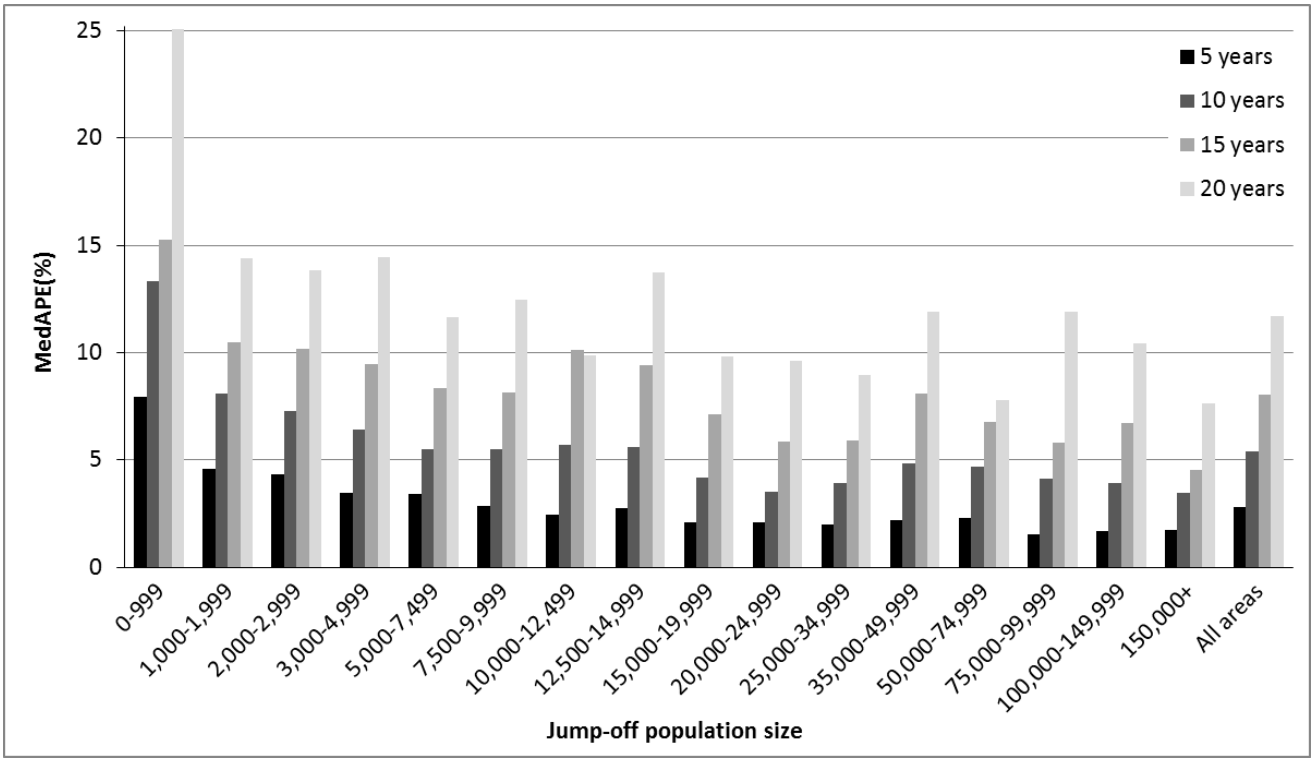

Figure 1: Median Absolute Percentage Errors of local area population forecasts by population size category and forecast horizons of 5, 10, 15 and 20 years

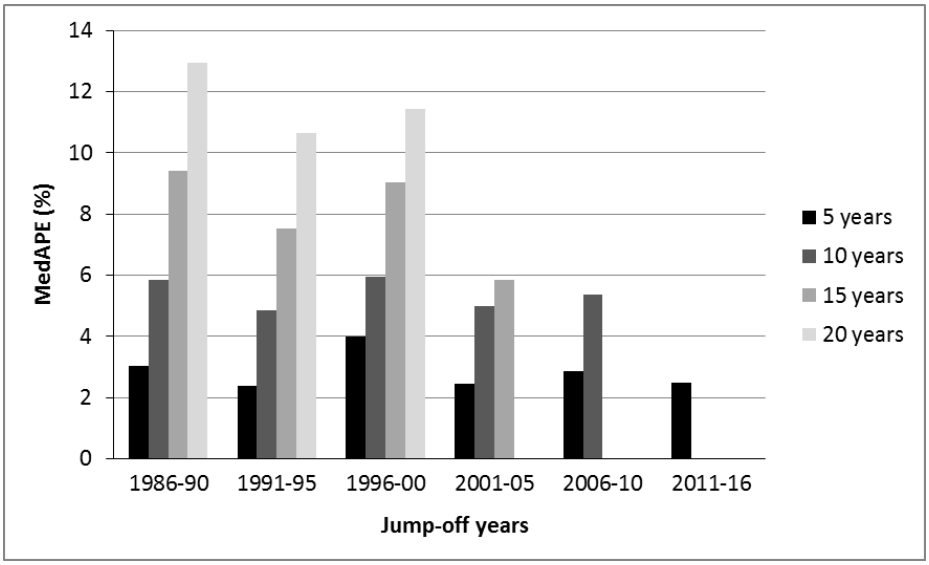

Figure 2: Median Absolute Percentage Errors of local area population forecasts jump-off years and forecast horizons of 5, 10, 15 and 20 years

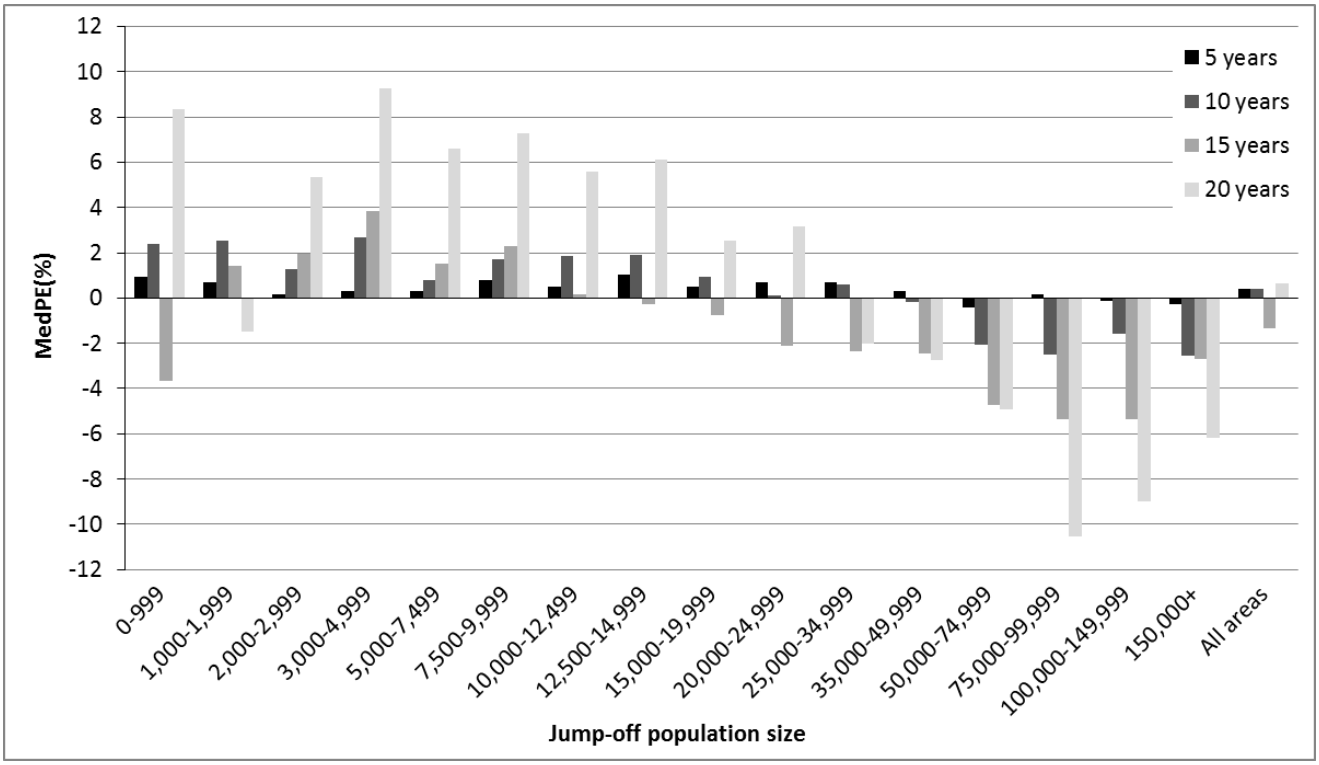

Figure 3: Median Percentage Errors of local area population forecasts by population size category and forecast horizons of 5, 10, 15 and 20 years

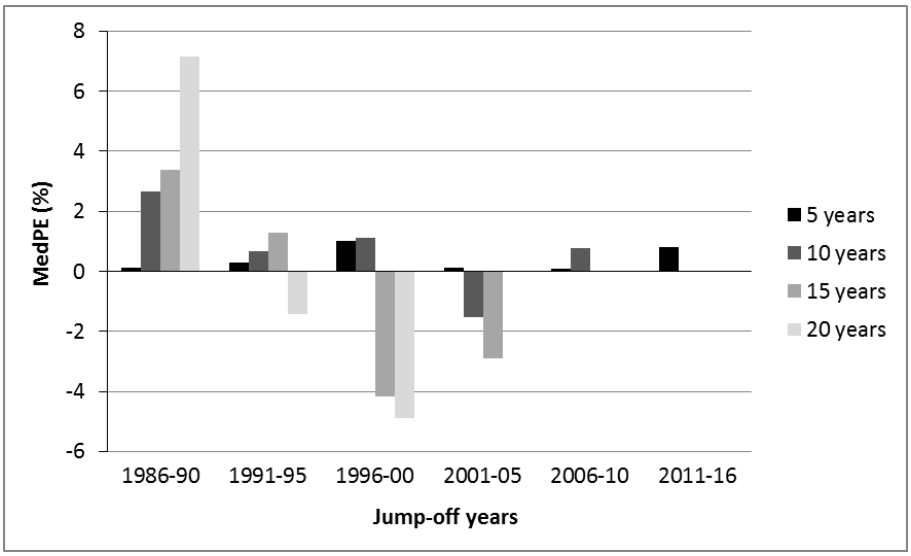

Figure 4: Median Percentage Errors of local area population forecasts jump-off years and forecast horizons of 5, 10, 15 and 20 years

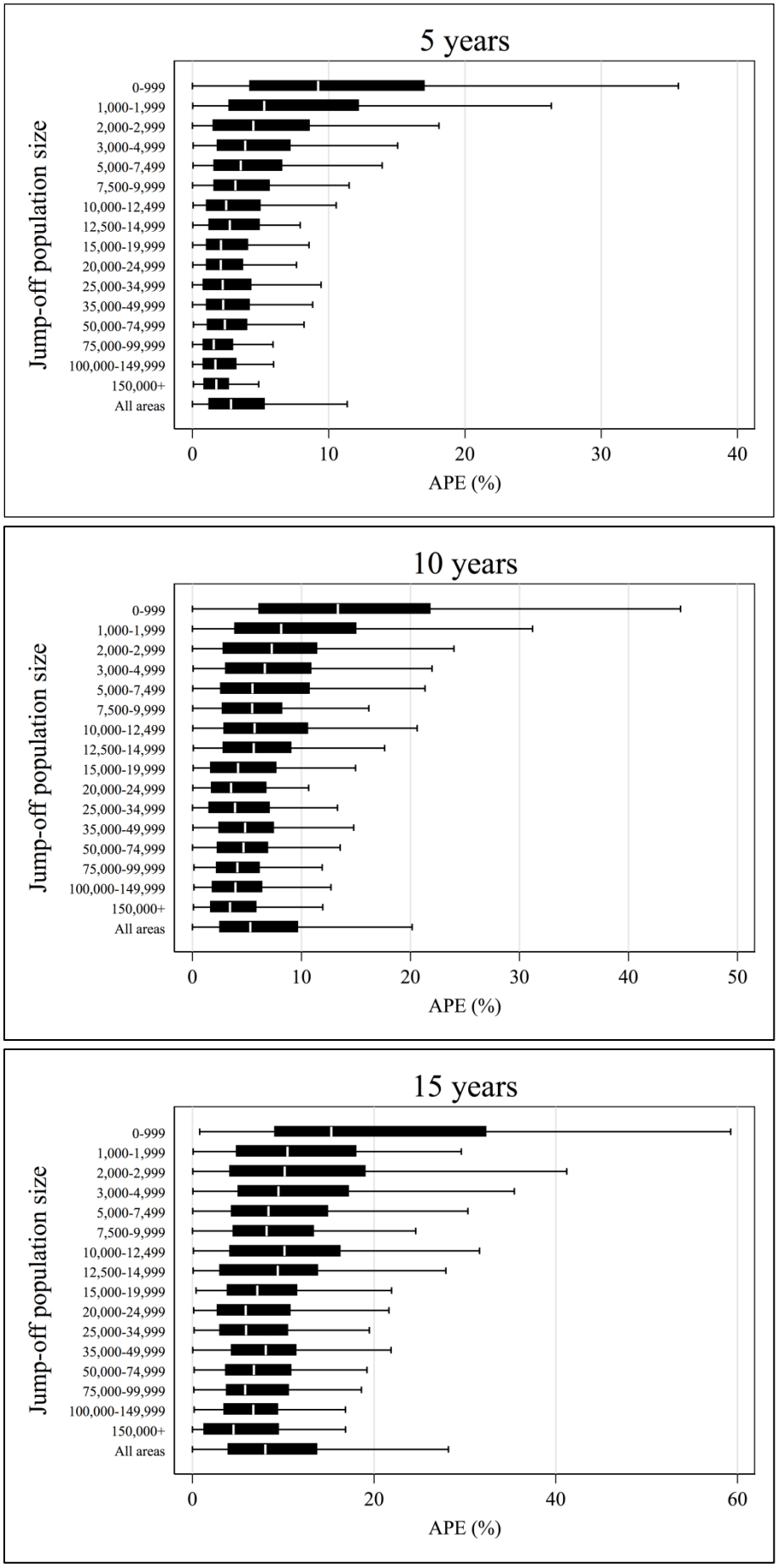

Figure 5: Distributions of APEs by population size category and forecast horizons of 5, 10, 15 and 20 years

Note: Horizontal axes vary in scale.

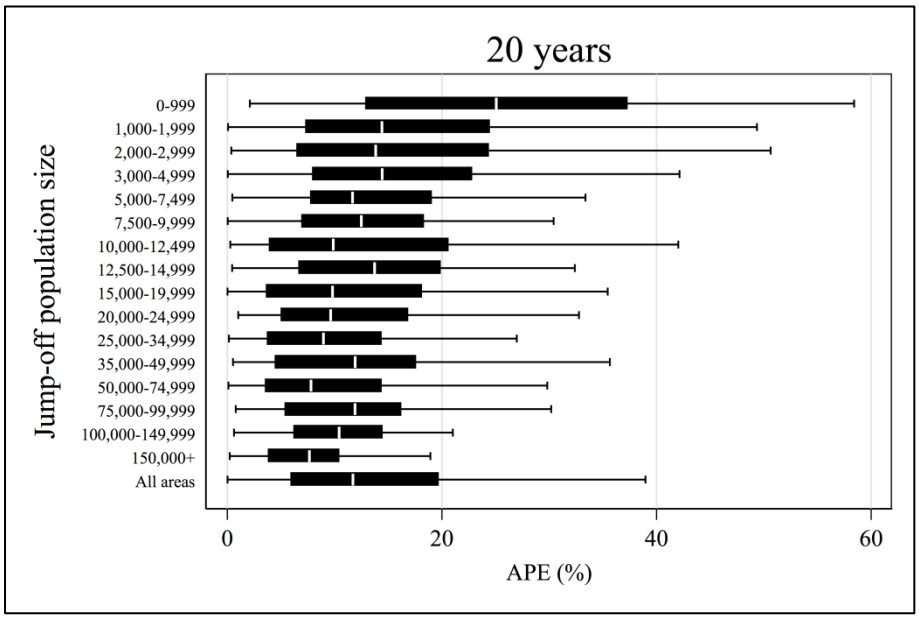

Figure 5 continued

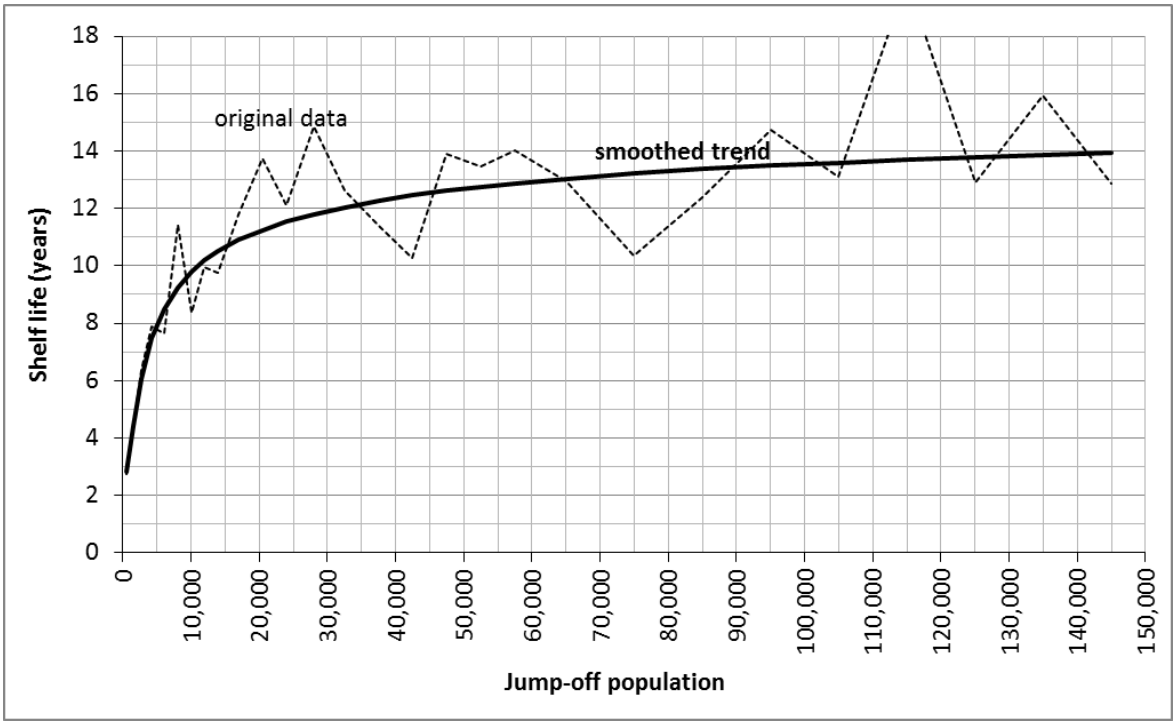

Figure 6: 'Shelf lives' of local area population forecasts showing the number of years into a forecast horizon 80% of local area population forecasts are likely to be within 10% APE

Note: The smoothed trend was obtained by interpolating the smoothed 80th percentile APE values shown in Figure 7.

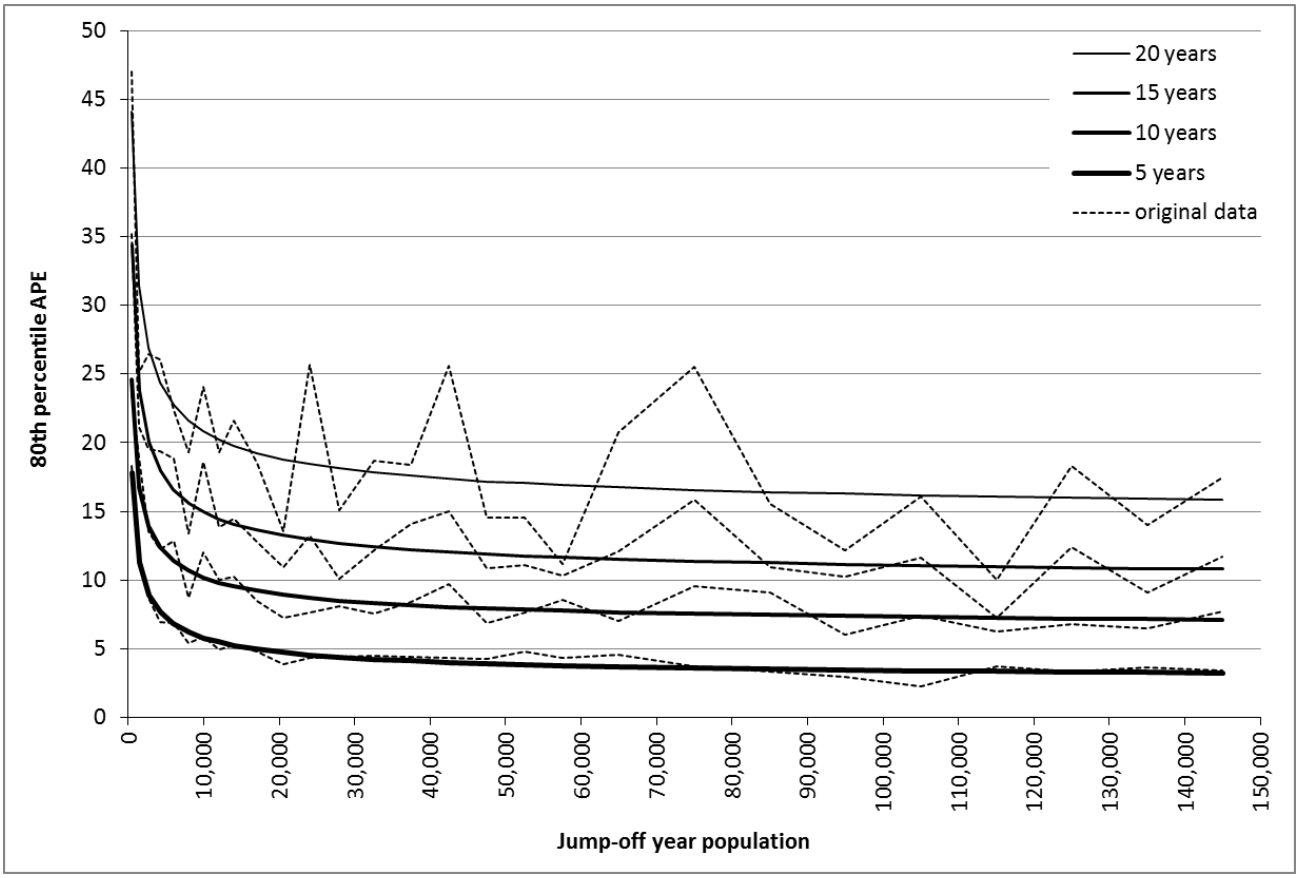

Figure 7: Smoothed and original 80th percentile values for Absolute Percentage Error by population size and forecast horizon

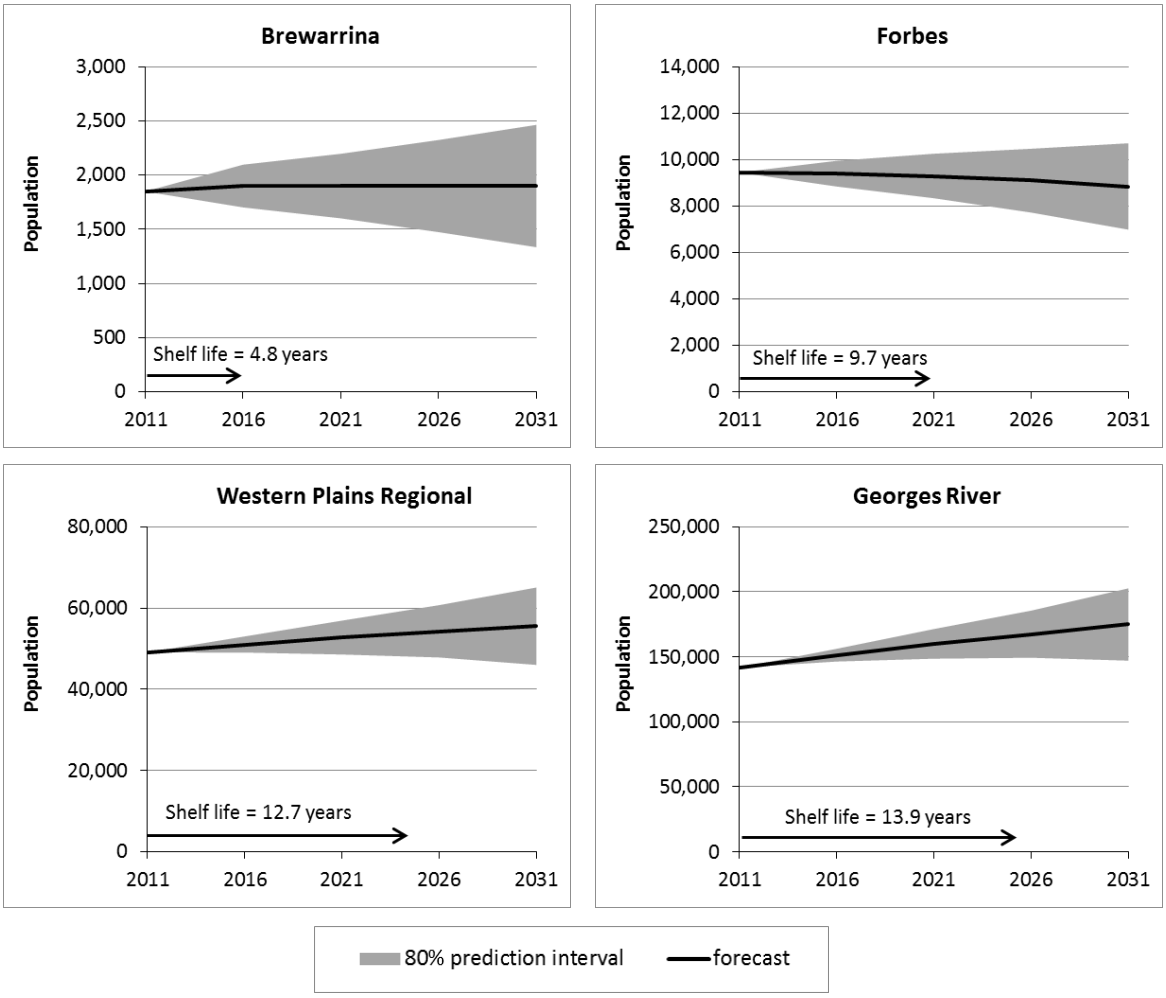

Figure 8: Population forecasts and 80% prediction intervals for four local government areas of New South Wales

Source: New South Wales Department of Planning & Environment (2016); authors' prediction intervals.