ON THE IMPORTANCE OF BEING UNCERTAIN: FORECASTING POPULATION FUTURES FOR AUSTRALIA

Heather Booth

Knowledge of population futures is central to a wide range of policy and planning issues. Population projections have traditionally been the source of this knowledge, but projections are merely what-if scenarios, based on unrealistic assumptions without claim as to their probability of occurrence. Stochastic population forecasts provide a best estimate of the future population with a prediction interval. Based on stochastic forecasts of population components, a population forecast allows for changes in the level and pattern of these components while correctly taking correlations between them into account. Stochastic forecasts provide consistent prediction intervals for all population indices. These advantages of stochastic forecasts are illustrated with respect to the Australian population, in particular with respect to ageing and migration.

Ever since Malthus, demographers have been concerned with the estimation of the future size and structure of human populations. Today, population projection is by far the most common application of demographic knowledge. Population projections are extensively used in social and economic planning as well as in population planning itself. Projections may refer to any geographical level from small areas to the entire world.

In contrast to other disciplines, demographers draw a distinction between a projection and a forecast: a population projection is the calculation of what will occur if certain assumptions are made, while a population forecast is a best estimate of future outcomes. According to the Dictionary of Demography, ‘Demographers have in theory insisted that, given the inherent unpredictability of human behaviour, they can make only projections’. Bourgeois-Pichat went so far as to distinguish between four ‘views’ of population futures. According to him, a projection is the formulation of future consequences of past developments, or a view of the future in which the past is reflected. A prospect is a view of the future that is considered possible without judgement about the probability of its actual occurrence. A forecast implies judgement that the stated outcome will probably occur. Finally, a prediction implies that the stated outcome is considered certain. While prediction is best left to soothsayers, forecasting is clearly the ideal to which demographers should aspire.

In practice, the distinction between projections and forecasts is often blurred by the producers of projections and not fully appreciated by their users. Thus, official projections are often regarded as forecasts in business and government decisions, and in policy formulation. This paper points to the limitations of population projections and argues for the adoption of population forecasting based on statistical methods.

POPULATION PROJECTIONS: LIMITATIONS

Demographers traditionally use population projections based on the cohort-component method to represent population futures. The projection is a scenario indicating what will happen if certain assumptions are made about the future of the three components of population dynamics, namely mortality, fertility and migration. The projection calculation is straightforward demographic
accounting. Typically, up to three assumptions are made for each component. These are combined in different ways to produce several projection scenarios — usually three — labelled High, Medium and Low. These projections are deterministic in the sense that each scenario is precisely specified with no possibility of variation.

Though widely used, population projections have several limitations. Importantly, they lack both objectivity and an indication of how likely they are to occur. While some considerable thought and consultation may inform assumptions about future vital rates, they remain subjective decisions devoid of probabilities of occurrence. This means, for example, that there is no reason why the Medium scenario should coincide with any statistical measure of central tendency, such as the mean, median or mode, of a distribution of possible futures. Indeed, there is no quantifiable relationship between them. Neither do the High and Low scenarios provide an indication of how likely it is that the future population will fall within their range, and they bear no relation to probabilistic prediction intervals. In practice, the High and Low scenarios are often ignored by users while the Medium scenario is almost universally regarded as a representation of what will occur. In other words, the Medium projection is used as a forecast.

A further limitation relates to the rigidity of the assumptions. In traditional population projections, a component is assumed to remain on a fixed trajectory through time, whether constant, increasing or decreasing. For example, if fertility is assumed to be on the high trajectory, it remains on this high trajectory for the entire projection period. The possibility that fertility may change between trajectories over the projection period, say from high to low, is simply not allowed. Where migration is concerned, this limitation is particularly problematic because of its sensitivity to both political and economic factors.

Limitations also arise from the need to be selective in combining assumptions to form scenarios. If three assumptions are made for each component, then 24 of the 27 possible combinations are rejected when only High, Medium and Low population scenarios are required. While some combinations can be more easily rejected as unlikely to occur, this exercise inevitably involves subjective choice. Moreover, it sets perfect correlations between levels of components that may be quite extreme while disallowing combinations involving more moderate levels that may be more realistic. The 2002 official population projections for Australia has two mortality, three fertility and three migration assumptions, giving a possible 18 combinations from which three are chosen. The High projection (Series A) is the result of combining the highest fertility with the highest life expectancy and highest net migration. Conversely, the Low projection (Series C) is the outcome of the combination of lowest fertility, lowest life expectancy and lowest net migration. But how likely are such combinations to occur? Might it not be more realistic, given the predominance of the notion of replacement migration and economic concerns about dependency ratios to combine low fertility with high migration and high life expectancy?

The number of possible combinations is commonly reduced (whether intentionally or not) by restricting the number of assumptions, often by including only one mortality assumption on the grounds that differences in mortality have little effect on the population outcome. In Australia, the 1999 and several earlier official population projections included only one
mortality assumption. The 1973 National Population Inquiry also took this approach. Recent research suggests, however, that differences in mortality can easily have a greater effect on population size than differences in fertility. Moreover, this effect is at older ages where most deaths occur. By disallowing differences in future mortality, this approach ignores the significant role of mortality in determining the size of the older population, which is of increasing concern.

In practice, projections are also limited with respect to the age pattern embodied in future age-specific rates of mortality, fertility and net migration. Often, assumptions are essentially restricted to the level of these components, that is life expectancy, total fertility and total net migration, while the age pattern is taken to be constant or undergoing only uni-dimensional change. The complexities of changing age patterns are not taken into account.

An important consequence of these limitations is that the ranges of indices derived from the population structure of the High, Medium and Low scenarios may be misleading. This applies in particular to ratios, such as the dependency ratio, where the combination of structural effects in the High and Low scenarios may produce a narrower (or possibly wider) range than would be produced if multiple intermediate projections were also examined. In other words, the limits defined according to population size may not be appropriate for population structure.

It might be argued by proponents of traditional population projections that the care taken in formulating the component assumptions is sufficient to give population projections the status of forecasts in the sense that the ‘stated outcome will probably occur’. However, the above battery of limitations goes far beyond the decisions involved and is sufficient to bring into question even the most careful population projection exercise. In order to overcome these limitations, more sophisticated methods are required.

**STOCHASTIC POPULATION FORECASTS: ADVANTAGES**

Recent developments in the assessment of demographic futures have included the adoption of a range of stochastic or probabilistic forecasting methods. Such methods are used to produce a forecast of the future population with a probabilistic prediction interval. Stochastic population forecasts overcome most, if not all, of the limitations of scenario-based population projections and are advantageous in several important respects.

First, the forecast itself represents the median of a distribution of future populations. The percentiles of the distribution provide information about uncertainty. It is therefore possible to specify with a given level of certainty the future range of population size. Rather than the High, Medium and Low scenarios of population projections, a forecast is a statistically-derived best estimate with a prediction interval. Similarly, probability statements can be made about other indices relating to the forecast population and its structure.

Second, the components of a population forecast are themselves stochastic forecasts. For each component, the forecast and its variability are derived from past experience using statistical methods. The subjective assumptions of traditional population projections are thus replaced by a statistical best estimate and its distribution. Further, since the methods normally involve modelling age-specific rates, changes in the age patterns of components form an integral part of the forecast.

Third, the distribution of possible future populations is generated by Monte
Carlo simulation, whereby simulated future populations are based on repeatedly randomly generated values of the mortality, fertility and migration parameters. In each of the several thousand simulations, each component follows a trajectory selected stochastically according to its estimated distribution. Thus, changes in component levels are a feature of stochastic forecasting. Further, the need to select combinations of assumptions does not exist: each simulation is a random combination of components, and correlations between components are correctly taken into account.

Fourth, and arguably the most important, stochastic forecasts, generated as described above, produce consistent probabilistic medians and prediction intervals for all indices. In other words, for each index the forecast is the median of its distribution and its prediction interval contains the appropriate percentage of population trajectories. The ranking of population trajectories from the Monte Carlo simulation differs between indices. Thus, the population trajectory for median population size is unlikely to be exactly the same population trajectory for the median of another index such as the proportion in a specific age group or the dependency ratio. Neither are the population trajectories defining a given prediction interval the same. This means that a population index derived from the High, Medium and Low population projections (defined according to population size) may bear little relation to the forecast index and its prediction interval.

A STOCHASTIC POPULATION FORECAST FOR AUSTRALIA
Recent research into the future of the Australian population has produced the first national stochastic population forecast. A similar stochastic forecast is used in this paper in order to illustrate the advantages of stochastic forecasting. This forecast is based on stochastic mortality and fertility components but excludes migration. The discussion and Figures 3 and 4 thus refer to the future ‘ex-migration’ population.

The mortality forecast, shown in Figure 1, embodies a more rapid increase in life expectancy than that assumed by the Australian Bureau of Statistics (ABS) in the corresponding official projections. Forecast fertility, shown in Figure 2, falls between the ABS assumptions: total fertility declines from 1.75 in 2000 to 1.64 in 2013 and remains in the range 1.63-1.64 until 2050. It is seen that variability in the forecast is considerably greater than that embodied in the ABS assumptions.

Figure 3 shows that, based on these mortality and fertility forecasts and in the absence of further migration, forecast population size would increase to 21.1 million in 2032 and then begin to decline. However, it is by no means certain that this would occur. The 95 per cent prediction limits show that ex-migration population size could begin to decline by 2024, or continue to increase for the foreseeable future (see the 95 per cent prediction limits). The 50 per cent limits show that there is a 50:50 chance that population size will peak between 2028 and 2038 at 20.6-21.9 million. By 2050, there is a 50 per cent probability that ex-migration population size will be between 19.2 and 21.7 million and a 95 per cent probability that it will be between 18.2 and 24.5 million.

Figure 3 also shows the two scenarios (labelled here as High and Low in line with the relevant fertility assumption) from the 1999 ABS population projections. It is seen that the population forecast is roughly equal to that portrayed by
Figure 1: Forecast life expectancy with 50 and 95 per cent prediction intervals and Australian Bureau of Statistics (ABS) assumption, 2000 to 2050

Source: Author’s calculations, *Population projections, Australia, 1999 to 2101*, Cat. no. 3222.0, ABS, Canberra, 2000
Figure 2: Forecast total fertility with 50 and 95 per cent prediction intervals compared with ABS assumptions, Australia, 2000 to 2050

Source: Author’s calculations, *Population projections, Australia, 1999 to 2101*, Cat. no. 3222.0, ABS, Canberra, 2000

Figure 3: Forecast population size with 50 and 95 per cent prediction intervals compared with ABS scenarios, Australia, 2000 to 2050 (excluding migration)

Source: Author’s calculations, *Population projections, Australia, 1999 to 2101*, Cat. no. 3222.0, ABS, Canberra, 2000

*People and Place*, vol. 12, no. 2, 2004, page 6
the High population projection. Since forecast fertility is lower than the High assumption, this is entirely due to lower mortality in the forecast.

The major difference between the forecast and official projections, however, is in the degree of variation. The range represented by the High and Low scenarios is by 2050 only a fraction of the 95 per cent prediction interval, and is considerably narrower than the 50 per cent prediction interval. The much wider prediction interval is the result of uncertainty in forecast fertility\(^\text{23}\) and to a lesser extent in forecast mortality.\(^\text{24}\) This difference between projected variation and forecast uncertainty is of crucial importance for planning and policy formulation in a wide range of applications.

THE STOCHASTIC FORECAST IN APPLICATION

The advantages of stochastic population forecasts over traditional deterministic scenario-based population projections are further illustrated with respect to two areas of current concern: ageing and migration.

Ageing Australia

One of the most fundamental and pervasive challenges currently facing developed societies, including Australia, is population ageing. Unprecedented changes to population structure are taking place largely as a result of past changes in fertility and past migration, but also because of significant increases in life expectancy. While the effects of the past are inevitable, the size of the elderly population over the next 50 years is subject to uncertainty arising from future mortality.\(^\text{25}\)

Accuracy in forecasting mortality is essential for forecasting the elderly population. Experience has shown that in most, if not all, developed countries the mortality assumptions used in official population projections have consistently underestimated the increases in life expectancy that have in fact occurred.\(^\text{26}\) This conservatism is the result of subjective assumption-making, and is based on widespread belief that some upper limit to life expectancy will soon be reached.\(^\text{27}\) Recent US research suggests that forecasts are much more reliable.\(^\text{28}\)

In Australia, official mortality assumptions have been similarly conservative.\(^\text{29}\) The effect of such conservatism is to underestimate the size of the elderly population. Figure 4 compares the forecast population aged 85+ with the projected population of that age based on the single mortality assumption in the 1999 official projections (as in Figure 3, migration is excluded). Since fertility does not affect the elderly population during the period of the forecast, the difference between these population futures is entirely due to lower forecast than assumed mortality. It is seen that not only does the projected population fall short of the forecast, but by 2021 it falls below the lower 50 per cent prediction limit and by 2050 it is only just within the 95 per cent prediction interval. Further, this single measure provides no indication of the likely range.

A similar forecast by Booth and Tickle\(^\text{30}\) suggests that the 1999 official Medium projection underestimates the population aged 65+ in 2031 by about five per cent and the population aged 85+ by about 22 per cent. Such underestimation has serious implications for policy formulation and planning in the areas of health and aged care and social security, as well as for the demand for housing and consumer goods and services designed for the elderly. Recent longevity forecasts for older Australian cohorts also demonstrate longer life expectancies than...
Migration and Australia

In Australia, many of the concerns about population, including ageing and sustainability, are focussed on migration. In that it increases population size, migration is seen as a major threat by environmentalists. On the other hand, positive net migration is promoted by the proponents of economic growth for its role in augmenting both the labour force and the domestic market. Migration is also often held to be a solution to structural ageing caused by past declines in fertility; though it has been shown that modest levels of net migration will achieve this effect, higher levels are subject to diminishing returns.32

While the forecasts presented above do not include migration, they are of relevance to the migration issue. By excluding migration they demonstrate Australia’s underlying population future against which a forecast that includes migration may be compared. The inclusion of (positive) net migration in the population forecast would, of course, increase population size; it would also increase the width of the prediction interval.

The incorporation of migration in population forecasting necessitates a stochastic forecast of net migration (and preferably separate stochastic forecasts of immigration and emigration). Many

Figure 4: Forecast population aged 85+ with 50 and 95 per cent prediction intervals compared with the ABS scenarios, Australia, 2000 to 2050 (excluding migration)

Source: Author’s calculations, Population projections, Australia, 1999 to 2101, Cat. no. 3222.0, ABS, Canberra, 2000
observers maintain that stochastic methods are inappropriate for migration because of the degree of unpredictability inherent in its political and, to a lesser extent economic, determinants. Forecasters point out, however, that this is precisely the situation where stochastic forecasting is most relevant and most needed.33 Not only is the forecast based on past equally-unpredictable experience, but the incorporation of variability represents a major advance on assumptions of constant migrant numbers (or rates) for the duration of the forecast. A current research need is to forecast migration for Australia and to produce a fully stochastic population forecast.34

The benefits of such an approach are widespread. For example, it would significantly increase the utility of studies such as that by Foran and Poldy, who use three population projection scenarios to examine the linkages between population and resource depletion.35 These scenarios differ only in terms of net migration: zero for the Deep Green scenario, 70,000 per annum for the Policy scenario and 0.67 per cent of population size per annum for the Business scenario; accordingly, population size in 2050 ranges from 20 to 25 to 32 million. This scenario-based approach suffers from the limitations discussed above: it is devoid of probabilities of occurrence, its fixed trajectories are unrealistic, and since no variation is permitted in mortality and fertility, correlations between these components and migration are not taken into account.36 Thus the approach precludes appreciation of the likely range of possibilities. While the range represented by the three migration assumptions is unrealistically wide, variation arising from mortality and fertility is absent. In particular, the Deep Green scenario, which is comparable with the ex-migration forecast (Figure 3), is about two million higher in 2050 than the lower 95 per cent prediction limit of the forecast. As the lowest scenario, it thus fails to adequately represent the lower limit of population size. Similar problems arise at the upper limit. It is evident that population-scenario-based studies are of limited use as a basis for policy.

ADOPTING STOCHASTIC FORECASTING

The clear advantages and technical superiority of stochastic population forecasts over conventional population projections raises the question as to why population forecasting has not been adopted by official agencies both in Australia and elsewhere. Among the obstacles are the complexities involved, the relative recency of the methods, a lack of resources, the momentum of tradition,37 and a belief (by many statisticians in particular) that users are unable to appreciate the sophistication of forecasts, especially their prediction intervals. While most of these obstacles are irrefutable, but not insurmountable, the last is open to debate.

The understanding of the nature of conventional population projections often differs between producer and user. For the producer, the projections are what-if scenarios demonstrating the results of certain assumptions, without concern for the specification of probabilities of occurrence nor indeed for the likelihood that they would ever occur. The Australian Bureau of Statistics makes this clear in the opening sentences of its 2002 projections: ‘These projections are not intended as predictions or forecasts, but as illustrations of growth and change in the population which would occur if certain assumptions … were to prevail’.38 Similarly, McDonald states: ‘The purpose is not to predict future population but to provide the evidence of
what would happen to Australia’s population if certain scenarios were to prevail’. In other words, the projections are not intended to represent what will probably occur in the future; they are not intended as a serious basis for planning.

Yet users are obliged to base plans and policies on these scenario-based projections. For them, the Medium projection represents what is most likely to occur: in other words, the projection is used as a forecast, despite the qualifications that producers seek to impose. Users often perceive the High and Low projections to be of little relevance, despite exhortations by producers to take them into account. If users have been unable or unwilling to take on board the range of possible futures embodied in the High and Low scenarios, as statisticians have claimed, it may be because such scenarios are not entirely rational: they are, after all, usually based on fixed assumptions or combinations of assumptions that are judged unlikely to occur. Might not users have realised some of the limitations of population projections and be simply making the best use of a poor product?

The adoption of stochastic forecasting would resolve these differences in understanding and enhance both use and usefulness in practice. The forecast would indeed be the most likely eventuality, given the data and method on which it is based, bringing producer and user understanding into align. Based on historic and recent experience, at least as far as mortality is concerned, the forecast would also be likely to be more accurate than the Medium projection. Moreover, the range of population futures available to the user — defined by prediction intervals rather than High and Low scenarios — would be given meaning by its known probability of occurrence. Indeed, multiple prediction intervals, that have the advantage of being consistent with each other, can be made available to meet user needs. The concept of probability is not so complex that users are unable to appreciate its meaning. Indeed, the provision of well-explained probabilistic prediction intervals would in all probability be welcomed by users: at last they would have a range of population futures that are rational, consistent and valid for use.

CONCLUDING REMARKS

Population futures are central to a wide range of pressing concerns in Australia today and underlie many aspects of social, economic and physical planning. The ability to incorporate both greater accuracy and uncertainty in population futures into macro-economic and other models would represent a major advance in decision-making and planning.

In his opening speech to a recent international conference on population ageing and health, the Federal Minister for Employment and Workplace Relations called for further research to support evidence-based policy: ‘we need greater predictive capacity to enable us to make policy decisions now to shape the future social and economic potential of our respective nations’. Clearly such evidence should be as accurate as possible. It should also correctly encompass the uncertainty involved.

Acknowledgments
I am grateful to Rob Hyndman, Peter McDonald, Len Smith, Leonie Tickle and Chris Wilson for useful comments.

References
1 In fact, demography appears to be alone among disciplines in its insistence that its projections are somehow not forecasts of the future.
5 For national projections, the subject of this paper, only international migration is of interest. For sub-national projections internal migration must also be taken into account.
6 A prediction interval is a confidence interval for a forecast.
7 Population Projections, Australia, 2002-2101, Cat. no. 3222.0, Australian Bureau of Statistics (ABS), Canberra, 2003. At the state and territory level, internal migration and the spatial distribution of international migration results in further possible combinations.
12 Numbers of migrants are often used instead of rates but the argument still applies.
14 The forecast is of course dependent on the particular method adopted and data used.
15 The cohort-component method applies as in deterministic population projections.
16 It is also possible, though more complicated, to use a stochastic Leslie matrix to generate the mean forecast and its variance. See R. D. Lee and S. Tuljapurkar, op. cit., 1994.
17 Ibid.
18 Indeed, the median population at a given time is not necessarily the median population at any other time.
20 This forecast is based on single-year age-specific mortality and fertility rates for 1968 to 2000. The base period for mortality is determined using methodology in H. Booth, J. Maintosh and L. Smith, op. cit., 2002. For fertility, the same period was used for this illustrative analysis; further research is needed to determine the optimum fertility forecast. Research on the stochastic forecasting of migration has yet to be comprehensively addressed.
21 The forecast does not correspond to the 2002 official projections because of data availability problems for recent years (the forecast is based on data by single years of age). This does not detract from the illustrative value of the comparison.
22 As there is no migration and only one mortality assumption, the entire variation is obtained from the two fertility assumptions (see Figure 2).
especially after 2020 when the number of women of childbearing age begins to have a stochastic basis.


25 In Australia, migration is largely restricted to the younger age groups; see Population Flows: Immigration Aspects, Department of Immigration, Multicultural and Indigenous Affairs (DIMIA), Canberra, 2001, p. 7. Over the 50-year period, future migration would increasingly add to uncertainty in the older age groups.


34 Research in this area is planned as part of the Population Modelling and Forecasting Initiative at the Australian Centre for Population Research, ANU. See also Wilson and Bell, op. cit., 2003.

35 B. Foran and F. Poldy, ‘Between a rock and a hard place: Australia’s population options to 2050 and beyond’, People and Place, vol. 11, no. 1, 2003, pp. 1-15. The three scenarios are all based on a total fertility rate of 1.74 in 2001 declining to 1.65 in 2011 and constant thereafter, and life expectancy increasing at a rate of one year in every ten to 77 years for males and 83 for females in 2061 and constant thereafter.

36 In addition, the Foran and Poldy scenarios involve the emigration of large numbers of older people — a highly unlikely eventuality arising from not taking the complexities of changing age patterns into account.

37 Not least involving a range of interested parties, among whom consensus on assumptions is sought.


40 Multiple prediction intervals, such as the 50 and 95 per cent intervals presented in this paper, relate to the same forecast and distribution of possible futures. While multiple scenarios can also be produced, these bear little relation to each other unless only one assumption is varied.