Ways to Improve Population Forecasting: What Should Be Done Differently in the Future?

DENNIS A. AHLBURG
WOLFGANG LUTZ
JAMES W. VAUPEL

TO IMPROVE POPULATION forecasting in the future, demographers should conduct more thorough assessments of the accuracy of past projections. Research should also focus on making greater use of: (1) models that include marriage, divorce, cohabitation, morbidity, and other demographic events that influence fertility, mortality, and migration as well as models that break populations down by educational achievement, employment status, and other variables; (2) models that take account of economic, social, and environmental dynamics, including integrated structural models and models with constraints; and (3) forecasting approaches that systematically quantify uncertainty. A further area that requires rethinking is the appropriate use of expert judgment in population projections. Finally, new ways need to be developed for distributing software for making population forecasts and for disseminating the results of alternative forecasts.

The accuracy of past projections

Although it is the single most important property of a forecast, accuracy has been treated in a cavalier fashion in the past. Most agencies that produce forecasts do not systematically investigate the accuracy of their earlier forecasts. This is wrong-headed because such studies could improve the accuracy of future forecasts. And improved forecasts could improve the quality of decisionmaking about important policy issues.

Many demographic projections have been grossly inaccurate. As Nathan Keyfitz (1996: xii) wrote about population projection: "The best demographers do it, but none would stake their reputation on the agreement of their forecasts with the subsequent realization—in fact some of the most eminent demographers have been the authors of the widest departures."

Improved forecasts could help produce better policy decisions. Consider, for instance, the enormous financial implications of alternative trends of future population aging for social security expenditures. Small differences in the projected size of the population above retirement age correspond to billions of dollars of future government expenditures under specific pension schemes. Inaccurate forecasts may even lead to the bankruptcy of certain funds. Better knowledge of the path of future aging could help policymakers to design more efficient and cheaper adjustment strategies.

We have three general suggestions concerning research on forecast accuracy. First, demographers need to study the past performance of earlier projections to uncover persistent patterns of errors, as Keilman has done in this volume. Forecasters have made systematic errors in all three components: fertility, mortality, and migration. A clearer recognition of recurring biases—such as the repeated underestimation of rates of mortality improvement at older ages—should help correct a major source of forecast inaccuracy.

Second, researchers should use criteria for model assessment and measures of forecast accuracy that are appropriate to the context. When there is a well-defined decisionmaker confronted with a problem, it may be useful to evaluate a forecasting method on the basis of the loss function of the forecast user. Forecasting methods that are suitable for projecting the population of a small school district over the next few years may be ill-suited for projecting the population aged 80 and older nationwide over the next several decades. A forecast may turn out to be accurate along one dimension—total population size, for example—but highly inaccurate on other dimensions. The number of younger people may be overestimated, for instance, while the number of older people is underestimated. Hence it is important that assessments of forecasting accuracy pay meticulous attention to errors in projected fertility, mortality, and migration arising at various ages, as well as to errors in the estimated size of the initial population.

Finally, research on forecast accuracy should include comparative analyses of forecasting models at different levels of structural complexity and of models based on different methodological approaches. The objective of comparative analysis should not be to find the best method—this, we believe, would be a futile exercise. Rather, the goal should be to gain a deeper understanding of what kinds of models perform best in what kinds of contexts. It may well be that forecasts based on some average or combination of forecasts from several different but plausible models will prove to be more accurate—or, at least, less wildly inaccurate—than forecasts based on a single approach. Such gains may be greatest when the forecasts one combines are based on different model specifications and sets of information. Results from the few studies that have been carried out to date on combining demographic forecasts are encouraging.

If it turns out that forecasts can be improved by combining forecasts based on different approaches, then it becomes all the more important to encourage different groups of forecasters at different institutions to develop alternative methods. The key question should be whether or not a new method helps to produce more accurate forecasts when averaged or combined with other methods.

Dimensionality and disaggregation

Population forecasts are being used today in an increasing number of fields ranging from small-area marketing research to global carbon dioxide emissions. Hence users of forecasts are demanding more population forecasts with special features. Cohort-component methods in their multi-state form can handle considerable expansion, as is shown for household forecasts and forecasts by educational achievement in this volume. Not only can these models provide users with detailed forecasts, they can also be used for highly informative policy simulations. However, there is often a tradeoff between increasing complexity on the one hand and diminishing possibility of evaluating the plausibility of all specific assumptions involved on the other. In fact, complexity may not necessarily lead to greater accuracy. As demographers build more detailed multi-state models, the relationship between complexity and accuracy needs to be investigated. These comments hold both for the macrosimulation models presented in the chapters in this volume and for microsimulation models. Microsimulation, which is particularly useful in projecting large numbers of individual characteristics, is likely to become more and more popular as the power of computers increases.

Another dimension that needs to be reconsidered is the length of forecast horizon that is feasible. Is it in fact the case that forecasts beyond 20 to 30 years are of questionable usefulness, as seems to be the common wisdom in demography? Or have new developments in methods and the explicit treatment of uncertainty lengthened the useful forecast horizon? There are no clear answers. Again, much depends on the purpose of the projections and the needs of the users. Even in the extreme case of climatologists demanding scenarios to the year 2200 as input to their models of global warming, demographers should not refuse to perform the calculations (which would mean leaving the job to less competent people), but they should make explicit the associated uncertainties and assumptions.

Structural complexity

We recommend that the practice of relying almost exclusively on stand-alone population forecasting models be fundamentally rethought. Research has shown that models that embed population forecasts in a broader social, economic, and environmental framework can produce more accurate forecasts than purely demographic models (see, e.g., Sanderson's chapter in this volume). And even if these models are not always more accurate than purely demographic models—because of the larger number of uncertain trends included—the combined forecasts of the different models are more likely to present a comprehensive picture of future trends than strictly demographic models are. However, the earlier comments referring to models of high dimensionality apply here as well: the larger the number of parameters and necessary assumptions, the more difficult it is to assess the plausibility of a forecasting model. The trick is to find the right level of structural complexity.

Models can be structurally integrated (as in the cases discussed by Sanderson). Alternatively, individual forecasts of different types of models can be compared with each other (as discussed by Cohen). Such a comparison can be made, for instance, of various food production or water consumption models—to shed light on possible constraints on population growth. Such an approach to making projections may not be easy to implement because interdisciplinary knowledge is required. Ultimately, the choice of model will have to be based on the judgment of experts from different fields about whether or not trends in other sectors are likely to have a sufficiently significant impact on population variables to warrant explicit treatment. The nature of such interdependencies may also depend on the time horizon: the issue of food security may not be sufficiently relevant to require explicit consideration within a 10-year horizon but it may become highly relevant for a 50-year horizon. What is needed as a first step is more interdisciplinary discussions among those who produce population projections.

Uncertainty

With the increasing concern about the accuracy of population forecasts, more attention is being given to different ways of dealing explicitly with the inherent uncertainty of future trends and our imperfect knowledge of the determinants of the three main demographic components: fertility, mortality, and migration. Although a large proportion of users seem to be satisfied with a "best-guess forecast," a growing number of users would prefer information about the range of uncertainties. In meeting this demand it is obvious that uncertainty in all three components needs explicit consideration. In many contexts, uncertainties concerning each of these components are so large that it would be inappropriate to disregard any of them.

There are three major ways of dealing with uncertainty in forecasting future demographic trends: scenarios, variants, and fully probabilistic projections. Scenarios are clear "if-then" statements that may correspond to specific consistent stories (including the assumed consequences of specific policies), but they are not associated with probabilities of any sort. In contrast, fully probabilistic forecasts (whether they are based on past errors, time-series models, expert judgment, or a combination thereof) provide users with probability distributions for all population parameters and all intervals of the forecast period.

Because such probabilistic forecasts require explicit assumptions that producers of population forecasts have been reluctant to make, and because they are a bit more complex in data handling, the variants approach has become a very popular compromise. Producers do not give up the safe ground of "if-then" statements, but at the same time they can provide some sort of plausible range of future population trends. Clearly, this is a compromise between two internally consistent but opposing approaches; the question is whether it is a good or a bad compromise.

We think that the current variants approach is a bad compromise for the various reasons stated in the contributions by Lee and by Lutz, Sanderson, and Scherbov in this volume. One key weakness is that it is difficult to know what to make of the low and the high variants. Typically, they are based on low and high fertility assumptions, respectively, combined with a middle assumption for mortality and migration. How can they be interpreted in terms of a plausible range of possible population outcomes? Any attempt to remedy this serious problem moves the variant approach toward either the scenario approach or the probabilistic approach. For this reason we suggest that agencies producing forecasts should phase out the variants approach and replace it with a mixture of probabilistic projections and selected scenarios, as well as employing more complex multi-dimensional or structural models. Instead of a single, deficient compromise, we think it would be better to offer users an array of forecasts.

The variants approach is a bad compromise between the scenario and probabilistic approaches. However, it may be possible to develop a more serviceable compromise, and we would suggest one approach toward this goal. Two (or more) equally likely fertility variants (one low and one high, for example) might be devised, one capturing the most likely trend in fertility if fertility is low and the other capturing the most likely trend if fertility is high. Conditional on each of these fertility variants, two (or more) equally likely mortality variants could then be formulated—a low and high mortality variant, say, for each of the fertility variants. Finally, conditional on the four fertility-mortality variants, two (or more) equally likely migration variants could be specified. The conditional nature of these variants is crucial. The basic idea is that certain conditions may be favorable to, for instance, both high fertility and low mortality, and this combination may tend to lead to low net in-migration. Also central to this approach is the stipulation that the variants are equally likely or have other specified probabilities. And the variants should capture the most likely (or, even better, median) trends under the assumed demographic regime. In the simple example above, the eight combined variants would each have a probability of 0.125. Ranked from low to high population growth, the eight combined variants would provide users with some idea of the probability distribution of likely population outcomes. Such an approach—which might be called SCOPE forecasting (for structured conditional-probability estimation)—needs, however, extensive research and discussion before it can be widely applied in demographic practice.

More generally, *all* probabilistic methods of population forecasting require more research and discussion. There is no one broadly accepted way of doing probabilistic forecasts. The two contributions to this volume by Lee and by Lutz, Sanderson, and Scherbov illustrate the difference between demographers who tend to rely more on time series models and those who give more weight to expert judgment. There is also still disagreement about whether one should rely for the actual calculations on randomly drawn (piece-wise linear) scenarios (which represent a more gradual evolution from the current practice of individual scenarios and variants) or on random walks (which are closer to real-world short-term fluctuations).

We do not view this as being a question of either-or. If probabilistic scenarios give a good enough approximation of the outcomes of short-term fluctuations and if they are more acceptable to the institutions producing forecasts, then this may be a good solution in some cases; in other cases one may choose the random walk. It is obvious that no matter which approach is chosen, experts will always play an indispensable role in defining the forecast model. And it is equally obvious that assumptions about the future will have to be based on the observed past patterns of fertility, mortality, and migration together with information about the errors of past projections and informed guesses concerning possible future structural discontinuities. The remaining differences of opinion (among experts) seem to be natural signs of an emerging methodology. Further scientific discussion is clearly needed, but this is likely to promote the production of probabilistic population projections in an increasing number of settings. Furthermore, as argued earlier, a combination of alternative methods should prove to be superior to reliance on a single approach.

The role of experts

An important question—not only in the discussion of probabilistic forecasts—is how great a role should be assigned to expert judgment. In all areas discussed so far, experts are the ones who ultimately determine the most appropriate way of studying the accuracy of past projections. And they decide on the basis of certain more or less explicit criteria how many dimensions a model should have, what the structural equations should be, if any, and how one should go about combining different forecasts.

But no matter how deeply they may have studied certain demographic issues, experts still have their personal biases and blind spots, and they tend to be overly impressed by the most recent trends. Hence it is preferable to rely on objective research findings rather than individual subjective judgments. Sometimes, however, expert judgments are the only information available. To monitor and improve expert judgments used in forecasts, it is important to encourage peer review of publications, substantive discussions at scientific meetings, and research on the accuracy of forecasts based on various kinds of expert input.

If individual experts are to define the specific assumptions of a consistent population projection, then intensive interaction among them is necessary. It may be helpful to invite experts with dissenting views to argue with each other about substantive issues. More research is needed on how best to obtain and consolidate expert knowledge. Users will demand this because they want to know what they can count on. More generally, as noted earlier, more research is needed on when and how expert judgments can be combined with objective research findings to improve forecasts.

Dissemination of software and results

New computer software and the Internet will make it increasingly easy for people without much experience in demography to produce their own population forecasts with a few keystrokes. This anticipated democratization of population forecasts carries with it both opportunities and dangers. It will make forecasts possible for specific areas or population subgroups for which no forecasts whatsoever have hitherto been carried out by producing agencies. It will also allow people with innovative ideas or knowledge of some specific area to challenge the established agencies. On the negative side, users may get confused if there are too many forecasts floating around and if they have no means to judge whether a forecast is based on a serious research effort or just a careless calculation. Hence the reputation of the producer may end up being the decisive criterion for the acceptance of a projection. But a name is not necessarily a guarantee of accuracy. A well-argued, published justification of the choice of model and the specific assumptions underlying it may be a better criterion for users.

Single-state population projection software has become popular. The dissemination of more powerful software, however, is not yet widespread enough to meet policy and research needs. As indicated in two contributions to this volume, the great potential of multi-state population projection is still broadly underestimated, because of a lack of awareness and the scarcity of easily implementable software. The same can be said of microsimulation and of some structural models. Here we see a clear need for information, training, and software dissemination.

As to the publication of forecast results by established agencies, the tendency must clearly be directed away from printing tons of thick tomes. This mode of dissemination will become unsustainable in any case if agencies move toward probabilistic projections or larger sets of alternative scenarios. An obvious solution is to publish only the key results in printed form and to put the rest in a database from which users can retrieve specific information via the Internet. An important condition for the ultimate success of this trend, which already seems to be underway, is that all the assumptions are well-documented so that users are informed about what the forecasts are based on.

Demographers need to recognize that we are experiencing times not only of rapid demographic change but also of significant institutional changes. The competitive nature of a globalized multi-disciplinary scientific community implies that the field of demography will not thrive if demographers confine themselves to cozy, esoteric niches. Demographers should pay more attention to the most important justification of the discipline in the eyes of the public, which is to produce reliable and useful information about future population trends. If we are to live up to this obligation, we must rethink population forecasting along the several dimensions that are outlined in this volume.

Reference

Keyfitz, Nathan. 1996. "Foreword," in Wolfgang Lutz (ed.), *The Future Population of the World: What Can We Assume Today?* Revised Edition. London: Earthscan.

Library of Congress Cataloging-in-Publication Data
Frontiers of population forecasting / Wolfgang Lutz, James W. Vaupel,
Dennis A. Ahlburg, editors.
p. cm.
Includes bibliographical references (p.).
ISBN 0-87834-096-3 (pbk. : alk. paper)
1. Population forecasting. I. Lutz, Wolfgang. II. Vaupel, James
W. III. Ahlburg, Dennis A.
HB849.53.F76 1999 99-21220
304.6'01'12--dc21 CIP

ISSN 0098-7921

ISBN 0-87834-096-3

Printed in the United States of America.