Knowledge Can Improve Forecasts: A Review of Selected Socioeconomic Population Projection Models

Warren C. Sanderson


Stable URL:
http://links.jstor.org/sici?sici=0098-7921%281998%2924%3E8%3C88%3E200.C0%3E2-J


Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/popcouncil.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact jstor-info@umich.edu.
Knowledge Can Improve Forecasts: A Review of Selected Socioeconomic Population Projection Models

WARREN C. SANDERSON

In 1982 NATHAN KEYFITZ wrote an article entitled “Can knowledge improve forecasts?” Its first paragraph ends with this striking assertion: “[The] rapid increase in knowledge, as presented in scores of journals and hundreds of research monographs, has not paid off in forecasting” (Keyfitz 1982: 729). In the decade and a half since his article appeared, the answer to Keyfitz’s question has changed. It is now “Yes, knowledge can improve forecasts.”

To understand the question and its answer, we have to spend a moment discussing terminology. It is clear that Keyfitz does not maintain that forecasts made in complete ignorance are likely to be as good as those made by experts in the field. By “knowledge,” he means knowledge of the socioeconomic determinants of population change that can be encapsulated in equations or computer code and applied systematically to the production of forecasts. In this chapter, we refer to forecasts that incorporate this sort of knowledge as “causal forecasts.”

By “improve forecasts” Keyfitz means doing better than national and international agencies are presently doing on the basis of noncausal methods. To keep this chapter manageable, we restrict our attention to areas that have not yet finished their demographic transitions. Since the United Nations is the most prominent producer of population forecasts for developing countries, we expand Keyfitz’s question into: “Can the use of formal models of fertility, mortality, and migration that include socioeconomic variables enable us to produce better forecasts than those published by the United Nations?” The answer is “Yes, they can.”

We begin the justification of our new answer with a discussion of the advantage of combining forecasts. We show that combining less accurate with more accurate forecasts can often produce results that are more accurate than either of them. This leads to two criteria for evaluating forecasts.
The weak criterion for forecast improvement is met when the causal forecasts combined with those from the United Nations are more accurate than the UN forecasts alone. The strong criterion is met when the causal forecasts themselves are more accurate than those produced by the United Nations.

In the following three sections, we show how knowledge has been employed in particular cases and whether the resulting forecasts meet either criteria for forecast improvement. First, we discuss the population of the world, focusing on the World3 model, which was the basis of the widely read books *The Limits to Growth* (Meadows et al. 1972) and *Beyond the Limits* (Meadows, Meadows, and Randers 1992). We show that the World3 model was not calibrated correctly in 1972. Nevertheless, it predicted world population growth slightly more accurately than did the United Nations. When the two forecasts are averaged they predict the world's population growth almost exactly. We then narrow our focus from the whole world to developing countries. Here, we consider the econometric model of David Wheeler. Wheeler's forecasts, based on a causal model, lead to improvements in forecasting that meet the strong criterion. In other words, Wheeler's forecasts were more accurate than those of the United Nations. Also, combining Wheeler's forecasts with those of the United Nations produces forecasts that are even more accurate than Wheeler's. We narrow our perspective further and consider the application of knowledge in detailed case studies of the Philippines and Kenya. Here we find causal forecasts that usually meet the strong or weak criteria. In these three sections, we demonstrate how the application of causal models can improve demographic projections. In the concluding section, we present some ideas on how to put this knowledge to even more productive use.

**The advantage of combining forecasts**

Suppose that we had two sets of forecasts in front of us, one from the United Nations and another produced by some causal model. What criteria could we apply to assess whether the use of the causal model led to forecast improvement? One obvious answer is that there are a variety of criteria in the literature such as root mean squared error, mean absolute error, and mean absolute proportional error that can be used to measure the accuracy of a set of forecasts. If the set of causal forecasts had lower values on those indicators than the UN forecast, we could say that they represented a forecast improvement. Clearly, this is not the only answer. A less accurate set of causal forecasts could also lead to forecast improvement. A new set of forecasts, derived by averaging a less accurate causal set with more accurate ones from the United Nations, could be even more accurate than the UN forecasts. Combining population forecasts has been advocated previously (see, for example, Ahlburg and Land 1992: 294; Willekens 1992:
309–310) and the advantage of combining forecasts in general is well known (see, for example, Bunn 1988; Mahmoud 1984; Makridakis et al. 1982).

Of course, not every less precise forecast can be combined with a more precise one to produce the most accurate forecast of the three. To see this more clearly, let us consider two sets of forecasts for a group of $N$ countries and a third forecast that is the average of the first two. The variable being forecast does not need to be specified here. It could be a life expectancy, total fertility rate, rate of population growth, or any other magnitude of interest. We call the ex post observed value of the error in predicting the variable for country $i$, $E_{UN,i}$ for the UN forecasts, $E_{S,i}$ for the causal forecasts, and $E_{C,i}$ for the combined set.

We can always express those errors as the sum of two terms, an average error for the group and a country-specific deviation from the average error.

$$
E_{UN,i} = \alpha_{UN} + \varepsilon_{UN,i}
$$

$$
E_{S,i} = \alpha_{S} + \varepsilon_{S,i}
$$

$$
E_{C,i} = \alpha_{C} + \varepsilon_{C,i}
$$

for $i = 1,2,\ldots,N$ (1)

where the $\alpha$'s are the average errors and therefore, the sum of the $\varepsilon$'s within each forecast group is zero.

The root mean squared errors for the UN and causal forecasts can be written as:

$$
RMSE_j = \sqrt{\alpha_j^2 + \text{var}(\varepsilon_j)}
$$

for $j = UN, S$ (2)

and the root mean squared error for the combined forecast as:

$$
RMSE_C = 0.5 \cdot \sqrt{(\alpha_{UN} + \alpha_{S})^2 + \text{var}(\varepsilon_{UN}) + \text{var}(\varepsilon_{S}) + 2 \cdot \text{cov}(\varepsilon_{UN}, \varepsilon_{S})}
$$

We will use decreases in the root mean squared error here as our indicator of forecast improvement. There are other means of judging forecasts, but, generally speaking, the conclusions that we reach will hold for those as well. Now we can define two criteria for judging forecast improvement. We will say that the causal forecasts meet the strong criterion for forecast improvement if the root mean squared error of the forecasts based on a causal model is smaller than that of the UN forecasts. We will say that the causal forecasts meet the weak criterion for forecast improvement, if when they are averaged with the UN set, the resulting forecasts have a lower root mean squared error than the UN forecasts. The crucial point to note here is that a set of forecasts can meet the weak criterion for forecast improvement even if it is worse than the first set in terms of both its average error and its variance.

In order to provide some intuition about how a worse forecast can be used to improve forecast accuracy, we have combined observed and hypothetical information in Table 1 for eight countries. We use these particular
countries because we encounter the same countries later when we discuss Wheeler's forecasts. Column 1 contains the observed percentage natural increase in the population from 1975 to 1995. Column 2 contains the prediction of that rate based on the UN 1978 assessment of population prospects (United Nations 1979). Columns 3, 4, and 5 contain three hypothetical forecasts of varying degrees of accuracy. Columns 6, 7, and 8 show the forecasts that result when those in columns 3, 4, and 5 respectively are averaged with the UN predictions.

The 1978 UN prediction has an average error of 10.9 percentage points. This indicates that, on average, the United Nations predicted more natural increase in these eight countries than actually occurred. This is not the average absolute error. That was 12.6 percentage points. Since the average natural increase in the eight countries is 60.7 percent, that average absolute error is 20.8 percent of the average natural increase. The variance of the UN 1978 forecasts is 132.1. Applying equation (1) to the average error and the forecast variance, it is easy to determine that the root mean squared error is 15.7 percentage points.

Column 3 contains a worse prediction, called Hypothetical Forecast 1. It is worse in the sense that the average error is larger than the average error of the UN forecast, the forecast variance is larger, and therefore, the root mean squared error is larger. The average of this worse prediction and the UN prediction is given in column 6. Clearly, the result is a lower root mean squared error. Hypothetical Forecast 1, then, is an example of a forecast that meets the weak criterion of forecast improvement. When combined with the UN forecast, it produces a forecast that is more accurate than the UN's.

Column 4 contains another worse prediction, Hypothetical Forecast 2. The average error there is -11.7 percentage points compared to 10.9 percentage points for the UN forecast and the forecast variance is 225.0 compared to 132.1 for the UN forecast. Nevertheless, when averaged with the UN forecast, the result is an even lower root mean squared error than was obtained in column 6. This arises largely because the average error of Hypothetical Forecast 2 has the opposite sign of the UN's forecast. When forecasts with average errors of opposite sign are combined, improvements in accuracy can occur, even when the causal forecast has a much larger forecast variance. Clearly, Hypothetical Forecast 2 also meets the weak criterion for forecast improvement.

Column 5 contains a set of hypothetical predictions that are more accurate than those of the United Nations. The average error is somewhat smaller and the forecast variance much smaller. As a result, the root mean squared error is nearly half that of the UN projections. Therefore, Hypothetical Forecast 3 meets the strong criterion for forecast improvement. Combining Hypothetical Forecast 3 with the UN predictions yields a smaller root mean squared error than the United Nations predictions, but not a smaller root mean squared error than Hypothetical Forecast 3. Beginning
<table>
<thead>
<tr>
<th>Country</th>
<th>Observed percentage natural increase—1975–95 (1)</th>
<th>UN predicted percentage natural increase (2)</th>
<th>Hypothetical Forecast 1 (3)</th>
<th>Hypothetical Forecast 2 (4)</th>
<th>Hypothetical Forecast 3 (5)</th>
<th>Average of UN and Hypothetical Forecast 1 (2)+(3)/2 (6)</th>
<th>Average of UN and Hypothetical Forecast 2 (2)+(4)/2 (7)</th>
<th>Average of UN and Hypothetical Forecast 3 (2)+(5)/2 (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico</td>
<td>64.1</td>
<td>92.6</td>
<td>85.4</td>
<td>33.5</td>
<td>69.1</td>
<td>89.0</td>
<td>63.0</td>
<td>80.8</td>
</tr>
<tr>
<td>Brazil</td>
<td>46.7</td>
<td>70.2</td>
<td>35.2</td>
<td>44.2</td>
<td>55.7</td>
<td>52.7</td>
<td>57.2</td>
<td>62.9</td>
</tr>
<tr>
<td>Nigeria</td>
<td>72.2</td>
<td>91.3</td>
<td>88.9</td>
<td>78.3</td>
<td>79.0</td>
<td>90.1</td>
<td>84.8</td>
<td>85.2</td>
</tr>
<tr>
<td>Egypt</td>
<td>61.7</td>
<td>57.7</td>
<td>69.4</td>
<td>39.3</td>
<td>67.4</td>
<td>63.6</td>
<td>48.5</td>
<td>62.5</td>
</tr>
<tr>
<td>Pakistan</td>
<td>83.7</td>
<td>80.8</td>
<td>106.5</td>
<td>52.0</td>
<td>88.7</td>
<td>93.7</td>
<td>66.4</td>
<td>84.8</td>
</tr>
<tr>
<td>India</td>
<td>49.1</td>
<td>52.8</td>
<td>74.0</td>
<td>51.4</td>
<td>65.9</td>
<td>63.4</td>
<td>52.1</td>
<td>59.3</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>62.1</td>
<td>75.6</td>
<td>81.7</td>
<td>44.8</td>
<td>72.6</td>
<td>78.6</td>
<td>60.2</td>
<td>74.1</td>
</tr>
<tr>
<td>Indonesia</td>
<td>45.9</td>
<td>51.2</td>
<td>43.2</td>
<td>48.7</td>
<td>45.8</td>
<td>47.2</td>
<td>49.9</td>
<td>48.5</td>
</tr>
<tr>
<td>Average error()</td>
<td>10.9</td>
<td>12.4</td>
<td>-11.7</td>
<td>7.3</td>
<td>11.6</td>
<td>-0.41</td>
<td>9.09</td>
<td>4.8</td>
</tr>
<tr>
<td>Forecast variance</td>
<td>132.1</td>
<td>156.3</td>
<td>225.0</td>
<td>25.0</td>
<td>62.4</td>
<td>103.1</td>
<td>41.6</td>
<td>14.0</td>
</tr>
<tr>
<td>Covariance of two forecasts</td>
<td>-19.34</td>
<td>27.5</td>
<td>4.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>15.7</td>
<td>17.5</td>
<td>18.7</td>
<td>8.7</td>
<td>14.0</td>
<td>9.03</td>
<td>11.0</td>
<td>11.0</td>
</tr>
</tbody>
</table>

\(pni = \left[1 + \frac{CBR_i - CDR_i}{1000}\right]^t\),

where \(pni\) is the percentage natural increase of country \(i\) over the period 1975–95, \(CBR_i\) is the crude birth rate in country \(i\) in five-year time period \(t\), and \(CDR_i\) is the crude death rate in country \(i\) in five-year time period \(t\).

**SOURCES:** Observed percentage natural increases are computed from crude birth and death rates in United Nations (1995). UN predicted percentage natural increases are computed from crude birth and death rates in United Nations (1979).
with Hypothetical Forecast 3 as our base, we can see that the UN forecasts do not meet the weak criterion for forecast improvement. Averaging them with Hypothetical Forecast 3 does not yield a forecast with a lower root mean squared error than that for Hypothetical Forecast 3. Combining forecasts with somewhat worse forecasts can often lead to forecast improvements, but combining forecasts with much worse forecasts (the forecast variance of the UN forecast is over five times that of Hypothetical Forecast 3) does not always do the trick.

An important feature of the three Hypothetical Forecasts in Table 1 is that the covariances between the errors in the causal forecasts and the errors in the UN forecasts are small. In reality, this is typically the case. The UN forecasts are based on expert opinion, while the causal forecasts are based on sets of explicitly specified and estimated equations. These two methodologies are sufficiently dissimilar that it would be very unusual indeed if their forecast errors were highly correlated.

Model-based predictions do not have to be more accurate than existing forecasts for them to lead to forecast improvement according to the weak criterion. Indeed, as in the cases of Hypothetical Forecasts 1 and 2, they can be considerably worse, although there is a limit on how bad they can be. If knowledge cannot lead to improvements in forecast accuracy even according to the weak criterion then knowledge not only cannot help us to make more accurate projections, it cannot even keep us from making absolutely atrocious ones. Is our knowledge base really that weak?

The World3 model

One of the most famous computer models ever written is World3, which forms the basis of the books *The Limits to Growth* and *Beyond the Limits*. It must immediately be said that the World3 model is not primarily a vehicle for making world population projections. It is a model that looks at the sorts of outcomes that arise on a world scale because of the interactions between population growth, economic growth, and environmental change. In focusing here on just the population sector of the model, we are using World3 in a way that was not intended by its authors. Nevertheless, World3 does contain a detailed and thoughtful application of knowledge to population forecasting and is therefore worth reviewing here.

The structure of the population sector of the World3 model

Fertility

The population in World3 is divided into four age groups, 0–14, 15–44, 45–64, and 65 and over. Births in any year depend on the number of women in the
reproductive age group (half of the population in the 15–44 age group) and the total fertility rate in that year. The total fertility rate in turn depends on three variables: (1) physiological maximum fertility, (2) desired fertility, and (3) the fertility control effectiveness rate. In equation terms, the relation is:

$$TFR_t = \text{MIN}\left[MTF_i, MTF_i \cdot (1 - FCE_i) + FCE_i \cdot DTF_i\right],$$  \hspace{1cm} (4)

where

$TFR_t$ is the total fertility rate in period $t$,  
$MTF_i$ is the physiological maximum total fertility in period $t$,  
$FCE_i$ is fertility control effectiveness in period $t$,  
$DTF_i$ is the desired level of the total fertility rate in period $t$.

Equation (4) says that if the desired level of the total fertility rate is above the physiological maximum rate, then the total fertility rate is equal to the physiological maximum rate. If desired fertility is below the physiological maximum rate then the total fertility rate is equal to a weighted average of the desired rate and the physiological maximum rate, where the weighting factor is the fertility control effectiveness rate. If fertility control effectiveness is 1.0, then the total fertility rate is equal to the desired rate. If fertility control effectiveness is 0.0, the total fertility rate is the physiological maximum rate.

Each of the three factors, the physiological maximum total fertility rate, the desired fertility rate, and the fertility control effectiveness rate, itself has determinants. The physiological maximum total fertility rate is the product of two terms. The first is a constant called the normal maximum total fertility rate, which is set to 12.0. The second is a multiplier that is a nonlinear, increasing, and concave function of life expectancy. When life expectancy is 40, the multiplier is 0.8, so the maximum total fertility rate is 9.6. When life expectancy is 80, the multiplier is 1.1, so the maximum total fertility rate in that case is 13.2.

The desired total fertility rate is the product of two components: (1) desired completed family size, in terms of surviving children, and (2) a multiplier that compensates for the mortality of children. This multiplier depends on the perceived level of life expectancy, which in turn depends on the actual level around 20 years earlier. When perceived life expectancy is 40, the multiplier is 1.3; when it is 80, the multiplier is 1.0.

Desired completed family size is the product of three factors, a baseline completed family size, which is a parameter set equal to 4.0, a multiplier called “family response to social norm,” and another called “social family size norm.” The “social family size norm” is a multiplier that is a downward sloping function of industrial output per capita around 20 years earlier. When lagged per capita industrial output is $200^9$ (in 1968 US dol-
lars), the multiplier is 1.0, so the baseline completed family size times the “social family size norm” is 4.0. When lagged per capita industrial output rises to $400, the multiplier is 0.9, so its product with the baseline completed family size is 3.6. The third multiplier depends on the current rate of growth of per capita industrial output. The faster the growth rate, the higher the multiplier. When per capita industrial output is not growing, the multiplier is 0.7. Therefore, if per capita industrial output were $400 and not growing, desired completed family size would be 2.5. Desired total fertility would be higher because of the compensation for infant and child mortality.

The level of desired total fertility will not be reached unless the fertility control effectiveness rate is 1.0. The fertility control effectiveness rate depends on per capita allocations to fertility control around 20 years earlier. Those expenditures, in turn, were determined by the past levels of service output per capita and the share of that service output that was allocated to family planning. The share allocated to family planning depended on the “need for fertility control” around 20 years earlier, and that was a function of the ratio of the physiological maximum total fertility rate to the desired total fertility rate at that time.

This formulation of the determinants of fertility takes into account infant and child mortality, the difference between the long-run and short-run relationship between income and fertility, and the lagged effect of spending on family planning facilities and knowledge, among other things. In the World3 model, education has no effect on fertility.

Mortality

Life expectancy at birth is specified to be the product of a constant baseline life expectancy, and four multiplier functions, which relate the baseline life expectancy to changes due to: (1) food availability, (2) health service availability, (3) persistent pollution, and (4) crowding. In equation form, we have:

\[
LE_t = LEN \cdot LMF_t \cdot LMHS_t \cdot LMC_t \cdot LMP_t,
\]

where

\(LE_t\) is life expectancy at birth in period \(t\),
\(LEN\) is the baseline life expectancy (a parameter set equal to 28),
\(LMF_t\) is the lifetime multiplier for food in period \(t\),
\(LMHS_t\) is the lifetime multiplier for health services in period \(t\),
\(LMC_t\) is the lifetime multiplier for crowding in period \(t\), and
\(LMP_t\) is the lifetime multiplier for pollution in period \(t\).

The food availability multiplier is an increasing nonlinear function of the ratio of food availability per capita to subsistence-level food consump-
tion per capita. When the ratio is 1.0, the multiplier is 1.0. When the ratio is 5.0, the multiplier is 1.4. Therefore, the difference between subsistence and affluence in food availability causes a 40 percent increase in life expectancy.

The lifetime multiplier for health services is somewhat more complex. First, a nonlinear increasing function is used to determine the amount of world service output that is allocated to health services. When world service output per capita is $500 (in 1968 US dollars), the amount spent on health services is $50 per capita. When world service output per capita rises to $1,000, the amount spent on health services is $140. These expenditures are smoothed and assumed to influence “effective health services” with a 20-year delay. An increase in effective health services per capita increases the lifetime multiplier for health services. The functional form of this increase is assumed to be different before and after 1940. After 1940, the same level of effective health services per capita translates into higher levels of the lifetime multiplier for health services.

The lifetime multiplier from crowding is 1.0 minus the product of two terms: (1) the fraction of the population that lives in urban areas, and (2) a crowding multiplier from industrialization. The fraction of the population that lives in urban areas is assumed to be a nonlinear increasing function of population size. For example, a world of 6 billion people, close to the world’s population today, would be 50 percent urban. A world of 10 billion people would be 65 percent urban. The crowding multiplier from industrialization is U-shaped. When industrial output per capita is low, this multiplier is positive and decreases with increasing industrial output. At a per capita industrial output of slightly over $200, the multiplier becomes negative. The multiplier’s lowest point comes when per capita industrial output is $400. From that point it rises slowly, becoming positive again at a level of around $850 per capita. The change of sign in the crowding multiplier for industrialization means that increases in urbanization at low and high levels of industrial output per capita decrease life expectancy, while increases in urbanization increase life expectancy when world per capita industrial output is between around $200 and $850.

The lifetime multiplier from pollution decreases as an index of persistent pollution increases. Thus, life expectancy depends on factors like the availability of food, lagged spending on health services, the interaction between the proportion urban and the level of industrial output, and the extent of pollution. Once life expectancy at birth was determined, the survival rates for the four age groups were taken from model life tables in United Nations (1956).

The World3 population projection

We are concerned here primarily with the original version of World3, World3-72 for short (Meadows et al. 1972), because it produced a popula-


<table>
<thead>
<tr>
<th>Projection/year</th>
<th>UN 1998 assessment</th>
<th>UN 1968 assessment</th>
<th>World3-72</th>
<th>World3-92</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1965</td>
<td>3.34</td>
<td>3.29</td>
<td>3.33</td>
<td>3.49</td>
</tr>
<tr>
<td>1970</td>
<td>3.70</td>
<td>3.63</td>
<td>3.64</td>
<td>3.81</td>
</tr>
<tr>
<td>1975</td>
<td>4.07</td>
<td>4.02</td>
<td>3.92</td>
<td>4.16</td>
</tr>
<tr>
<td>1980</td>
<td>4.44</td>
<td>4.46</td>
<td>4.24</td>
<td>4.53</td>
</tr>
<tr>
<td>1985</td>
<td>4.84</td>
<td>4.93</td>
<td>4.58</td>
<td>4.94</td>
</tr>
<tr>
<td>1990</td>
<td>5.27</td>
<td>5.44</td>
<td>4.93</td>
<td>5.37</td>
</tr>
<tr>
<td>1995</td>
<td>5.67</td>
<td>5.96</td>
<td>5.28</td>
<td>5.83</td>
</tr>
<tr>
<td>Crude birth rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1965–70</td>
<td>33.8</td>
<td>33.8</td>
<td>36.8</td>
<td>33.3</td>
</tr>
<tr>
<td>1970–75</td>
<td>30.9</td>
<td>33.2</td>
<td>35.4</td>
<td>32.1</td>
</tr>
<tr>
<td>1975–80</td>
<td>28.3</td>
<td>32.1</td>
<td>34.0</td>
<td>30.9</td>
</tr>
<tr>
<td>1980–85</td>
<td>27.4</td>
<td>30.9</td>
<td>32.6</td>
<td>29.6</td>
</tr>
<tr>
<td>1985–90</td>
<td>26.6</td>
<td>29.1</td>
<td>31.2</td>
<td>28.2</td>
</tr>
<tr>
<td>1990–95</td>
<td>23.9</td>
<td>27.0</td>
<td>29.7</td>
<td>26.8</td>
</tr>
<tr>
<td>Crude death rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1965–70</td>
<td>13.4</td>
<td>14.0</td>
<td>20.2</td>
<td>15.7</td>
</tr>
<tr>
<td>1970–75</td>
<td>11.6</td>
<td>12.8</td>
<td>19.1</td>
<td>14.6</td>
</tr>
<tr>
<td>1975–80</td>
<td>11.0</td>
<td>11.6</td>
<td>18.1</td>
<td>13.5</td>
</tr>
<tr>
<td>1980–85</td>
<td>10.3</td>
<td>10.5</td>
<td>17.3</td>
<td>12.3</td>
</tr>
<tr>
<td>1985–90</td>
<td>9.6</td>
<td>9.6</td>
<td>16.5</td>
<td>11.3</td>
</tr>
<tr>
<td>1990–95</td>
<td>9.3</td>
<td>8.7</td>
<td>15.9</td>
<td>10.4</td>
</tr>
</tbody>
</table>

**SOURCES:**
All United Nations numbers refer to the medium variant.
World3-72 and World3-92: Author’s runs of the World3 model programmed in VENSIM. The VENSIM version of World3-92 is available at this writing on the worldwide web at: http://www.std.com/vensim/vensim.htm.
VENSIM is a registered trademark of Ventana Systems Inc. World3-72 results were obtained by changing seven parameters back to their original levels. A description of these seven changes can be found in Meadows, Meadows, and Randers (1992): 243–251.

The projection for the world that can be compared with one published by the United Nations (1971, 1973) at about the same time. Table 2 contains estimates and projections of the population of the world, the world’s crude birth rate, and the world’s crude death rate from four sources. The first is the 1998 revision of the United Nations population estimates and projections (United Nations 1998). We take these figures as the benchmarks against which to judge the other projections. The second is the United Nations 1968 set of population estimates and projections (United Nations 1971, 1973). The third comes from the World3-72, and the last from the World3 model with the revised parameters used in *Beyond the Limits* (Mead-
ows, Meadows, and Randers 1992). That version of the model is called World3-92.

Although the World3 model technically is initialized on 1900 data, the World3-72 parameters are set so as to approximate known time trends in variables from that date until the model was constructed in 1970/71. For the most part, the latest data that the authors of World3-72 had available to them came from the late 1960s. Therefore, it is appropriate to think about the period from 1900 until the late 1960s as the calibration period, and the period from the late 1960s to 1995 as the projection period. The latest citation in Meadows et al. (1974) to a projection of the world’s population was to United Nations (1971). That projection has a base year of 1968. Fortuitously, then, we have two projections, one from Meadows et al. (1972) and the other from United Nations (1971) and United Nations (1973), that have virtually the same base year.

According to United Nations (1998) the population of the world grew by 53.2 percent between 1970 and 1995. The predicted population growth from the United Nations 1968 assessment (UN 1971, 1973) was 64.2 percent, and from World3-72, 45.1 percent. In other words, the United Nations 1968 assessment predicted population growth that was 11.0 percentage points too high, while the World3-72 model predicted population growth that was 8.1 percentage points too low. By accident, the average of United Nations and the World3 forecasts would have been almost exactly correct.

We can get a better perspective on these differences in predicted growth by looking at the crude birth and death rates. The estimate in United Nations (1971, 1973) of the world’s crude birth rate in 1965–70 was 33.8. This was the number known to the authors of World3-72, which, for that period, produced a crude birth rate of 36.8, quite a bit higher.

United Nations (1971, 1973) estimated the world’s crude death rate in 1965–70 to be 14.0. The World3-72 figure was 20.2. Obviously, World3-72 overestimated both the crude death rate and crude birth rate at the beginning of the projection period. The overestimation of the crude death rate was especially large. Indeed, the expectation of life at birth (both sexes) in 1965–70 was 53.1 years according to United Nations (1973). The same life expectancy according to World3-72 was 44.4 years, 8.7 years too low. The authors of World3-72 knew of these problems (Meadows et al. 1974: 147) and decided not to fix them because of the complexity involved and because the qualitative features of the model, not precise quantitative features, were the primary interest of World3’s authors.

In 1992, some of the authors of the original World3 model revisited it. They made seven changes in parameters or table functions, three of them in the population sector. The baseline total fertility rate was decreased from 4.0 to 3.8 and the lifetime multipliers for food and for health services were shifted upward. The result of the seven changes is the projection seen in
the column labeled World3-92. With hindsight, population growth between 1970 and 1995 now matches the observed growth to within one percentage point, although the crude birth and death rates are still too high.

Table 3 shows the age structure estimates from the same four sources used in Table 2. This table may shed some light on why death rates in the World3 model were set so high. For simplicity, we have collapsed the 15–44 and 45–64 year old age groups. Look, for a moment, at the percentage distributions of the population in 1970. In United Nations (1971) 5.2 percent of the population was 65 years old or above. In the UN 1998 assessment (United Nations 1998) that figure is 5.5 percent. In the 1972 parameterization of World3, the percentage 65 and above was 7.1 percent. So even with a life expectancy that was 8.7 years too low, World3 obtained too many people in the upper age group. Any attempt to increase life expectancy would also increase the percentage of the population 65 and older and therefore make the error in estimating the number of people in that age group even larger. This is exactly what we see when we look at the 1992 parameterization of World3, where the life expectancy is higher.

TABLE 3  World age structure estimates and projections, percentage of the population in each age group, 1970–95

<table>
<thead>
<tr>
<th>Projection/year and age group</th>
<th>UN 1998 assessment</th>
<th>UN 1968 assessment</th>
<th>World3-72</th>
<th>World3-92</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–14</td>
<td>37.4</td>
<td>37.0</td>
<td>36.7</td>
<td>35.0</td>
</tr>
<tr>
<td>15–64</td>
<td>57.1</td>
<td>57.8</td>
<td>56.2</td>
<td>56.7</td>
</tr>
<tr>
<td>65+</td>
<td>5.5</td>
<td>5.2</td>
<td>7.1</td>
<td>8.3</td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–14</td>
<td>35.2</td>
<td>36.7</td>
<td>35.4</td>
<td>33.8</td>
</tr>
<tr>
<td>15–64</td>
<td>58.9</td>
<td>57.8</td>
<td>56.8</td>
<td>57.1</td>
</tr>
<tr>
<td>65+</td>
<td>5.9</td>
<td>5.5</td>
<td>7.8</td>
<td>9.1</td>
</tr>
<tr>
<td>1985</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–14</td>
<td>33.5</td>
<td>36.3</td>
<td>34.6</td>
<td>33.2</td>
</tr>
<tr>
<td>15–64</td>
<td>60.6</td>
<td>58.2</td>
<td>57.2</td>
<td>57.3</td>
</tr>
<tr>
<td>65+</td>
<td>5.9</td>
<td>5.5</td>
<td>8.2</td>
<td>9.5</td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–14</td>
<td>32.4</td>
<td>35.5</td>
<td>33.8</td>
<td>32.5</td>
</tr>
<tr>
<td>15–64</td>
<td>61.4</td>
<td>58.9</td>
<td>57.6</td>
<td>57.5</td>
</tr>
<tr>
<td>65+</td>
<td>6.2</td>
<td>5.6</td>
<td>8.6</td>
<td>10.0</td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–14</td>
<td>31.2</td>
<td>34.3</td>
<td>33.0</td>
<td>31.7</td>
</tr>
<tr>
<td>15–64</td>
<td>62.2</td>
<td>59.8</td>
<td>58.0</td>
<td>57.8</td>
</tr>
<tr>
<td>65+</td>
<td>6.6</td>
<td>5.9</td>
<td>9.0</td>
<td>10.5</td>
</tr>
</tbody>
</table>

SOURCES: See Table 2.
The reason for the overestimation of the 65+ population is an error in demographic accounting. The annual death rate for the population 65 and above is computed as the reciprocal of life expectancy at age 65. The problem with this is that the assumption of a constant death rate, independent of age, does not work well for the population 65 and older. It would not be difficult to make a better approximation of the death rates of those 65 and above and put in into the model.

Wheeler's forecasts

David Wheeler (1984) presents population forecasts for developing countries based on a careful econometric study of the interrelations between economic change, human resource change, and demographic change. His estimates are based on an international cross-section of changes over time. The core model has eight equations. The dependent variables are changes in: (1) economic output, (2) nutrition, (3) literacy, (4) life expectancy, (5) investment, (6) the general fertility rate, (7) family planning effort, and (8) the crude death rate. The data used in the study come from 88 developing countries, although not all the countries are used in each equation. These data, which have been collected by the World Bank, refer to three time points, 1960, 1970, and 1977. The time periods over which each equation has been estimated vary and the details are provided below.

Wheeler combines his econometric results with some demographic accounting to produce a simulation model of changes in the population, in the economy, and in human resources. The model has two exogenous variables, the ratio of population to the number of doctors, and the secondary school enrollment ratio. The ratio of population to doctors is an exogenous factor influencing death rates. The secondary school enrollment ratio is taken as a proxy for the institutional capacity of the country and appears in the family planning equation. In addition to the time paths of these exogenous variables, the evolution of the endogenous variables (and functions of the endogenous variables) depends on the initial conditions in each country. The simulation model requires initial conditions for the general fertility rate, the crude death rate, population size, the size of the adult population, the investment rate, calorie sufficiency, life expectancy, literacy, and per capita income. Once the values of the estimated coefficients of the model, the initial conditions, and the exogenous variables are known, a population projection can be made.

The equations

The change in the general fertility rate (the ratio of births to women 15–49) and the change in family planning effort equations are estimated si-
mutually with three-stage least squares, using 62 observations of changes over the period 1960 to 1977. The estimated change in the general fertility rate equation is:

\[
\Delta GFR = 187.338 \cdot \Delta w^{15-49}_{25-34} + 9.63614 \cdot \Delta cdr - 0.2470 \cdot \Delta(cdr)^2
\]

\[-15.4159 \cdot \frac{\Delta y}{y_{t-1}} - 1.3122 \cdot plan\]  
(6)

The change in the general fertility rate, GFR, depends on the change in the fraction of women in the 15–49 age group who are 25 to 34 years old, \( \Delta w^{15-49}_{25-34} \), the change in the crude death rate \( \Delta cdr \), the change in the square of the crude death rate, \( \Delta(cdr)^2 \), the percentage change in per capita output, \( \frac{\Delta y}{y_{t-1}} \), and the level of the Mauldin-Berelson family planning effort index (Mauldin and Berelson 1978) at the end of the period, plan. Although education, in the form of the literacy rate, is in Wheeler’s model, it is not included in the fertility equation.

The change in family planning effort is a function of the secondary school enrollment rate, taken as a proxy for the nation’s institutional capacity and the change in the general fertility rate.

A separate equation is estimated for the change in the crude death rate over the period 1960–77 using data from 64 countries.

\[
\Delta cdr = 6.6523 - 1.06705 \cdot cdr_{t-1} + 0.0193 \cdot cdr^2_{t-1}
\]

\[+ 0.000024956 \cdot med_{t-1} + 19.3986 \cdot \Delta p_{0-14} + 19.8577 \cdot \Delta p_{50+}\]  
(7)

The crude death rate change, \( \Delta cdr \), is influenced by the initial level of the crude death rate, \( cdr_{t-1} \), and the initial value of the crude death rate squared, \( cdr^2_{t-1} \), the initial level of population per doctor, \( med_{t-1} \), the change in the percentage of the population in the 0–14 year age group, \( \Delta p_{0-14} \), and the change in the percentage of the population in the 50 and above age group, \( \Delta p_{50+} \).

The equations determining the percentage change in real GDP, the percentage change in calorie sufficiency, and the literacy change equation are estimated together using three-stage least squares to take into account the possibility of simultaneous causation. They are estimated over a pooled set of observations on changes, 39 of which come from the period 1960–70, and 43 of which come from the period 1970–77.

The percentage change in output depends on the percentage change in capital, labor, calorie sufficiency (a proxy for the effect of nutrition),
and the absolute increase in the percentage of the adult population literate. The percent change in calorie sufficiency depends on the past level of calorie sufficiency, and on the level and change in per capita output. The absolute change in the literacy rate depends on the percent change in the adult population, the percent change in output per capita, and past primary school enrollment rates.

The percent change in capital in the output growth equation is derived from an investment growth equation based on 39 observations and estimated over the period 1960–77. The absolute change in investment depends on the initial level of investment, the growth of per capita income, the absolute change in life expectancy, and the absolute change in literacy.

The percentage change in life expectancy equation is estimated separately. It uses 63 observations on life expectancy changes over the period 1970–77. The independent variables in that equation are the initial level of life expectancy, the percent change in per capita income, the past percent change in life expectancy, the past percent change in calorie sufficiency, the initial values of literacy, and population per doctor, and past primary school enrollment rates. It is interesting to note in passing that although the percentage change in life expectancy equation, which is used in the output/human resources block of the model, contains the literacy rate as an independent variable, the separately estimated change in crude death rate equation, which is used in the demographic block of the model, does not.

The estimated equations are tied together with a very simple demographic accounting structure. The number of births is equal to the general fertility rate multiplied by the number of women aged 15–49 years. Only one death rate, the crude death rate, is applied to people at each age to determine the number of deaths.

The results

Wheeler (1984) uses the simulation model to make a number of different forecasts. Here we consider only the projections he makes for eight developing countries with relatively large populations, because these projections include the detail on crude birth and death rates that we need for our evaluation below. To make a projection, Wheeler’s model requires just the initial conditions for the country involved and assumptions about the time paths of the two exogenous variables, the secondary school enrollment rate and the ratio of population per doctor. Thus, it would be easy to use this framework to make many more projections.

The first step in evaluating the results of the Wheeler model is to find a United Nations projection that is as comparable as possible. Wheeler’s equations are estimated on absolute or percentage changes between one
date and another. The latest period over which changes are estimated is 1970 to 1977. Some of the data used in those equations were published after 1977. For example, the Mauldin–Berelson index of family planning effort was published in 1978, although the figures refer to an earlier period. Therefore, it seems plausible to ascribe 1978 as the base year for Wheeler’s projections. Fortunately, there is a United Nations forecast with the same base year, United Nations (1979).

Table 4 provides the root mean squared errors of the 32 forecasts (eight countries and four five-year time periods) for the crude birth rate, crude death rate, and the rate of natural increase. The root mean squared error of the crude birth rate forecasts is 3.78 for Wheeler’s forecasts and 5.12 for the United Nations 1978 assessment (United Nations 1979). Therefore, according to that criterion, the Wheeler crude birth rate forecasts performed better than the UN forecasts. If the Wheeler crude birth rate figures are averaged with those from the UN 1978 assessment, the result is a combined forecast that does considerably better than the UN’s alone. The root mean squared error of the combined forecast is 3.53. Wheeler’s crude birth rates meet both the strong and weak criteria for forecast improvement.

The UN 1978 assessment does better than Wheeler in predicting the crude death rate. The root mean squared error of the UN forecasts of the crude death rates is 1.33, while for Wheeler’s forecasts it is 1.63. The root mean squared error of the combined forecasts of the crude death rates is 1.34, just slightly worse than the UN’s. Therefore, Wheeler’s crude death rate forecasts do not meet the strong criterion for forecast improvement and narrowly miss the weak criterion. It is not surprising that Wheeler’s framework does poorly in predicting mortality changes. Wheeler’s model has only a rudimentary age structure.

It is clear that the root mean squared errors of the crude birth and death rates from the combined forecast are only slightly worse than those from the better forecasts and much better than the worse ones. The forecasts of the crude birth rates and the crude death rates are correlated, and

<table>
<thead>
<tr>
<th></th>
<th>Wheeler</th>
<th>UN 1978 assessment</th>
<th>Wheeler and UN 1978 assessment combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude birth rate</td>
<td>3.78</td>
<td>4.82</td>
<td>3.86</td>
</tr>
<tr>
<td>Crude death rate</td>
<td>1.63</td>
<td>1.33</td>
<td>1.34</td>
</tr>
<tr>
<td>Rate of natural increase</td>
<td>3.79</td>
<td>5.12</td>
<td>3.78</td>
</tr>
</tbody>
</table>

NOTE: Errors are measured from the data in United Nations (1998).
from the perspective of forecasting future population size, it is the error in the rate of natural increase that is most important. The third row of Table 4 gives the root mean squared errors of the rates of natural increase. Here Wheeler’s forecasts do better than those of the United Nations 1978 assessment. The root mean squared error of the Wheeler forecasts of the rate of natural increase is 0.379 percent per annum compared with 0.512 percent per annum for the UN. Here the combined forecast does better than either. The combined forecast has a root mean squared error of 0.378 percent a year. Clearly, Wheeler’s predicted rates of natural increase meet both the strong and weak criteria for forecast improvement.

Using the criterion of root mean squared error of the forecasts of the rate of natural increase, we would prefer Wheeler’s forecasts to those of the UN 1978 assessment, but would prefer the combination of the two to Wheeler’s alone. Within countries, errors in forecasts of the rate of natural increase are likely to be correlated, so before we make a final assessment of Wheeler’s forecasts versus those from the UN 1978 it is useful to consider the predicted rates of natural increase over the entire 20-year period. This is done in Table 5.

Panel A of Table 5 shows the predicted and observed rates of natural increase for the eight-country sample over the period 1975 to 1995. The data are given in percents, so the figure for Mexico under UN 1998 indicates that without any international migration the Mexican population would have grown by 64.1 percent from 1975 to 1995. We focus here on Panel B, which shows the forecast errors in percentage points. For example, Wheeler’s forecast of the percent increase in the Nigerian population was 1.0 percentage points higher than the observed increase. The UN 1978 forecast of the Nigerian population was 19.1 percentage points too high and the combined forecast was 9.9 percentage points too high.

On average Wheeler’s forecasts were too low. The average of the errors is –3.0 percentage points. Wheeler’s largest errors are a forecast that is 17.9 percentage points too low for Egypt and one that is 13.6 percentage points too low for Pakistan. The UN 1978 assessment tended to err on the high side. Its average error is 10.9 percentage points. Its largest errors are a 28.5 percentage point overestimate for Mexico and a 23.5 percentage point overestimate for Brazil. In the eight-country sample, Wheeler’s forecasts have a smaller average error, average absolute error, and root mean squared error than the UN forecasts. The combined forecasts do considerably better than the UN estimates in terms of average absolute and root mean squared errors, and have a considerably smaller average error.

Table 5 shows that, on the criteria used here, the combined forecasts are worse than Wheeler’s and better than the UN’s and that the Wheeler predictions are to be preferred to those of the United Nations. Clearly, the knowledge incorporated in the Wheeler estimates improved forecasts.
TABLE 5  Twenty-year predicted and observed rates of natural increase

Panel A. 20-year rates of natural increase (in percent)

<table>
<thead>
<tr>
<th>Country</th>
<th>UN 1998</th>
<th>Wheeler</th>
<th>UN 1978</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico</td>
<td>64.1</td>
<td>68.5</td>
<td>92.6</td>
<td>80.2</td>
</tr>
<tr>
<td>Brazil</td>
<td>46.7</td>
<td>57.9</td>
<td>70.2</td>
<td>63.9</td>
</tr>
<tr>
<td>Nigeria</td>
<td>72.2</td>
<td>73.2</td>
<td>91.3</td>
<td>82.1</td>
</tr>
<tr>
<td>Egypt</td>
<td>61.7</td>
<td>43.8</td>
<td>57.7</td>
<td>50.6</td>
</tr>
<tr>
<td>Pakistan</td>
<td>83.7</td>
<td>70.1</td>
<td>80.8</td>
<td>75.4</td>
</tr>
<tr>
<td>India</td>
<td>49.1</td>
<td>41.7</td>
<td>52.8</td>
<td>47.2</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>62.1</td>
<td>64.9</td>
<td>75.6</td>
<td>70.2</td>
</tr>
<tr>
<td>Indonesia</td>
<td>45.9</td>
<td>41.0</td>
<td>51.2</td>
<td>46.1</td>
</tr>
</tbody>
</table>

Panel B. Errors (in percentage points)

<table>
<thead>
<tr>
<th>Country</th>
<th>Wheeler</th>
<th>UN 1978</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico</td>
<td>4.4</td>
<td>28.5</td>
<td>16.1</td>
</tr>
<tr>
<td>Brazil</td>
<td>11.2</td>
<td>23.5</td>
<td>17.2</td>
</tr>
<tr>
<td>Nigeria</td>
<td>1.0</td>
<td>19.1</td>
<td>9.9</td>
</tr>
<tr>
<td>Egypt</td>
<td>-17.9</td>
<td>-4.0</td>
<td>-11.1</td>
</tr>
<tr>
<td>Pakistan</td>
<td>-13.6</td>
<td>-2.9</td>
<td>-8.3</td>
</tr>
<tr>
<td>India</td>
<td>-7.4</td>
<td>3.7</td>
<td>-1.9</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>2.8</td>
<td>13.5</td>
<td>8.1</td>
</tr>
<tr>
<td>Indonesia</td>
<td>-4.9</td>
<td>5.3</td>
<td>0.1</td>
</tr>
<tr>
<td>Average error</td>
<td>-3.0</td>
<td>10.9</td>
<td>3.8</td>
</tr>
<tr>
<td>Average absolute error</td>
<td>7.9</td>
<td>12.6</td>
<td>9.1</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>9.6</td>
<td>15.7</td>
<td>10.7</td>
</tr>
</tbody>
</table>

For sources and computation method see notes to Tables 1 and 4.

The Bachue models

Bachue is the name of a Colombian goddess of love, fertility, and harmony between people and the environment. During the 1970s and 1980s four large scale "Bachue" models of the interrelationships between demographic, economic, and human resource changes were constructed at the International Labour Organisation (ILO). Two of these studies were published in book form, Bachue–Philippines (Rodgers, Hopkins, and Wéry 1978) and Bachue–Kenya (Anker and Knowles 1983), and are considered here. Like World3, the Bachue models were not designed primarily to produce population projections. They were created to conduct policy analysis on a variety of topics including population, human resources, and income inequality. For this purpose, the models are large and detailed, each one determining several thousand endogenous variables. Roughly speaking, the Bachue models are an order of magnitude larger than the World3 model.
and two orders of magnitude larger than the Wheeler model. An evaluation of them simply on their performance in forecasting population size is necessarily highly incomplete. Nevertheless, the Bachue models contain a thoughtful approach to the use of knowledge to produce population forecasts and therefore are worth considering here.

The structures of the demographic-sector models

The Bachue–Philippines model was the first in the Bachue series. It projects population by age groups (0–1, 1–4, 5–9, ...60–64, and 65 and over), sex, urban/rural location, and three education categories. Together there are 152 cells in each time period. Bachue–Philippines was one of the first models to incorporate such a four-dimensional multi-state population projection. The Bachue–Kenya model has the same four-dimensional structure, but the population is divided into single-year age groups. This results in a population divided into 792 categories.

In both models, people are born, die, migrate, marry, receive education, and form households. Economic behavior, such as consumption and labor force participation, depends on the characteristics of those households.

Fertility

In the Bachue–Philippines model the gross reproduction rate is computed from a regression using data from 47 developing countries with a population of 500,000 or above in 1969, after level corrections.

\[
GRR_k(t) = b_k - 0.0064 \cdot R_k(t-1) + 0.0106 \cdot I_k(t-1) \\
- 0.0446 \cdot E_{0,k}^0(t-1) + 0.0059 \cdot L_{a}(t-1)
\]  

(8)

The equation relates the gross reproduction rate in period \( t \) and region \( k \) (rural or urban) linearly to the labor force participation rate for women 15 to 44 in that region in the previous period, \( R_k(t-1) \), the percent of the adult population in the region in the previous period who were illiterate, \( I_k(t-1) \), life expectancy at birth in the previous period, \( E_{0,k}^0(t-1) \), and the percentage of the labor force in agriculture, \( L_{a}(t-1) \). The same equation is applied separately in both rural and urban areas with different independent variables generated for each region, with the exception of the percentage of the labor force in agriculture, which is the same in both equations. The constant terms, the \( b_k \), are adjusted to give the 1965 region-specific gross reproduction rates. Age-specific fertility rates are derived from these gross reproduction rates.

Instead of using international cross-sectional data, fertility in Bachue–Kenya is estimated from the analysis of two household-level surveys, one from 1969 and the other from 1974. The equation for the urban area is:
\[ NB_u = K_u + 0.3017 \cdot EDW2_u - 0.3338 \cdot EDW3_u \\
+ 2.3516 \cdot \frac{1}{SURVIVE3_u} + 0.7988 \cdot FMIG_u - 0.0379 \cdot STERILE_u \] (9)

In the urban area \( NB_u \), an index of fertility derived initially from equations on the number of children ever born to women, depends positively on the proportion of women in the 15–49 age group in the urban area who have primary education, \( EDW2_u \), negatively on the proportion of those women with more than primary education, \( EDW3_u \), negatively on the proportion of births in urban areas surviving to age 3, \( SURVIVE3_u \), positively on the proportion of rural immigrants among urban women 15 to 49 years old, \( FMIG_u \), and negatively on the proportion of sterile women in the urban area, \( STERILE_u \).

Many of the same variables operate to determine rural fertility:

\[ NB_r = K_r - 0.4446 \cdot EDW2_r - 1.8783 \cdot EDW3_r \\
+ 0.985 \cdot \frac{1}{SURVIVE3_r} - 0.0379 \cdot STERILE_r + 0.3202 \cdot YPERAD_r \] (10)

There are three differences to note between the urban and rural equations. First, primary education is negatively related to fertility in rural areas. Second, the migration term does not appear in the rural equation. Third, rural fertility is positively related to rural household income per adult, \( YPERAD_r \), while income does not appear in the urban equation. All the independent variables are generated endogenously in the model. The authors of the Bachue–Kenya model note: "The estimated fertility relationships embodied in Bachue–Kenya imply that fertility rates in Kenya are not very sensitive to socio-economic change. . . ." 21

Total fertility rates in Bachue–Kenya are assumed to move proportionally with the \( NB \) indices above. The initial total fertility rates are for the year 1969. Age-specific rates are determined using a fixed age-specific rate pattern.

Mortality

In Bachue–Philippines, age-specific mortality rates are determined using Coale–Demeny West model life tables. The life table used was chosen to be consistent with a life expectancy at birth derived from an equation estimated from international cross-section data. The equation, after a level adjustment, is:
\[ e_{0,k}^0 = 89.5 - 3389 \cdot (A/Y_k) + 76880 \cdot (A/Y_k)^2 - 36.47 \cdot G_k \] (11)

Life expectancy at birth in region \( k \) (rural or urban), \( e_{0,k}^0 \) is determined by an equation in the reciprocal of location-specific post-tax per capita income, \( Y_k \), the reciprocal of the square of location-specific post-tax per capita income, and the location-specific Gini coefficient of inequality in household income, \( G_k \). \( A \) is a constant used for accounting purposes.

In Bachue–Kenya, the determination of age-specific mortality rates is based on an equation estimated on 1974 household data, after adjustments. That equation determines the region-specific (urban or rural) probability of surviving from birth to age 3.

\[
SURVIVE3_r(L) = K(L) + 0.0572 \cdot [EDW2_r(L) + EDW3_r(L)] + 0.0236 \cdot YPECU_r(L) - 0.001444 \cdot YPECU_r(L)^2 + 0.0012 \cdot t
\] (12)

The probability of surviving from birth to age 3 in region \( L \) at time \( t \) depends on the proportion of women in the 15–49 age group who have completed standard 5 in primary school or beyond, \( EDW2_r(L) + EDW3_r(L) \), income per adult-equivalent consumer (children 14 and younger count as one-half an adult consumer), \( YPECU_r(L) \), income per adult-equivalent consumer squared, and a time trend. The coefficients in the equation are the same for the two locations, but the values of the independent variables and the constant terms differ. This survival rate is then used to determine an appropriate set of Coale–Demeny North model life tables.

**Migration**

In Bachue–Philippines net migration flows are determined by separately computing the rural to urban flow and the urban to rural flow. All those rates are specific for age, sex, and education groups. The factors that influence migration rates include age, age-specific marriage rates, education, wage rates, the income distributions, and the proportion of the population that is already urban. The full migration flow specification includes six equations. Because of space limitations, those equations and the ones from Bachue–Kenya are not presented here.

Net rural to urban migration in the Bachue–Kenya model is determined as the product of a total migration rate and various age, sex, and education adjustments. The total net rural–urban migration rate depends on the distribution of education in the rural area, the ratio of rural to urban household income after taxes and adjusted to a per adult-equivalent basis (children are counted as one-half an adult), and urban modern-sec-
tor earnings per worker. Age-, sex-, and education-specific net migration rates are determined by multiplying the total migration rate by adjustment factors. The age-specific adjustment factors are derived from a third degree polynomial in age. Terms are added to the age-specific adjustment factors for sex and education based on survey data.

This is only a partial presentation of the important features of the demographic sectors. In particular, we have not discussed the determination of marriage rates, education rates, and the number of households. Marriage rates do not directly influence fertility in either model, but they do influence migration rates directly or indirectly and so have an indirect effect. Useful discussions of the determination of the educational distribution of the population and the number of households can be found in the Bachue volumes themselves (Rodgers, Hopkins, and Wéry 1978; Anker and Knowles 1983) and in Wéry and Rodgers (1980).

The results

Both Bachue models are very large and were designed to study a number of issues dealing with the interaction of economic, demographic, and human resource changes. They were not designed primarily as a tool for making demographic forecasts. Nevertheless, with this caveat in mind, it is still instructive to compare the population forecasts from the Bachue models with those made by the United Nations at about the same time.

Here we compare the reference runs from Bachue–Kenya and Bachue–Philippines with UN projections. The reference runs are the baselines to which other alternatives are compared. They are essentially the authors' view of the most likely outcome and therefore the appropriate ones to use here. The reference runs are dated 1965 in the case of Bachue–Philippines and 1969 in the case of Bachue–Kenya. In both cases, though, information available after those dates was used in the model. For example, the fertility equation in Bachue–Philippines used international cross-sectional data from 1969, and the migration equations used data from the 1968 National Demographic Survey of the Philippines. These data were not used to calibrate the model, but rather to establish functional relationships. It seems plausible, in light of this, to compare the predictions of the Bachue–Philippines model with the 1968 United Nations assessment of future population prospects or perhaps even an earlier one. In the case of Bachue–Kenya, the fertility and migration equations use data collected in a 1974 household survey, but again these data were used to determine patterns of relationships, not for calibration. We could compare the Bachue–Kenya population forecasts with those from the 1968 United Nations assessment because of the 1969 starting date or with those from the 1973 United Nations assessment of future population trends (United Nations 1977) because of the household survey data.
Neither Rodgers, Hopkins, and Wéry (1978) nor Anker and Knowles (1983) give much demographic information that can be used to assess the accuracy of their forecasts. For Bachue–Kenya, forecasts are presented for 1984 and 1999, years 15 and 30 respectively in the model. For Bachue–Philippines forecasts are presented for 1985 and 2000. They are 20- and 35-year forecasts. For both models, we can discuss the accuracy of the earlier forecast with more confidence than we can the later one. We assess both models relative to the 1968 and 1973 United Nations assessments. The latest period covered by the 1968 assessment is 1980–85, so only the 1973 assessment can be used to appraise the longer forecasts.

Panel A of Table 6 shows the predictions for the Philippines in 1985. The crude birth rate and the life expectancy predicted from the Bachue–Philippines model both are closer to the actual (i.e., UN 1998 assessment) data than those predicted in either the 1968 or 1973 UN assessments. According to United Nations (1998), the crude birth rate in the Philippines in 1985 was 34.3. The Bachue–Philippines forecast was 33.0, while the United Nations 1968 assessment (United Nations 1973) predicted 39.1 and the United Nations 1973 assessment (United Nations 1977) predicted 36.8. The forecast from the Bachue–Philippines model was the most accurate of the three. Similarly, the prediction of life expectancy in the Bachue–Philippines model was more accurate than either the 1968 or the 1973 United Nations assessment. Averaging the Bachue–Philippines forecasts with those from the United Nations 1968 assessment results in better forecasts than those of the United Nations. The same is true when the Bachue–Philippines forecasts are combined with the United Nations 1973 assessment figures.


The evaluation of the forecasts for the years 1999 and 2000 is more problematic. First, we must compare the early forecasts not with what actually happened, but with a later forecast. Here we chose the medium variant of the UN 1998 assessment. Second, the UN 1968 assessment (United Nations 1971, 1973) only provides figures up to 1985, so it cannot be used. Third, the UN 1973 assessment does not provide information for 1999 or 2000. The latest rates that are presented there are for the period 1995–2000. We resolve this by assuming that the annual rate of change from the...
TABLE 6  Comparison of the demographic forecasts from Bachue–Kenya and Bachue–Philippines with those of the 1968, 1973, and 1998 UN assessments

Panel A. Philippines, 1985

<table>
<thead>
<tr>
<th>Measure</th>
<th>UN 1998 assessment</th>
<th>Bachue–Philippines</th>
<th>UN 1968 assessment</th>
<th>UN 1973 assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude birth rate</td>
<td>34.3</td>
<td>33.0</td>
<td>39.1</td>
<td>36.8</td>
</tr>
<tr>
<td>Life expectancy (both sexes)</td>
<td>63.0</td>
<td>63.2</td>
<td>66.1</td>
<td>63.8</td>
</tr>
</tbody>
</table>

Panel B. Kenya, 1984

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude birth rate</td>
<td>47.4</td>
<td>49.6</td>
<td>45.1</td>
<td>46.2</td>
</tr>
<tr>
<td>Crude death rate</td>
<td>12.6</td>
<td>13.5</td>
<td>11.4</td>
<td>12.0</td>
</tr>
<tr>
<td>Rate of natural increase (in %)</td>
<td>3.49</td>
<td>3.61</td>
<td>3.37</td>
<td>3.42</td>
</tr>
</tbody>
</table>

Panel C. Philippines, 2000

<table>
<thead>
<tr>
<th>Measure</th>
<th>UN 1998 assessment</th>
<th>Bachue–Philippines</th>
<th>UN 1973 assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude birth rate</td>
<td>27.1</td>
<td>29.3</td>
<td>27.0</td>
</tr>
<tr>
<td>Life expectancy</td>
<td>69.1</td>
<td>68.2</td>
<td>69.0</td>
</tr>
</tbody>
</table>

Panel D. Kenya, 1999

<table>
<thead>
<tr>
<th>Measure</th>
<th>UN 1998 assessment</th>
<th>Bachue–Kenya</th>
<th>UN 1973 assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude birth rate</td>
<td>33.4</td>
<td>46.3</td>
<td>40.0</td>
</tr>
<tr>
<td>Crude death rate</td>
<td>13.0</td>
<td>9.9</td>
<td>7.7</td>
</tr>
<tr>
<td>Rate of natural increase (in %)</td>
<td>2.04</td>
<td>3.64</td>
<td>3.23</td>
</tr>
</tbody>
</table>

NOTES and SOURCES:
All UN data are from the medium variant, except the data for the Philippines in 2000. Those are from the high variant.
UN 1968 assessment: Figures are extrapolations from United Nations (1973). The extrapolations assume that the figure for 1975–80 can be assigned to 1977.5 and that the figure for 1980–85 can be assigned to 1982.5. The figures in the table are linear extrapolations to 1984.5 or 1985.5. The crude birth rate data for the Philippines come from Table A.9, p. 151. The crude birth and death rate figures come from Table A.9, p. 153. Life expectancy data for the Philippines come from Table A.8.1, p. 142.
1984 or 1985: The figures in the table are linear interpolations. The data for 1980–85 and 1985–90 are assigned to the midpoints of their periods. The crude birth rate data for the Philippines come from Table 33, p. 111. Life expectancies for the Philippines come from Table 41, p. 143. The crude birth rate data for Kenya come from Table 33, p. 110, and crude death rates for Kenya come from Table 34, p. 113.
1999 or 2000: The figures in the table are linear extrapolations computed analogously to the 1984 or 1985 figures derived from in the 1968 United Nations assessments, as discussed above.
Bachue–Philippines: Rodgers, Hopkins, and Wéry (1978): Table VIII.6, p. 310 and Table VIII.7, p. 311.

Panel C shows the forecasts from Bachue–Philippines and the projections from the UN assessments from 1998 and 1973. According to the UN 1998 assessment the crude birth rate will be 27.1 in the year 2000. According to the Bachue model, it will be 29.3, and according to the UN 1973 assessment it will be 27.0. The UN 1973 assessment is considerably closer. The UN 1973 assessment underpredicts life expectancy by 0.1 years, while the Bachue–Philippines model underpredicts it by 0.9 years. In this case, taking the average of the two would not have produced the most accurate forecast. The averaged forecast would have underpredicted life expectancy by 0.5 years. Still, given that we do not know what life expectancy in the Philippines will be in the year 2000, the combined forecast may yet prove to be the best.

Panel D gives the 1999 projections for Kenya. Here the UN 1998 assessment middle variant was chosen. The UN 1973 assessment produces a much better forecast of the crude birth rate than the Bachue–Kenya model. In the 1970s the United Nations anticipated that fertility would begin to fall more rapidly in Kenya near the end of the 1980s. This more rapid decline is not predicted in the Bachue–Kenya model. On the other hand, the UN 1973 assessment sees a more rapid decline in the crude death rate as well and this does not seem to be occurring. The Bachue–Kenya model seems likely to be closer to the 1999 crude death rate than the UN 1973 assessment. The rate of natural increase in 1999 according to the UN 1998 assessment would be 2.04 percent per year. According to the UN 1973 assessment it would be 3.23 percent, and according to the Bachue model it would be 3.64 percent. In this case, combining the two rates of natural increase would result in a worse prediction.

Does knowledge help? For the earlier two forecasts, 20 years ahead in the case of the Philippines and 15 years ahead in the case of Kenya, knowledge improves forecasts. In the case of the Philippines, the crude birth rate and life expectancy predictions of the Bachue model are more accurate than those of the United Nations. In the case of Kenya, the forecasts are of comparable accuracy, and combining the Bachue and UN rates does even better. In the longer term, it is not clear that the use of the Bachue models would aid predictive accuracy. The prediction to 2000 made by Bachue–Philippines is less accurate than the UN 1973 forecast. Combining the forecasts seems unlikely to produce more accurate rates of population growth.

The 1999 Bachue–Kenya prediction is clearly worse than that from the UN 1973 assessment. The combined forecast would have a smaller absolute error for the crude death rate than the UN 1973 assessment forecast, but a larger absolute error in the crude birth rate and rate of natural increase.

The Bachue models are large and rich. They were not designed as population forecasting models, but as aids in economic and social policy
formation. Nevertheless, the two models reviewed here did well in terms of population forecasting. Population projections for a period around two decades into the future that averaged the UN and the Bachue predictions would have been more accurate than the UN predictions alone.

Conclusions

Since Keyfitz (1982), and perhaps even earlier, there has been a presumption that causal modeling has little to contribute to demographic forecasting. We argue in this chapter that this presumption should now be reversed. In general, combining forecasts from causal models with those of the United Nations produced more accurate predictions than those of the UN. This occurs because the forecast errors from the causal models and those of the United Nations are not highly correlated and because the forecasts from the causal model are not less accurate (Wheeler’s are more accurate) than those of the United Nations. For causal modeling to be useless would require that it produced much less accurate forecasts than the United Nations. This was not the case in the models reviewed here.

We considered three very different types of causal models. World3 is a medium-sized simulation model of the world that considers demographic, economic, and environmental interactions. It is not primarily a tool for population projection and indeed its authors did not take its demographic calibration very seriously. Nevertheless, it made a better 25-year forecast of the world’s population than the United Nations did. Wheeler developed an econometric model of the interactions between population, economic, and human resource growth. It was applied to 88 developing countries and, on average, produced more accurate 20-year forecasts of natural increase than did the United Nations in the eight countries for which Wheeler presented his results. The two Bachue models are both very large and detailed models for single countries. They use quite different approaches to demographic specification than either of the other two models. In particular, both include internal migration and estimate fertility and mortality separately in urban and rural areas. The 20-year forecasts for fertility and mortality from the Bachue–Philippines were more accurate than those of the United Nations. The 15-year forecasts from Bachue–Kenya were about as accurate as those of the United Nations, but even in this case combining the two sets of estimates produced the best results. Clearly, the value of causal modeling is not restricted to a particular approach.

The thesis of this chapter is not that causal modeling is generally superior to the UN methodology. We have too few observations to make this argument. On the other hand, the evidence presented here certainly is inconsistent with the presumption that knowledge cannot improve forecasts. In the small number of cases considered, knowledge did improve forecasts. The thesis instead is that the addition of causal modeling to the forecasters’
toolbox is likely to be productive. The reason is simple. The average of two uncorrelated, equally accurate forecasts produces better predictions than either taken separately.

If causal modeling is to become part of a forecaster’s toolbox, it is natural to ask how it can be improved. Among the models considered here, only Wheeler’s had as one of its goals the creation of population forecasts and it produced the most accurate ones. There seem to be several immediate studies that can be done based on the Wheeler framework. First, Wheeler considered population projections for eight particular countries. He could easily have produced many more. One study is to compute population forecasts for other countries and to investigate whether the accuracy of the forecasts is due simply to the particular countries that Wheeler considered. Another is to introduce a more detailed demographic accounting system into Wheeler’s model. Of all the models reviewed here, Wheeler’s has the simplest demographic structure. With that accounting system in place, new projections could be made and then tested against what was observed. Currently, much more information is available on developing countries than there was when Wheeler did his work. A third study could take advantage of the extended database and new econometric techniques. A fourth could be the application of Wheeler’s techniques to post-transitional populations. There are certainly many more possibilities.

World3, Bachue—Philippines, and the Wheeler model all use international cross-sectional data in forecasting fertility, while Bachue—Kenya uses national cross-district data. Since fertility is predicted better in the other models than in the Bachue—Kenya model, the international cross-section approach may be preferable. At least this is one hypothesis that we can keep in mind as we learn how to make even better projections.

Notes

1 To answer Keyfitz’s question for post-transitional countries is much more difficult and would require a separate analysis.

2 These were the only models that we could find that (1) made significant use of socioeconomic variables to predict future demographic variables for the world, regions, or developing countries, (2) had an accessible list of equations, and (3) had published forecasts. Mesarovic and Pestel (1974) used only one socioeconomic variable and that only for one region of the world. Herrera et al. (1976) produced forecasts of regional population sizes, but no equations were available. Demographic equations and forecasts were not available for Bachue—Yugoslavia and Bachue—Brazil, but there is still a good deal of interesting information published about those models. See Wéry and Rodgers (1980); Macura and Popovic (1984); Hopkins and van der Hoeven (1984).

3 The α’s are not the average absolute errors. They can be interpreted roughly as the “bias” of the forecast of N countries. For example, if α is positive, it means that, on average, the forecasts are too high.

4 The United Nations updates its population projections periodically. Each one has two associated dates. The first is the base or jumping-off date of the forecast and the second is
the date on which it was published. In this chapter, when we write, for example, "the United Nations 1978 assessment of population prospects," we are referring to the base date of the forecast. When the date appears in parentheses, such as in "(United Nations 1979)," we are referring to a particular publication that appeared in 1979. In this case, the UN 1978 assessment was published in 1979 and can be found in United Nations (1979).

5 There are many reviews of the World3 model. See, for example, Cole et al. (1973); Nordhaus (1992); and Sanderson (1994).

6 Donella Meadows's chapter, "Population sector," in Meadows et al. (1974) is 164 pages long, contains over 150 references, and has 113 figures and tables.

7 The number of births is set equal to the number of women of childbearing age multiplied by one-thirtieth of the total fertility rate. The fertility specification in World3-72 is discussed in Meadows et al. (1974): 95–134.

8 The 20-year figure is a parameter that can easily be changed.

9 This is roughly the world's per capita industrial output in 1965, again in 1968 US dollars. Meadows et al. (1974): 220.

10 The mortality specification in World3-72 is discussed in Meadows et al. (1974): 57–95.


12 We use "estimates" from UN (1998) as our benchmarks. In the past, analogous "estimates" have been changed over time by the UN. These "estimates" may also change in the future.

13 It is technically more accurate to speak of the parameter set used in Meadows et al. (1974) instead of Meadows et al. (1972). A few changes were made between the two dates (see Meadows et al. [1974]: viii).


15 Discussing the performance of the population-sector model, Meadows wrote:

The model produces a fair quantitative fit, given the uncertainties in the real-world data and the simplifying assumptions in the equations. The most serious quantitative error is in the death rate, which is too high in the model year 70, thus making the calculated net population growth rate too low. This discrepancy could be eliminated by altering any one of the lifetime multiplier tables. . . . If the death rate were lowered in the model, however, the calculated total population in year 70 would be too high. (Meadows et al. 1974: 147–148)

16 Wheeler's projection model can be found in Wheeler (1984): Figure 12, pp. 72–73.

17 A review of the Bachue–Philippines model can be found in Sanderson (1980).


19 In Bachue–Philippines, fertility is estimated using data from a cross-section of developing countries. In the Wheeler model, a cross-section of fertility changes is used.


23 The coefficient of the linear term is positive and that of the squared term negative. Given the sizes of those coefficients, the effect of income on fertility decreases as income increases.

24 We use life expectancy instead of crude death rate here because predictions for the crude death rate were not published in Rodgers, Hopkins, and Wéry (1978).

25 The data from the United Nations sources are derived by linear interpolation or extrapolation. See the notes to Table 6.

26 The UN 1998 crude death rate for Kenya in 1999 reflects the effects of HIV infection. Twenty years ago, no model took this into consideration.
References


