International Journal of Forecasting 8 (1992) 495-508
North-Holland

Forecasting state and household populations

Evaluating the forecast accuracy and bias of alternative population projections for states

Stanley K. Smith
Department of Economics, University of Florida, Gainesville, FL 32611, USA

Terry Sincich
Department of Information Systems and Decision Sciences, University of South Florida, Tampa, FL, USA

Abstract: Many different techniques can be used for making population projections. Most fall into four general categories: trend extrapolation, ratio extrapolation, cohort-component and structural. Techniques within these categories differ considerably in terms of their complexity and sophistication. A common perception among producers (and users) of population projections is that complex and/or sophisticated techniques produce more accurate forecasts than simple and/or naive techniques. In this paper we test the validity of that perception by evaluating the forecast accuracy and bias of eight commonly used projection techniques drawn from the four categories mentioned above. Using data for state population projections from a number of different time periods, we find no evidence that complex and/or sophisticated techniques produce more accurate or less biased forecasts than simple, naive techniques.

Keywords: Demographic projections and forecasts, Population forecasts

1. Introduction

Many different techniques can be used to make population projections [for further discussion, see Irwin (1977), Land (1986), Ahlburg (1987), Long and McMillen (1987), Murdock and Ellis (1991)]. Most fall into four general categories: (1) trend – historical trends in total population are extrapolated into the future using mathematical formulas or statistical techniques; (2) ratio – smaller-area population (e.g. state) is expressed as a proportion of larger-area population (e.g. nation) and historical trends in proportions are extrapolated and applied to projections of the larger-area population; (3) cohort-component – births, deaths and migration are projected separately for each age–sex cohort in the population; (4) structural – causal models relating population change to economic and/or other variables are used to project future population. Techniques from different categories can be combined, as when structural models are used to project components of growth in cohort-component projections.

Techniques in these four categories may differ considerably in terms of their complexity and sophistication. Statistical and mathematical techniques may be simple or complex, while theoreti-

Correspondence to: S.K. Smith, Department of Economics, University of Florida, Gainesville, FL 32611, USA.

cal models may be naive or sophisticated. We define a naive model as one in which future population values depend solely on past population values, whereas a sophisticated model is one in which population change is expressed as a function of changes in economic and/or other variables. Under this definition, trend and ratio techniques are naive but their mathematical forms may be either simple (e.g. linear extrapolation) or complex (e.g. ARIMA time series models). Cohort-component techniques involve complex disaggregations of population change, but they may be either sophisticated or naive, depending on whether they use structural models for projecting fertility, mortality and/or migration. Structural models are by definition sophisticated, but they may be either stochastic or deterministic and the models themselves may vary considerably in terms of complexity.

A common perception among producers (and users) of population projections is that complex techniques produce more accurate forecasts than simple techniques and sophisticated techniques produce more accurate forecasts than naive techniques [e.g. Birch (1977), Irwin (1977), Morrison (1977), Pittenger (1980), Keyfitz (1981), Beaumont and Isserman (1987)]. Is this perception valid? To our knowledge, there is no solid empirical evidence showing complex or sophisticated techniques to produce more accurate population forecasts than simple or naive techniques.[1] On the contrary, several studies have found simple and/or naive techniques to produce forecasts that are just as accurate as more complex and/or sophisticated techniques [e.g. Siegel (1953), White (1954), Kale et al. (1981), Murdock et al. (1984), Smith (1984)].

The research performed to date on this topic has been quite limited, however [Beaumont and Isserman (1987)]. The few studies that have been done covered only a few projection techniques and/or a limited number of time periods. In the present study we try to overcome these limitations by evaluating forecast accuracy and bias for a wide variety of projection techniques and – perhaps more important – techniques applied in five different time periods. This gives us a large sample of observations from which to draw conclusions and allows us to track changes in accuracy and bias over time. The analysis focuses on projections of total population for states in the United States; it does not consider projections of demographic characteristics.

In this paper a population projection is defined as the future population value produced by a particular projection technique and set of base data. Forecast error refers to the percentage difference between a population projection and the 'true' population enumerated or estimated for the same year. In other words, we treat population projections as if they were forecasts (or predictions) of future population. We use the following terminology to describe population projections:

(1) Base year: the year of the earliest observed population size used to make a projection.

(2) Launch year: the year of the latest observed population size used to make a projection.

(3) Target year: the year for which population size is projected.

(4) Base period: the interval between the base year and the launch year.

(5) Projection horizon: the interval between the launch year and the target year.

All the projections evaluated in this paper have explicit launch years, target years and projection horizons. For some techniques, however, there are no clearly defined base years or base periods (e.g. the cohort-component projections).

2. Projections to be evaluated

2.1. Trend

We evaluated three trend techniques. First was linear extrapolation (LINE), which assumes that a population will increase (decrease) by the same number of persons in each future year as

[1] The only empirical study we know of that found simple techniques to be less accurate than more complex techniques was a study of national forecast errors [Keyfitz (1981)]. This study reported larger errors for exponential extrapolations than for more complex projections produced by professional demographers. However, this study tested only one simple technique (exponential extrapolation) and used only a single launch year (1955). In addition, it used a 5-year base period for the exponential extrapolations, a base period shown in another study to produce less accurate forecasts than 10- or 20-year base periods [Smith and Sincich (1990)]. Consequently, the results do not provide sufficient evidence to conclude that simple techniques generally produce less accurate forecasts than more complex or sophisticated techniques.

the average annual increase (decrease) during the base period:

$$\hat{P}_t = P_o + x/y(P_o - P_b)\,, \tag{1}$$

where $\hat{P}_t$ is the state population projection for the target year, P_o is the state population in the launch year, P_b is the state population in the base year, x is the number of years in the projection horizon and y is the number of years in the base period.

The second technique was exponential extrapolation (EXPO), which assumes that a population will increase (decrease) at the same annual percentage rate during the projection horizon as during the base period:

$$\hat{P}_t = P_o \exp(rx)\,, \tag{2}$$

where r is the average annual growth rate during the base period, calculated as

$$r = \frac{\ln(P_o/P_b)}{y}\,. \tag{3}$$

The first two trend techniques were very simple; the third was a complex autoregressive, integrated, moving average (ARIMA) time series model of population change. Voss et al. (1981) demonstrated that an ARIMA (1, 1, 0) model (i.e. a model of first differences with a first-order autoregressive equation and no moving average component) outperformed a number of other time series models in forecasting state populations. This is the model we tested. In its simplest algebraic form, the model can be expressed as

$$\hat{P}_t = a_0 + (1 + a_1)P_{t-1} - a_1 P_{t-2} \tag{4}$$

where $\hat{P}_t$ is the state population projection in target year t; P_{t-1} and P_{t-2} are the state's population in the two years prior to the target year; and a_0 and a_1 are regression-like parameters estimated by fitting the model to annual data collected during the base period.[2]

2.2. Ratio

We evaluated two ratio techniques. First was shift-share (SHIFT), which assumes that each state's share of the national population will change by the same annual amount during the projection horizon as the average annual change during the base period:

$$\hat{P}_t = \hat{P}_{jt}[P_o/P_{jo} + x/y(P_o/P_{jo} - P_b/P_{jb})]\,, \tag{5}$$

where $\hat{P}_{jt}$ is the national population projected for the target year, P_{jo} is the national population in the launch year and P_{jb} is the national population in the base year.

The second ratio technique was share-of-growth (SHARE), in which each state's share of national population growth during the projection horizon is projected to be the same as during the base period:

$$\hat{P}_t = P_o + [(P_o - P_b)/(P_{jo} - P_{jb})](\hat{P}_{jt} - P_{jo})\,. \tag{6}$$

Both the SHIFT and SHARE techniques require independent projections of national population ($\hat{P}_{jt}$). We created such projections by applying the LINE and EXPO techniques to the US population and taking the average as a national projection.

We constructed all trend and ratio projections using as base data a series of annual intercensal population estimates for states [US Bureau of the Census, (1956, 1965, 1971, 1984)]. For the ARIMA projections we used all the data from 1900 to the launch year. For the other four techniques we used only the ten years prior to the launch year; a previous study found ten years of base data were generally sufficient to achieve the highest level of accuracy for these techniques [Smith and Sincich (1990)].

2.3. Cohort-component

Many different techniques can be used to apply the cohort-component method of population projection. They differ according to the structure of the demographic model, sources of data and the development of assumptions regarding mortality, fertility and migration. In this

[2] Ideally, one would fit separate models for all 48 × 5 = 240 state/launch year combinations to be evaluated. However, our objective in this study was not to find the 'best' model for each individual time series, but rather to test a single model that would provide a reasonable fit for all state/launch year combinations. Consequently, we used the ARIMA (1, 1, 0) model developed by Voss et al. (1981). On the basis of plots of the sample autocorrelations and sample partial autocorrelations, as well as estimates and statistical tests of the autoregressive lag parameter (a_1), we have concluded that this model provides a reasonable fit of the data in most instances.

study we evaluated the projections produced by the US Bureau of the Census. We chose these projections because the Census Bureau has been a leader in developing the cohort-component methodology, has been making state projections since the mid-1950s and follows a consistent methodology in all states. In addition, the Census Bureau's projections are by far the most widely used in the United States. We believe an evaluation of the Census Bureau's projections provides a good test of the cohort-component method.

The Census Bureau typically publishes several series of projections rather than a single series. Each series represents a different combination of assumptions regarding fertility, mortality and migration trends. The Census Bureau does not indicate which series is most likely to provide an accurate forecast of future population; that is, it does not specify a single forecast for each state. Consequently, we evaluated each series of Census Bureau projections published between the mid-1950s and early 1980s, except for those that assumed no interstate migration.[3] A description of the assumptions underlying these projections can be found in each Census Bureau publication [US Bureau of the Census (1957, 1966, 1972, 1979, 1983)].

2.4. Structural

Structural models use causal analysis to relate population change to changes in economic and/or other variables. These models may be relatively simple (e.g. tying net migration to exogenous projections of employment) or very complex (e.g. hundreds or even thousands of equations relating demographic and socioeconomic variables). Although large, complex structural models have been developed [e.g. the ECESIS model described in Beaumont (1989)], they have been used primarily for simulation and policy analysis rather than for population forecasting. To our knowledge, no population forecasts from such models have been published for all states for any of the time periods covered by the present study. This in itself reflects the difficulty of using complex, sophisticated models for forecasting the populations of a large number of places.

We therefore confined our analysis to two simple structural models which relate migration to projections of employment: the OBERS model developed by the Bureau of Economic Analysis [US Bureau of Economic Analysis (1974, 1981, 1985)] and the economic-based model developed by the National Planning Association [National Planning Association (1972, 1976, 1982)]. Both models view migration as depending on projected differences between the supply and demand for labor. Supply depends on factors such as the age–sex structure of the population and labor force participation rates; demand depends on the projected employment in various industries or occupations. Projections from both models have been widely used in the United States. Detailed descriptions of the projection methodology can be found in the references cited above.

2.5. Summary

In this study we evaluated the forecast accuracy and bias of four simple, naive techniques in which data from two points in time are used to project total population (LINE, EXPO, SHIFT, SHARE); one complex, naive technique in which time series data are used to project changes in total population (ARIMA); one complex, naive technique which projects components of growth by age–sex cohort (Census Bureau, or CB); and two simple, sophisticated techniques in which projections of migration are based on projections of employment (OBERS, NPA). We did not evaluate any complex, sophisticated models because – to our knowledge – none has been used explicitly for population forecasting in a large number of states.

The Census Bureau projections are a bit difficult to classify. The methodology is clearly complex, since it incorporates detailed accounting procedures covering mortality, fertility and migration rates for each age–sex cohort in the population. But is the methodology sophisticated or naive? It uses professional judgment in developing assumptions regarding mortality, fertility and migration, but does not explicitly incorporate structural models. We therefore classify the

[3] We did not evaluate projections based on the hypothetical assumption that there would be no interstate migration because that assumption was intended to be purely illustrative [US Bureau of the Census (1979, p. 1)].

Census Bureau projections as complex but naive. The reader should be aware, however, that some degree of causal analysis implicitly lies behind these projections. Exhibit 1 identifies the launch years, target years, publication years (if any) and names of all the projections evaluated in this study.

3. Methodology and results

For each technique we compared the projections of total population with decennial population counts or intercensal estimates for each target year [US Bureau of the Census (1971, 1984, 1991)]. Forecast error (F_t) was calculated

Exhibit 1
Summary of projections.

Name	Launch year	Publication year	Target years
LINE	1955	–	1960, 1965, 1970, 1975
EXPO	1955	–	1960, 1965, 1970, 1975
ARIMA	1955	–	1960, 1965, 1970, 1975
SHIFT	1955	–	1960, 1965, 1970, 1975
SHARE	1955	–	1960, 1965, 1970, 1975
CB-1	1955	1957	1960, 1965, 1970
CB-2	1955	1957	1960, 1965, 1970
CB-3	1955	1957	1960, 1965, 1970
CB-4	1955	1957	1960, 1965, 1970
LINE	1965	–	1970, 1975, 1980, 1985
EXPO	1965	–	1970, 1975, 1980, 1985
ARIMA	1965	–	1970, 1975, 1980, 1985
SHIFT	1965	–	1970, 1975, 1980, 1985
SHARE	1965	–	1970, 1975, 1980, 1985
CB-1B	1964	1966	1970, 1975, 1980, 1985
CB-2B	1964	1966	1970, 1975, 1980, 1985
CB-1D	1964	1966	1970, 1975, 1980, 1985
CB-2D	1964	1966	1970, 1975, 1980, 1985
LINE	1970	–	1975, 1980, 1985, 1990
EXPO	1970	–	1975, 1980, 1985, 1990
ARIMA	1970	–	1975, 1980, 1985, 1990
SHIFT	1970	–	1975, 1980, 1985, 1990
SHARE	1970	–	1975, 1980, 1985, 1990
CB-1C	1970	1972	1975, 1980, 1985, 1990
CB-1E	1970	1972	1975, 1980, 1985, 1990
NPA	1970	1972	1975, 1980, 1985
OBERS	1971	1974	1980, 1985, 1990
LINE	1975	–	1980, 1985, 1990
EXPO	1975	–	1980, 1985, 1990
ARIMA	1975	–	1980, 1985, 1990
SHIFT	1975	–	1980, 1985, 1990
SHARE	1975	–	1980, 1985, 1990
CB-2A	1975	1979	1980, 1985, 1990
CB-2B	1975	1979	1980, 1985, 1990
NPA	1975	1977	1980, 1985, 1990
OBERS	1978	1981	1985, 1990
LINE	1980	–	1985, 1990
EXPO	1980	–	1985, 1990
ARIMA	1980	–	1985, 1990
SHIFT	1980	–	1985, 1990
SHARE	1980	–	1985, 1990
CB-0	1980	1983	1990
NPA	1980	1982	1990
OBERS	1983	1985	1985, 1990

as the percent difference between the projection ($\hat{P}_t$) and the 'true' population for the same year (P_t):

$$F_t = [(\hat{P}_t - P_t)/P_t]100\,. \quad (7)$$

We assumed that the population estimates and counts published by the Census Bureau reflected the 'true' populations; that is, no attempts were made to adjust for estimation or enumeration error. Except for ARIMA, the analysis covered all 50 states for projections produced after 1960; prior to 1960, the Census Bureau did not produce projections for Alaska and Hawaii. The ARIMA projections covered only the 48 contiguous states because annual intercensal estimates back to 1900 were not available for Alaska and Hawaii.

There is no consensus in the literature regarding the most appropriate summary measures of forecast error [e.g. Ahlburg (1992)]. Consequently, we have chosen to use a number of measures, each with a somewhat different focus. We believe this variety strengthens the validity of our analysis and conclusions.

Mean absolute percent error (*MAPE*) is the average error when the direction of error is ignored; root mean square percent error (*RMSPE*) is a measure giving a heavier weight to large errors; and the 90th percentile error (*90PE*) is the absolute percent error larger than exactly 90% of all other absolute percent errors. These are measures of accuracy, or how close the projections were to population counts or estimates for the same year, regardless of whether they were too high or too low.

Mean algebraic percent error (*MALPE*) is the average percent error when the direction of error is accounted for; this is a measure of bias. A positive error indicates a tendency for projections to be too high and a negative error indicates a tendency for projections to be too low. Since a few extreme errors in one direction can disproportionately affect the sign of the *MALPE*, the proportion of positive errors (*%POS*) was used as another measure of bias.

We present an overview of forecast accuracy and bias in Exhibit 2. In this exhibit we grouped errors by length of projection horizon and aggregated over all launch years for each technique; all Census Bureau projection series were averaged together and are identified as CB. In instances where projection horizons did not exactly match 5, 10, 15 and 20 years (e.g. Census Bureau projections with launch year 1964 and OBERS projections with launch year 1971), we included them with the most similar horizon (e.g. four-

Exhibit 2
Measures of forecast accuracy and bias: Averages covering all launch years.

Measure	Technique	Length of Projection Horizon (years)			
		5	10	15	20
MAPE	LINE	3.5	6.0	8.0	11.3
	EXPO	3.9	7.0	10.6	16.2
	ARIMA	3.3	6.3	9.1	11.5
	SHIFT	3.8	6.4	9.2	13.4
	SHARE	3.6	6.0	8.2	11.7
	CB	3.7	6.1	8.3	12.4
	NPA	4.3	6.8	8.4	–
	OBERS	4.0	6.5	9.1	12.8
RMSPE	LINE	5.1	8.2	10.8	14.3
	EXPO	6.3	11.7	20.2	33.0
	ARIMA	4.6	8.2	11.7	14.8
	SHIFT	5.5	9.3	13.2	18.7
	SHARE	5.2	8.4	11.3	15.2
	CB	5.0	8.2	10.7	15.1
	NPA	5.3	8.5	10.3	–
	OBERS	5.8	8.8	11.6	15.2
90PE	LINE	7.7	11.8	16.4	22.3
	EXPO	8.6	13.6	21.3	32.0
	ARIMA	7.2	13.6	18.9	23.6
	SHIFT	8.1	13.1	19.5	27.7
	SHARE	7.8	12.1	17.9	23.4
	CB	8.1	13.2	17.5	24.7
	NPA	8.3	13.4	17.9	–
	OBERS	9.7	14.1	18.3	26.1
MALPE	LINE	0.1	−0.5	−1.1	−1.9
	EXPO	1.2	2.4	4.3	7.8
	ARIMA	−1.1	−2.8	−4.4	−6.0
	SHIFT	0.4	0.2	−0.2	−0.8
	SHARE	0.4	0.2	0.2	0.4
	CB	−0.7	−1.1	−0.4	2.4
	NPA	−2.4	−0.9	−0.6	–
	OBERS	1.7	−3.6	−2.6	−4.9
%POS	LINE	51.3	46.7	47.5	44.7
	EXPO	59.0	60.4	61.5	60.7
	ARIMA	40.6	40.0	36.0	34.7
	SHIFT	54.0	51.6	51.5	46.7
	SHARE	54.0	51.2	54.0	49.3
	CB	44.4	46.0	50.3	55.7
	NPA	33.0	46.0	49.0	–
	OBERS	64.0	34.0	43.0	40.0

and six-year horizons were included with five-year horizons).[4]

A number of results are noteworthy. First of all, accuracy levels were very similar for all eight techniques, with the exception of EXPO (and, to a lesser extent, SHIFT) for longer forecast horizons. For 10-year horizons, *MAPE* values ranged only from 6.0 to 7.0; *RMSPE* values from 8.2 to 11.7; and *90PE* values from 11.8 to 14.1. For 20-year horizons the ranges (excluding EXPO and SHIFT) were 11.3 to 12.8 for *MAPE*, 14.3 to 15.2 for *RMSPE* and 22.3 to 26.1 for *90PE*. EXPO and SHIFT produced a few very large errors for longer projection horizons; other than that, levels of accuracy were about the same for all projection techniques.

For all eight techniques and all three measures of accuracy, errors increased steadily with the length of the projection horizon. This is a common finding in the literature [e.g. White (1954), Kale et al. (1981), Keyfitz (1981), Smith (1987)]. For *MAPE*, the increases were approximately linear. This has been noted before for several trend and ratio techniques [Smith and Sincich (1991)]; here it is found for cohort-component and structural models as well.

The results for *MALPE* and *%POS* showed an upward bias for EXPO projections for all four projection horizons; ARIMA and NPA had a downward bias for all horizons; and OBERS and LINE had downward biases for all horizons longer than five years. CB, SHIFT and SHARE had mixed results. Other than EXPO and ARIMA, (and to a lesser extent, OBERS) none of the techniques displayed large biases in either direction; SHIFT and SHARE, in particular, exhibited very little bias. We will return to the issue of bias later in the paper.

There is no indication from the results shown in Exhibit 2 that complex or sophisticated techniques produced more accurate or less biased forecasts than simple, naive techniques. In fact, LINE and SHARE often performed better than any other technique. Only EXPO displayed a tendency to produce larger errors than the other techniques, and these differences were substantial only for longer projection horizons and – as will be seen later – were never statistically significant.

The results summarized in Exhibit 2 were based on averages covering several launch years. The number of sets of projections and the launch years included in these averages were not the same for all techniques. In addition, the different series of Census Bureau projections were based on different sets of demographic assumptions and the OBERS projections did not always fit neatly into 5-, 10-, 15- and 20-year horizons. Consequently, it is important to observe the results for each individual technique and each launch year before drawing any conclusions. Such results are shown in Exhibits 3–7. For brevity, we report only the results for *MAPE* and *MALPE*. Results for *RMSPE*, *90PE* and *%POS* were similar to those reported here and are available from the authors upon request.

These exhibits show that no single technique or group of techniques consistently produced smaller (or larger) *MAPE* values than any other technique or group of techniques. In Exhibit 3, *MAPE* values for the four Census Bureau series of projections were slightly smaller than for LINE, SHARE and ARIMA and considerably smaller than for EXPO and SHIFT. In Exhibit 4, *MAPE* values at each horizon were similar for all projections except EXPO and CB-2B, which were somewhat larger than the others. In Exhibit 5, *MAPE* values for the trend and ratio projections were slightly smaller than *MAPE* values for the cohort-component and structural model projections (except for NPA). In Exhibit 6, *MAPE* values were generally smaller for the simple trend and ratio projections than for the ARIMA, cohort-component and structural model projections. In Exhibit 7, *MAPE* values were similar for all techniques except two: EXPO had the largest errors for both horizons and OBERS the smallest. The OBERS projections, however, covered only 2- and 7-year horizons rather than 5- and 10-year horizons.

Although some differences in *MAPE* values by technique can be seen in Exhibits 3–7, in most instances they are relatively small. Were these differences caused by true differences in accuracy or simply by random variation? To answer this question, we conducted formal statistical tests of hypotheses. For each combination

[4] Only for OBERS did this create a problem. For comparisons by length of horizon, we treated OBERS projections with launch year 1978 as having a launch year of 1975 and projections with launch year 1983 as having a launch year of 1985.

Exhibit 3
Forecast accuracy and bias: Launch year 1955.

Technique	Launch year	*MAPE* or *MALPE* in target year			
		1960	1965	1970	1975
Accuracy (MAPE)					
LINE	1955	3.9	6.1	8.0	12.1
EXPO	1955	4.5	8.3	15.1	22.8
ARIMA	1955	3.8	6.6	8.5	10.9
SHIFT	1955	4.4	7.3	11.7	17.0
SHARE	1955	3.9	6.2	8.5	12.8
CB-1	1955	3.4	5.8	7.8	–
CB-2	1955	3.1	5.6	6.8	–
CB-3	1955	3.1	5.3	7.0	–
CB-4	1955	3.3	5.9	7.0	–
Bias (MALPE)					
LINE	1955	−1.3	−2.2	−1.4	−3.4
EXPO	1955	0.2	2.3	7.8	11.5
ARIMA	1955	−2.6	−4.6	−4.8	−7.5
SHIFT	1955	−1.1	−1.7	−0.7	−2.7
SHARE	1955	−1.0	−1.4	0.1	−1.1
CB-1	1955	−0.7	−1.0	1.1	–
CB-2	1955	−1.5	−2.4	−0.7	–
CB-3	1955	−1.2	−1.7	0.3	–
CB-4	1955	−2.0	−3.9	−4.0	–

Exhibit 4
Forecast accuracy and bias: Launch years 1964 and 1965.

Technique	Launch year	*MAPE* or *MALPE* in target year			
		1970	1975	1980	1985
Accuracy (MAPE)					
LINE	1965	2.9	5.0	8.2	10.3
EXPO	1965	4.0	6.6	10.8	14.7
ARIMA	1965	2.4	5.4	9.1	11.5
SHIFT	1965	3.3	5.5	8.9	11.5
SHARE	1965	3.2	5.1	8.4	10.8
CB-1B	1964	3.9	5.5	9.2	12.7
CB-2B	1964	4.0	5.9	9.7	14.1
CB-1D	1964	3.4	5.0	8.4	11.0
CB-2D	1964	3.6	5.1	8.5	11.1
Bias (MALPE)					
LINE	1965	1.8	0.4	−0.2	0.6
EXPO	1965	3.1	3.6	5.4	9.3
ARIMA	1965	−0.2	−3.0	−5.3	−5.9
SHIFT	1965	2.1	1.2	1.0	2.2
SHARE	1965	2.2	1.4	1.6	3.4
CB-1B	1964	2.3	1.9	3.2	6.4
CB-2B	1964	2.5	2.5	4.1	7.9
CB-1D	1964	0.7	−1.3	−2.1	−1.1
CB-2D	1964	0.9	−0.8	−1.2	0.3

Exhibit 5
Forecast accuracy and bias: Launch years 1970 and 1971.

Technique	Launch year	*MAPE* or *MALPE* in target year			
		1975	1980	1985	1990
Accuracy (MAPE)					
LINE	1970	4.5	8.3	10.7	11.5
EXPO	1970	4.3	7.8	10.4	11.0
ARIMA	1970	4.2	8.3	10.7	12.0
SHIFT	1970	4.5	8.4	10.9	11.8
SHARE	1970	4.4	8.2	10.6	11.3
CB-1C	1970	4.8	8.6	11.4	12.8
CB-1E	1970	5.0	9.1	11.8	12.8
NPA	1970	4.3	7.6	10.0	–
OBERS	1971	–	9.3	11.9	12.8
Bias (MALPE)					
LINE	1970	−2.2	−3.6	−3.4	−2.7
EXPO	1970	−1.4	−1.6	0.0	2.6
ARIMA	1970	−2.6	−4.9	−5.6	−4.7
SHIFT	1970	−2.0	−3.2	−2.8	−1.9
SHARE	1970	−1.9	−2.9	−2.3	−1.0
CB-1C	1970	−2.0	−1.5	1.1	3.9
CB-1E	1970	−2.7	−4.1	−3.8	−3.3
NPA	1970	−1.7	−1.5	−0.3	–
OBERS	1971	−2.2	−5.7	−5.7	−4.9

of launch year and target year, we tested the null hypothesis that *MAPE* values are identical for all techniques:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_n, \tag{8}$$

where μ_1 is the true *MAPE* for LINE, μ_2 is the true *MAPE* for EXPO,... and μ_n is the true *MAPE* for the last technique in each set.

Using a data set similar to that used in this study for the trend and ratio techniques, Smith and Sincich (1988) showed that the distribution of algebraic percent errors tended to be normal, but that absolute percent errors had skewed (non-normal) distributions truncated at zero. Consequently, the normality assumption required for traditional one-way analysis-of-variance F-tests for *MAPE* is highly likely to be violated. We therefore tested H_0 using a distribution-free non-parametric test, the Kruskal–Wallis H test.

For all launch/target year combinations except two, the test was non-significant; that is, there was insufficient evidence to reject H_0 at a conservative significance level of 0.10. Except for these two combinations, there was no evidence of significant differences among *MAPE* values for the different forecasting techniques.

The only exceptions were errors for the launch/target year combinations of 1975/80 and 1975/85. We investigated these errors further, using a Bonferroni multiple comparison-of-means procedure designed to control the experiment-wise Type I error rate. Conservatively, we selected an overall significance level of 0.10. For 1975/80, the *MAPE* for CB-2A (4.66) was significantly larger than the *MAPE* for EXPO (2.78) and SHIFT (2.75). For 1975/85 the *MAPE* for CB-2A (6.08) was significantly larger than the *MAPE* for OBERS (3.66). No other pairs of *MAPE* values in any other sets of projections were found to be significantly different from each other. Thus, there is no statistical evidence in this data set that simple or naive techniques produce less accurate forecasts than complex or sophisticated techniques.

Exhibits 3–7 also provide information on the direction of forecast errors. Two sets of projections had a downward bias for most techniques and horizons (launch years 1955 and 1970/71); two sets had an upward bias for most techniques and horizons (launch years 1964/65 and 1980/

Exhibit 6
Forecast accuracy and bias: Launch years 1975 and 1978.

Technique	Launch year	*MAPE* or *MALPE* in target year		
		1980	1985	1990
Accuracy (MAPE)				
LINE	1975	2.9	4.4	5.3
EXPO	1975	2.8	4.3	6.1
ARIMA	1975	3.6	6.0	8.0
SHIFT	1975	2.8	4.2	5.1
SHARE	1975	2.8	4.3	5.4
CB-2A	1975	4.7	6.1	6.3
CB-2B	1975	3.7	4.3	5.5
NPA	1975	4.3	6.0	6.7
OBERS	1978	2.9	3.7	6.3
Bias (MALPE)				
LINE	1975	−1.1	−0.6	0.6
EXPO	1975	−0.3	1.3	4.1
ARIMA	1975	−2.0	−2.8	−1.9
SHARE	1975	−0.8	0.0	1.7
SHIFT	1975	−0.8	−0.1	1.6
CB-2A	1975	−3.7	−3.4	−2.5
CB-2B	1975	−3.0	−2.2	−0.7
NPA	1975	−3.1	−2.7	−0.9
OBERS	1978	−1.1	−1.4	0.5

Exhibit 7
Forecast accuracy and bias: Launch years 1980 and 1983.

Technique	Launch year	*MAPE* or *MALPE* in target year	
		1985	1990
Accuracy (MAPE)			
LINE	1980	2.7	6.1
EXPO	1980	3.6	8.0
ARIMA	1980	2.3	5.2
SHIFT	1980	3.1	6.9
SHARE	1980	2.9	6.4
CB	1980	–	6.5
NPA	1980	–	6.7
OBERS	1983	2.7	4.0
Bias (MALPE)			
LINE	1980	1.4	3.5
EXPO	1980	2.5	6.4
ARIMA	1980	0.0	1.1
SHIFT	1980	1.8	4.5
SHARE	1980	1.6	4.2
CB	1980	–	3.8
NPA	1980	–	1.6
OBERS	1983	1.4	1.7

83); and one set had mixed results (launch years 1975/78). All projection techniques had some positive *MALPE* and some negative *MALPE* values. We believe that the projection techniques evaluated in this paper (with two possible exceptions) are unbiased in the sense that, if applied at a number of different points in time, they would produce projections that prove to be too high about as often as they produce projections that prove to be too low. There is no way to know in advance whether any given set of projections will have an upward or downward bias (unless obviously unrealistic assumptions are used).

The only possible exceptions to this conclusion are the EXPO and ARIMA techniques. EXPO was found to have an upward bias (occasionally quite strong) in four of the five sets of projections, and ARIMA was found to have a downward bias (occasionally quite strong) in four of the five sets. We believe an upward bias is a general characteristic of EXPO projections for places that grew rapidly during the base period [Smith (1987)]. The downward bias observed for the ARIMA projections, however, reflects either the specification of the model or the historical time period covered by the analysis, rather than an inherent characteristic of the technique itself.

4. Discussion

Beaumont and Isserman (1987) questioned whether sufficient empirical evidence exists to conclude that simple and/or naive techniques can produce population forecasts that are as accurate as those produced by more complex and/or sophisticated techniques. We believe we have taken a major step toward answering that question. In this study we evaluated the forecast accuracy and bias of state population projections produced by a number of different techniques for launch years in the 1950s, 1960s, 1970s and 1980s. We found that differences in complexity and sophistication had no consistent or statistically significant impact on forecast accuracy. In some instances, simple techniques were more accurate than complex techniques and naive techniques more accurate than sophisticated techniques; in other instances the opposite was true. Rarely were the differences large and they were almost never statistically significant. We conclude that this study provides ample empirical evidence that – to date – the sophistication of structural models and the complexity of time series and cohort-component techniques have not led to greater accuracy in forecasting total population than can be achieved with simple, naive techniques.

This study did not cover every possible projection technique, of course. We evaluated only eight of the huge number of techniques or models that potentially could be used for state population projections. In particular, we did not evaluate any techniques that were both complex and sophisticated. Our conclusions must therefore be viewed as preliminary.

The techniques we did evaluate, however, are those most commonly used for state and local projections. They included several trend and ratio extrapolation techniques, an ARIMA time series model, the Census Bureau's cohort-component model and the structural models developed by the Bureau of Economic Analysis and the National Planning Association. The analysis thus included both naive and sophisticated models and techniques that vary considerably in terms of complexity. Yet we found no consistent or significant differences in forecast accuracy. We therefore believe the burden of proof lies with those who would claim that more complex or sophisticated techniques are capable of producing more accurate forecasts than simple, naive techniques. Before we can accept this claim as valid, we must be shown supporting empirical evidence.

Why did the complex and the sophisticated techniques evaluated in this study not produce more accurate forecasts than the simple, naive techniques? We believe there is a certain irreducible level of uncertainty regarding the future. No projection technique – no matter how complex or sophisticated – can consistently improve forecast accuracy beyond that level. Based on the evidence to date, it appears that the relatively small amount of historical information contained in simple trend and ratio extrapolation techniques provides as much guidance to this uncertain future as does the much larger amount of information contained in more complex or sophisticated techniques. Why might this be true?

We believe the cohort-component projections were no more accurate than the trend and ratio projections because forecasting fertility, mortality and migration rates is just as difficult as forecasting changes in total population (perhaps more so). This difficulty counteracts the advantages of disaggregation and the stability the age–sex structure adds to cohort-component projections. Will the application of time series techniques or the development of new data sources (e.g. annual gross migration data from IRS records) improve the forecast accuracy of current and future cohort-component projections? We do not think so. These techniques are still based on the extrapolation of past trends, and those trends are highly correlated with those underlying the simple trend and ratio projections. We believe it is unlikely that more complex approaches to extrapolating past trends in cohort-component models will lead to any significant improvements in forecast accuracy.

A comparison of the various series of Census Bureau projections for each launch year is instructive. For all launch years except 1980, the Census Bureau published several series of projections based on alternative assumptions regarding fertility, mortality and migration. Although *MALPE* values varied among the different series (indicating differences in bias), *MAPE* values were quite similar for all series with the same launch year. For example, *MAPE* values for 15-year horizons ranged only from 6.8 to 7.8 for the four series with launch year 1955; from 8.4 to 9.7 for the four series with launch year 1964; from 11.4 to 11.8 for the two series with launch year 1970; and from 5.5 to 6.3 for the two series with launch year 1975. Differences in assumptions may have led to substantially different errors for any given state, but those differences were largely wiped out when averaged over all states. Even if one were to choose retrospectively the Census Bureau series with the smallest errors, the errors still would not be consistently (or significantly) smaller than for the simple, naive techniques. Aggregate results were simply not very sensitive to moderate differences in assumptions. Improvements in forecasting future components of growth would therefore have to be very large before they would lead to significant across-the-board reductions in average forecast error.

What about structural models? Those evaluated in this study were simple deterministic models tying net migration to exogenous changes in employment. Would more complex stochastic models lead to improvements in forecast accuracy? Again, we are skeptical. Our understanding as demographers of the determinants of population change is far from perfect. Thus we cannot construct models that completely explain population change. Even if we could, we could not be sure that past relationships between demographic and socioeconomic variables would remain constant in the future. More critical yet, even if those relationships were to remain constant, the future values for the socioeconomic variables themselves would still be unknown. (Is there any reason to believe that socioeconomic variables can be projected more accurately than demographic variables?) Given all these uncertainties, it is not surprising that projections from simple structural models have been no more accurate than those from naive trend or ratio techniques; we are not hopeful that more complex structural models will perform any better.

Have demographers improved over time in their abilities to forecast future populations? An evaluation of Exhibits 3–7 would seem to say 'no'. *MAPE* values for the trend and ratio techniques vary by launch year, but show no distinct upward or downward trends. These techniques are purely mechanical, of course; since they contain no applications of judgment or changes in methodology, variations in *MAPE* values for these techniques occur solely because of differences in launch years and random variation in the underlying population dynamics.

The Census Bureau projections, on the other hand, are affected by methodological changes and the application of professional judgment in making assumptions about mortality, fertility and migration. Do these projections indicate any trends in accuracy? For launch year 1955, *MAPE* values for 10-year horizons for the CB techniques ranged from 5.3 to 5.9%; for 1964, from 5.0 to 5.9%; for 1970, from 8.6 to 9.1%; for 1975, from 4.5 to 6.3%; and for 1980, the single CB projection had a *MAPE* of 6.5%. No clear trend in accuracy is apparent from these results. The relative performance of the CB techniques compared with the simple extrapolation techniques also shows no clear trends over time. If Census Bureau demographers can be taken as representative of professional demographers in general, it does not appear that demographers have improved in their abilities to forecast future state populations.

Exhibits 3–7 show that it is risky to base general conclusions regarding forecast accuracy and bias on projections from a single launch year. Launch year 1975 showed relatively small *MAPE* values for all techniques and all target years, whereas launch year 1970 showed relatively large *MAPE* values. Launch year 1980 showed an upward bias for all techniques and target years, whereas launch year 1970 showed a downward bias for most techniques and target years. The EXPO technique had larger *MAPE* values than any other technique for launch year 1965 and smaller *MAPE* values than most techniques for launch year 1970. It appears that a valid assessment of forecast accuracy and bias (including comparisons of alternative techniques) can be made only when several different sets of projections are evaluated.

Several caveats regarding the findings reported in this study should be mentioned. First, the analysis covered only horizons of 5–20 years. It is possible that some approaches or techniques will perform significantly better (or worse) than others for horizons of shorter than 5 years or longer than 20 years. Second, the analysis covered states. Results could be different for other levels of geography. Finally, the analysis did not investigate potential differences in accuracy and bias for places with different demographic or socioeconomic characteristics. One technique or category of techniques could perform significantly better (or worse) than another for places with specific characteristics (e.g. very high or very low growth rates). Although we have no reason to suspect that any of these factors would lead to results different from those reported in this paper, they all reflect the need for further research.

5. Conclusions

Population projections can play a number of different roles. They can be used to evaluate the impact of different economic or demographic assumptions on population size and characteristics; to analyze individual components of population growth; to illustrate the potential outcomes of recent trends; to provide support for specific points of view; and to provide forecasts (i.e. predictions) of future population change. The last is the most common for the majority of users of population projections, even when the authors of those projections do not intend for them to be used as forecasts. In this study we have demonstrated that for a large number of projections covering different launch years and horizons, complex techniques did not produce more accurate forecasts of total population than simple techniques and sophisticated techniques did not provide more accurate forecasts than naive techniques. Although there are many reasons why complex or sophisticated techniques may be more useful than simple, naive techniques, the evidence to date clearly suggests that greater forecast accuracy is not one of them.

References

Ahlburg, D.A., 1987, "Population forecasting", in: S. Makridakis and S. Wheelwright, eds., *The Handbook of Forecasting* (John Wiley, New York).

Ahlburg, D.A., 1992, "Error measures and choice of a forecast method: Lessons for forecasters", *International Journal of Forecasting*, 8, 99–100.

Beaumont, P.M., 1989, *ECESIS: An Interregional Economic-Demographic Model of the United States* (Garland Publishing Company, New York).

Beaumont, P.M. and A.M. Isserman, 1987, "Comment", *Journal of the American Statistical Association*, 82, 1004–1009.

Birch, D.L., 1977, "Comments on long-term population and employment forecasting" in: *Reports of the Conference on Economic and Demographic Methods for Projecting Population* (American Statistical Association, Washington DC), 148–152.

Irwin, R., 1977, "Guide to local area population projections", Technical Paper 39, US Bureau of the Census, Washington, DC.

Kale, B., P. Voss, C. Palit and H. Krebs, 1981, "On the question of errors in population projections", unpublished paper presented at the Annual Meeting of the Population Association of America, Washington, DC.

Keyfitz, N., 1981, "The limits of population forecasting", *Population and Development Review*, 7, 579–593.

Land, K.C., 1986, "Methods for national population forecasts: A review", *Journal of the American Statistical Association*, 81, 888–901.

Long, J.F. and D.M. McMillen, 1987, "A survey of Census Bureau population projection methods", *Climatic Change*, 11, 141–177.

Morrison, P. 1977, "Forecasting population of small areas: An overview", in: *Population Forecasting for Small Areas* (Oak Ridge Associated Universities, Oak Ridge, TN), 3–13.

Murdock, S.H., F.L. Leistritz, R.R. Hamm, S.S. Hwang and B. Parpia, 1984, "An assessment of the accuracy of a regional/economic–demographic projection model", *Demography*, 21, 383–404.

Murdock, S.H. and D.R. Ellis, 1991, *Applied Demography* (Westview Press, Boulder).

National Planning Association, 1972, *Regional Demographic Projections 1960–85*, Report 72-R-1, Washington, DC.

National Planning Association, 1976, *Regional Economic Projections: 1960–1990*, Report 76-R-1, Washington, DC.

National Planning Association, 1982, *Regional Economic Projection Series*, Report 81-R-1, Washington, DC.

Pittenger, D., 1980, "Some problems in forecasting population for government planning purposes", *The American Statistician*, 34, 135–139.

Siegel, J.S., 1953, "Forecasting the population of small areas", *Land Economics*, 29, 72–88.

Smith, S.K., 1984, *Population Projections: What Do We Really Know?*, Monograph No. 1, University of Florida, Bureau of Economic and Business Research.

Smith, S.K., 1987, "Tests of forecast accuracy and bias for county population projections", *Journal of the American Statistical Association*, 82, 991–1003.

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

Smith, S.K. and T. Sincich, 1990, "The relationship between the length of the base period and population forecast errors", *Journal of the American Statistical Association*, 85, 367–375.

Smith, S.K. and T. Sincich, 1991, "An empirical analysis of the effect of length of forecast horizon on population forecast errors", *Demography*, 28, 261–274.

US Bureau of the Census, 1956, "Estimates of the population of states: 1900 to 1949", *Current Population Reports*, Series P-25, No. 139, Washington, DC.

US Bureau of the Census, 1957, "Illustrative projections of the population, by state, 1960, 1965, and 1970", *Current Population Reports*, Series P-25, No. 160, Washington, DC.

US Bureau of the Census, 1965, "Revised estimates of the population of states and components of change: 1950 to 1960", *Current Population Reports*, Series P-25, No. 304, Washington, DC.

US Bureau of the Census, 1966, "Illustrative projections of the population of states: 1970 to 1985", *Current Population Reports*, Series P-25, No. 326, Washington, DC.

US Bureau of the Census, 1971, "Preliminary intercensal estimates of states and components of population change: 1960 to 1970", *Current Population Reports*, Series P-25, No. 460, Washington, DC.

US Bureau of the Census, 1972, "Preliminary projections of the population of states: 1975 to 1990", *Current Population Reports*, Series P-25, No. 477, Washington, DC.

US Bureau of the Census, 1979, "Illustrative projections of state populations by age, race, and sex: 1975 to 2000", *Current Population Reports*, Series P-25, No. 796, Washington, DC.

US Bureau of the Census, 1983, "Provisional projections of the population of states by age and sex: 1980 to 2000", *Current Population Reports*, Series P-25, No. 937, Washington, DC.

US Bureau of the Census, 1984, "Estimates of the population of states: 1970 to 1983", *Current Population Reports*, Series P-25, No. 957, Washington, DC.

US Bureau of the Census, 1991, "Preliminary intercensal estimates of the population of states: 1981 to 1989", unpublished table, Washington, DC.

US Bureau of Economic Analysis, 1974, *Area Economic Projections, 1990*, Washington, DC.

US Bureau of Economic Analysis, 1981, *BEA Regional Projections*, Washington, DC.

US Bureau of Economic Analysis, 1985, *BEA Regional Projections*, Washington, DC.

Voss, P., C. Palit, B. Kale and H. Krebs, 1981, "Forecasting state populations using ARIMA time series techniques", Applied Population Laboratory, University of Wisconsin, Madison.

White, H.R., 1954, "Empirical study of the accuracy of selected methods of projecting state population", *Journal of the American Statistical Association*, 49, 480–498.

Biographies: Stanley K. SMITH is a Professor of Economics and Director of the Bureau of Economic and Business Research at the University of Florida. He received his Ph.D. in Economics from the University of Michigan in 1976, with a specialization in Economic Demography. His major research interests include the methodology and analysis of small-area population estimates, the evaluation of population forecast errors, and population trends in Florida.

Terry SINCICH is Assistant Professor of Information Systems/Decision Sciences at the University of South Florida. He received his Ph.D. in Statistics from the University of Florida in 1980. His research interests include applied statistical modeling in the areas of population projections, cost effectiveness of HMO physicians, and job turnover.