ON THE UTILITY OF POPULATION FORECASTS*

JEFF TAYMAN AND DAVID A. SWANSON

Many customers demand population forecasts, particularly for small areas. Although the forecast evaluation literature is extensive, it is dominated by a focus on accuracy. We go beyond accuracy by examining the concept of forecast utility in an evaluation of a sample of 2,709 counties and census tracts. We find that forecasters provide "value-added" knowledge for areas experiencing rapid change or areas with relatively large populations. For other areas, reduced value is more common than added value. Our results suggest that new forecasting strategies and methods such as composite modeling may substantially improve forecast utility.

Evaluations of population forecasts generally have overlooked the concept of utility. Instead, forecasts typically are judged by their accuracy as based on comparisons of ex post facto forecasts with known targets, most often the census. In this paper we aim to broaden our understanding of the utility of population forecasts by addressing two primary questions. First, given the time and other resources needed to construct a forecast, how much "value-added" knowledge do we gain over and above that provided by a low-cost or no-cost and readily available alternative? Second, can useful generalizations be made about forecast utility in specific forecasting situations such as size, direction and rate of change, and type of geographic area? Our exploration of these two questions is empirically based; we analyze forecast utility through a "value-added" criterion proposed by Swanson and Tayman (1995) and called PRE, or proportionate reduction in error measure.

Given the many uses of population forecasts and the need to provide a warning about the size of their errors (Keyfitz 1972), it is not surprising that population forecasts have been evaluated frequently (e.g., Smith 1987; Tayman 1996). These evaluations have created an atmosphere in which the "stand-alone" accuracy of a given forecast, by itself or relative to competing forecasts, is the dominant criterion and consequently the normative expectation for judging adequacy (Yokum and Armstrong 1996). Forecasters and their audiences tend to overlook other aspects of a forecast's utility, and forecasts often are prepared to support clients' decision-making needs. In our investigation we examine this "value-added" aspect of forecast utility.[1]

MEASURING FORECAST UTILITY (PRE)

A proportionate reduction in error measure proposed by Swanson and Tayman (1995) is used to quantify the "value-added" component of forecast utility. Costner's (1965) conceptualization of PRE, although placed in the context of measures of association, remains the most general in its application and most parsimonious in its presentation:

$$\text{PRE} = (\text{error by rule (b)} - \text{error by rule (a)}) \,/\, \text{error by rule (b)}. \qquad (1)$$

Costner imposed a restriction such that error by rule (a) ≤ error by rule (b), which limits PRE to a value between 0 and 1.0 inclusive. For our purpose, this restriction is removed and PRE is multiplied by 100 so that it is expressed as a percentage. With these modifications, PRE has theoretical limits of minus infinity to 100% (Swanson and Tayman 1995).

What constitutes rule (a) and rule (b) in our approach? Rule (a) is a population forecast resulting from a technique such as the cohort component or land use model; rule (b) is based on data already available as the result of an existing count. In this study, rule (b) is the 1980 census count.[2] Thus PRE transfers the concept formulated by Costner to the evaluation of a forecast by establishing a forecast error based on existing information. It evaluates the (presumed) reduction of error found by using the "actual" forecast method (rule (a)) rather than the error of a lower-cost and readily available but "naive" alternative (rule (b)).

If the error of the "actual" forecast is less than the error of the alternative forecast, PRE is positive, and a larger value

*Jeff Tayman, San Diego Association of Governments, 401 B. Street, Suite 800, San Diego, CA 92101; e-mail: jta@sandag.cog.ca.us. David A. Swanson, Center for Population Research and Census, Portland State University. The authors are grateful for comments from a number of colleagues and from the reviewers. An earlier version of this paper was presented at the 1995 meetings of the Population Association of America.

1. Several important issues related to utility are not addressed in this paper. For example, we provide no information on the costs involved in generating the forecasts evaluated here; we simply note that these forecasts cost more than the alternative used in constructing the PREs and are less timely. Nor do we discuss the cost of making bad decisions—to either a client or a forecaster—or the level of accuracy that is "just sufficient" to support a correct decision. These and other issues, such as optimizing accuracy, timeliness, cost minimization, and using alternative definitions of "value-added," form the outline for a broad research agenda on forecast utility.

2. We take the view that the low-cost or no-cost alternative is the last census. This is reasonable because census data are generally available to a wide range of users at very low cost or for nothing. There are alternatives to the last census however. For example, one could take a "stale" forecast that was made before the last census and now is immediately available at no cost. Another possibility is a set of current post censal estimates.

indicates a greater gain of information. Conversely, greater error in the "actual" forecast results in a negative PRE, indicating that the "actual" forecast provides less information about the future (and at a greater production cost) than the readily available alternative. To illustrate, a PRE of 60 indicates that the forecaster reduces the error by 60% using the "actual" rather than the "naive" method. Similarly, a value of −150 shows that the forecaster increases the error by 150% over the error produced by the "naive" method.

Swanson and Tayman (1995) suggest that the PRE be examined by using several measures of forecast error to provide a broader context for evaluating forecast utility. For this purpose, the PRE is computed for the two measures most often used to evaluate forecast precision and bias: the mean absolute percentage error (MAPE) and the mean algebraic percentage error (MALPE) (e.g., Smith 1987; Tayman 1996). The MAPE (average percentage error over all geographic areas, ignoring the direction of the error) measures forecast precision; the MALPE (average percentage error taking into account the direction of the error) measures forecast bias.

DATA AND FORECAST METHODS

We made ex post facto comparisons of 1990 population forecasts and 1990 census counts using a "purposive and convenience" sample of counties from seven states and census tracts from three metropolitan areas.[3] We obtained county forecasts from "official" state demographic centers in Arkansas, Florida, Iowa, North Dakota, Ohio, Washington, and Wyoming. Census tract forecasts were prepared by the Council (or Association) of Governments in the San Diego, Dallas-Fort Worth, and Detroit-Ann Arbor metropolitan areas.[4]

We analyze subcounty areas because of the increasing interest in small-area forecasts as well as consumers' demand, and because the provision of reliable forecasts on this scale is a major challenge to those in this field. Also, the geodemographic methods on which these forecasts are based offer a different approach than is typically used for developing small-area forecasts (e.g., extrapolation or shift-share), and this class of models has received little formal evaluation.

We consciously sought geographic areas that covered the gamut in population size and growth rate (the "purposive" aspect of our sample) and used forecasts that were available not only to us, but also to decision makers and the general public at the time of their release (the "convenience" aspect of the sample). As discussed below, this sample represents a wide range of size and growth characteristics that have been shown to influence forecast error (Smith 1987; Smith and Shahidullah 1995; Tayman 1996) and thus provides a comprehensive assessment of forecast utility.

Table 1 provides a description of the sample. Three states (Iowa, Wyoming, and North Dakota) lost population between 1980 and 1990, and most counties in these states also lost population. A relatively stable population growth characterized Ohio and Arkansas: The great majority of their counties showed a change between −9.9% and 9.9%. Washington (17.8%) and Florida (32.7%) grew much faster than the other five states in the sample: 25% of Washington's counties grew by 20% or more, while 73% of Florida's counties grew at this rate. The average 1980 county population ranged from 12,300 in North Dakota to 145,500 in Florida.

Detroit was the only metropolitan area to lose population between 1980 and 1990; the population declined in six out of 10 census tracts. Conversely, Dallas-Fort Worth (25%) and San Diego (36%) grew rapidly during the 1980s, although their growth patterns varied considerably. Twenty-eight percent of Dallas-Fort Worth's census tracts grew by at least 20% and 43% lost population. San Diego had the smallest percentage of declining census tracts (18%) and the largest percentage of those which grew fastest (51%). Census tracts in Dallas-Fort Worth and San Diego were similar in size, with average 1980 populations of 4,600 and 4,850 respectively. Census tracts in Detroit averaged 3,400 persons.

All of the county forecasts were based on the cohort-component method and used a top-down approach, whereby the counties are "controlled" to an independent state forecast. The subcounty forecasts also used this approach, which involved several forecasting techniques. Forecasts for the overall metropolitan areas were developed using the cohort-component method integrated with an econometric model. These forecasts were disaggregated, typically in two stages, to subcounty geographic levels, using spatial-interaction land use models. The first stage produced a forecast for 150 to 200 zones that were based on aggregations of census tracts. For this first stage, San Diego used the Projective Land Use Model (Tayman and Kunkel 1989), Dallas-Fort Worth used DRAM/EMPAL (Putman 1983), and Detroit used an in-house model. The zonal forecasts were then allocated to traffic analysis zones in Detroit and Dallas-Fort Worth and to areas generally equivalent to census blocks in San Diego.[5] For the latter two areas, we aggregated these forecasts to census tracts for use in this study.

ACCURACY AND UTILITY OF FORECASTS

To provide a reference point for evaluating forecast utility, we first address the usual question asked in these types of studies: How well did the forecast predict the population, as measured by the MAPE and the MALPE? Then we discuss the PRE that describes how these results compare with the error of a "naive" forecast based on the 1980 census.

3. Most of these forecasts were released two to three years after the 1980 census and are themselves informed (trended in accordance with) post censal estimates.

4. For Detroit, the forecasts are organized by traffic analysis zones, which include census tracts divided for transportation modeling. Some census tracts are excluded from the analysis. In San Diego, forecasts were not prepared for the 11 census tracts that represent military bases or for the six census tracts that cover the eastern half of the metropolitan area, which contained less than 1% of the 1980 population. For Dallas-Fort Worth and Detroit, subcounty forecasts were made only for the urbanized part of the metropolitan area, which contained approximately 90% of the 1980 populations.

5. Little has been written about the models used in the second-stage allocation. A discussion of the San Diego model is found in Tayman (1996).

TABLE 1. POPULATION SIZE AND GROWTH RATE CHARACTERISTICS OF THE SAMPLE

Geographic Area	1980 Census	1990 Census	1980–1990 % Change[a]	Distribution of 1980–1990 Percentage Change: < 10.0 %	–9.9–0%	0–9.9%	10–19.9%	20+%	Avg. 1980 Population
Census Tract[b]									
Detroit (1,377)	4,674,728	4,586,155	–1.9	30%	31%	15%	9%	15%	3,395
Dallas-Fort Worth (525)	2,417,016	3,022,127	25.0	22%	21%	17%	12%	28%	4,604
San Diego (363)	1,761,047	2,400,641	36.3	3%	15%	18%	13%	51%	4,851
County[b]									
Iowa (99)	2,913,808	2,776,755	–4.7	49%	44%	6%	1%	0%	29,432
Wyoming (23)	469,557	453,588	–3.4	39%	22%	26%	4%	9%	20,416
North Dakota (53)	652,735	638,800	–2.1	64%	25%	9%	2%	0%	12,316
Ohio (88)	10,797,603	10,847,115	0.5	5%	41%	44%	9%	1%	122,700
Arkansas (75)	2,286,357	2,350,725	2.8	15%	33%	39%	9%	4%	30,485
Washington (39)	4,132,180	4,866,692	17.8	3%	13%	46%	13%	25%	105,953
Florida (67)	9,746,324	12,937,926	32.7	0%	2%	10%	15%	73%	145,468

[a]The 1980–1990 percentage change is based on the larger unit of geography. For example, for Iowa counties it is the state of Iowa; for San Diego census tracts it is the San Diego metropolitan statistical area. The order of the areas represents an ascending sort based on growth rate.

[b]Sample size in parentheses.

Accuracy

As observed in Table 2, the MALPE indicates that bias is low for all areas except Detroit's census tracts and counties in Wyoming and North Dakota. The level of bias is not related to whether the area is a county or a census tract; the range of the MALPE is similar in either case. Washington, Florida, and San Diego, which have the highest growth rates, show a negative MALPE, indicating that forecasters were unable to capture the full extent of the population increase. The largest positive MALPEs, however, indicate that forecasters in Detroit, Wyoming, and North Dakota were optimistic about growth and did not anticipate the population losses that occurred.

Unlike the MALPE, the MAPE differs in size between counties and census tracts. County forecasts are more precise than forecasts for the census tracts. Counties in every state have a lower MAPE than census tracts in every metropolitan area, with the exception of North Dakota, whose MAPE exceeds that of Dallas-Fort Worth. Moreover, counties in five of the seven states have a MAPE of less than 10, while the smallest MAPE for the census tracts is 19.

Utility

As we discuss the PRE scores in Table 2, we must keep in mind the value of a given error measure when using the 1980 census to predict the 1990 population. If the value is small, there is not much room for reducing error and a great deal of room for increasing it. If the value is large, there is a great deal of room for reducing error and not much room for increasing it.

The PRE shows that in six of 10 comparisons, forecasters not only fail to decrease forecast bias from that obtained by using the 1980 census, but actually increase it. This poor showing is not restricted to census tracts; it also occurs consistently in the counties. Iowa is the only sample area that lost population between 1980 and 1990 in which the forecast shows less bias than does the census, with a PRE of 21%. The large increase in the bias for Arkansas (PRE –400%) shows that because of the relatively small error resulting from the census-based forecast, it is difficult for forecasters to improve their prediction. In areas with moderately high growth (Washington) and very high growth (Florida and San Diego), forecasters substantially reduce forecast bias, with MALPE PREs greater than 75%. Dallas-Fort Worth, however, has a MALPE PRE of –76%, but its census-based MALPE of –3.3 also leaves little room for improvement.

The picture is somewhat less gloomy with respect to the forecasters' precision. Seventy percent of the MAPE PREs are positive and range from 3% to 61%. Detroit, North Dakota, and Wyoming are the only areas where the 1980 census outperforms the forecaster, with PREs ranging from –33% to –6%. It is clear that the greatest "value added" related to precision is found in the counties. The highest PRE for the census tracts is 26%, compared with 61% for the counties. The counties in four states have PREs greater than 30%.

How well did the forecasters in view of the opportunity (or lack of it) offered by the census-based forecasts? Florida, Dallas-Fort Worth, and San Diego show large errors in terms of the census-based forecasts. Forecasters in these areas took advantage of this "room for improvement" and posted reductions in error, especially for Florida. North Dakota and De-

TABLE 2. FORECAST BIAS, PRECISION, AND UTILITY, BY GEOGRAPHIC AREA

	MALPE			MAPE		
Geographic Area	Forecast[a]	Naive Forecast[b]	PRE	Actual Forecast[a]	Naive Forecast[b]	PRE
Census Tract						
Detroit	20.1	9.4	–114%	28.5	23.3	–22%
Dallas-Fort Worth	5.8	–3.3	–76%	18.6	25.3	26%
San Diego	–3.6	–18.9	81%	20.9	21.5	3%
County						
Iowa	7.4	9.4	21%	8.0	10.0	20%
Wyoming	15.1	6.0	–152%	15.1	14.2	–6%
North Dakota	19.2	12.8	–50%	19.2	14.4	–33%
Ohio	–1.4	–1.1	–27%	3.2	5.0	36%
Arkansas	3.5	0.7	–400%	4.1	7.6	46%
Washington	–2.0	–9.0	77%	4.3	11.1	61%
Florida	–5.3	–24.8	79%	9.8	24.8	60%

[a]Forecasts prepared in the early 1980s using cohort-component and land use models.
[b]Based on the 1980 census.

troit, where the population declined, also are consistently high in error measures based on the 1980 census. In these cases, however, forecasters not only fail to reduce error but actually increase it. Ohio and Arkansas are consistently low in terms of the error measures based on the 1980 census because they changed least among the seven states. These states provide little "room for improvement"; nevertheless, forecasters managed to increase their precision.

We also applied the PRE measure to forecasts made for places of varying size and different growth rates. These results are not shown, but we summarize them as follows:[6] In both the counties and the census tracts, a strong direct relationship emerges between population size and the "value added" by the forecast, as measured by the MAPE. The 1980 census outperforms the forecasters for census tracts with fewer than 5,000 persons; then the MAPE PRE rises steadily and reaches 38% in census tracts with at least 7,500 persons. Forecasters also create a higher MAPE for the smallest counties (<10,000 population), but the MAPE PRE reaches 64% for the largest counties (≥40,000 population).

Forecasters substantially reduce the MAPE for areas experiencing rapid change. The fastest-growing counties (≥15% change) and census tracts (≥50% change) have the highest MAPE PREs: 66% and 42% respectively. Forecasters generally did not fare well with slow, stable, and declining growth rates. All four categories in which census tracts grew by less than 20% yield more precise census-based forecasts. A similar pattern occurs for counties that grew by less than 5%.

6. The detailed tables by size and growth rate are available from the authors.

DISCUSSION

We now turn to the first question posed in the introduction: Given the time and other resources needed to construct a forecast, how much do we gain in "value-added" knowledge over and above a "naive" alternative that is readily available at virtually no cost to a client? Swanson and Tayman (1995) offer guidelines for judging the PRE; Table 3 shows the distribution, according to this criterion, for every comparison of area, size, and growth rate.

These forecasters generally do not provide significant improvements over a "naive" alternative, and the gains are much less for the census tracts than for the counties. Seventy-one percent of the census tract comparisons show bad to poor results (PRE < 26%), compared with 55% of the counties. For comparisons showing good to excellent improvement (PRE > 50%), the corresponding figures are 18% and 30% respectively for the census tracts and the counties. This finding is not surprising because census tracts are much smaller than counties and their growth rates vary more widely.

The MALPE PRE distributions contain more extreme values than the MAPE distributions. Forecasters are either very good or very bad at predicting forecast bias, and our data show that they are more often very bad. Almost three-quarters of the MALPE PREs fall into the lowest and the highest categories for both the census tracts and the counties. The corresponding figures for the MAPE PREs are 57% and 29%. Moreover, the only excellent scores (PRE > 75%) occur in predictions of bias. For the census tracts the best results indicate only average reductions in absolute error, compared with the "naive" forecasts.

TABLE 3. DISTRIBUTION OF PROPORTIONATE REDUCTION IN ERROR MEASURE

Geographic Area	Poor < 0%	Bad 0–25%	Average 26–50%	Good 51–75%	Excellent 76–100%	Number of Comparisons
All Measures						
Census tract	57%	14%	11%	11%	7%	28
County	38%	17%	15%	15%	15%	34
MALPE						
Census tract	57%	7%	0%	21%	15%	14
County	47%	18%	0%	6%	29%	17
MAPE						
Census tract	57%	21%	22%	0%	0%	14
County	29%	18%	29%	24%	0%	17

Our second question is this: Can useful generalizations be made about gains in "value-added" knowledge in specific types of forecasting situations? We find that "value" is added by forecasters where change occurs. Forecasters, however, tend to place growth everywhere and are not adept at identifying areas that do not change or that undergo decline.

As a working hypothesis we suggest that the reasons why forecasters tend to view change predominantly in the form of population increase involve both the technical and the normative environments in which forecasts are made. An example of the technical aspect is the tendency to use "top-down" forecasting, which allocates growth everywhere in the system when (as is usually the case) growth is forecast for the area as a whole. An example of this normative environment is found in the situation for Detroit. The forecasters preparing the metropolitan area forecast initially showed that it declined between 1980 and 1990. This outcome, however, was not viewed as acceptable by makers of policy and other decisions because of its effect on attracting business and activity to the area. Thus the forecast was altered; the result was a large upward bias in comparison with the 1990 census.

We also find that forecasters and the "last census" apparently have a common characteristic: Both are better at predicting the future for large places than for small places. In such cases, forecasters offer improvements over the fairly accurate "predictions" already made by the last census. Clearly both have the ability to "see" the future of large places better than small places. Unlike the situation in the 1980 census, however, forecasters tend to steadily increase their ability to foresee the demographic future as size increases. The inability to forecast small places accurately, although known for decades (Schmitt 1954), is probably a more pressing issue today than 10 or 15 years ago. Clients' needs and legislative actions such as the Clean Air and Intermodal Surface Transportation Efficiency Acts are increasing the demand for, and use of, small-area forecasts.

CONCLUSION

Our experience, conversations with colleagues, and the literature suggest the existence of knowledge, although somewhat fragmented, of where and how forecasting techniques and processes go wrong. However, not much progress has been made in improving them. Possibly it is difficult to develop generalizations for activities such as forecasting because the problems of applied demography are posed largely by customers, and characteristically have unique solutions and involve unique methods (Swanson, Burch, and Tedrow forthcoming). Yet because of this collective demand on the part of customers, demographers and others must find new methods and solutions for improving our forecasts.

We suggest that, in addition to the issues stated in footnote 1, the research agenda for population forecasting should include projects aimed at improving the ability to incorporate constancy and decline into the forecasting process. The evidence presented in this paper strongly suggests that utility, in the form of "value-added" accuracy, would improve markedly if we could identify, in a timely and cost-efficient manner, areas that are likely to remain stable or decline and could understand what differentiates them from areas that are likely to grow. One way of doing this may be to examine the history of areas and develop generalizations in the form of typologies, protocols, or models (Perry and Voss 1996). "Composite" modeling, a related avenue of research, also should be pursued (Smith and Shahidullah 1995). This approach suggests that using different techniques for places with different characteristics may improve forecast accuracy. We could greatly enhance the usefulness of forecasts and provide additional incentive for decision makers to use and purchase them by adding sufficient value above the last census or some other readily available low-cost or no-cost alternative.

REFERENCES

Costner, H. 1965. "Criteria for Measures of Association." *American Sociological Review* 30:341–53.

Keyfitz, N. 1972. "On Future Population." *Journal of the Ameri-*

can Statistical Association 67:347–62.

Perry, M. and P. Voss. 1996. "Using Geo-Demographic Methods for Improving Small Area Estimates." Presented at the annual meetings of the Population Association of America, New Orleans.

Putman, S. 1983. *Integrated Urban Models*. London: Pion.

Schmitt, R. 1954. "A Method for Projecting the Population of Census Tracts." *Journal of the American Institute of Planners* 20:102.

Smith, S. 1987. "Tests of Accuracy and Bias for County Population Projections." *Journal of the American Statistical Association* 82:991–1003.

Smith, S. and M. Shahidullah. 1995. "An Evaluation of Population Projection Errors for Census Tracts." *Journal of the American Statistical Association* 90:64–71.

Swanson, D., T. Burch, and L. Tedrow. Forthcoming. "What Is Applied Demography?" *Population Research and Policy Review*.

Swanson D. and J. Tayman. 1995. "Between a Rock and a Hard Place: The Evaluation of Demographic Forecasts." *Population Research and Policy Review* 14:233–49.

Tayman, J. 1996. "The Accuracy of Small Area Population Forecasts Based on a Spatial Interaction Land Use Modeling System." *Journal of the American Planning Association* 62:85–98.

Tayman, J. and S. Kunkel. 1989. "Improvements to the Projective Land Use Model for Producing Small Area Forecasts." Presented at the annual meetings of the Population Association of America, Baltimore.

Yokum, J. and J. Armstrong. 1996. "Beyond Accuracy: Comparison of Criteria Used to Select Forecasting Models." *International Journal of Forecasting* 11:591–97.