Popul Res Policy Rev (2007) 26:163–184
DOI 10.1007/s11113-007-9030-0

Population forecast accuracy: does the choice of summary measure of error matter?

Stefan Rayer

Received: 6 June 2005 / Accepted: 21 February 2006 / Published online: 27 March 2007

Abstract Population projections are judged primarily by their accuracy. The most commonly used measure for the precision component of accuracy is the mean absolute percent error (MAPE). Recently, the MAPE has been criticized for overstating forecast error and other error measures have been proposed. This study compares the MAPE with two alternative measures of forecast error, the Median APE and an M-estimator. In addition, the paper also investigates forecast bias. The analysis extends previous studies of forecast error by examining a wide range of trend extrapolation techniques using a dataset that spans a century for a large sample of counties in the US. The main objective is to determine whether the choice of summary measure of error makes a difference from a practitioner's standpoint. The paper finds that the MAPE indeed produces error values that exceed the robust measures. However, except for situations where extreme outliers rendered the MAPE meaningless, and which are rare in real world applications, there was not a single instance where using an alternative summary measure of error would have led to a fundamentally different evaluation of the projections. Moreover, where differences existed, it was not always clear that the values and patterns provided by the robust measures were necessarily more correct than those obtained with the MAPE. While research into refinements and alternatives to the MAPE and mean algebraic percent error are worthwhile, consideration of additional evaluation procedures that go beyond a single criterion might provide more benefits to producers and users of population forecasts.

Keywords Forecast accuracy · MAPE · Error measures · Population projections

S. Rayer (✉)
Bureau of Economic and Business Research, University of Florida, 221 Matherly Hall, P.O. Box 117145, Gainesville, FL 32611-7145, USA
e-mail: stefanr@bebr.ufl.edu

Introduction

Population projections are first and foremost judged by their accuracy (Yokum & Armstrong, 1995). This study focuses on the two main components of forecast accuracy, precision and bias.[1] Examinations of projection precision have traditionally used the mean absolute percent error (MAPE) as the summary measure of error of choice (Isserman, 1977; Murdock, Leistritz, Hamm, Hwang, & Parpia, 1984; Smith, 1987; Smith & Shahidullah, 1995; Smith & Sincich, 1988, 1990, 1991, 1992; Tayman, Schafer, & Carter, 1998; White, 1954). In recent years, the ubiquitous use of the MAPE has been challenged. In particular, it has been argued that the MAPE tends to overstate forecast error, and that other measures are more suited and should be used instead to evaluate the precision of population projections (Coleman & Swanson, 2004; Swanson, Tayman, & Barr, 2000; Tayman & Swanson, 1999).

Many of the studies that have evaluated projection errors did so using a limited set of data both with regard to space and time (see, e.g., Isserman, 1977; Murdock et al., 1984; Smith, 1987; Smith & Sincich, 1991; Tayman & Swanson, 1999; Tayman et al., 1998), or were based on simulations (Coleman & Swanson, 2004). In particular, there exists a paucity of sub-state projection evaluations that examine more than a few decades of data, and which are nationally representative. This is unfortunate because small area population projections are most frequently used for actual planning purposes. Moreover, in order to make generalizations about the characteristics of the various summary measures of error it is advisable to have sample data that span many decades and cover a large number of geographic areas. The present study attempts to fill this gap by analyzing a range of commonly applied trend extrapolation techniques using population data from every decennial census from 1900 to 2000 for all counties in the continental US for which comparable data are available.

The main objective of this study is to determine whether the choice of summary measure of forecast error makes a difference from a practitioner's standpoint. Specifically, the paper examines whether using a robust measure of forecast error would lead to different conclusions about the precision and bias of the projections vis-à-vis using the MAPE/mean algebraic percent error (MALPE). For that purpose, the analysis compares results obtained with the most popular summary measure of forecast precision, the MAPE, with robust measures such as the Median APE and an M-Estimator, Tukey's biweight. In addition to the focus on precision, the paper also examines projection bias, by comparing results obtained with the most commonly used summary measure—MALPE—to more robust alternative measures. The study starts with investigating the various error measures in the aggregate. This is followed by a more detailed analysis of different base period lengths and county growth characteristics. Using 100 years of data, almost 2,500

[1] Demographic convention considers *projections* to be conditional statements about the future, which reflect the outcome of particular assumptions, while *forecasts* are unconditional statements that reflect what the analyst believes most likely to happen in the future (Smith, Tayman, & Swanson, 2001). The terms projection and forecast are used interchangeably in this article, because most people consider projections to be forecasts, whatever the intention of the producer.

counties, ten trend extrapolation techniques, and three error measures for assessing forecast precision and bias, the paper extends previous projection evaluations and helps both users and practitioners of population projections make better informed decisions.

Data and techniques

Throughout the twentieth century many counties in the US experienced changes to their boundaries that make a comparison of population figures from one census to the next problematic. In order to preserve comparability, the analysis was limited to those counties that did not experience significant boundary changes over the study period. To determine which of the changes were significant, this study largely follows Forstall, who identified the census date since which each ''county has had no significant territorial change, that is, a boundary change large enough to have a significant effect on the county's population as of the preceding census'' (Forstall, 1996, p. 8). This resulted in a total number of 2,482 counties that did not experience significant boundary changes between 1900 and 2000, which amounts to 79.0% of all the counties in Census 2000. In a companion study, it has been shown that this restricted sample can be considered to be representative of the nation at large (Rayer, 2004).

The terminology used follows Smith et al. (2001): base year refers to the year of the earliest data used to make a projection; launch year refers to the year of the most recent data used to make a projection; target year refers to the year for which a population is projected; base period refers to the interval between the base year and the launch year of the projection; and projection horizon refers to the interval between the launch year and the target year of a projection.

Covering census data for the entire twentieth century, the analysis involves 125 projection horizon/base period combinations spanning a range between 10 and 50 years. Specifically, the analysis includes 35 10-year, 30 20-year, 25 30-year, 20 40-year, and 15 50-year projection horizons and base periods. The first target year for which projections were made was 1920, a 10-year projection with a 1910 launch year and a 1900–1910 base period. In addition, three projections were made for 1930, 6 for 1940, 10 for 1950, 15 for 1960, 19 for 1970, 22 for 1980, 24 for 1990, and 25 for 2000.

For each of these 125 projection horizon/base period combinations, a total of ten projection techniques were applied, including seven primary techniques and three averages. The primary techniques include linear (LIN), modified linear (MLN), share-of-growth (SHR), shift-share (SFT), exponential (EXP), constant-share (COS), and constant (CON). In addition, the study calculated three average projections comprising all seven trend extrapolation techniques (AV7), excluding the highest and the lowest projection (AV5), and excluding the two highest and two lowest projections (AV3). The seven primary techniques were calculated as follows:

LIN: In the LIN extrapolation technique, it is assumed that the population will increase (decrease) by the same number of persons in each future decade as the average per decade increase (decrease) observed during the base period:

$$P_t = P_l + x/y(P_l - P_b),$$

where P_t is the population in the target year, P_l the population in the launch year, P_b the population in the base year, x the number of years in the projection horizon, and y is the number of years in the base period.

MLN: The MLN extrapolation technique initially equals the LIN method but in addition distributes the difference between the sum of the linear county projections and the independent national projection proportionally by population size at the launch year:

$$P_{it} = \text{LIN} + P_{il}/P_{jl}\left(P_{jt} - \Sigma\text{LIN}\right),$$

where i represents the county, j the nation, and LIN is the previously calculated linear county projection for the target year.

The MLN method, as well as the SHR, SFT, and COS techniques, require an independent national projection for the target year population. Although population projections for the nation have been available for quite a long time (see, e.g., Bonynge, 1852; Pritchett, 1891; Whelpton, 1928), there exists no satisfactory set that covers all the target years used in this study. Instead, a new set was produced by applying the LIN and EXP trend extrapolation techniques to the national population, using the appropriate base period and target year. For example, county projections for the year 2000 with a 1970–1990 base period were calculated using a national projection for 2000 derived from extrapolating population change during the 1970–1990 base period. To flatten out the discrepancies between the LIN and EXP methods, an average of the two techniques was then calculated and used for the four ratio methods.[2]

SHR: In the SHR technique, it is assumed that the county's share of population growth will be the same over the projection horizon as it was during the base period:

$$P_{it} = P_{il} + \left[(P_{il} - P_{ib})/\left(P_{jl} - P_{jb}\right)\right](P_{jt} - P_{jl}).$$

SFT: In the SFT technique, it is assumed that the average per decade change in each county's share of the national population observed during the base period will continue throughout the projection horizon:

$$P_{it} = P_{jt}\left[P_{il}/P_{jl} + (x/y)\left(P_{il}/P_{jl} - P_{ib}/P_{jb}\right)\right].$$

EXP: In the EXP technique, it is assumed that the population will grow (decline) by the same rate in each future decade as it did, per decade, during the base period:

[2] The ratio methods were also run using actual census data for the target years instead of the inputs from the projection averages. The results were very similar, which shows that the projections made with the ratio techniques are not very sensitive to the choice of a national projection. Because regular ex ante ratio method population projections do not have actual census data for the nation available for the target years when the projections are made, and require an independent national projection, the ex post analysis presented here likewise uses projected rather than census values.

$$P_t = P_l\, e^{rx}, \quad r = \frac{\ln(P_l/P_b)}{y},$$

where e is the base of the natural logarithm and ln is the natural logarithm.

COS: In the COS technique, it is assumed that the county's share of the national population will be the same in the target year as it was in the launch year:

$$P_{it} = \left(P_{il}/P_{jl}\right)P_{jt}.$$

CON: In the CON technique, it is assumed that the county population in the target year is the same as in the launch year:

$$P_t = P_l.$$

Summary measures of forecast accuracy

This analysis focuses on the precision of population projections and on their bias. According to the National Research Council (1980), any summary measure of error should meet the criteria of measurement validity, reliability, ease of interpretation, clarity of presentation, and support of statistical evaluation. In addition, sensitivity and relationship to decision making can be important considerations (Armstrong & Collopy, 1992). With respect to precision, the most popular error measure in population forecasting is the mean absolute percent error or MAPE (see, e.g., Ahlburg, 1995; Isserman, 1977; Smith, 1987; Smith & Sincich, 1988, 1990, 1992). It is calculated as follows:

$$\mathrm{MAPE} = \frac{\Sigma|\mathrm{PE}_t|}{n}, \quad \mathrm{PE}_t = [(F_t - A_t)/A_t] \times 100,$$

where PE represents the percent error, t the target year, F the population forecast, A the actual population, and n is the number of areas.

Projections that are completely precise result in a MAPE of zero. The MAPE has no upper limit—the larger the MAPE, the lower the precision of the projections. The MAPE is popular, because it meets most of the above described desired criteria, although it has been criticized particularly with respect to its reliability and validity (Coleman & Swanson, 2004; Swanson et al., 2000; Tayman & Swanson, 1999). Being an arithmetic mean, the MAPE is susceptible to outliers, and because the average percent errors of population projections are right-skewed, it has been argued that it tends to overstate forecast error.

To overcome the problem with the effect of outliers, the Median APE is sometimes used as a robust alternative measure of error (Armstrong & Collopy, 1992; Tayman, 1996). The Median APE represents the error that falls in the middle of the distribution: half of the absolute percent errors are larger and half are smaller. The median, being essentially an extreme trimmed mean, is not influenced by the presence of a few extreme values. However, it neglects some of the information contained in the sample and responds to the presence of small errors in the centermost observations that may result from grouping (Rosenberger & Gasko, 1983, p. 302). Error measures based on a median rather than a mean are therefore better suitable for large sample data.

Another alternative was proposed by Tayman, Swanson, and Barr (1999) and Swanson et al. (2000). They advance a method based on a modified Box-Cox transformation, from which they calculate an average measure of error termed MAPE-R. The transformation works by creating a symmetrical distribution of the APEs from which a summary measure of accuracy can be calculated that reduces the impact of outliers. MAPE-R has been applied to evaluate forecast errors in state projections by Census Bureau staff (Campbell, 2002). Using MAPE, the ratio of MAPE to Median, and MAPE-R, Campbell found that MAPE overstated forecast errors and that ‘‘a different conclusion would have been drawn if the original error distribution were not corrected for skewness and asymmetry’’ (2002, p. 9). However, the differences between the three error measures were not that great. While the MAPE did produce the largest projection errors throughout, the Median APE and MAPE-R were rather similar, and using the median-based measure would have led to essentially the same conclusions as with MAPE-R.

A final summary measure of error considered here refers to the minimization (M) estimators. Tayman and Swanson (1999) advocate using M-estimators as robust measures of forecast accuracy. M-estimators assign different weights to values depending on their distance from the center of the distribution, thus minimizing the impact of outliers. Four M-Estimators were introduced by Goodall (1983): Huber’s H, redescending Hampel, Andrews’ waves, and Tukey’s biweight. Using differing tuning constants, the M-estimators vary by their resistance to outliers and their efficiency for heavy-tailed distribution, with Huber’s H being the least resistant and Andrews’ waves and Tukey’s biweight being the most resistant. For a more detailed discussion of the application of M-estimators in the context of population forecasts see Tayman and Swanson (1999). Hoaglin, Mosteller, and Tukey (1983) provide the theoretical background.

Using ex post facto comparisons of 1990 population projections and 1990 census counts for seven states at the county level, Tayman and Swanson (1999) found MAPE to overstate forecast error between 20 and 100%, with the M-estimators, especially Andrews’ waves and Tukey’s biweight having a higher level of validity. However, while M-estimators may have higher validity than MAPE, they lack the intuitive and interpretative qualities of the latter, which makes their widespread adoption problematic.

For the purpose of this study, three summary measures of forecast precision will be compared: MAPE, Median APE, and Tukey’s biweight.[3] Most of the literature regarding error measures for population projections deals with precision. Investigations of forecast bias are less frequent, and they are generally limited to the

[3] Although there are some differences between the four M-estimators, they are not great. Tayman and Swanson (1999), following Goodall (1983), recommend the use of Tukey’s biweight, because it is the most popular and robust M-estimator.

application of the mean algebraic percent error or MALPE (see, e.g., Smith, 1987; Smith & Shahidullah, 1995; Smith & Sincich, 1988, 1990; Tayman, 1996). The MALPE can be calculated analogously to the MAPE, though using algebraic rather than absolute percent errors. Negative values on the MALPE indicate a tendency for projections to be too low, while positive values indicate a tendency for projections to be too high. There has been very little research devoted to refinements of the MALPE along the lines of the robust alternatives to the MAPE. When bias is analyzed in more detail, it is generally through the application of another indicator for bias, such as the proportion of positive errors, rather than a refined version of the MALPE (see, e.g., Smith & Sincich, 1992). To investigate whether the findings obtained from the comparison of the various summary measures of forecast precision are also applicable to an analysis of forecast bias, the study will conclude with a discussion of the MALPE vis-à-vis potential alternative measures.

Analysis

The analysis starts by examining the three summary measures of error for the ten trend extrapolation methods in the aggregate using the entire century of county data. Table 1 reports forecast precision for projection horizons ranging from 10 to 50 years in addition to the overall average. Within each horizon, the data represent the average of all target years and base period lengths. For each projection horizon there are three rows of data representing the three chosen summary measures of forecast precision: MAPE, Median APE, and Tukey's biweight.

The data in Table 1 lend initial support to the notion that MAPE overstates forecast error (Swanson et al., 2000). For every technique and horizon length, the MAPEs are consistently higher than either the Median APEs or the M-estimator. For most techniques the MAPE exceeds the Median APE and the M-Estimator by about 30–40%. These results are largely in line with what Tayman et al. (1999) found by looking at the MAPE to Tukey-M ratio for counties and census tracts from a select number of states for 1990. The differences between the MAPE and the robust measures are greatest for EXP, where even for the shortest projection horizon the MAPE exceeds both the Median APE and the M-estimator by more than 50%, and smallest for SFT, COS, and CON, ranging from 10 to 37%, depending on the length of the horizon. For LIN, MLN, SHR, AV5, and AV3 the MAPE to Median ratio stays fairly constant between 30 and 43%, irrespective of horizon length.

While the MAPE exceeds the Median APE and the M-estimator for every projection technique and projection horizon, the differences are most pronounced for EXP—and AV7 by extension—for which the MAPE takes on truly enormous values for the longer horizons. This highlights a core problem with the MAPE, which is bounded on the left but unbounded on the right. In general, the amount by which the MAPE exceeds the robust measures depends on the degree of asymmetry in the error distribution. Because population projections are usually not allowed to go negative, the maximum error for under-projecting a population is fixed at 100%, whereas there is no upper limit for projections that are too high. The EXP method is particularly prone to produce extreme values on the high end, especially for longer

Table 1 Comparison of MAPE, Median APE, and Tukey's biweight by projection horizon and projection technique

N	Horizon	Error measure	LIN	MLN	SHR	SFT	EXP	COS	CON	AV7	AV5	AV3
35	10	MAPE	11.0	11.0	11.4	14.3	13.0	14.2	11.7	10.6	10.6	10.9
35	10	Median APE	8.1	8.1	8.4	10.9	8.3	12.0	8.9	7.7	7.8	8.0
35	10	Tukey's biweight	7.9	7.8	8.2	10.7	8.0	12.0	8.8	7.5	7.6	7.8
30	20	MAPE	20.4	20.3	21.9	30.3	47.1	27.6	19.7	22.6	19.5	20.1
30	20	Median APE	15.3	15.1	16.4	23.7	15.7	23.3	15.8	14.4	14.6	15.1
30	20	Tukey's biweight	15.1	14.8	16.0	23.5	15.1	23.4	15.7	14.1	14.3	14.8
25	30	MAPE	31.2	31.0	34.2	48.6	1,371.2	42.0	26.8	219.7	29.8	30.7
25	30	Median APE	24.1	23.4	26.2	40.8	24.3	34.7	22.4	22.3	22.7	23.5
25	30	Tukey's biweight	23.7	23.0	25.8	42.1	23.3	34.5	22.6	21.9	22.4	23.1
20	40	MAPE	40.9	41.3	46.0	64.1	147,654.4	58.4	33.5	21,124.6	39.9	40.8
20	40	Median APE	31.4	30.6	34.8	57.3	32.0	46.2	28.7	28.8	29.7	30.5
20	40	Tukey's biweight	31.0	30.2	34.7	56.0	30.1	45.4	29.1	28.0	29.3	30.2
15	50	MAPE	51.1	52.5	58.9	80.0	15,255,460.9	74.3	39.5	2,179,391.3	51.5	52.2
15	50	Median APE	37.8	37.3	42.6	71.5	39.3	56.0	34.1	34.8	36.1	37.1
15	50	Tukey's biweight	37.3	36.5	41.8	65.0	36.1	54.2	34.6	33.3	35.0	36.2
125	All	MAPE	26.9	27.1	29.7	40.8	1,854,569.2	37.3	24.3	264,959.2	26.1	26.8
125	All	Median APE	20.3	19.9	22.2	34.6	20.8	30.0	20.2	18.9	19.3	19.9
125	All	Tukey's biweight	20.0	19.6	21.9	33.8	19.7	29.6	20.3	18.4	18.9	19.5

horizons, which is why we see the large differences between the MAPE and the robust measures for this technique. This can be problematic for areas with small population bases that sometimes experience rapid population changes over short periods of time. For example, the population of Crane County, Texas, increased from 37 persons in 1920 to 2,221 persons in 1930, a growth of about 6,000% for the decade. Basing any projection on such an increase will lead to questionable results at best, but when used with the EXP method and for a long projection horizon, it should come as no surprise that astronomical projection values, and consequently absolute percent errors, are the outcome. In practice, nobody should produce a projection knowing these input parameters without making reasonable adjustments. While clearly an extreme case, it nicely illustrates the vulnerability of mean-based summary measures of error to outliers.

The two robust measures of forecast precision—Median APE and Tukey's biweight—provide essentially identical results. For most projection horizons and techniques the Median APE exceeds the biweight by a small amount, with the greatest differences coming from EXP, which, as demonstrated above, is particularly susceptible to extreme forecast errors. But even here the differences are not great, amounting to less than 10% for each horizon. These results are echoed by the analysis of forecast bias (data not shown).

As demonstrated in Table 1, the MAPE consistently reported the highest forecast errors of the three measures. But would have using the MAPE rather than a robust measure of error led to different conclusions regarding the precision of the projections, apart from the overall level of reported error? For most techniques, the results are rather similar with respect to the three error measures. The clear exception, however, are the projections obtained with the EXP method. Here, the MAPE takes on truly enormous values, especially for horizons exceeding 20 years. This is a function of a relatively small number of counties, usually with very small populations, which experienced rapid increases in population at some point in time, mostly in the early decades of the twentieth century. The above given example of Crane County, Texas, is a case in point. Having just a few of these obvious outliers in the analysis greatly affects any mean-based summary measure of error, and the outcome can clearly be seen in Table 1. There are some other differences in interpretation of the results between the MAPE on the one hand, and the two robust measures of error on the other hand—such as the MAPE identifying CON as the primary projection technique with the greatest precision for the 20-year projections compared to MLN using the robust measures—but overall the three summary measures of forecast error lead to similar conclusions, apart from the overall level of precision reported.

Having compared the three error measures in the aggregate, the analysis turns now to a more disaggregate level. First, different base period lengths within each projection horizon are investigated. This is followed by an analysis of county growth patterns. Tables 2–4 display the forecast error for each projection horizon differentiated by the length of the base period, ranging from 10 to 50 years. Table 2 shows the MAPEs, Table 3 the Median APEs, and Table 4 results obtained with Tukey's biweight.

Table 2 MAPE by projection horizon, base period, and projection technique

N	Horizon	Base	LIN	MLN	SHR	SFT	EXP	COS	CON	AV7	AV5	AV3
9	10	10	12.1	12.0	12.5	14.3	16.9	14.9	11.7	11.6	11.4	12.0
8	10	20	10.5	10.5	10.9	13.0	11.7	14.3	11.7	10.1	10.2	10.4
7	10	30	10.5	10.4	10.9	13.7	11.8	13.9	11.7	10.2	10.2	10.3
6	10	40	10.8	10.7	11.2	14.7	11.6	14.2	11.7	10.5	10.4	10.6
5	10	50	11.0	10.8	11.5	16.5	11.7	13.4	11.7	10.7	10.6	10.7
8	20	10	21.5	21.5	23.0	29.2	104.8	28.6	19.7	30.7	20.2	21.3
7	20	20	19.7	19.7	21.1	28.2	27.2	27.8	19.7	19.2	18.9	19.4
6	20	30	20.1	20.1	21.6	30.3	27.4	28.0	19.7	19.9	19.3	19.8
5	20	40	20.5	20.3	21.9	31.9	24.9	27.3	19.7	20.0	19.7	20.1
4	20	50	19.7	19.3	21.2	34.0	23.8	24.8	19.7	19.3	18.9	19.2
7	30	10	33.2	33.1	36.5	49.0	4,731‘.2	44.3	26.8	699.9	31.1	32.7
6	30	20	31.0	31.0	34.0	46.9	79.7	43.8	26.8	35.0	29.5	30.5
5	30	30	31.0	31.0	34.0	48.3	70.7	42.3	26.8	34.3	29.8	30.5
4	30	40	30.1	30.0	32.9	48.7	49.7	39.6	26.8	30.9	29.1	29.6
3	30	50	28.7	28.1	31.9	51.6	43.8	35.6	26.8	29.1	27.7	28.1
6	40	10	44.9	45.3	50.8	68.0	491,359.2	63.6	33.5	70,226.8	43.1	44.9
5	40	20	40.7	41.3	45.5	62.0	560.2	60.1	33.5	111.0	39.6	40.6
4	40	30	39.2	39.8	43.9	61.9	386.0	56.7	33.5	85.9	38.7	39.2
3	40	40	37.5	37.9	42.1	62.1	131.2	52.5	33.5	48.6	37.2	37.6
2	40	50	37.4	37.6	42.8	65.3	97.1	50.6	33.5	43.8	37.3	37.5
5	50	10	56.5	57.7	65.1	85.9	45,756,386.3	79.7	39.5	6,536,668.7	56.0	57.5
4	50	20	49.5	51.0	56.7	76.3	9,278.7	74.6	39.5	1,364.3	49.8	50.5
3	50	30	48.1	49.6	55.4	77.5	3,756.4	71.6	39.5	575.2	49.1	49.3
2	50	40	47.2	48.9	54.3	76.3	611.9	69.2	39.5	125.8	48.5	48.5
1	50	50	46.9	48.2	55.3	79.5	374.2	64.9	39.5	91.5	48.5	48.3

The data in Table 2 make clear that the length of the base period has only a minor impact on the precision of the projections for most techniques, using the MAPE as summary measure of error.[4] The projections produced with the EXP technique form an exception. Here, just lengthening the base period from 10 to 20 years results in significantly enhanced precision for all projection horizons, with the improvement most pronounced for the longer horizons. However, even using base periods of up to 50 years still produces projections with the lowest precision of any technique for the 40 and 50-year horizons. AV7, which includes all seven primary projection techniques, and which is heavily influenced by the extreme outliers associated with EXP, displays a similar dynamic. For all other methods, differences in forecast precision due to varying base period lengths within each projection horizon are

[4] Because CON holds the population constant at the launch year value there are no differences in the error measures among the five base periods. COS holds the county's share of the national population constant at the launch year's value, but different base periods come into play in the form of the national projections used in their calculation.

Table 3 Median APE by projection horizon, base period, and projection technique

N	Horizon	Base	LIN	MLN	SHR	SFT	EXP	COS	CON	AV7	AV5	AV3
9	10	10	8.6	8.5	8.9	10.4	8.8	12.6	8.9	7.8	8.2	8.5
8	10	20	7.7	7.7	8.0	9.8	7.8	12.1	8.9	7.4	7.5	7.6
7	10	30	7.8	7.7	8.1	10.6	8.0	11.7	8.9	7.5	7.5	7.7
6	10	40	8.2	8.1	8.5	11.6	8.2	12.1	8.9	7.9	7.9	8.1
5	10	50	8.4	8.3	8.8	13.0	8.5	11.4	8.9	8.1	8.1	8.3
8	20	10	15.7	15.6	16.7	22.0	16.4	24.5	15.8	14.1	14.7	15.5
7	20	20	14.6	14.4	15.7	22.0	15.1	23.6	15.8	13.7	14.0	14.4
6	20	30	15.3	14.9	16.3	23.9	15.4	23.4	15.8	14.5	14.6	15.0
5	20	40	15.8	15.6	16.9	25.6	15.9	22.8	15.8	15.3	15.2	15.5
4	20	50	15.4	14.8	16.5	27.0	15.4	20.5	15.8	14.7	14.5	14.9
7	30	10	24.7	24.2	26.9	39.1	25.3	37.4	22.4	22.0	22.9	24.1
6	30	20	23.5	23.0	25.7	39.2	23.9	36.2	22.4	21.9	22.2	23.0
5	30	30	24.4	23.6	26.3	41.1	24.3	34.6	22.4	22.7	23.0	23.7
4	30	40	23.9	23.3	25.8	42.1	23.8	32.3	22.4	22.6	22.8	23.2
3	30	50	23.6	22.4	25.8	46.0	23.6	28.9	22.4	22.5	22.4	22.9
6	40	10	33.3	32.8	37.2	59.1	34.1	51.5	28.7	29.6	31.4	32.5
5	40	20	31.0	30.2	34.4	55.1	31.3	47.7	28.7	28.5	29.3	30.1
4	40	30	30.4	29.9	33.3	54.5	30.8	44.1	28.7	28.3	29.1	29.7
3	40	40	30.0	29.0	33.2	57.2	30.7	40.8	28.7	28.5	28.7	29.1
2	40	50	30.5	29.1	33.7	63.0	31.6	39.5	28.7	28.7	29.0	29.4
5	50	10	40.7	40.0	46.0	74.6	42.3	61.1	34.1	36.2	38.4	39.8
4	50	20	36.4	35.9	40.7	66.9	37.2	55.9	34.1	33.7	34.7	35.5
3	50	30	36.3	36.1	40.7	71.2	37.7	53.0	34.1	34.0	35.1	35.9
2	50	40	36.0	36.0	40.8	70.6	38.3	51.6	34.1	34.8	35.2	35.8
1	50	50	36.8	35.6	42.2	76.6	39.3	48.2	34.1	35.1	35.4	35.9

insignificant compared to the differences between the projection horizons. Extending the base period from 10 to 20 years yields some improvement in precision, with mixed results thereafter.

Table 3 is structured analogously, showing the Median APEs by projection horizon, base period length, and projection technique. In contrast to the results obtained with the MAPE, EXP and AV7 now reveal very competitive results vis-à-vis the other techniques. Once again, the length of the base period has only a minor impact on precision, though for most methods, and especially for longer horizons, some improvement is apparent when the base period is lengthened from 10 to 20 years. In accordance to the aggregate results shown in Table 1, using the median rather than the mean generally reduces forecast error by about 30–40%. However, beyond the overall level of precision, and apart from the suppression of the impact of outliers, which manifests itself most clearly in the EXP method and AV7, the two error measures lead to similar conclusions. The same is true when using the M-estimator (Table 4), which once more provides results very much in line with the median. Results for bias are largely comparable (data not shown).

Table 4 Tukey's biweight by projection horizon, base period, and projection technique

N	Horizon	Base	LIN	MLN	SHR	SFT	EXP	COS	CON	AV7	AV5	AV3
9	10	10	8.3	8.2	8.6	10.0	8.4	12.4	8.8	7.6	7.9	8.2
8	10	20	7.5	7.5	7.8	9.6	7.6	12.0	8.8	7.2	7.3	7.4
7	10	30	7.6	7.5	7.9	10.4	7.7	11.7	8.8	7.3	7.4	7.5
6	10	40	8.0	7.9	8.3	11.4	8.0	12.1	8.8	7.7	7.8	7.9
5	10	50	8.4	8.1	8.7	12.9	8.3	11.4	8.8	8.1	8.0	8.2
8	20	10	15.3	15.1	16.3	21.6	15.7	24.5	15.7	13.8	14.4	15.1
7	20	20	14.4	14.2	15.4	21.9	14.5	23.7	15.7	13.5	13.7	14.2
6	20	30	15.0	14.6	15.9	23.9	14.8	23.7	15.7	14.2	14.2	14.6
5	20	40	15.6	15.3	16.5	25.4	15.4	23.0	15.7	15.0	15.0	15.3
4	20	50	15.3	14.7	16.4	27.1	14.9	20.7	15.7	14.6	14.4	14.7
7	30	10	24.2	23.6	26.3	39.9	24.1	37.2	22.6	21.7	22.7	23.7
6	30	20	23.2	22.5	25.2	40.0	22.8	36.1	22.6	21.5	22.0	22.6
5	30	30	23.9	23.3	25.9	42.5	23.2	34.4	22.6	22.4	22.7	23.2
4	30	40	23.6	23.1	25.5	43.8	23.2	32.1	22.6	22.4	22.6	23.1
3	30	50	23.5	22.1	25.8	48.3	22.8	28.7	22.6	22.1	22.1	22.5
6	40	10	33.2	32.5	37.5	57.3	31.9	50.9	29.1	28.7	30.9	32.4
5	40	20	30.4	29.8	33.9	53.6	29.2	46.9	29.1	27.3	28.7	29.5
4	40	30	29.7	29.1	32.9	54.0	29.0	43.4	29.1	27.5	28.4	29.0
3	40	40	29.7	28.8	32.9	56.8	29.4	39.6	29.1	28.0	28.5	28.9
2	40	50	30.3	28.9	34.1	61.0	30.1	38.3	29.1	28.3	28.9	29.3
5	50	10	40.4	39.6	45.5	66.7	38.5	60.1	34.6	34.2	37.0	39.1
4	50	20	35.7	34.9	39.7	61.6	34.1	54.5	34.6	32.1	33.5	34.6
3	50	30	35.8	35.0	40.0	64.6	35.0	50.5	34.6	32.9	34.3	34.9
2	50	40	35.7	35.0	40.0	65.4	35.5	49.0	34.6	33.6	34.3	34.9
1	50	50	36.2	34.3	41.1	70.4	36.2	45.4	34.6	33.5	34.4	34.9

Following the comparison of the three error measures in the aggregate (Table 1), and by base period (Tables 2, 3, 4), the analysis continues by investigating these measures by county characteristics. The attributes most commonly examined for this purpose include population size and the rate of population growth or decline, which have been found to be important and consistent determinants of projection accuracy (Isserman, 1977; Murdock et al., 1984; Smith, 1987; Smith & Sincich, 1988; Tayman, 1996; Tayman et al., 1998; White, 1954). As a general rule, projections made for larger places tend to be more accurate than those for smaller places, and projections made for slow to moderately growing places tend to be more accurate than those for fast growing and declining places. To conserve space, only results by county growth rate are presented here. The results by county size were very similar with respect to the performance of the three summary measures of error.

As discussed above, the length of the base period had only a minor impact on forecast precision. Lengthening the base period from 10 to 20 years resulted in a modest improvement in precision, with very limited gains thereafter. The remainder of the analysis will therefore be restricted to projections using 20-year base periods. Furthermore, to facilitate exposition, only five of the ten techniques are presented (LIN, SFT, EXP, CON, and AV7) and the discussion will focus on 10- and 20-year projection horizons, which are most commonly used in practice.

The data in Fig. 1a–c show projection precision by population growth rate over the base period for the three error measures. Ten categories of population growth (and decline) are investigated: less than −15, −15 to −10, −10 to −5, −5 to 0, 0–5, 5–10, 10–15, 15–25, 25–50%, and more than 50% growth per decade over the base period. As can be seen in Fig. 1a, according to the MAPE, each of the five projection techniques exhibits a u-shaped relationship between forecast precision and rate of population change. Precision is lowest for counties that had declining or rapidly growing populations over the base period, and highest for counties with little to no population change. The five projection techniques differ with respect to whether growing or declining populations produce larger forecast errors. SFT performs worst for counties with rapidly declining populations, a category for which LIN also shows large errors. In contrast, EXP performs poorly for rapidly growing counties, but has among the lowest errors for counties that declined in population over the base period. For every projection technique, longer projection horizons are associated with larger forecast errors.

Figures 1b, c are structured analogously but measured using the Median APE and the M-estimator. Overall, the three measures are very similar, all showing the u-shaped relationship between forecast precision and population change, in addition to highlighting similar patterns at both ends of the growth spectrum. Differences exist with respect to the level of precision, which is generally lowest using the MAPE and highest using the M-estimator. Results obtained with the Median APE fall in between, though they are closer to the M-estimator than to the MAPE. In fact, for most methods, horizons, and growth categories, Fig. 1b, c are virtually identical. For the longer projection horizons the MAPE tends to diverge further from the two robust measures, but the overall relationship between forecast precision and base period population change remains essentially the same (data not shown).

Because both robust summary measures of error exhibited very similar results throughout the study, the following analysis of bias will be limited to a comparison between the MALPE and the Median ALPE, substituting an additional error measure with a somewhat different focus, percent positive, for the M-estimator. It can be argued which measure is preferable, and other robust measures such as MAPE-R could also be investigated, but for practical purposes the results are too similar to warrant a focus on the small differences where they exist.[5] The median-based measure was chosen because it is well known, easy to calculate, and easy to interpret. There might be situations where using a median-based summary measure of error are not advised, e.g., for small samples, but these do not apply here.

Figures 2a–c display data for the same five projection techniques, base periods, projection horizons, and growth categories as were shown in Fig. 1a–c, but this time focusing on bias. Figure 2a presents results obtained with the MALPE. One can see that SFT and LIN are most biased for fast declining counties, and EXP is most

[5] Tayman et al. (1999) also found results obtained with MAPE-R to be very similar to Tukey's biweight.

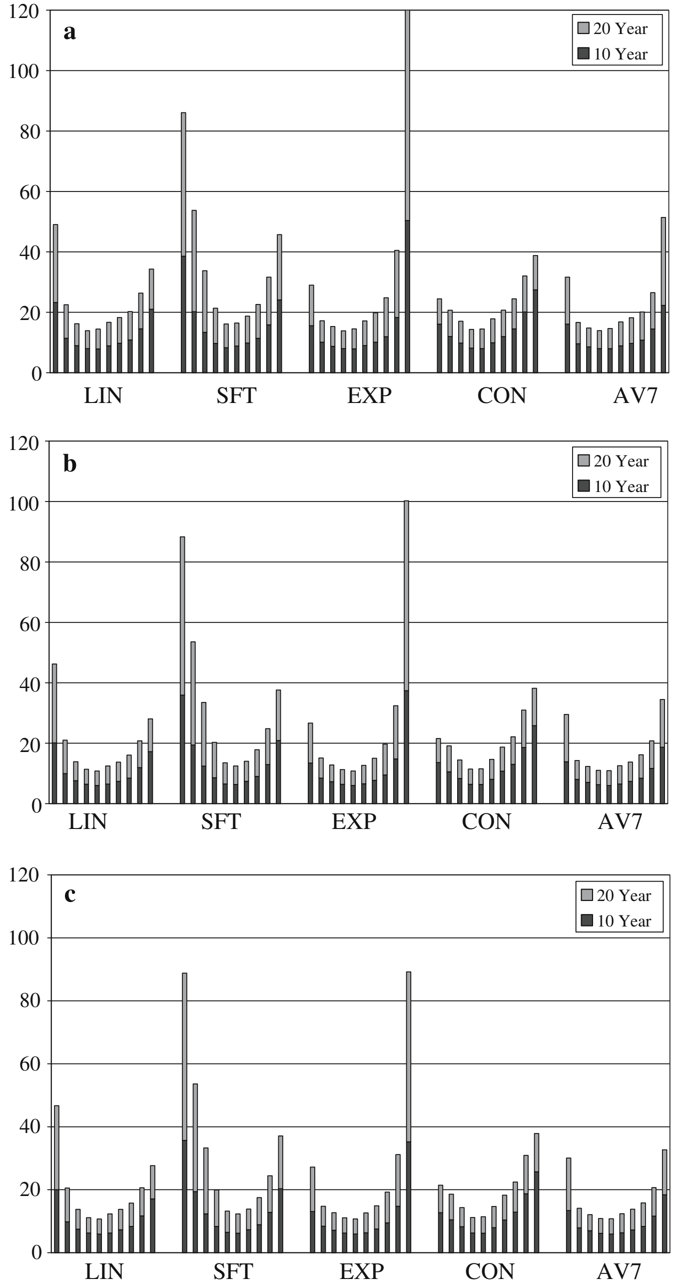

Fig. 1 **a** MAPE by population growth, 20-year base period, 10 and 20 years projection (<−15, −15 to −10, −10 to −5, −5 to 0, 0–5, 5–10, 10–15, 15–25, 25–50, >50%). **b** Median APE by population growth, 20-year base period, 10 and 20 years projection (<−15, −15 to −10, −10 to −5, −5 to 0, 0–5, 5–10, 10–15, 15–25, 25–50, >50%). **c** Tukey's Biweight by Population Growth, 20-year base period, 10 and 20 years projection (<−15, −15 to −10, −10 to −5, −5 to 0, 0–5, 5–10, 10–15, 15–25, 25–50, >50%)

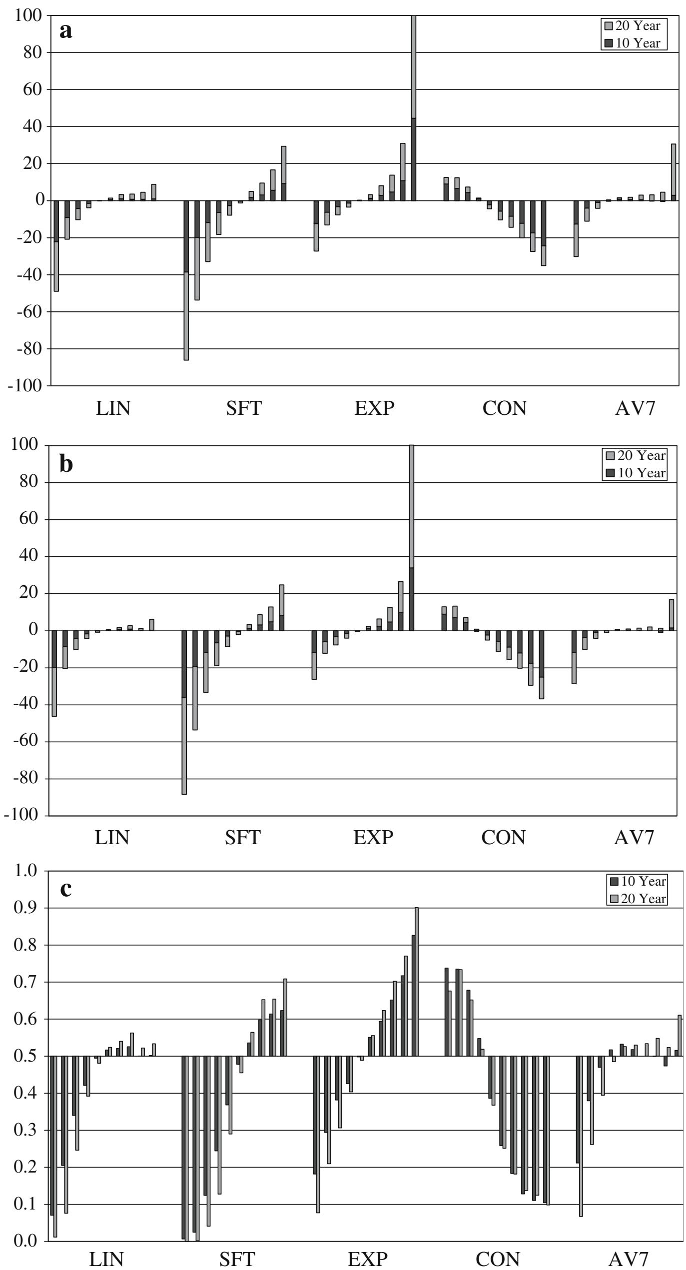

Fig. 2 **a** MALPE by population growth, 20-year base period, 10 and 20 years projection (<−15, −15 to −10, −10 to −5, −5 to 0, 0–5, 5–10, 10–15, 15–25, 25–50, >50%). **b** Median ALPE by population growth, 20-year base period, 10 and 20 years projection (<−15, −15 to −10, −10 to −5, −5 to 0, 0–5, 5–10, 10–15, 15–25, 25–50, >50%). **c** Percent positive by population growth, 20-year base period, 10 and 20 years projection (<−15, −15 to −10, −10 to −5, −5 to 0, 0–5, 5–10, 10–15, 15–25, 25–50, >50%)

biased for fast growing counties. This mirrors the results presented in Fig. 1a where the same methods also exhibited the lowest precision for counties with these characteristics. LIN, SFT, EXP, and AV7 show counties with declining populations over the base period tend to get under-projected while those with growing populations are generally projected too high for the target year. The CON technique reveals the opposite pattern. The same relationships appear when bias is measured using the Median ALPE (Fig. 2b). Any differences between the measures once again relate to the level of error and not the overall pattern, and in general both measures provide very similar results.

Whereas the MALPE and Median ALPE focus on the magnitude of bias, the percent positive measure (Fig. 2c) simply records what proportion of all projection errors are positive. Percent positive (or negative) thus highlights a different aspect of bias. The data in Fig. 2c demonstrate that the general relationship between forecast bias and base period population change remains the same for all methods. However, some projection techniques appear differently biased compared to the evaluation that used the MALPE or Median ALPE. For example, CON exhibited only modest amounts of bias according to Fig. 2a, b. Yet, as Fig. 2c shows, the vast majority of counties in growth categories as moderate as 5–10% positive change over the base period tend to get under-projected by this method. Thus, while CON avoids large errors, the method is quite biased in that it under-projects a significant proportion of all counties with particular growth regimes. This serves as a reminder that one should look at more than one summary measure of error before making a determination regarding the choice of the appropriate projection technique for the task at hand.

Discussion and conclusion

What should we conclude from the above analysis? Does the MAPE really overstate forecast error, as Tayman and Swanson (1999) and others suggest? Are the perceived problems with the MAPE grave enough to consider it invalid for evaluating population projections? And, if so, which measures of projection accuracy should be used instead?

This analysis of county population projections compared the performance of three summary measures of error with respect to precision and bias. Using a database that covered the entire twentieth century and all counties in the continental US for which comparable data were available, the study provided a comprehensive look at the performance of projection evaluation measures for a wide range of projection methods. The study confirmed that the commonly used MAPE can be problematic in that it has a tendency to go to extremes in the presence of outliers. Compared to the Median APE and an M-Estimator, the MAPEs were on average higher by about 30–40% for most methods and projection horizons. With respect to bias, the MALPE also sometimes yielded patterns that diverged from alternative measures. The differences were most pronounced for the EXP method, which can produce truly astronomical projections especially for counties with small populations and for long forecast horizons, resulting in occasionally greatly inflated mean-based summary measures of forecast error.

However, it can be argued that although the extremely elevated MAPEs sometimes limit the measure's usefulness for comparative purposes, they also expose problems with particular projection models.[6] While it is true that the EXP method generally produced quite competitive population projections—indeed for some county growth categories projections obtained with EXP were among the most precise and least biased of any projection technique—for counties with certain characteristics the method is completely unsuited, a fact that almost escapes detection if only the robust measures are used. In that sense, the MAPE (or the MALPE) does not really overstate error—after all, the error is present in the projections—but the mean-based measure does contain conceptual features that make its application for comparable purposes problematic, because a few outliers can obscure the potentially high accuracy and low bias of most other units in the analysis. Thus, when extreme outliers are present, the MAPE should not be used, or should be supplemented with other evaluative measures, unless forecasters want to perpetuate the perception that population projections are inaccurate and untrustworthy (Swanson & Tayman, 1995; Tayman & Swanson, 1996).

Previous studies of forecast accuracy at the sub-state level were limited by investigating only a restricted geographic area and/or a few decades of data. Having broad temporal and geographic scope, and using a wide range of projection techniques, the present analysis extended those earlier evaluations and, in addition, compared the performance of three different summary measures of error. This facilitated making generalizations about the findings, but because of its scope the study also involved projection scenarios that are rare in a real world setting, and which may have influenced the comparison of the performance of the error measures.

The paper concludes by evaluating projection accuracy for one set of actual 20-year projections for counties for the State of Florida. Figure 3 shows the absolute percent errors of the medium series population projections for Florida counties, produced by the Bureau of Economic and Business Research at the University of Florida in 1981 for the target year 2000.[7] The Florida projections involve a combination of cohort-component methods at the state level and ratio-share procedures at the county level (Smith & Mandell, 1981). In addition to the APEs for all 67 Florida counties, Fig. 3 also presents the MAPE, the Median APE, and Tukey's biweight as summary measures of projection precision.

As was true for the various trend extrapolation techniques for the nation described above, the Florida projections show a MAPE that exceeds both robust error measures, though the differences are less pronounced. With respect to bias the

[6] This analysis focused on trend extrapolation techniques, because the extensive temporal and geographic scope precluded the use of alternative projection models. While extreme outliers are particularly apparent with the exponential technique, cohort-component and other more complex projection methods are not immune to them. After all, if an area grew very rapidly over the base period, this would be reflected somewhere, e.g., the migration rates or structural variables used in more complex projection models. Thus, the substantive findings reported here with respect to the performance of the various summary measures of error apply to cohort-component and structural models as well.

[7] The 1981 series was produced before many of the detailed data from 1980 Census became available (e.g., the projections use 1965–1970 migration data). They were chosen here because they provide the only 20-year projection horizon that can be compared to actual census data.

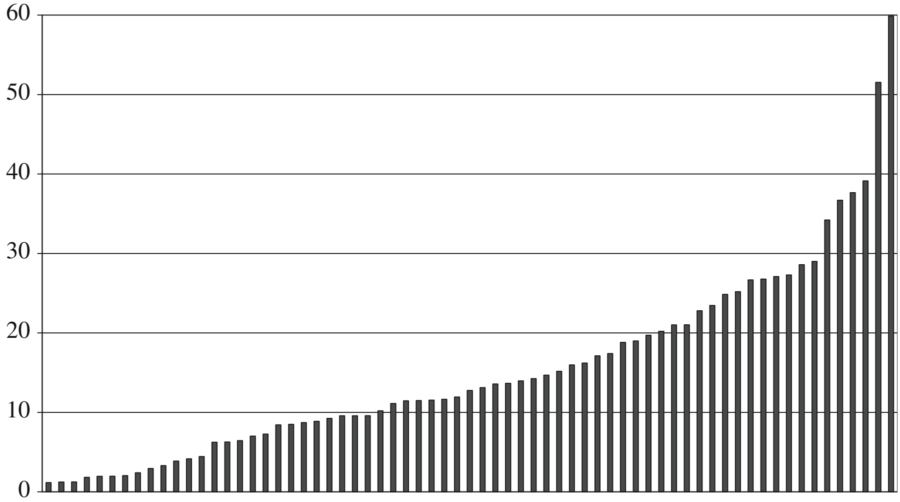

Fig. 3 APEs, Florida county projections for 2000, produced July, 1981 (MAPE = 15.47, Median APE = 12.76, Tukey's biweight = 12.52)

differences are smaller still; here a MALPE of -12.87 contrasts with a Median ALPE of -11.56 and an M-estimator of -11.53 (see Fig. 4). For both precision and bias, the two robust measures of error are very similar. This demonstrates that even a sample as small as 67 counties does not necessarily preclude the use of a median-based summary measure of error. When looking at the pattern of individual APEs shown in Fig. 3, one notices that the distribution is fairly smooth up to the six counties with the highest error values, and only two counties are real outliers. Figure 3 also shows that a relatively small number of counties (less than 10%) are mainly responsible for the MAPE exceeding the robust error measures. Even though none of the six counties that were off by more than 30% did take on extreme values, they nevertheless pulled the MAPE upwards.

Knowing the distribution of the individual county APEs and ALPEs, which of the error measures represents the ''true'' overall accuracy of the projections? This is hard to determine, because there does not exist a ''true'' summary measure of forecast error. Rather, each error measure reflects a different ''truth.'' There is no question that the counties whose projections for the year 2000 differed from Census 2000 by more than 30% are primarily responsible for the MAPE exceeding the Median APE and Tukey's biweight. Yet at the same time, the error is real, and should be reflected in the summary measure. The Median does not account for the large errors and neither, it appears, does the M-estimator.[8] In that sense, just as the MAPE can be accused of overstating forecast error, the robust alternative measures can be charged for understating it.

[8] The four M-estimators vary in their objective functions and tuning constants, resulting in different weighting patterns and, consequently, their resistance to outliers. The Huber's and Hampel's M-estimators are less resistant to outlying observations than are Tukey's and Andrews'. Using Huber's would have produced an absolute percent error value of 13.37 whereas Hampel's M-estimator shows a value of 13.29. Both are now higher than the Median APE, but still lower than the MAPE.

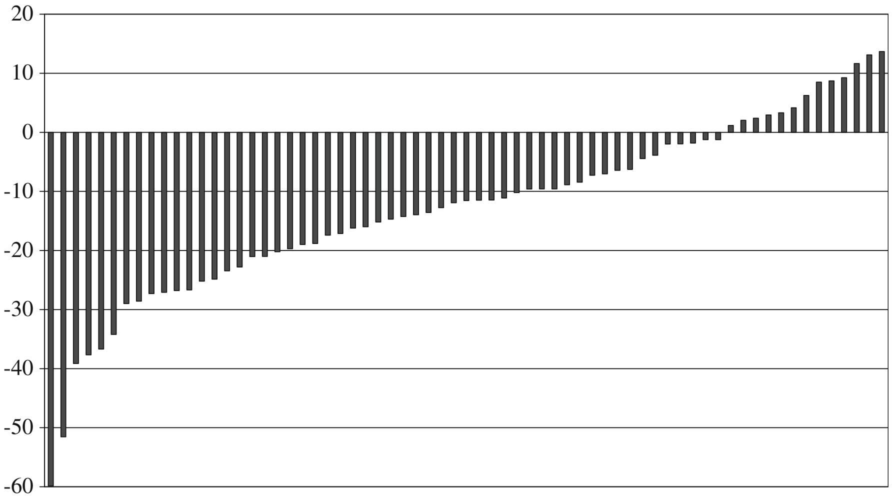

Fig. 4 ALPEs, Florida county projections for 2000, produced July, 1981 (MALPE = −12.87, Median ALPE = −11.56, Tukey's biweight = −11.53)

The findings of this study lead to the realization that it might be worthwhile to take a step back from striving to find the most refined summary measure of forecast error. There are situations in which a particular measure is clearly unsuited for the task at hand (e.g., the MAPE when dealing with extreme outliers, and the Median APE when dealing with very small samples), but in practice—for typical national, state, and county projections for short to medium horizons produced with standard techniques—most summary measures of error lead to similar conclusions. Except for situations where extreme outliers rendered the MAPE and MALPE meaningless, and which are rare in real world applications, there was not a single instance in this analysis where using an alternative summary measure of error would have led to a fundamentally different evaluation of the projections. And where there were differences, it was not at all clear that the values and patterns provided by the robust measures were necessarily more correct than those obtained with the MAPE and MALPE.

Therefore, for practical purposes, a more fruitful approach might lie in complementing an evaluation of projections with additional procedures that go beyond the various refinements of the MAPE and the MALPE. Figure 2c suggested one possible alternative. The proportion of all projection errors that are positive provides a different look at bias by discounting the magnitude of the errors and focusing instead on the overall tendency of the projections to come out too high or low. As was shown, some projection techniques were biased in a different way than was apparent when looking at either the MALPE or its robust counterparts. Another alternative is to investigate the actual distribution of the projection errors in addition to calculating summary measures or error. Looking at Fig. 3, different observers would probably come away with more than one conclusion about which of the three summary measures best represents the underlying distribution of projection errors. But whether one leans toward the MAPE, the Median APE, or another error measure, the pattern displayed in Fig. 3 provides an additional perspective that helps to interpret a set of population projections that no single summary measure can offer.

Many other procedures for determining forecast accuracy are possible. Tayman et al. (1998) looked at the variability of forecast error by means of the standard deviation of the MAPE and the MALPE, in addition to using the coefficient of variation (or CV, calculated as the ratio of the standard deviation over the mean, multiplied by 100). They found that as the size of the base population increased, so did the stability of forecast bias. A similar approach could be applied for deciding among projection methods that show otherwise comparable levels of forecast error. To give an example from the dataset used in this study, for 30-year projection horizons and 20-year base periods the following techniques are the most precise: AV7, AV5, and CON with Median APEs of about 22, and LIN, MLN, EXP, and AV3 with Median APEs of about 23–24 (see Table 3). Of these, CON exhibits the lowest coefficient of variation (9.6) and EXP the highest (15.6), with the other four techniques having CVs of about 13 (data not shown). Yet while it appears that CON might be the best performer because its projection errors are the least variable, this technique also exhibits the greatest bias (−9.8) whereas EXP has the lowest (−1.8). This further demonstrates that decisions about model choice should be made by carefully evaluating a whole range of indicators.

Finally, it needs to be reiterated that precision and bias, however measured, are not the only criteria by which forecasts should be judged. Timeliness, cost savings, ease of interpretation, ease of use, and more generally the ''utility'' of the projections are often important considerations as well (Swanson & Tayman, 1995; Tayman, 1996; Tayman & Swanson, 1996; Yokum & Armstrong, 1995).

Ahlburg (1992, p. 99), in a commentary on error measures, asked whether the choice of an error measure to identify the most accurate forecasting method was a question of personal taste. He suggested that this appeared to be the case and pointed out that in a survey of various papers dealing with population forecasts none of the authors justified their choice of error measure. This paper does not advocate that it does not matter which summary measure of forecast accuracy is chosen when evaluating a set population projections. There often exist valid theoretical and/or practical concerns against a particular measure and, when the circumstances warrant it, error measures that are problematic should not be used. However, this study found that for most practical purposes simple and easy to understand measures such as the MAPE and the Median APE offer all the relevant information a summary measure of error can provide, and that additional evaluation procedures that go beyond a single criterion should also be considered. Doing so will lead to a more thorough understanding of the underlying structure of the projections and, it is to be hoped, to better projections in the future.

References

Ahlburg, D. A. (1992). Error measures and the choice of a forecast method. *International Journal of Forecasting, 8*, 99–100.

Ahlburg, D. A. (1995). Simple versus complex models: Evaluation, accuracy, and combining. *Mathematical Population Studies, 5*, 281–290.

Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting, 8*, 69–80.

Bonynge, F. (1852). *The future wealth of America*. New York: Author.

Campbell, P. R. (2002). Evaluating forecast error in state population projections using census 2000 counts. Population Division Working Paper Series No. 57. Washington, DC: US Census Bureau.

Coleman, C. D., & Swanson, D. A. (2004). On MAPE-R as a measure of estimation and forecast accuracy. Annual meeting of the Southern Demographic Association, Hilton Head Island, South Carolina.

Forstall, R. L. (1996). *Population of states and counties of the United States: 1790 to 1990*. Washington, DC: US Census Bureau.

Goodall, C. (1983). M-estimators of location: An outline of the theory. In D. C. Hoaglin, F. Mosteller, & J. Tukey (Eds.), *Understanding robust and exploratory data analysis* (pp. 339–403). New York: Wiley.

Hoaglin, D., Mosteller, F., & Tukey, J. W. (Eds.). (1983). *Understanding robust and exploratory data analysis*. New York: Wiley.

Isserman, A. (1977). The accuracy of population projections for subcounty areas. *Journal of the American Institute of Planners, 43*, 247–259.

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

National Research Council. (1980). *Estimating population and income for small places*. Washington, DC: National Academies Press.

Pritchett, H. S. (1891). A formula for predicting the population of the United States. *Quarterly Publication of the American Statistical Association, 2*, 278–286.

Rayer, S. (2004). Assessing the accuracy of trend extrapolation methods for population projections: The long view. Annual meeting of the Southern Demographic Association, Hilton Head Island, South Carolina.

Rosenberger, J. L., & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In D. C. Hoaglin, F. Mosteller, & J. Tukey (Eds.), *Understanding robust and exploratory data analysis* (pp. 297–338). New York: Wiley.

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., & Mandell, M. (1981). 1982–2020 projections of Florida population by county. Bulletin No. 56, Population Studies, Bureau of Economic and Business Research. Gainesville FL: University of Florida.

Smith, S. K., & Shahidullah, M. (1995). An evaluation of population projection errors for census tracts. *Journal of the American Statistical Association, 90*, 64–71.

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

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

Smith, S. K., & Sincich, T. (1992). Evaluation the forecast accuracy and bias of alternative population projections for states. *International Journal of Forecasting, 8*, 495–508.

Smith, S. K., Tayman, J., & Swanson, D. A. (2001). *State and local population projections: Methodology and analysis*. New York: Kluwer Academic/Plenum Publishers.

Swanson, D. A., & Tayman, J. (1995). Between a rock and a hard place: The evaluation of demographic forecasts. *Population Research and Policy Review, 14*, 233–249.

Swanson, D. A., Tayman, J., & Barr, C. F. (2000). A note on the measurement of accuracy for subnational demographic estimates. *Demography, 37*, 193–201.

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., & Swanson, D. A. (1996). On the utility of population forecasts. *Demography, 33*, 523–528.

Tayman, J., & Swanson, D. A. (1999). On the validity of MAPE as a measure of population forecast accuracy. *Population Research and Policy Review, 18*, 299–322.

Tayman, J., Schafer, E., & Carter, L. (1998). The role of population size in the determination and prediction of population forecast errors: An evaluation using confidence intervals for subcounty areas. *Population Research and Policy Review, 17*, 1–20.

Tayman, J., Swanson, D. A., & Barr, C. F. (1999). In search of the ideal measure of accuracy for subnational demographic forecasts. *Population Research and Policy Review, 18*, 387–409.

Whelpton, P. K. (1928). Population of the United States, 1925 to 1975. *American Journal of Sociology, 34*, 253–270.

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

Yokum, J. T., & Armstrong, J. S. (1995). Beyond accuracy: Comparison of criteria used to select forecasting methods. *International Journal of Forecasting, 11*, 591–597.