

UC Riverside

UC Riverside Previously Published Works

Title

Between a Rock and a Hard Place: The Evaluation of Demographic Forecasts

Permalink

<https://escholarship.org/uc/item/1tw6g8pv>

Journal

Population Research and Policy Review, 14

Authors

Swanson, David A
Tayman, Jeff

Publication Date

1995

Peer reviewed

Between a rock and a hard place: The evaluation of demographic forecasts

DAVID A. SWANSON¹ & JEFF TAYMAN²

¹*Arkansas Institute for Economic Advancement, University of Arkansas at Little Rock, and NIMH Center For Rural Mental Healthcare Research, University of Arkansas For Medical Sciences, Little Rock, Arkansas, USA;* ²*San Diego Association of Governments, San Diego, California, USA*

Abstract. Forecasting, in general, has been described as an unavoidable yet impossible task. This irony, which comprises the 'rock' and the 'hard place' in the title, creates a high level of cognitive dissonance, which, in turn, generates stress for those both making and using forecasts that have non-trivial impacts. Why? Because the forecasted numbers that are invariably accorded a high degree of precision inexorably reveal their inevitable imprecision when the numbers forming the actuality finally take place and the numbers comprising the forecast's errors are precisely measured. The current state of the art in demography for dealing with the 'rock' and the 'hard place' is a less-than-successful strategy because it is based on an acceptance of accuracy as the primary evaluation criterion, which is the source of cognitive dissonance. One way to reduce cognitive dissonance is to change the relationship of the very cognitive elements creating it. We argue that forecast evaluations currently focused on accuracy and based on measures like RMSE and MAPE be refocused to include utility and propose for this purpose the 'Proportionate Reduction in Error' (PRE) measure. We illustrate our proposal with examples and discuss its advantages. We conclude that including PRE as an evaluation criterion can reduce stress by reducing cognitive dissonance without, at the same time, either trivializing the evaluation process or substantively altering how forecasts are done and presented.

Key words: Cognitive dissonance, Proportionate-reduction-in-error, Utility

The rock and the hard place: Forecasting's irony

The highly-respected demographer, Nathan Keyfitz, frames well the two-part irony confronting all of us who want to provide a description of the future. The first part (the rock) is that 'Numbers provide the rhetoric of our age . . . but to forecast in the sense of making an estimate that will turn out to coincide with what is actually going to happen is beyond human capacity' (Keyfitz 1987: 235). The second part of this irony (the hard place) is that 'Forecasting (is) impossible yet unavoidable' (Keyfitz 1987: 236).

Forecasts must be done in the modern world and, moreover, must be in the form of numbers. Yet, the forecasted numbers invariably turn out to be different than the numbers that actually happen. Perhaps the real irony is found in the fact that the social 'authority' imparted to these forecasts by users is itself constantly threatened by the presence of the very numbers that give it this authority. Why? Because the forecasted 'numbers' that are

invariably accorded a high degree of precision inexorably reveal their inevitable imprecision when the 'numbers' forming the actuality finally take shape and the 'numbers' comprising the forecast's errors are precisely measured.

Of course, forecasters know that their forecasts will have errors and that there are different levels of accuracy expectations for different types of forecasts. The expectations among demographers depend mostly on the size of the population and the length of the forecast horizon (Murdock et al. 1984; Murdock et al. 1989; Smith 1987; Smith & Sincich 1990, 1992; Stoto 1983; Swanson & Beck 1994; Swanson & Tayman 1994; Tayman 1996). To a large degree this holds true for informed users of population forecasts as well. Nonetheless, the vast majority of the forecast evaluation literature is focused on accuracy as the major criterion and we argue that this has contributed to what is, in fact, the normative expectation for judging the adequacy of a given forecast – its accuracy (Starr 1987).

Unfortunately, this expectation has also served to support the demand on the part of many users that forecasts meet standards of accuracy that exceed those commonly accepted as reasonable by experienced forecasters. As an analogue, consider what has happened in the USA in regard to the decennial census between 1970 and 1990. The evaluations of these census counts focused on net undercount errors by geographic and socio-demographic strata (Robinson et al. 1991; US Bureau of the Census 1973, 1982). Given that these census counts have been very accurate and, moreover, are generally believed to have improved in accuracy between 1970 and 1990, it is clear that what are perceived to be 'acceptable' levels of error among professional demographers are not 'acceptable' for many 'stakeholders'. To a large degree this highly-charged situation is also the result of the attention given to accuracy as a major 'evaluation criterion' – the US Bureau of the Census's diligent research reports on census errors largely served to focus public attention on error levels at the expense of other important evaluation criteria. The current level of contentiousness and dispute affecting the 'accuracy' of the decennial census represents a situation that we believe is important for forecasting to avoid. We argue, therefore, that the use of a criterion like 'utility' should be included in forecast evaluations along with accuracy criteria. In this paper we attempt to show that adding this criterion will change the normative expectations regarding what forecasts should achieve in a manner that enhances forecasting, reduces stress levels on forecasters, and avoids trivializing the evaluation process.

Forecasting and job-related stress

For demographers, the full import of forecasting's two-part irony is not always immediately grasped, partly because errors can be measured only a few years after a census or a current estimate is available and partly because the membership in the 'significant user' audience changes frequently. Yet, even without grasping the full import, most demographers who forecast

quickly experience the stress that stems from it when their forecasts are used in making significant decisions. This stress is not something to be taken lightly. The potential adverse consequences of 'wrong' forecasts have been well documented. They include losing one's professional credibility (Dorn 1950), being sued (D'Allesandro 1987), and being fired.

Harold Dorn (1950: 314–315), gives an idea of the professional credibility crisis faced by demographic forecasters in the late 1940s when it became apparent that they had missed the Baby Boom, by quoting the remarks of one disgruntled user: 'I am ashamed that, like most of my fellow social scientists, I have so long accepted the conclusions of the population specialists with naive faith'

An 'op-ed' piece by D'Allesandro (1987) nearly thirty years later in *Applied demography* posed the question 'Should applied demographers take out liability insurance?' and advised demographers to take extra steps to protect themselves from claims of negligence.

Stress on the producers of forecasted information also stems from a lack of understanding of the limitations and difficulties of quantifying future events by the consumers of forecasts. Forecast consumers often place demands for more detailed information that cannot be produced within any realistic and reasonable degree of precision. However, such forecasts are routinely prepared in order to: (1) to maintain a competitive advantage; (2) provide evidence that their organization is on the cutting edge of technology; (3) to maintain credibility and remain responsive to users; and (4) to satisfy requirements of many federal, state and local programs (e.g., ISTEA and Clean Air acts). It is ironic, however, that the 'advantages' for doing these forecasts rest on a very shaky house of cards built on unrealistic expectations of precision.

Stress is not only limited to those demographers who produce forecasts. It also occurs among the users of these numbers and may be greater than the stress on producers. Producers are held directly accountable for their numbers, but users are directly held accountable for decisions based on these numbers. Forecast users generally have direct 'on the line' accountability for decision-making while producers tend to have indirect and more diffused accountability. Bad decisions concerning, for example, whether to build or not build schools, roads, shopping centers or residential communities usually have a quicker effect on a decision-maker's career longevity than do poor descriptions of the future provided by a forecaster. However, note that forecasting, as part of 'Applied Demography', may be on the road toward becoming a decision-making science in its own right (Burch, Swanson & Tedrow 1996). If this comes to fruition then forecasters will themselves have more direct accountability.

Reducing job-related stress: The current state of the art

Demographers have developed a number of strategies for dealing with the

'irony' of forecasting. To a large degree, these same strategies can be viewed as devices designed to reduce the stress inherent in carrying out an 'impossible' task. One of these 'stress-reduction' strategies is to use the term 'projection' in place of 'forecast'. Keyfitz (1972) and Pittenger (1978) point out that demographers prefer to make 'projections' while the user audience wants 'forecasts'. Smith & Bayya (1992: 4) note, in this regard: 'Historically, demographers have preferred to make population projections rather than forecasts. This reluctance to predict is not surprising, given the degree to which many past forecasts have been wide of the mark.' However, this stress-reducing strategy usually falls far short of its mark as pointed out by Smith & Bayya (1992: 4): '. . . users will generally interpret projections as forecasts, regardless of the author's intentions and whatever terminology or disclaimers might be used. A basic fact of life for demographers is that as soon as their projections reach the public, they become forecasts.'

Another stress-reducing strategy has been to present forecasts along with some idea of the uncertainty inherent in them. There are two ways in which this is done, informally and formally. The informal approach has been used for quite a number of years. It is usually in the form of high, medium, and low projections (US Bureau of the Census 1992). The formal approach is usually done in the form of confidence intervals (Tayman et al. 1994; Espenshade & Grummer-Strawn 1991; Smith & Sincich 1988; Stoto 1983; Keyfitz 1981; Sykes 1969).

'Normative' forecasting, although infrequent in demographic applications, is another approach to reducing stress among the producers of forecasts (Moen 1984). This active approach to forecasting involves first deciding what future outcomes are desirable and then designing policies and actions to achieve them. Along similar lines, Romaniuc (1994) argues that it is less important that forecasts are right or wrong than whether and how they are used in decision-making – forecasting as a proactive tool that becomes an instrument for *creating* rather than simply *discovering* the future. In both interpretations, Moen's and Romaniuc's, the 'proactive' approach calls for the recognition of human volition. As such, it can be viewed as a stress-reduction strategy in that forecasting is explicitly acknowledged as a tool for something other than 'discovering' the future in advance.

Involving and educating major constituents and users throughout the forecasting process is another technique used by forecasters to help achieve credibility and acceptance of a forecast and, hence, reduce stress. If users think that a forecast comes from a methodological 'black box' and do not understand the nature and role of assumptions, then the forecast may be criticized, regardless of its empirical accuracy (Rainford & Masser 1987). Ascher (1978: 19) and Pittenger (1978) both argue that the acceptance and accuracy of forecast assumptions is more important than the form and complexity of the mathematical model used to generate the forecast. Although, as pointed out by Keyfitz (1982), the ability to present projections as conditional and resting on specific assumptions, is often blocked from the start by a lack of adequate data.

The major device that users adopt to reduce stress is to distance themselves from the forecast numbers and point to the producer if things go badly. Firms hire consultants not only for their expertise, but also to have someone at which to point the finger if necessary (Armstrong 1982–83). The relationship between users and producers of demographic forecasts is similar. This is not necessarily a bad thing, but blind and uncritical acceptance of demographic forecasts can lead to situations of considerable stress for both the users and producers.

Do the strategies just outlined effectively deal with stress? Our answer is 'No'. Users of forecasts most often expect an unencumbered, 'single' description of the future in spite of efforts by producers to qualify forecasts as 'multiple' descriptions or something other than information providing any such description. Most users have neither the time nor the desire to become specialists in forecasting's technical aspects, which is what the current set of stress-reduction strategies requires. In general, users are 'bottom line' oriented and require readily-grasped information for decision-making. That is why, for example, when confidence intervals or ranges are presented most users pay little attention to the magnitude of the intervals and use the 'point estimate' as the forecast. Information demonstrating consistency between past and current forecast errors is also of limited use in reducing stress. First, this information is neither readily available or understandable to most forecast users, especially for subcounty geographic areas. Second, if the errors are viewed as 'too large' – as would often be the case – it does not matter if they are consistent over time. In addition, the strategies do not work because they do not effectively deal with the fact that stress is a direct result of cognitive dissonance (Festinger 1957) and the reduction of stress can only be accomplished by reducing the level of cognitive dissonance. Before offering our proposal for reducing stress, we first turn to a discussion of cognitive dissonance and present the background information that forms the basis for our proposal.

When prophecy fails: Cognitive dissonance

It is of more than passing interest that Leon Festinger was led to the development of the theory of cognitive dissonance by research he had done on the failures of forecasting. Here, however, it was not technical forecasting such as that found in demography and economics that he studied. Rather, it was the effects of inaccurate doomsday forecasting made by certain cults on the behavior, values, beliefs, and attitudes of cult members and documented in the book, *When prophecy fails* (Festinger, Riecken & Schachter 1956).

What is cognitive dissonance? A full answer to this question is beyond the scope of this paper and, in any event, is found, for the most part, in Festinger (1957). Basically, cognitive dissonance exists when there are two relevant cognitive elements and the obverse of one follows from the other (Festinger 1957: 13). One of many examples Festinger (1957: 13) gives of this is in the

form of describing a debt-bound person who buys a new car. Here, the two elements are relevant and the obverse of one follows from the other: owing money and purchasing a new car. That is, purchasing a new car would not be expected to follow from a condition of indebtedness. Had the person in the example not purchased a new car, Festinger would describe the two relevant elements as being consonant. That is, one would expect a debt-ridden person not to purchase a new car.

For purposes of this paper, it is easy to apply this concept to the task of forecasting: It is unavoidable, yet impossible, and its impossibility is measured by comparing the forecast with the actuality when it finally arrives. Forecasting and inaccuracy are, thus, two relevant cognitive elements, one of which is the obverse of the other. That is, if we forecast and others use them, the forecast is expected to be within acceptable and reasonable error limits. Unfortunately, the normative expectations of forecasters do not match those of the users of forecast information: The users expect greater accuracy than the forecasters. But note here also that if the magnitude of error in many forecasts was consistently low, then accuracy would not have assumed such a prominent position in evaluations. Unfortunately, the magnitude of error is not consistently low (Keyfitz 1981; Smith & Shahidullah 1995; Teyman, Schafer & Carter 1994), and, moreover, generally exceeds the normative expectations of many users. This, in part, has caused accuracy to become such a dominant issue in evaluations and the normative expectation by which the adequacy of a forecast is judged. This dominance and expectation has contributed to job-related stress as well as credibility and other problems that can – and should – be minimized.

Also, pertinent to this paper, Festinger (1957: 261) argues that 'dissonance almost always exists after a decision has been made between two or more alternatives'. Further, the magnitude of the dissonance is a function of the importance of the elements (Festinger 1957: 16) and its presence gives rise to pressures to reduce or eliminate it, where the strength of the pressures to reduce it are a direct function of its magnitude (Festinger 1957: 18). Again, we can see the immediate application to forecasting. There is always a choice between alternatives when considering the future and, further, when the forecasts are used for non-trivial purposes (such as developing operating budgets committing funds to capital facilities, etc.), it is clear that the two elements (the forecast and its accuracy) have great importance. Thus, in these cases, the magnitude of the dissonance is high, as are the pressures to reduce it.

One of the consequences of cognitive dissonance that is apparent to forecasters is the cynical view of a given forecast manifested by users. Forecasters often hear comments such as: 'these forecasts are useless and cannot possibly be true'; 'I could have done better forecasts on the back of an envelope'; or 'if you torture a number long enough, it will tell you anything'. In some instances, these comments are warranted and justifiable. However, many times these comments are made because the forecast does not depict a

particular view of the future desired by a user. One of the authors often has to present demographic forecasts of municipalities to their locally elected officials. When forecasts are brought down to this level there are many competing groups that have a vested interest in a particular forecast outcome. This can become a very hostile environment in which to conduct an honest and straight-forward discussion of the forecast and its implications. Slight degrees of error in a forecast are often used as a reason to invalidate them. For example, a housing forecast made in 1985 for one municipality was high by 50 units from the 1990 census count of 2,750 units, an error of 1.8 percent. The elected officials used this fact to argue that their 2010 forecast of 6,500 units was too high by 1,000 units. Their rationale was that the error of 50 units would occur every year for 20 years. Given all of this, what can be done to reduce cognitive dissonance? For our answer, we turn to one of the ways that Festinger says dissonance can be reduced.

Where is there breathing room?

The thesis inherent in our proposal follows naturally from Festinger's observation (1957: 21) that one way to reduce the total magnitude of cognitive dissonance is add new cognitive elements to produce a new cognitive consonance for the relationship of the elements.

We have argued that the high magnitude of dissonance (and hence stress) in forecasting stems directly from the fact that forecasts are inherently 'inaccurate'. That is, we have two relevant and important cognitive elements, one of which is the obverse of the other. We propose to change this relationship by using the concept of utility as a basis for forecast evaluation. We argue that this offers a way of presenting evaluations of demographic forecasts in a way that does not diminish the importance and use of the forecasts themselves.

The main thrust of this paper, then, is to examine a way of presenting demographic forecasts that still makes them useful but reduces the magnitude of cognitive dissonance associated with them and, consequently, the level of stress on both those who produce them and those who use them. Our argument in this paper is that by adding utility to the list of forecast evaluation criteria this can be accomplished. This argument stems from Tayman (1993) and Swanson & Beck (1994).

Accuracy versus utility

Our comments here have bearing, although with slightly different emphasis, on demographic forecasters in both the profit and nonprofit sectors. Not all of these forecasters have the time and resources required to justify the usefulness of their product. Consequently, many rely on the academic sector

for studies concerning usefulness. For the most part, these studies focus on a very specific definition of accuracy (i.e., an *ex post facto* comparison between the forecasted number and a census count) in evaluating various combinations of methods and data and do not address the larger issue of utility (Swanson 1986). This has created an atmosphere in which accuracy tended to become the dominant issue and the normative expectation for judging the adequacy of a demographic forecast (Starr 1987), although this has been tempered in different ways for the nonprofit and profit sectors.

By focusing on 'ex post facto' accuracy, demographic forecasters – and, perhaps more importantly, their audiences – have tended to overlook other dimensions. This has been noticed by others (Brettschneider & Gorr 1992; Makridakis & Hibon 1979). At least two other dimensions that are worthy of consideration are timeliness and cost (Swanson 1986). Please note here that we use the term 'tended to have been overlooked'. This means that timeliness and cost as well as others have not been ignored, only that they have not been placed in a 'context' in which a better understanding of the utility of population forecasting can be gained.

'Utility' is not a new concept to the field of population forecasting, as the following examples show. Nearly forty years ago, Hajnal (1955) presented a case for 'utility' when he argued that the value of forecasting is in the analytical insights it may provide rather than its numerical accuracy. Musham (1965) argued that cost functions should be used in conjunction with alternative estimates of future population. More recently, Kintner & Swanson (1994), Murdock et al. (1984), Murdock & Leistritz (1980), and Swanson, Tayman & Beck (1995) have suggested that the 'utility' of population forecasts not be judged solely on the basis of their accuracy.

Another important dimension of 'utility' relates to the additional information furnished by a forecast. Does the forecast 'add value' – provide information that helps make better decisions? This criterion may be the most important factor in deciding whether a forecast is good or not. Wheelwright & Makridakis (1980: 331–332) suggest that one aspect of a successful forecasting application involves forecasters who view the forecast as a way to improve decision making. Small-area forecasts, even with an average errors of 20% or more, were an integral part of many important and correct decisions regarding infrastructure expenditures involving millions of dollars (Tayman 1993).

Measuring utility: PRE

A statistical concept known as proportionate reduction of error (PRE) (Costner 1965; Agresti 1990: 24) offers a way to measure the gain of information from a forecast. As described by Reynolds (1977: 32), 'PRE measures rest on a simple concept of association. Imagine a game in which one randomly draws people from a population and guesses their scores on Y, the

dependent variable. The predictions can be made in either of two ways: first, knowing nothing at all about the individuals or, second, knowing their scores on another, independent variable, X. Whatever rule is followed one will surely guess wrong at least some of the time. But if Y depends on X, then knowledge of X categories should reduce the error.'

In many ways, Costner's (1965: 344) original conceptualization of PRE, although placed in the context of measures of association, remains the most general in its application and parsimonious in its presentation:

$$\text{PRE} = \frac{\text{Error by rule(b)} - \text{Error by rule (a)}}{\text{Error by rule(b)}}$$

Thus, estimating or predicting some value (such as population), is done by two methods, 'rule(a)' and 'rule(b)'. The error arising from each of the two methods is defined and measured and the proportionate reduction in error found by using rule (a) as opposed to rule(b) is determined by placing both error measures in the preceding formula. Since Costner examined PRE in the context of measures of association (e.g., r^2 , Spearman's rho, Goodman & Kruskal's gamma, Kendall's Tau), some boundaries were implicit in his formulation. For our purposes, the most important of these boundaries was that Error by rule(a) \leq Error by rule(b), which results in the following theoretical and practical limits on Costner's formulation of PRE:

$$0.0 \leq \text{PRE} \leq 1.0.$$

We remove Costner's restriction that Error by rule(a) \leq Error by rule(b) for the purpose of measuring the utility of population forecasts and allow Error by rule(a) to be less than, equal to, or greater than Error by rule(b). We also multiply PRE by 100 so that it has a 'percentage' interpretation. Thus, in our approach, PRE has the following theoretical limits, ($-\infty$) \leq PRE \leq (100). The practical lower limit of PRE in our formulation is, of course, much more restricted than the theoretical lower limit of negative infinity, as will be shortly apparent.

What constitutes rule(a) and rule(b) in our approach? Rule(a) is the prediction of population resulting from some projection technique such as the cohort-component method while rule(b) is the prediction of population resulting from data already at hand through an existing 'count', such as the last census. We include under the rubric 'projection technique' the judgment and data required in its operation.

Thus, we transfer the PRE concept as formulated by Costner (1965) and his successors (Agresti 1990; Reynolds 1977) to the evaluation of forecasts by establishing a forecast error based on existing information, most likely a prior census. This 'naive' forecast - rule(b) - represents the theoretically (and most often, the practical) maximum error for the prediction based on no new knowledge. By using the PRE formula we evaluate the (presumed) reduction of error found by using the 'actual' projection method - rule(a) -

Table 1. The utility of projections as measured by PRE for two states with different growth patterns for 1980 to 1990: Very low (Ohio) and moderately high (Washington)

	Ohio State	Washington State
No. of counties	88	39
Average population per county	123,263	124,787
Annual growth rate, 1980–1990*	0.05%	1.7%
1990 RMSE* using:		
1980 Census (RMSEc)*	15,105.0	48,050.7
Projection (RMSEPr)*	6,163.8	16,977.3
PRE*	59.2%	64.7%

* See text for definition.

over the error in the ‘naive’ forecast – rule(b). Thus, our PRE shows the reduction in error or gain in ‘knowledge’ due to the particular method (and its judgments and input data) under evaluation. Forecast evaluations using this formulation of PRE have been conducted using MAPE and RMSE and other measures of forecast precision (Swanson & Beck 1994; Tayman 1993).

Example application of the PRE method of forecast evaluation

We selected two states for our example application, Ohio and Washington. They represent very different patterns of population change between 1980 and 1990. Using a standard demographic model for measuring the annual rate of population change between two points in time (which in this case is $r = [\ln(P_{1990})/P_{1980}/10]$, we find that Washington had a moderate of growth, about 1.7 percent annually while Ohio experienced virtually no change. Its annual growth rate was 0.05 percent. Ohio has over twice as many counties as Washington, 88 compared to 39, and over twice the population, 10,847,115 compared to 4,866,692, as measured in the 1990 census.

Table 1 contains data results for the two states comprising our illustrative application of PRE as way to place the evaluation of forecasting on a ‘utility’ basis and, thus, reduce cognitive dissonance. Before turning to a discussion of Table 1, we first describe the way in which we specifically operationalized our formulation of PRE using Root Mean Square Error.

$$PRE = \{[(1990 \text{ RMSEc}) - (1990 \text{ RMSEPr})] / (1990 \text{ RMSEc})\} * 100$$

where RMSEc (Root Mean Square Error resulting from using the 1980 census to predict the 1990 population by county)

$$= \sqrt{[\sum(P_{i80} - P_{i90})^2 / N]}$$

where P_{i80} = 1980 census population of county i ; P_{i90} = 1990 census population of county i ; N = Number of counties and RMSEPr (Root Mean Square

Error resulting from using the Cohort-Component Method to predict the 1990 population by county)

$$= \sqrt{[\sum(\text{Pr}_{i90} - P_{i90})^2/N]}$$

where Pr_{i90} = Projected 1990 population of county i ; P_{i90} = 1990 census population of county i ; N = Number of counties.

Turning to Table 1, we see RMSEc values for each state as found for ex post facto predictions of their respective 1990 county populations from the county specific populations measured in the 1980 census. The errors summarized in each state's RMSEc are measured against the reported 1990 census counts for each county. For Ohio, its RMSEc score indicates that, on average, the 1980 census counts led to a 1990 county population prediction that is off by 15,105, while for Washington, it is much higher at 48,050.7. The relative values of the RMSEc scores reflect, in part, the differences in population size and geographic concentration, the number of counties, and the level of population growth during the 1980s. It is not surprising that Washington, the state experiencing a faster rate of growth, has by far the largest RMSEc. The RMSEc score for Ohio is actually very respectable in terms of prediction accuracy, particularly given the size of its overall 1990 population and the average number of persons per county in 1990 (123,263). The RMSEc score for Washington would generally considered to be very poor, given its overall 1990 population and an average of 124,787 persons per county.

What improvement in accuracy, if any, occurs in each state by using county populations forecasted by a 'projection method' for 1990 instead of the 1980 census populations? In answering this question, we selected Cohort-component method projections done by each state's 'official' demographic center. The Cohort-component method is the technique most widely employed by demographers (Murdock et al. 1989). Thus, the sets used here were done according to standard practice. They were also done with the benefit of 1980 census data under the guidance of trained demographers, well-experienced in the Cohort-component method and knowledgeable about demographic trends in their respective states.

For Ohio, the 1990 county forecasts are taken from a report published by Ohio Data Users Center (1985). The RMSEPr for Ohio is 6,163.8 and the PRE is 59.2%. Thus, the use of the projections to predict 1990 county populations instead of the 1980 census counts reduces Root Mean Square Error by 59.2%, going from 15,150 as measured by RMSEc to 6,163.8, as measured by RMSEPr.

For Washington, the 1990 county forecasts are taken from a report published by the state's Office of Financial Management (1986). Here, we see in Table 1 that RMSEPr for Washington is 16,997.3 and PRE is 64.7%. Thus, the projections reduced Root Mean Square Error by 64.7% over that found by using the 1980 census counts.

In both cases, the 'utility' of using forecasts over the 1980 census to

predict 1990 county populations is supported in that there is an approximate reduction in error of 60 percent.¹

Discussion

In using the illustrative example just presented, we are not arguing that the cohort-component method should always be preferred over alternatives. We selected Ohio and Washington as the example states because we were familiar with the demographic centers in both states and considered them to be representative of the many states with strong demographic centers. Both demographic centers like their sister agencies elsewhere, typically use the cohort-component method to produce 'official' forecasts and consider the method to be the one of choice. Clearly, there are alternative approaches to demographic forecasting and these could have just as easily been used in examples involving other states or substate areas. Similarly, as stated earlier, we are not arguing that RMSE is the only measure of forecast accuracy worth using. There exist alternatives and they could have just as easily been used.

What we demonstrate in the example is that 'accuracy' as it is typically defined in demographic forecasting evaluations (i.e., *ex post facto* comparison of a forecast with a census number) is limited and that it is possible to present a readily-grasped and meaningful evaluation criteria that goes beyond the concept of accuracy and gets directly at the broader concept of utility.

We also argue that PRE is useful because it is compatible with two other readily-grasped evaluation criteria, cost and timeliness. In terms of 'cost', the inherent utility of the PRE measure could be extended by linking it with the price of information. In this context, 'cost' would represent different things to different players. In regard to the producers of forecasts, cost would probably be defined in terms of the labor and other resources required to generate and distribute forecasts to users. For users, cost would probably involve the price of less accurate information in making some decision.

As an example of the idea of 'producer' cost, we asked the principals involved in producing the Ohio and Washington forecasts to provide us with an estimate of the professional person-hours it took to generate the forecasts. We requested that they include only the time spent on assembling and entering data required to run the projections and running scenarios and making decisions about the final assumptions. Recall that Ohio has over twice the counties and population of Washington but otherwise many of the procedures and activities of the two demographic centers would be very similar. For Ohio, we received an estimate of 2,000 person-hours; for Washington, an estimate of 1,000 person-hours. It makes intuitive sense that Ohio's is twice that of Washington's because a great deal of time is spent on each individual county and Ohio has just over twice as many as Washington.

By taking the ratio of cost/PRE for each state, we find that, on average,

Ohio required about 34 professional person-hours for each percent reduction in error while Washington required 16. It is this type of analysis that we believe can be added to evaluations of forecast utility. This type of measurement could, in turn, be used to evaluate the demographic centers themselves. Ohio, for example, supports a forecasting activity that has demonstrated a substantial reduction in error (59%) in its 1990 forecasts over using a 'free' resource (the 1980 census). The cost for each percent reduction was about 34 professional person-hours for a state with 88 counties, an average population of 123,263 per county, and a 1980-90 average annual growth rate of 0.05 percent. Washington also supports a forecasting activity that demonstrated a substantial reduction in error (65%) in its 1990 forecasts over using the 1980 census as a free resource. The cost for each percent reduction was about 16 professional person-hours for a state with 39 counties, an average population of 124,787 per county, and an average annual growth rate of 1.7 percent from 1980 to 1990.

In terms of timeliness, the linkage with the PRE measure is at least as straight-forward as that found for cost. The most recently available census figures are, generally speaking, the ultimate in timeliness in that they can be obtained in the time it takes to acquire a printed report or electronic file. A user may, however, have to wait on the production of a 'current' forecast, one that takes advantage of the most recently available census, birth, and death data for example.

In addition to the relationship between the PRE measure, cost, and timeliness, other issues come to mind in terms of evaluating utility. For example, instead of the last census as the 'naive' projection, one could use a post-censal estimate and then calculate PRE on this basis in a subsequent evaluation. Here, however, one would have to take care in that it is often the case that Cohort-component forecasts are themselves informed (tended in accordance with) post-censal estimates. This is the case in both Washington and Ohio as well as many other states.

Another issue is in regard to evaluating the PRE measures themselves. For example, is a PRE of 66% 'excellent', or is it 'fair'? Is one of 15% 'poor'? What we need are some guidelines in terms of ranking. These guidelines need to be developed empirically, and are likely to be specific to the historical context, size, and growth rate of the geographic area under consideration. As a starting point, we suggest that something like the following be considered in evaluating PRE scores in the absence of empirical guidelines: less than zero, bad; 0-25%, poor; 26-50% average; 51-75%, good; 76-100%, excellent.

As the tone of our discussion indicates, we strongly suggest that PRE scores not be used in isolation. In this regard, we include not only cost and timeliness but also accuracy as criteria that could be used in conjunction with the PRE Measure. In terms of accuracy, much like the necessity of having both a measure of central tendency and dispersion in descriptive statistics, it is important to know the levels of error associated with naive and other forecasts as well as the PRE scores involving them. That is, in the same way

that it is virtually impossible to get an r^2 score that is regarded as 'good', when the standard deviation for a given 'dependent' variable is close to zero, it is unlikely that PRE would even be 'average' if the RMSEc score is extremely low. Along similar lines, it is of interest to note that in a state with very low change (like the example of Ohio, 1980 to 1990, used here), one would expect, *ceteris paribus*, that the 'last census' would be quite accurate in terms of a forecast. In fact, this is the case for Ohio. We observe, in this regard, that the PRE score associated with using the Cohort-Component projection (Ohio Data Users Center 1985) instead of the last census indicates that it is remarkably accurate. This type of finding may be common in low- or no-change areas.

Conclusion

As we know, forecasts invariably turn out to be different than reality even though the forecast is usually afforded an unwarranted degree of precision when it is produced. Therefore, some have described forecasting as an impossible task, yet forecasts cannot be avoided. This creates considerable stress on both the producers and users of forecast information. Current strategies for dealing with this stress (e.g., developing confidence intervals or multiple forecast series) are unsuccessful because they do not address the fundamental problem: stress is a direct result of cognitive dissonance.

In this paper we propose a new way of evaluating forecasts that acknowledges the demand by users for a straight-forward description of the future but reduces the level of stress on both those who produce and use them by introducing criteria that fit within the values and beliefs that shape the world in which forecasters and their 'customers' exist. We argue that this can be accomplished by adding the PRE measure to the set of forecast evaluation criteria. Moreover, we argue that the PRE measure is also useful because it easily accommodates related criteria involving other dimensions of utility such as timeliness, cost, and 'value added'.

The major advantage of adding utility as an evaluation criterion is that it provides another way to determine the usefulness of forecasts and suggests dimensions that may be at least as important as accuracy in their evaluation. Once again, note that we are not arguing that 'ex post facto' accuracy should be ignored, but that it should not be the sole criteria for judging a forecast's worth. Because accuracy has acquired such a prominent role, Demographers and other producers of forecast information have learned not 'shoot themselves in the foot' by showing evaluations based on RMSE, MAPE, and other measures of accuracy to uniformed users, especially for substate or subcounty level forecasts, because they often fail to meet user expectations for accuracy. Providing an additional perspective offers a more positive way to approach this subject and could lead to a wider exchange of information and ideas than is happening now. It may also represent a way in which the

field of population forecasting can avoid the problems afflicting the field of population enumeration, as we discussed earlier in regard to modern US census activities.

Using the statistical concept of proportionate reduction of error (PRE), we developed and illustrated a way to measure the gain of information provided by a forecast or its utility. We suggest that this approach to evaluation should be considered in any forecasting area where there is a high level of cognitive dissonance due to unrealistic standards of forecast accuracy. By introducing a measure like PRE, we believe that the consciousness of both producers and users can be redirected toward the more constructive concept of forecast utility.

Note

1. It is appropriate and useful to examine the PRE using other error measures such as the MAPE or MALPE. This would provide a broader context in which to judge a forecast's utility. Along these lines we have completed analyses using the MAPE and found PRE values of 35.7 percent for Ohio and 61.4 percent for Washington using the same data underlying the use of RMSE as reported here. It is of interest to note that the PRE scores using either RMSE or MAPE are very similar for Washington while the MAPE-based PRE score for Ohio shows a much smaller reduction in error compared to the RMSE-based PRE score. These findings suggest further research in terms of different error measures under different conditions of population change (Tayman & Swanson 1995).

Acknowledgments

This is a revision of a paper presented at the 14th International Symposium on Forecasting, 12–15 June 1994, Stockholm, Sweden. An excerpt was presented at the 1994 Annual Meeting of the Federal-State Cooperative Program for Population Projections, 4 May 1994, Miami, Florida, USA.

The authors wish to thank Mary McGehee, Hallie Kintner, and Stan Smith for reading earlier drafts and providing thoughtful comments. Participants at the meetings where we presented the ideas found in this paper also provided useful comments for which we are grateful. In addition, we acknowledge the anonymous reviewers and the comments we received through the review process that improved this paper. Finally, we are indebted to Barry Bennett and Theresa Lowe for providing data used in the examples for Ohio and Washington, respectively.

References

- Agresti, A. (1990). *Categorical data analysis*. New York: John Wiley.
- Armstrong, J. (1982–83). The 15 most common pitfalls and how to avoid them, *Journal of Business Forecasting* 1(6): 12–15.

- Ascher, W. (1978). *Forecasting: An appraisal for policy-makers and planners*. Baltimore, MD: The Johns Hopkins Press.
- Brettschneider, S. & Gorr, W. (1992). Alternatives to forecast error based evaluation: Communicability, manipulability, credibility, and policy relevance, pp. 114–123, in: *Proceedings of the Federal Forecasters Conference*, 1992.
- Burch, T., Swanson, D.A. & Tedrow, L. (1996). What is applied demography? What can it be? How can we get there from here?, *Population Research And Policy Review*, forthcoming in a special issue (Spring, Vol. 15).
- Costner, H. (1965). Criteria for measures of association, *American Sociological Review* 30: 341–353.
- D'Allesandro, F. (1987). Should applied demographers take out liability insurance?, *Applied Demography* 3(Fall): 1–3.
- Dorn, H. (1950). Pitfalls in population forecasts and projections, *Journal of the American Statistical Association* 45: 311–334.
- Espenshade, T. & Grummer-Strawn, L. (1991). Evaluating the accuracy of US population projection models. Paper presented at the Population Association of America Conference, Washington, DC.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.
- Festinger, L., Riecken, H., & Schachter, S. (1956). *When prophesy fails*. Minneapolis, MN: University of Minnesota Press.
- Hajnal, J. (1995). The prospects for population forecasts, *Journal of The American Statistical Association* 50: 309–322.
- Keyfitz, N. (1972). On future population, *Journal of the American Statistical Association* 67: 347–363.
- Keyfitz, N. (1981). The limits of population forecasting, *Population and Development Review* 7: 579–594.
- Keyfitz, N. (1982). Can knowledge improve forecasts?, *Population and Development Review* 8: 729–751.
- Keyfitz, N. (1987). The social and political context of population forecasting, pp. 235–258, in: W. Alonso & P. Starr (eds.), *The Politics of Numbers*. New York: Russell Sage Foundation.
- Kintner, H. & Swanson, D. (1994). Forecasting health benefit populations. Paper presented at the 14th International Symposium on Forecasting, Stockholm, Sweden.
- Makridakis, S. & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation, *Journal of the Royal Statistical Society, Series A*: 142: 97–145.
- Moen, E. (1984). Voodoo forecasting: Technical, political, and ethical issues regarding the projection of local population growth, *Population Research and Policy Review* 3: 1–25.
- Murdock, S. & Leistritz, L. (1980). Selecting socio-economic assessment models: A discussion of criteria and selected models, *Journal of Environmental Management* 10: 1–12.
- Murdock, S., Hamm, R., Fannin, D., Pecotte, B. & Voss, P. (1989). *Evaluating small-area population estimates and projections*. Applied Community Research Monograph E3. Alexandria, VA: American Chamber of Commerce Researchers Association.
- Murdock, S., Leistritz, F.L. Hamm, R., Hwang, S. & Parpia, B. (1984) An assessment of the accuracy of a regional economic-demographic projection model, *Demography* 21: 383–404.
- Musham, H.V. (1965). The use of cost functions in making assumptions for population forecasts, *Proceedings of the World Population Conference*. New York: United Nations.
- Ohio Data Users Center (1985). *Population projections, Ohio and counties by age and sex: 1980 to 2000*. Columbus, OH: Ohio Department of Development.
- Pittenger, D. (1978). The role of judgement, assumptions, techniques, and confidence limits in forecasting population, *Socio-Economic Planning Sciences* 12: 271–276.
- Rainford, P. & Masser, I (1987). Population forecasting and urban planning practice, *Environment and Planning A* 19: 1463–1475.
- Reynolds, H.T. (1977). *Analysis of nominal data*. Beverly Hills, CA: Sage.
- Robinson, J.G., Ahmed, B., Das Gupta, P. & Woodrow, K. (1991). Estimating coverage of

- the 1990 United States census: Demographic analysis. Paper presented at the Annual Meeting of The American Statistical Association, Atlanta, Georgia.
- Romaniuc, A. (1994). Reflection on population forecasting: From prediction to prospective analysis, *Canadian Studies in Population* 21(2): 165–180.
- Smith, S. (1987). Tests of forecast accuracy and bias for county population projections, *Journal of the American Statistical Association* 82: 991–1003.
- Smith, S. & Bayya, R. (1992). An evaluation of population forecasts for Florida and its counties, *Applied Demography* 7 (Spring): 1–5.
- Smith, S. & Shahidullah, M. (1995). An evaluation of population projection errors for census tracts, *Journal of The American Statistical Association* 90(1): 69–71.
- Smith, S. & 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. & Sincich, T. (1992). Evaluating the forecast accuracy and bias of alternative projections for States, *International Journal of Forecasting* 8: 495–508.
- Stoto, M. (1983). The accuracy of population projections, *Journal of the American Statistical Association* 78: 13–20.
- Starr, P. (1987). The sociology of official statistics, pp. 7–58, in: W. Alonso & P. Starr (eds.), *The politics of numbers*. New York: Russell Sage Foundation.
- Swanson, D.A. (1986). Evaluating population estimates and short-term projections, *Applied Demography* 2 (November): 5–6.
- Swanson, D.A. & Beck, D. (1994). New short-term county population projection method. *Journal of Economic and Social Measurement* 20: 1–26.
- Swanson, D.A. & Tayman, J. (1994). Measuring the utility of population projections. Paper presented at the Annual Meeting of the Ohio Academy of Science, Toledo, Ohio.
- Swanson, D.A., Tayman, J. & Beck, D. (1995). On the utility of lagged ratio-correlation as a short-term county population projection method: A case study of Washington State, *Journal of Economic and Social Measurement* 21: 1–16.
- Sykes, Z. (1969). Some stochastic versions of the matrix model for population dynamics, *Journal of the American Statistical Association* 44: 111–130.
- Tayman, J. (1993). How accurately can we forecast small area population? Presented at the Annual Meeting of the American Statistical Association, San Francisco, California.
- Tayman, J. (1996). The accuracy of small area population forecasts based on a spatial interaction land use modelling system, *Journal of The American Planning Association* (forthcoming).
- Tayman, J., Schafer, E. & Carter, L. (1994). Confidence intervals for small area population forecast error: A repeated sampling approach. Paper presented at the Annual Meeting of the Population Association of America, Miami, Florida.
- Tayman, J. & Swanson, D. (1995). Alternative measures for evaluating population forecasts: A comparison of state, county, and subcounty areas. Paper presented at The Annual Meeting of the Population Association of America, San Francisco, California.
- US Bureau of the Census (1975). *Coverage of population in the 1970 census and some implications for public programs*. Current Population Reports, P-23, No. 56. Washington, DC: US Government Printing Office.
- US Bureau of the Census (1982). *Coverage of the national population in the 1980 census by age, sex, and race*. Current Population Reports, P-23, No. 115. Washington, DC: US Government Printing Office.
- US Bureau of the Census (1992). *Population projections of the United States. by age, sex, race, and Hispanic origin: 1992–2050*. Washington, DC: US Government Printing Office.
- Washington State, Office of Financial Management (1986). *Forecasts of the State and County populations by year for selected age groups: 1980–2000*. F86–11. Olympia, WA: Population Estimation and Forecasting Unit.
- Wheelwright, S. & Makridakis, S. (1980). *Forecasting methods for management*, 3rd ed. New York: John Wiley.

Address for correspondence: David A. Swanson, Arkansas Institute for Economic Advancement, University of Arkansas at Little Rock, 2801 S. University Avenue, Little Rock, AR 72204–1099, USA

Phone: (501) 569 8529; Fax: (501) 569 8538