The Status of Evaluating Accuracy of Regional Forecasts[*]

Carol T. West

Department of Economics, Warrington College of Business, University of Florida
224 Matherly Hall, Gainesville, FL, 32611-7140, e-mail: ctwest@ufl.edu

Abstract

Recent reviews of the regional growth and change literature conclude that the strand of research attempting to identify determinants of growth in general terms has been reasonably successful, but the strand attempting to provide public policy direction has been far less successful. A critical distinction between these two strands is that the former studies the past while the latter makes a forecast. One of the reasons we may not have been particularly effective in guiding public policy is that we haven't addressed the question, "Is the accuracy of our regional forecasting record acceptable?" This paper reviews the literature on evaluating regional forecast accuracy, discusses the question of "When is a regional forecast 'accurate'?" and outlines how accuracy analysis can be used to improve forecast precision. A concluding section suggests possible avenues for future research.

1. INTRODUCTION

The starting point for the series of articles in this volume was the observation that the literature on regional growth has two strands: First, that which explores the growth process in general terms of determinants of growth and second, that which attempts to specifically guide public policy in affecting the growth process. My colleagues suggest that we've met with reasonable success in the first endeavor, but attempts to provide public policy direction have been much less successful (Bartik, Boehm and Schlottmann 2003).

A critical difference between the two strands is that when we move from "exploring the regional growth process" to "providing public policy direction," we move from studying the past to making a forecast. And here we have to ask ourselves, "Is the accuracy of our regional forecasting record acceptable?" One of the reasons we may not have been particularly effective in guiding public policy is that we haven't asked this question often enough.

It's a tough question. Like "beyond a reasonable doubt," there is no clear-cut, quantitative, theoretically based standard that unambiguously defines "acceptable accuracy." Consequently, differing explicit (or implicit) standards of accuracy exist in the literature and in the case of regional forecast evaluation, that literature is spread over journals and publications from diverse disciplines—e.g., demography, forecasting, general economics, and public finance as well as regional science. Section 2 of this study surveys the literature on regional forecast evaluation, clarifying and comparing the various accuracy standards that have been considered and applied; and Section 3 critically summarizes the review, suggesting guidelines for future evaluations.

There is a natural follow-up question to "Is the record acceptable?" Namely, "Can the record be improved?" While there is a large body of regional literature addressing the latter through the introduction of new forecast modeling techniques, integrating basic evaluation of accuracy into an ongoing forecast improvement process has been less intensively examined. Both regional forecast evaluation and forecast improvement have generally been "model-focused," not "modeling-process focused."[1] The evaluation literature is "model-focused" in the sense that it is used to select between models or validate the predictions of a particular model. The improvement literature is "model-focused" in the sense that it proposes increasing accuracy by *substituting* one model for another. In contrast, "modeling-process focused" approaches use the accuracy analyses to *modify* or *enhance* an initial model, resulting in its *evolution* over time into a more accurate predictor of the future. Section 4 of the study considers use of regional accuracy analysis in modeling-process focused approaches to increase the precision of regional forecasts. A final section proposes agenda items for future research.

2. REVIEW OF THE LITERATURE ON EVALUATION OF THE ACCURACY OF REGIONAL FORECASTS

2.1 Scope of the Review

Like all reviews in a subject where one topic tangentially relates to many, practical considerations require specific delineation of the relevant literature. The scope of the current

---

[1] There are some notable exceptions discussed in the subsequent sections of the study.

study is limited by explicit definitions of "forecast" and "evaluation." In the policy guidance context that motivates this research, the policy maker is concerned with a *future* period. Consequently, the review is restricted to the literature based on "true *ex ante*" forecasts, predictions made at a point in time *prior to* the period being forecast. Specifically excluded in this review are the many model test studies that use the term "forecast" in reference to the following situation: The researcher in year $N + t_1, t_1 > 0$, has data for years 1 to N. In *estimating* the model parameters, he/she uses only data from 1 to $N - t_2$, $t_2 > 0$. As part of the model testing process, the estimated model, based on data from 1 to $N - t_2$, is used to "forecast" the period $N - t_2 + 1$ to N. The "forecast" and actual can immediately be compared. This is not a "true *ex ante* forecast*" since the prediction is being made in $N + t_1$, which post-dates the period of "forecast" that ends in N. This situation is more correctly referred to as a "post sample test." While certainly informative as part of a set of model diagnostics, such tests differ significantly from the *ex ante* context of interest here. As Nelson (1984) has emphasized, true *ex ante* implies complete lack of knowledge about the future. In addition, often the available data used at the time of a true *ex ante* forecast are preliminary and hence may contain considerable misinformation about the actual state of the economy compared with revised or final data. Structural models based on final data that truncate the period of estimation and reserve the most recent for the post sample test violate both these conditions—the test uses data not available at the time of "forecast" and, inevitably, knowledge of the post sample period has had some influence on specification.[2]

There are some instances in which the model estimated at time $N + t_1$, using data from period 1 to $N - t_2$, and then "forecast" from $N - t_2 + 1$ to N does not differ significantly from a true *ex ante* forecast made in period $N - t_2$. Two features characterize these exceptions. First, the modeling methodology *solely* relies on the past history of the variable being forecast. The forecaster makes no choice about additional variables to include in the model, thus avoiding the influence of information available since $N - t_2$ on model specification. This situation includes many simple deterministic methodologies (e.g., linear extrapolation of past trends) and *univariate* time series methodologies. Second, the model estimation is *based on data available at time $N - t_2$*, avoiding the problem of incorporating revised or final data that would not have actually been available in $N - t_2$. If *both* conditions are satisfied, use of a *univariate methodology based solely on data available at time $N - t_2$*, then arguably the forecast made in $N + t_1$ for $N - t_2 + 1$ to N differs inconsequentially from that which would actually have been made in $N - t_2$ for the same period. Application and empirical estimation of the methodology is invariant to whatever occurred subsequent to the "launch year," the final time period included in the estimation—in the example, $N - t_2$.[3] This situation, referred to as "*ex post* creation of *ex ante* forecasts*," in fact occurs in many of the studies considered below.

_____

[2] Certainly when the "post sample test" uses actual values of exogenous model variables, the deviation from true *ex ante* conditions becomes even larger. Similarly, accuracy analysis derived from *ex post* simulations over the period of model fit is of "limited value since, given sufficient effort, one can almost always obtain a reasonably good fit to a set of sample data" (Granger and Newbold 1977, p. 290).

[3]As noted by Lenze (2000), strictly speaking, these forecasts should not be classified as *ex ante.* They are what Fair and Shiller (1990) would consider "quasi- *ex ante.*" In the linear extrapolation example, the choice of base period to determine trend may be influenced by knowledge of the forecast horizon or simply by research published after the launch year. To eliminate the "quasi-*ex ante*" in addition to the "post sample tests" would reduce the number of

In addition, low frequency projections (ones forecasting over long time periods such as 5, 10, or 15 years) based on univariate methodologies may be reasonably close to "true *ex ante*" even if based on final data. As noted by McNees (1992), as the projection horizon lengthens, the distinction between preliminary and revised data in the launch year of the forecast becomes less important. The same is not necessarily true of high frequency (e.g., monthly or quarterly) forecasts based on final data versus high frequency forecasts based on preliminary data. Low frequency projections based on univariate methodologies are treated as "true *ex ante*" in the current review regardless of whether they technically use revised or final data in the launch year.

Finally, on a semantic point, the terms "forecast," "prediction," and "projection" are used interchangeably in this paper. Historically, demographers in particular have tried to distinguish the latter from the former two, but as Smith, Taymen, and Swanson (2001) aptly note (p. 301), "A basic fact of life for demographers is that their projections become forecasts as soon as they reach the public."

Having confined the review to *ex ante* forecasts, defined to include the near *ex ante* as well as true *ex ante* as discussed above, the remaining delineating dimension distinguishes "evaluation of accuracy" from the important complementary activity of simply "reporting accuracy." Evaluation in the current context, as opposed to simply reporting, implies comparison with an explicit (or implicit) *standard of accuracy*. As noted in the Introduction, defining that standard is a central problem of evaluation.

However, none of the proposed standards are specific numbers universally applicable to all variables and projection horizons. For example, none says a 2.0 percent error is "always acceptable" but a 2.8 percent error is "never acceptable." Proposed standards draw in some way from feasible forecasting alternatives; and what is "feasible," given the state of forecasting, may or may not be "acceptable" (or worth paying for) in any particular applied circumstance. Suppose the standard for a specific variable/projection horizon is a 50 percent error; then a forecasting record with a consistent error of 40 percent will be deemed "acceptable." However, the user needs to know the typical error is 40 percent. That knowledge will affect how he/she uses the forecast and whether he/she decides to pay for the forecast. For an alternative variable/ projection horizon combination, the standard may be a 0.4 percent error; so a forecasting record with a consistent error of 0.6 percent will be deemed "unacceptable" even though it may be a very useful input in many applied planning applications. In short, formal evaluation can never replace simply reporting forecasting errors and letting accuracy be, like beauty, in the eye of the beholder.

Indeed, such reporting constitutes critical due diligence, notifying users of product quality. It is widely recognized that the failure of the large macroeconomic forecasting firms in the mid-1970s to recognize the shortcomings of their "black box" outputs set the stage for their demise (*Business Week* 1981), and due diligence by regional forecasters can perhaps avoid a similar demise. To their credit, some on-going regional forecasting projects regularly or periodically report their forecast errors, including perhaps some discussion of what occurred in the economy that was unforeseen and may have contributed to reported inaccuracy. The reports contain

---

qualifying studies to so few the analysis would not be of interest. Consequently, "quasi-*ex ante*" is treated as *ex ante*.

measures of accuracy, but not a standard of accuracy for comparison. Examples include Economy.com (Thomas and Zandi 2002 is a sample of a regular annual report; Conway 1990; Regional Economic Analysis Division 1993; and Hammond 2000).[4]

However useful the raw reported prediction errors are, they can also be difficult for a user to interpret. "Is this as good as it gets?" is a natural question to ask. Providing some systematic comparative standard of accuracy can at least partially address that question. In the above examples, while the user needs to know the 40 percent and the 0.6 percent errors, he/she also benefits from knowing the former may be "as good as it gets" while the latter probably isn't. For the discipline of regional forecasting as a whole, evaluation is important as a measure of how effectively it is applying the current art and science of forecasting to its particular prediction issues. The existing reports of the type noted above provide potentially rich data sets for future systematic evaluations that this study hopes to encourage, evaluations that will ultimately provide guidelines for both scholarly research and practical application.

Given these two caveats on the scope of the review (*ex ante* forecasts and evaluation that goes beyond simple reporting of errors), there are two cohesive, and distinct, trends in the regional forecasting accuracy evaluation literature. These are analysis of population projection accuracy and analysis of state revenue forecast accuracy. A third more varied group covers accuracy of other economic predictions such as employment, income, and housing starts.

2.2. Regional Demography

Implicitly, regional demography has defined a forecast as "accurate" if it comes from the *generally most accurate methodology.* The literature is a quest for the latter with a history dating back more than three-quarters of a century. Siegel (1953) reports that Engelbardt (1925) compared the results of forecasts for a number of cities from the "component method"[5] with fitting a logistic curve. Chamberlain and Crawford (1932) sought to determine the most accurate among four methods for forecasting city population. While these early studies may lack modern standards of scientific methodology, they initiated the questions about projection technique that were later bluntly stated by Siegel (1953, p. 72):

> *Many demographers have preferred the elaborate "cohort-survival" method on* a priori *or logical grounds but, for all we now know, this may represent an unfortunate waste of time and effort where simpler methods would be just as accurate.*

White (1954) first systematically examined this question. She compared the accuracy of Thompson and Whelpton's projection of state populations for 1940 and 1950 using the cohort-survival method based on 1930 Census data with those of other methods "which did not involve extensive computing and for which the results would not be biased by the worker's knowledge of population trends since 1930" (pp. 480-81). In short, she compared the relative accuracy of

---

[4] The examples are illustrative, not exhaustive. Many similar relevant reports undoubtedly exist but are proprietary or not reported in the mainstream literature.

[5] The "component method" in this context is separate estimation of births, deaths, and migration but without taking age into account.

"complex" and "simple" projection techniques.[6]   Her findings were (pp. 484-491):   (1) the complex method (cohort-survival) did ***not*** yield "definitely superior results" to the simpler techniques; (2) "no one method was clearly superior to all other methods tested" and only one was "clearly inferior" (the one that put "too much emphasis on population change in relatively remote periods"); (3) "the errors of projections tend to increase almost directly as the length of the projection period increases"; (4) "the errors of projections tend to be larger for areas with smaller base populations"; and (5) "the errors of projections tend to be larger for areas with larger net migration rates in the recent past."

Comparing the relative accuracy of complex and simple methods has been a dominant theme and hotly debated issue in regional demography accuracy analysis, including recently an entire 1995 issue of *Mathematical Population Studies* (Vol V., No. 3) devoted to the topic. Many of these studies are summarized well in Chapter 14 of Smith, Taymen, and Swanson (2001). Almost five decades of empirically examining the relative accuracy of complex and simple methods has added ARIMA models and structural economic-demographic models to the debate and has generated a taxonomic controversy ("naïve" v. "sophisticated," "complex" v. "simple," "extrapolative" v. "causal," etc.), but has not fundamentally changed or added significant new knowledge to the five conclusions drawn by White in 1954.[7]

2.3 State Revenues

The literature on state revenue forecasting accuracy has had a very different set of focuses. First, relative to other areas of regional forecast analysis, it has been far more concerned with the issue of bias compared with absolute accuracy.[8] That concern derives from political modeling of budget decision making that implies bias toward underprediction (e.g., Larkey and Smith 1984, and Bretschneider and Schroeder 1985).

---

[6] The "simple" techniques were variants of (a) geometric: assume the state's historical average percent rate of growth will continue; (b) arithmetic: assume the state's historical average absolute change in population level will repeat; (c) ratio: assume the state's rate of change in national population share declines to zero in 50 years; (d) apportionment: distribute projected national population gains to states according to their realized shares of gain in an historical period. Methods (c) and (d) require a national projection, which was the same as the Thompson and Whelpton national projection. Variations of the methods included (1) whether the sum of states in (a)-(c) are forced to equal the predetermined national total and (2) the length of historical period used in determining averages and shares in (a), (b), and (d).

[7] An exception is the observation that launch year of the forecast can consistently affect accuracy across different methodologies, an observation White was unable to make since she studied only one launch year (1930).

[8] Absolute accuracy (or perhaps more precisely, absolute *inaccuracy*) is the *magnitude* of difference between actual and forecast values, and it may be measured in a variety of ways. Given a sample of T forecast and actual values of the variable X, X_t^f, and X_t^a respectively with $t = 1, 2, \ldots T$, common measures of absolute accuracy include

(a) mean absolute error, $\dfrac{1}{T}\sum_{t=1}^{T}\left|X_t^f - X_t^a\right|$; (b) mean absolute percent error, $\dfrac{1}{T}\sum_{t=1}^{T}\left|\dfrac{(X_t^f - X_t^a)}{X_t^a}\right|$; (c) root mean

square error, $\sqrt{\dfrac{1}{T}\sum_{t=1}^{T}\left(X_t^f - X_t^a\right)^2}$; and (d) root mean square percent error, $\sqrt{\dfrac{1}{T}\sum_{t=1}^{T}\left(\dfrac{X_t^f - X_t^a}{X_t^a}\right)^2}$. None take

into account the direction of error, just how far away actual is from forecast. Bias indicates whether a forecast is

Second, the revenue literature has a history of explicitly modeling the forecast error (both percent error and absolute percent error) as a function of economic, institutional, procedural, and political characteristics of states (e.g., Bretschneider et al. 1989; Cassidy, Kamlet, and Nogin, 1989; Bretschneider and Gorr 1992; Mocan and Azad 1995). Examples of economic variables include business cycle indicators (e.g., percent change in GNP); standard broad state economic indicators such as unemployment rate, employment, and per capita income; specific measures of the state economic base (e.g., oil-producing or not oil-producing); and national economic forecast errors (e.g., the percent error in the GNP forecast made by a major national forecasting firm). Institutional factors specify the legal environment of revenue projection and budgeting— e.g., whether there is allowance for deficit carryovers and/or gubernatorial authority to cut spending during the year. Procedural variables cover forecast technique, use of outside consultants, and whether a consensus estimating process is used. Relevant political characteristics are the party affiliations of the executive and legislative branches of government.

The formal multivariate framework contrasts with the analysis of forecast error determination in the demographic literature. The latter has observed some broad characteristics of regions that appear to systematically affect accuracy across projection methodologies—e.g., size and growth rate. The conclusions, however, derive from grouping regions by size or growth rate and then noting trends in mean absolute percent error (MAPE) across groups. This is essentially an observational simple correlation and the characteristics have not generally been analyzed in a multivariate framework that determines magnitude and significance of factors simultaneously.[9]

Third, in evaluating forecast accuracy, the revenue literature has focused on the notion of rationality or efficiency of the forecast—do the forecasts optimally incorporate all information available at the time the forecast was made? This is very different from the regional demography approach of comparing the outcome of one forecast methodology with that of another. Feenberg et al. (1989) formally present the efficiency concept, which is intuitively relatively simple. Suppose there is a sample T, $t = 1, 2, \ldots T$, of r period ahead forecasts of the variable X. Let X_{t-r}^{f} be the forecast of variable X at time t made r periods ago. Let X_{t}^{a} denote the actual outcome of X at time t. Variables $I_{1,t-r}$, $I_{2,t-r}, \ldots I_{n,t-r}$ are the relevant set of information available when the forecast was made. Such information might include recent data on the economy—both local and national—and other available forecasts. If the set of forecasts is *strongly rational*, then X_{t-r}^{f} should incorporate all relevant information in the set of n I_{t-r} variables; and the forecast errors, $X_{t}^{a} - X_{t-r}^{f}$, should hence be uncorrelated with these variables. The test is to estimate the regression:

$$(1) \quad \left(X_{t}^{a} - X_{t-r}^{f}\right) = \beta_{1}I_{1,t-r} + \beta_{2}I_{2,t-r} + \ldots \beta_{n}I_{n,t-r} + u_{t},$$

---

consistently high or low. Statistically, a set of forecasts is unbiased if $E\left(X_{t}^{f}\right) = X_{t}^{a}$ or intuitively, the set is

unbiased if $\sum_{t=1}^{T}\left(X_{t}^{f} - X_{t}^{a}\right)$ is near zero.

[9] An exception is the recent analysis by Lenze (2000) that estimates an empirical model of absolute errors in five-year forecasts of the population of Florida's 67 counties.

where $E(u_t \mid I_{1,t-r}, I_{2,t-r} \ldots I_{n,t-r}) = 0$ and then test the joint hypothesis all βs are zero. If the hypothesis is accepted, then the forecast error is uncorrelated with information available when the forecast was made—i.e., the forecast effectively incorporated all such information. If the hypothesis is rejected, then one or more of the I variables is significantly correlated with the forecast error, and the forecast could have been improved by better incorporating that information.[10]

An alternative concept is that of *weak rationality*. If the forecaster uses only a subset of the I variables in making the forecast but uses it efficiently in the sense that he/she gets the correct answer on average, then the forecast is weakly rational. This amounts to testing $E\left(X_{t-r}^f\right) = X_t^a$, which is the conventional test for bias. Run the regression:

(2) $X_t^a = \alpha_0 + \alpha_1 X_{t-r}^f + u_t$

and then test the joint hypothesis $\alpha_0 = 0$ and $\alpha_1 = 1$.

In running the test for strong rationality, there is clearly a problem of determining what to include in the information set I_{t-r}. In the case of testing the rationality of state revenue forecasts, obvious candidates include recent economic and demographic data on the state available at $t - r$, recent national data such as inflation rates known at $t - r$, last year's actual state revenues and various institutional and political factors in the state at time $t - r$. In general, the studies find that state revenue forecasts are not strongly rational, but whether they are weakly rational is less clear (e.g., Gentry 1989; Feenberg et al. 1989; and Mocan and Azad 1995).

Comparative methodologies receive far less attention in the state revenue analyses than in the regional demography accuracy studies, but they are not neglected. Dummy variables designating whether the revenue forecasting process uses judgmental, simple quantitative, or sophisticated quantitative methods typically appear in the regressions explaining model error. Results do not contradict the findings of the population literature—i.e., they are inconclusive. For example, Bretschneider et al. (1989) find that forecast accuracy increases when simple regression models and judgmental methods are used as opposed to univariate time series methods or econometric models; but Mocan and Azad (1995) find that use of judgmental forecasting significantly raises forecast error, while use of sophisticated econometric techniques significantly reduces it.

2.4. Other Regional Forecasts

Forecasts of other regional variables (e.g., employment, housing starts, income) do not have accuracy evaluation literatures as extensive as those of population and state revenues. There are three reasons for this situation. First, any plausible accuracy analysis needs a large number of *ex ante* forecasts as a sample of observations, but the variables in question are newcomers to the

---

[10] As a simple example, suppose all coefficients do not differ significantly from zero except that on $I_{j,t-r}$. Then $E\left(X_t^a - X_{t-r}^f\right) = \beta_j I_{j,t-r} \neq 0$. If the forecast had been $X_{t-r}^{f'} = X_{t-r}^f + \beta_j I_{j,t-r}$, then $E\left(X_t^a - X_{t-r}^{f'}\right) = 0$, a more accurate forecast than X_t^f.

regional forecasting field compared with population (which, as noted above, has been projected regionally at least since 1925) and state revenues (with forecasts dating back to the early post-World War II period). Systematic, ongoing *ex ante* regional forecasting of employment, housing starts, income, etc. has a much shorter history. Indeed, the *start* of regional time series on some of these variables post-dates the start of *ex ante* forecasting of population and revenues.

Second, for short-term quarterly forecasts, the history of their national macroeconomic counterparts makes clear that accuracy varies with the business cycle. Consequently, systematic evaluation of these regional forecasts is logically postponed until a full cycle of forecasts has been made—i.e., there is a forecasting record that spans all phases of the business cycle. Although methodological development of such regional forecasting models was rapid in the 1970s, few actual, ongoing regional forecasting programs for these variables predate the mid-1970s recession. Many also postdate the 1980 and 1981-82 recessions. By highlighting the differential growth paths of many regional economies compared with the national average, it was the "rolling recession" of the 1980s that escalated the demand for more economic forecasts specific to local markets. Consequently, few regional areas had a forecasting track record that included a recession period prior to the early 1990s. For longer-term quinquennial or decennial projections, there was the simple problem of having to wait an extended period to observe an "actual."

Third, accuracy analysis of short-term forecasts is particularly confounded by the large revisions to these regional variables that occur over time and render unclear how "actual" should be measured for purposes of quantifying forecast error. This is not a problem peculiar to regional forecast analysis, but one it shares with its national macroeconomic counterpart. McNees (1988) has shown how difficult it is to *ex post* incorporate these revisions into the evaluative process, and McNees (1992) illustrates how choice of "actual" can significantly affect conclusions on forecast accuracy. Note that the revenue forecasts were also short-term, but they did not have this problem of revision—actual revenues are primary data, not estimates based on samples or secondary techniques that give rise to the sequence of revisions observed in data on employment, housing starts, income, etc.

Early accuracy analyses of *ex ante* regional forecasts other than population and revenue include Kozlowski (1987) and Goldstein and Cruze (1987). Neither evaluates in the sense of applying a standard of accuracy, but each comprehensively examines issues pertinent to a specialized forecasting context. In the case of Kozlowski (1987), the context is regional indices of leading indicators and the issue addressed is their reliability in signaling regional turning points. In the case of Goldstein and Cruze (1987), the context is Bureau of Labor Statistics (BLS) state projections of occupational employment and the issue is identifying sources of error. More recently, a sequence of articles (West and Fullerton 1996; Fullerton and West 1998; Fullerton, Luevano, and West 2000; and Fullerton, Laaksonen, and West 2001) evaluates accuracy of state and metropolitan area quarterly forecasts of employment and housing starts. Basic data for the studies are *ex ante* projections for Florida and its MSAs that were published in *The Florida Outlook* over the decade 1985-1995.

All four borrow from a long history of national macroeconomic model accuracy analysis (e.g., Nelson 1972, 1984; McNees 1988, 1992; Ashley 1983, 1988) that focuses on the notion of

"benchmarking," comparing accuracy from a complex model with that produced by a simple univariate time-series approach. They are thus similar to the accuracy analyses of regional demography. The housing start studies tend to confirm the findings of the demographic literature—namely, little consistent difference in accuracy between the complex and simpler methodologies. The employment-based study, West and Fullerton (1996), is less clear on this issue. The more complex methods outperform a variety of simpler methodologies at short-term horizons (one to five quarters out), but differences tend to diminish at longer horizons (six to ten quarters out). These findings are difficult to compare directly with the demographic outcomes that typically do not include samples from very short-term, high frequency regional projections.

In contrast, Lenze (2000) follows the traditions of state revenue forecast accuracy in studying the precision of quinquennial projections of population, employment, income, and housing starts for the 67 counties of Florida. Accuracy is looked at in the sense of informational efficiency using the approach of Mork (1987), who proposes regressing algebraic forecast error on (1) a constant, (2) growth in the previous period, and (3) an independently prepared forecast. Specifically, let $G^a_{x,n,t,t-r}$ denote actual growth in variable *X* between period $t-r$ and period *t* for the region *n*, $G^f_{x,n,t,t-r}$ denote forecast growth for period $t-r$ to *t* made in period $t-r$, and $G^{f'}_{x,n,t,t-r}$ denote an alternative forecast for growth made independently of $G^f_{x,n,t,t-r}$. Noting that $G^a_{x,n,t-r,t-2r}$ is actual growth in *X* between periods $t-2r$ and $t-r$ for region *n*, the regression estimated is:

$$(3) \quad \left(G^a_{x,n,t,t-r} - G^f_{x,n,t,t-r}\right) = \alpha_0 + \alpha_1 G^a_{x,n,t-r,t-2r} + \alpha_2 G^{f'}_{x,n,t,t-r} + u_n.$$

Lenze (2000, p. 217) explains that a significant α_0 indicates bias, a significant α_1 indicates a tendency to smooth the forecast from the most recent observed value, and a significant α_2 indicates failure to incorporate information available to forecasters. Employment forecasts are found to be efficient. Population forecasts exhibit the desirable property of insignificant effect of lagged growth; but in contrast to the employment forecasts, they are significantly biased and ignore information available to forecasters at the time. Income forecasts are unbiased; but in contrast to both employment and population, the coefficient on lagged growth is significant.[11]

Also similar to the tradition of the state revenue forecasting accuracy literature, Lenze (2000) explicitly models forecast accuracy finding generally (pp. 210-11) that (1) size matters (larger regions are more accurately forecast); (2) age composition matters (counties with larger shares of retirees are more difficult to forecast); (3) industrial specialization does not matter; and (4) growth rate sometimes matters (is significantly positive in explaining population and housing start errors, but is not significant in explaining employment and income errors).

West (2003) also attempts to model accuracy of short-term metropolitan area employment projections and additionally expands that modeling to the important question of the *relative* accuracy of forecasts from two different methodologies (structural econometric model and univariate ARIMA time series). In contrast to the Lenze (2000) data set, which is a simple cross-

---

[11] Efficiency of housing start forecasts is not examined because of lack of an alternative projection.

section, the West (2003) data set is a sequence of one to six quarter ahead forecasts for 20 metropolitan areas. The model of forecast error therefore allows for the common observations that forecast accuracy is a function of the length of the projection horizon (step-length) and the particular point in time a forecast of a given length is made (see e.g., Theil 1967; McNees 1978; Charney and Taylor 1984; Taylor and Theil 1988). Subdividing the data set, the model of forecast error as a function of period effects (time the forecast was made) and regional characteristics is estimated separately for each projection horizon.

There are three particularly interesting outcomes: (1) there is little consistency with the Lenze (2000) study with respect to significant regional characteristics that impact forecast accuracy; (2) for a single projection methodology, the model has much better explanatory power at higher step-lengths than lower ones; and very disappointingly (3) the model is unable to explain *relative* forecast accuracy from two different projection techniques. Outcome (1) may reflect a problem that confounds regional control group analysis (Isserman 1995)—namely, there is no theory to draw upon in selecting pertinent regional characteristics to include in the model. Since there are essentially unlimited possibilities for creating measures of "regional characteristics," some process must limit the scope of variables considered. Lenze (2000) narrows a long list to a short list through a sequence of model estimations that look for the simplest specification satisfying a set of statistical criteria. In contrast, prior to model estimation, West (2003) combines related characteristics from a long list to create a short list that does not have significant multicollinearity problems. Outcome (1) may also reflect lesser variance in characteristics across 20 metropolitan areas of Florida compared with 67 counties in the state. Outcome (2) is interesting when considered in the context of the results in West and Fullerton (1996)—namely, results at one step-length do not necessarily carry over to another step-length.

3. WHEN IS A REGIONAL FORECAST "ACCURATE"?

This question is similar to the Partridge and Rickman (2003) one of "Do we know economic development when we see it?" And, like that analysis, there is no single answer to the question. The above review suggests three possible answers. The forecast is accurate:

(1) when it comes from a method that "beats" an alternative (or alternatives) in comparative accuracy;

(2) when it is efficient in the sense of not ignoring pertinent information available at the time the forecast was made;

(3) when it compares well with the accuracy of projections for a similar region (or regions), one with similar characteristics that importantly affect forecast accuracy.

There are two subcontexts to answer (1): (a) the forecast comes from a method that systematically "beats" alternatives in comparative *average* accuracy across a large number of regions and time periods or (b) the forecast comes from a method that systematically beats alternatives for the particular region-variable combination. Given the extensive empirical evidence to date, the methodology described by (a) simply does not exist. Similarly, (b) is

unlikely to be established. There are far too many potential "alternative" methodologies. If methodology A for forecasting variable *X* in region *N* over time period *R* yields *B*% accuracy, *some* alternative methodology can be found that yields greater than *B*% precision for the *X-N-R* combination. Response (3) is difficult to apply because of the problem of identifying with confidence the pertinent regional factors; in practice, a comparable regional forecast may not exist.

Accuracy in the sense of efficiency is perhaps the most attractive measure. Conceptually, a forecast is accurate if the forecaster has efficiently incorporated information available at the time the projection was made. In practice, it may be difficult to apply because a large sample of forecasts is needed to estimate regression (1). In addition, there appears to be an unlimited set of information available to be tested against; and one suspects that an extensive enough search will turn up some "neglected" item.[12] Even with this tendency to reject forecasts as being accurate, a decided benefit of the efficiency approach is that the testing process itself yields guidance on how to improve precision of projections (overlooked information).

The conceptual and/or practical shortcomings to all three approaches to determining when a forecast is accurate suggest the use of more than one in evaluation. From the benchmarking literature, there are reasonably standard univariate alternative models for (1) that range from the very simple Theil U (growth in the projection period will be the same as the immediate past period) to more sophisticated Holt-Winters smoothing and ARIMA. Similarly, plausible candidates for "available information" at the time the projection was made include recent past data on major regional aggregates (unemployment rate, growth in jobs, population, and income), similar summary national data, and forecasts of the regional and national variables. Passing comparative accuracy and efficiency tests from more than one standard provides convincing evidence of accuracy. To date, evaluation of regional forecast accuracy has not taken this broader perspective.

4. USE OF ACCURACY ANALYSIS TO IMPROVE PRECISION OF FORECASTS

In general, accuracy analysis in regional forecasting is still at a stage of being a tool to select between models or validate the predictions of a particular model. It is "model-focused," not "modeling-process-focused." The distinction can be illustrated by what has become a standard format for introducing new regional forecasting techniques in the literature—namely, exclude a set of recent data from the model estimation process and then use those data in a post sample test. The latter is typically presented as the capstone in the model verification procedure. A "modeling-process-focused" approach would have reserved *two* sets of recent data from the initial model estimation. The first would again be used for a post-sample test, but then analysis of the accuracy of that test would lead to model re-design that would be tested with the second set of reserved data. An on-going process of model development is described, not just one particular model.

Follow-up to the accuracy analysis includes (1) identifying (and correcting to the extent feasible) sources of error in the model and (2) incorporation of other forecasts/methods of

---

[12] In the literature, there is a tendency to distrust a positive outcome of the test for strong rationality and to suggest that maybe something else should also have been considered.

forecasting. Component (1) varies by methodology and context. For example, Goldstein and Cruze (1987) systematically identify sources of error in state occupational projections that lead to specific recommendations for improving accuracy. Charney and Taylor (1984) provide an error decomposition for structural econometric models that run as satellites to a national forecast (see Bolton 1985 for a general description of such models) and illustrate how the decomposition results can be used to improve forecast accuracy. For time series models, critical sources of error potentially include robustness of the model structure in the period of forecast versus the period of estimation and misleading information from recent "preliminary" data estimates. Clements and Hendry (1998, p. 168) provide a comprehensive decomposition framework. To the author's knowledge, no applications of such error analysis exists for *ex ante* regional econometric time series models, although the large number of specification changes in Fullerton, Luevano, and West (2000) and the magnitude of error in preliminary versus final regional employment data (Coomes 1992; Lenze 1992) suggest these issues should be investigated.

Identification of sources of error indicates modifications that could improve forecast accuracy. For example, substitute a more accurately forecast national variable for a poorly forecast one in a structural econometric model run satellite to a national model. Allowance for time-varying parameters may reduce specification instability between the period of estimation and that of forecast. Alternatively, derivation of the parameters from a broader set of data (e.g., a panel of regional data instead of a time series on one region) may provide a more robust relationship. Methods exist for "screening" preliminary data and taking its preliminary nature into account in the forecast process (e.g., Howrey 1978; Harvey et al. 1983; and Bordignon and Trivellato 1989). Finally, efficiency analysis may indicate particular variables that need to be included in the analysis, and these can be incorporated through specification changes.

Benchmarking and efficiency testing generate and/or utilize alternative forecasts. To the extent these alternatives contain information independent of that in the original forecast, projection accuracy can be improved by their incorporation into the forecasting process. Fair and Shiller (1989, 1990) provide tests of the informational content of alternative forecasts, and application in the regional context can be found in Lenze and West (2000). The latter illustrates one approach to incorporating the additional information—namely, enhancing model specification to include critical features of the alternative. Such improvement in basic model structure is the most satisfying approach, but may not always be feasible.

Alternatives may also be integrated into the process through direct combination of forecasts—e.g., weight forecast A by 60 percent and forecast B by 40 percent in determining the final forecast. The simplest weighting scheme is, of course, averaging. In regional forecasting, the derivation of a consensus forecast from averaging a panel of forecasts represents such a combination (e.g., Bank One Economic Outlook Center, *Western Blue Chip Economic Forecast,* various issues).

Despite an extensive general forecasting literature dating from the classic paper by Bates and Granger (1969), few regional forecast combination studies have attempted to determine non-equal weights based on the history of the relative accuracy of competing projections. Fullerton and West (1998) use regression-based weights to combine alternative housing start forecasts, but

the combination does not systematically improve accuracy.[13]   West (1996) investigates six approaches to combining ARIMA and structural econometric model forecasts of employment in 19 metropolitan areas.   Each approach is first applied individually to each series (where a "series" is a set of forecasts for a particular region and a particular projection horizon or step-length—e.g., six-quarter ahead forecasts for the Miami MSA).   None of the combination procedures improves on the option of using the most accurate forecast alone where "most accurate" is determined solely on comparative accuracy *prior* to the combination test period. However, when weights are determined in a broader context—specifically, for a given MSA/projection horizon, weight-determination takes into account comparative accuracy for *all* 19 MSAs at the given step-length and *all* ten step-lengths for the given MSA—there is systematic forecast accuracy improvement from combination.

The latter outcome is a manifestation of what has become broadly known as a *Stein effect* that dates from the seminal paper by Stein (1956) and the subsequent development of the James-Stein estimator (James and Stein 1961).   It has been concisely summarized by Casella (1985, p. 84): "Simply put, the Stein effect asserts that estimates can be improved by using information from all coordinates when estimating each coordinate."   In the above example, one coordinate is a particular MSA/projection horizon and its forecast is improved by using information from other coordinates, namely all MSAs at the given projection horizon and all projection horizons for the given MSA.   West and Theil (1991) use a formal model of Stein estimation that modifies individual regional growth projections by the forecast aggregate growth and thereby obtain modest improvements in average regional forecast accuracy.   In regional population forecasts, the observed "regression toward the mean over time" for extreme growth rates is also a Stein effect, and in that context it is generally applied to rule out certain forecast methodologies for particular regions (e.g., Smith, Taymen, and Swanson 2001).

5. FUTURE RESEARCH

Regional forecasts are going to continue to be made and used, and objective accuracy evaluations are critical to improving those projections.   Dominated by the regional demographic forecasting literature, the "improvement" aspect to date has been excessively focused on the quest for a definitive "most accurate" projection methodology.   Fifty years of research on the topic has definitively determined such a method does not exist.   It is time to move on.   New research needs to:

- Focus more on the notion of the *efficiency* of a forecast as opposed to its comparative accuracy with another forecast.   This includes drawing from the broader regional growth literature to identify plausible sets of key variables to use in conducting the efficiency test.

---

[13] Given two sets of forecasts for variable X over T time periods, X_t^{f1} and X_t^{f2} for $t = 1, 2, \ldots T$, the weights, α_1 and α_2, are determined by running the regression:

$$X_t^a = \alpha_1 X_t^{f1} + \alpha_2 X_t^{f2} + u_t$$

where X_t^a is actual X and u_t is the error term.   Note that the weights do not necessarily add to one.

- Move from a "model" focus to a "modeling process" focus. The literature is particularly scant in terms of using the accuracy analysis to improve the forecasts.

- Examine more extensively how a particular regional forecast can be improved by integrating into it outcomes from a broader set of regions. For empirical work, a key advantage to regional studies is that there are lots of regions. This advantage has been used to obtain more robust estimates from panel data in other modeling contexts, and the small amount of research on regional forecast combination (which is a problem of obtaining robust estimates of the weights) and Stein effects in regional forecasts suggests more systematic study of such integration in the regional context could improve accuracy.

- Utilize potentially rich regional panel data sets such as *The Western Blue Chip Economic Forecast* (Bank One Economic Outlook Center, various issues) to examine alternative forecast combination techniques, including system-based approaches suggested by West (1996). *The Western Blue Chip* provides a matrix of forecasts over time for each of ten western states with rows representing different forecasters and columns representing different variables. Alternative series-specific combination techniques to the reported consensus average would weight individual forecasts for a given state variable by their relative historical accuracies. A system-based approach would simultaneously take into account the relative historical accuracy of a given forecaster not only for the variable in question (e.g., employment growth), but also his/her relative accuracy on other regional variables (e.g., housing starts and retail sales).

- Conduct more accuracy evaluations of *ex ante* regional forecasts. The entire literature reviewed was small compared with the importance of the issue. If we do not honestly ask ourselves "Is the accuracy record of regional forecasting acceptable?" then we cannot expect to influence forward-looking policy.

REFERENCES

Ashley, R., 1983. "On the Usefulness of Macroeconomic Forecasts as Inputs to Forecasting Models," *Journal of Forecasting* 2, 211-23.

Ashley, R., 1988. "On the Relative Worth of Recent Economic Forecasts," *International Journal of Forecasting* 4, 363-76.

Bank One Economic Outlook Center, Various issues. *Western Blue Chip Economic Forecast*. Seidman Research Institute, Arizona State University College of Business: Tempe, AZ.

Bartik, T.J., T.P. Boehm, and Alan M. Schlottmann, 2003. "The Perplexing Literature on Growth and Change," *The Review of Regional Studies* 33(1), 1-16.

Bates, J.M. and C.W.J. Granger, 1969. "The Combination of Forecasts," *Operations Research Quarterly* 20, 451-68.

Bolton, R., 1985. "Regional Econometric Models," *Journal of Regional Science* 25, 495-520.

Bordignon, S. and U. Trivellato, 1989. "The Optimal Use of Provisional Data in Forecasting with Dynamic Models," *Journal of Business and Economic Statistics* 7, 275-86.

Bretschneider, S. and W. Gorr, 1992. "Economic, Organizational, and Political Influences on Biases in Forecasting State Sales Tax Receipts," *International Journal of Forecasting* 8, 457-66.

Bretschneider, S., W.L. Gorr, G. Grizzle, and E. Klay, 1989. "Political and Organizational Influences on the Accuracy of Forecasting State Government Revenues," *International Journal of Forecasting* 5, 307-19.

Bretschneider, S. and L. Schroeder, 1985. "Revenue Forecasting, Budget Setting and Risk," *Socio-Economic Planning Sciences* 19, 431-39.

Business Week, 1981. "Where the Big Econometric Models Go Wrong," *Business Week* (March 30) 70-71, 77.

Casella, G., 1985. "An Introduction to Empirical Bayes Data Analysis," *The American Statistician* 39, 83-7.

Cassidy, G., M.S. Kamlet, and D.S. Nagin, 1989. "An Empirical Examination of Bias in Revenue Forecasts by State Governments," *International Journal of Forecasting* 5, 321-31.

Chamberlain, L.M. and A.B. Crawford, 1932. *The Prediction of Population and School Enrollment in the School Survey.* Bureau of Social Service, College of Education, University of Kentucky: Lexington, KY.

Charney, A.H. and C.A. Taylor, 1984. "Decomposition of Ex Ante State Model Forecasting Errors," *Journal of Regional Science* 24, 229-47.

Clements, M.P. and D.F. Hendry, 1998. *Forecasting Economic Time Series.* Cambridge University Press: Cambridge, UK.

Conway, R.S., 1990. "The Washington Projection and Simulation Model: A Regional Interindustry Econometric Model," *International Regional Science Review* 13, 141-65.

Coomes, P.A., 1992. "A Kalman Filter Formulation for Noisy Regional Job Data," *International Journal of Forecasting* 7, 473-81.

Engelbardt, F., 1925. *Forecasting School Population.* Teacher's College, Columbia University: New York.

Fair, R.C. and R.J. Shiller, 1989. "The Informational Content of Ex Ante Forecasts," *Review of Economics and Statistics* 71, 325-31.

Fair, R.C. and R.J. Shiller, 1990. "Comparing Information in Forecasts from Econometric Models," *American Economic Review* 90, 375-89.

Feenberg, D.R., W. Gentry, D. Gilroy, and H.S. Rosen, 1989. "Testing the Rationality of State Revenue Forecasts," *Review of Economics and Statistics* 42, 429-40.

Fullerton, T.M., Jr. and C.T. West, 1998. "Regional Econometric Housing Start Forecast Accuracy in Florida," *Review of Regional Studies* 28, 15-42.

Fullerton, T.M., Jr., J.A. Luevano, and C.T. West, 2000. "Regional Single-Family Housing Start Forecast Accuracy," *Journal of Housing Research* 11, 109-20.

Fullerton, T.M., Jr., M.M. Laaksonen, and C.T. West, 2001. "Regional Multi-Family Housing Start Forecast Accuracy," *International Journal of Forecasting* 17, 171-80.

Gentry, W.M., 1989. "Do State Revenue Forecasters Utilize Available Information?" *National Tax Journal* 42, 429-39.

Goldstein, H.A. and A.M. Cruze, 1987. "An Evaluation of State Projections of Industry, Occupational Employment," *Monthly Labor Review* 110 (October), 29-38.

Granger, C.W.J. and P. Newbold, 1977. *Forecasting Economic Time Series*. Academic Press: New York.

Hammond, G.W., 2000. "Focus on Evaluating *West Virginia Economic Outlook Forecasts*," in *West Virginia Economic Outlook 2001*. Bureau of Business and Economic Research, College of Business and Economics, West Virginia University: Morgantown, WV.

Harvey, A.C., C.R. McKenzie, D.P.J. Blake, and M.J. Desai, 1983. "Irregular Data Revisions," in A. Zellner (ed.), *Applied Time Series Analysis of Economic Data*. U.S. Department of Commerce, Bureau of the Census: Washington, DC.

Howrey, E.P., 1978. "The Use of Preliminary Data in Econometric Forecasting," *The Review of Economics and Statistics* 60, 193-200.

Isserman, A.M., 1995. "The Recovery of Rural Economies From Military Base Closures: Control Group Analysis of Two Decades of Experience," paper presented at the 1995 Annual Meetings of the Southern Regional Science Association.

James, W. and C. Stein, 1961. "Estimation with Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press: Berkeley, CA.

Kozlowski, P.J., 1987. "Regional Indexes of Leading Indicators: An Evaluation of Forecasting Performance," *Growth and Change* 18 (Summer), 62-73.

Larkey, P. and R. Smith, 1984. "The Misrepresentation of Information in Government Budgeting," in L. Sproull and P. Larkey (eds.), *Advances in Information Processing in Organizations* 1. JAI Press: Greenwich, CT.

Lenze, D.G., 1992. "How Reliable are Regional Employment Statistics?" *Economic Leaflets* 51(12). Bureau of Economic and Business Research, University of Florida: Gainesville, FL.

Lenze, D.G., 2000. "Forecast Accuracy and Efficiency: An Evaluation of Ex Ante Substate Long-Term Forecasts," *International Regional Science Review* 23, 201-26.

Lenze, D.G. and C.T. West, 2000. "Merging Demographic and Economic Population Projections in the Context of Regional Forecasting Models." Bureau of Economic and Business Research, University of Florida: Gainesville, FL.

McNees, S.K., 1978. "The Rationality of Economic Forecasts," *American Economic Review* 68, 301-05.

McNees, S.K., 1988. "How Accurate are Macroeconomic Forecasts?" *New England Economic Review* (July-August), 15-36.

McNees, S.K., 1992. "How Large are Economic Forecast Errors?" *New England Economic Review* (July-August), 25-33.

Mocan, H.N. and S. Azad, 1995. "Accuracy and Rationality of State General Fund Revenue Forecasts: Evidence from Panel Data," *International Journal of Forecasting* 11, 417-27.

Mork, K.A., 1987. "Ain't Behavin': Forecast Errors and Measurement Errors in Early GNP Estimates," *Journal of Business and Economic Statistics* 5, 165-75.

Nelson, C.R., 1972 . "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy," *American Economic Review* 52, 902-17.

Nelson, C.R., 1984. "A Benchmark for the Accuracy of Econometric Forecasts of GNP," *Business Economics* 19, 52-8.

Partridge, M.D. and D.S. Rickman, 2003. "Do We Know Economic Development When We See It?" *The Review of Regional Studies* 33(1), 17-39.

Regional Economic Analysis Division, 1993. "Tracking the BEA Regional Projections, 1988-91," *Survey of Current Business* (June), 28-41.

Siegel, J.S., 1953. "Forecasting the Population of Small Areas," *Land Economics* 29, 72-88.

Smith, S.K., J. Tayman, and D.A. Swanson, 2001. *State and Local Population Projections: Methodology and Analysis*. Kluwer Academic/Plenum Publishers: New York.

Stein, C., 1956. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press: Berkeley, CA, 197-206.

Taylor, C.A. and H. Theil, 1988. "Modeling the Accuracy of Certain Regional Predictions," *Regional Science and Urban Economics* 18, 453-62.

Theil, H., 1967. *Economics and Information Theory.* North Holland: Amsterdam.

Thomas, C. and M.M. Zandi, 2002. "Forecast Accuracy," in S.G. Cochrane (ed.), *Regional Financial Review* XIII(4). Economy.com Inc.: West Chester, PA, 26-36.

West, C.T., 1996. "System-Based Weights versus Series-Specific Weights in the Combination of Forecasts," *Journal of Forecasting* 15, 369-83.

West, C.T., 2003. "Structural Regional Factors that Determine Absolute and Relative Accuracy of U.S. Regional Labor Market Forecasts," *Journal of Agricultural and Applied Economics* 35 (Supplement), 121-35.

West, C.T. and T.M. Fullerton, Jr., 1996. "Assessing the Historical Accuracy of Regional Economic Forecasts," *Journal of Forecasting* 15, 19-36.

West, C.T. and H. Theil, 1991. "An Empirical Analysis of Stein Effects in Regional Forecasts," *Empirical Economics* 16, 363-73.

White, H.R., 1954. "Empirical Study of the Accuracy of Selected Methods of Projecting State Populations," *Journal of the American Statistical Association* 49, 480-98.