Journal of Economic and Social Measurement 34 (2009) 119–132
DOI 10.3233/JEM-2009-0318
IOS Press

An evaluation of real-time forecasting performance across 10 western U.S. States

Keith R. Phillips* and Joaquin Lopez
Federal Reserve Bank of Dallas, San Antonio, TX 78204, USA

The recent financial crisis and economic downturn has emphasized the importance of accurate sub-national forecasting models. To judge which models work best, researchers have emphasized the importance of looking at the true real-time performance of models and not simply an analysis of out-of-sample results. In this study we utilize real-time forecasts from the Western Blue Chip Economic Forecast to analyze and evaluate a host of different forecasters and models across time and 10 U.S. states to see if some models and forecasters consistently outperform others. We use the forecast accuracy criteria established by the Blue Chip publication. To evaluate accuracy we develop a scoring procedure based on the number of years that the forecaster/model was closest to actual relative to what we would expect just by random chance. We also utilize standard measures such as the Root Mean Square Error and Theil's inequality coefficient and test the statistical significance of the best forecasts. We then take a closer look at one model that has proven to be very accurate.

Keywords: Econometric models, regional data, benchmarking, and seasonal adjustment

1. Introduction

The recent financial crisis and its uneven impact across U.S. states highlight the importance of having accurate regional forecasting models. For example, many states experienced jobs loses in the first half of 2008 while other states, such as Texas, continued to grow. For businesses in Texas it was very important to know if Texas was immune to the national recession or if changing conditions for energy prices and exports and the steepening of the global credit crisis in the second half of 2008 might drag Texas into recession in late 2008 or early 2009. The likely degree of decline across the states in 2009 is also important for the regional allocation of federal stimulus funds if those funds are meant to stem the decline in jobs.

Since the late 1980s the *Western Blue Chip Economic Forecast* (WBCF) has been publishing forecasts from a host of different forecasters for 10 Western U.S. states. We examine this real-time forecast data to see if there are any models or modelers that persistently perform better than others. If data were widely distributed and free

*Corresponding author: Joaquin Lopez was an economic analyst at the Federal Reserve Bank of Dallas when this article was written but currently is a Ph.D. student at the University of Chicago. Tel.: +1 210 978 1409; E-mail: keith.r.phillips@dal.frb.org. The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas or the Federal Reserve System.

and there were no benefit from using any given model or market experience then for any individual forecaster there would be little persistence in relative forecast accuracy over time and average forecast accuracy across forecasters in each state would be similar. Our results show that for most states there are forecasters that persistently and on average do much better than others. Based on an informal survey of forecasters it appears that many of the most accurate forecasters use a formal statistical model rather than pure judgment.

Based on our results, one of the most accurate models has been a model that the Dallas Fed has used to forecast job growth in Texas. The Dallas Fed Texas regional forecasting model, which is a simple transfer function used with a Texas leading index, has been closest to the actual in forecasting Texas job growth in eight out of the 14 years in the study, out of an average of 7.4 forecasters. The model also has the lowest RMSE and Theil's U for the five Texas forecasters that were in the panel for the entire 14-year period. Since we were able to obtain detailed information on the model and how it was used in real-time, we highlight the model and give some likely reasons for its level of accuracy.

2. Methodology

One issue that arises in forecast evaluations is what forecast horizon and variable to focus on. Many times a variety of forecast horizons and variables are chosen with a wide variety of differing results. In the case of the WBCF however, the criteria for judgment is set forth by the publication. Once a year, usually May, June or July, the state employment growth forecast published in January of the previous year is compared to the actual job growth for the previous year. Forecasters who are closest to the actual are then recognized. Since forecasters know in advance that they will be judged by this criterion it is expected that they will put the most effort into reaching this objective. While forecasts are published 10 times per year, many times forecasters do not change their initial forecast published in January even as data become available throughout the year. This is likely due to the forecasters focusing on the criteria which they will be judged. Because of this we focus our evaluation on the employment forecast published in January of the forecast year and use the data present at the time of the evaluation issue (May–July of the following year) to evaluate the accuracy of the forecasts.

The simplest criterion to determine accuracy of a given model/forecaster over time is to count the number of times its forecast was closest to the actual. This is how the WBCF determines the most accurate forecasters for a given year. However, across years this could not be a very precise indicator because the margin of error can vary from year to year.

If a model consistently outperforms others over time, one would expect that model's accuracy to be robust to different evaluation methods. Hence, we utilize three criteria to evaluate the performance of a sample of forecasters for the 10 U.S. states in the

WBCF. Because we needed a consistent sample that went as far back as possible, some of the current forecasters are not included. Given this constraint, the numbers of forecasters as well as the number of years available for each state are not always the same. The evaluation measures utilized are: Root Mean Square Error (RMSE), Theil's inequality coefficient (U), and a score that was built based on the probability of being the closest to the actual in the same number of years by random chance.

3. Forecast accuracy measures

One of the most common measures of forecast accuracy is the RMSE. The main shortcoming of using this measure as a means to evaluate forecasts is that it is only useful in a relative sense, that is, it is not reflective of how accurate a single model is because it does not have an upper bound.

A more useful measure to evaluate the predictive accuracy of a model is Theil's inequality coefficient (Pindyck and Rubinfeld [6]), which measures the root mean square error in relative terms, and is defined as

$$U = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t^s - Y_t^a)^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t^s)^2} + \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t^a)^2}}$$

Where Y_t^s is the forecasted value of the series, Y_t^a is the actual value and $n$is the number of periods of the forecast. The denominator imposes an upper bound to the U coefficient, which is bounded above by 1 and bounded below by 0, that is, $0 \leqslant \mathrm{U} \leqslant 1$. This is particularly useful since it gives a threshold to evaluate the accuracy of a model and not only compare it to other models. The closer to 0 the coefficient is, the more accurate the model is, while a coefficient equal to 1 indicates that the forecast performance of the model is as bad as it could be. The U coefficient can be decomposed into three proportions that provide useful additional information on the performance of the model.

Bias,

$$U^M = \frac{(\overline{Y}^s - \overline{Y}^a)^2}{\frac{1}{n}\sum_{t=1}^{n}(Y_t^s - Y_t^a)^2}$$

Variance,

$$U^S = \frac{(\sigma_s - \sigma_a)^2}{\frac{1}{n}\sum_{t=1}^{n}(Y_t^s - Y_t^a)^2}$$

Covariance,

$$U^C = \frac{2(1-\rho)\sigma_s\sigma_a}{\frac{1}{n}\sum_{t=1}^{n}(Y_t^s - Y_t^a)^2}$$

The bias proportion measures the systematic error of the forecast; it gathers the share of the simulation error that comes from bias, that is, the difference between the averages of the forecasted series and the actual series. The variance proportion is intended to provide a measure of how well our forecast replicates the volatility of the actual series. The covariance proportion offers a measure of the unsystematic error in the forecast. The ideal distribution of proportions for any non-zero inequality coefficient would be $U^M = U^S = 0$, and $U^C = 1$. The results for these proportions are also included.

While forecast accuracy is most often measured as the average or relative size of the error over time, we also use a measure that is based on having the forecast that is closest to the actual most often. In the Western Blue Chip, forecasters that are closest to the actual get their name and affiliation listed in the evaluation issue. It is possible for a forecaster to have the lowest mean squared error over time but never to be listed as a most accurate forecaster. If one were to seek the notoriety of being the most accurate forecaster then one might alter his forecast up or down to increase his chances. For example, if all the other forecasters were expected to forecast growth in the range of 2.5 to 3.0 percent and a forecaster's model predicted 1.0 percent, than the forecaster, hoping to win, would send in a forecast of 2.4 percent so that any actual growth less than 2.5 would result in a win for the forecaster.

Since there is an incentive in some cases to alter the forecast to increase the chance of being closest to the actual, we also score the models based on a how often a forecaster is closest to the actual relative to how often one would expect to be closest to the actual just by random chance. This score was built on the probability that the performance of a given model occurred out of luck. By performance we mean the number of years that the model was closest to actual. This can be a tricky exercise because the number of forecasters (N) varies from year to year; hence the probability of getting it right just by chance (1/N) varies across years as well. One way to get around this is, for any given model, to take the probability of having performed exactly as it did just by chance

$$P = \prod_{k=1}^{K_i}\left(\frac{1}{N_i}\right)\prod_{j=1}^{J}\left(1 - \frac{1}{N_j}\right)$$

Where Krepresents the total number of years the model was closest to actual, while J denotes the number of years the model was not closest to actual. Multiply P times the number of possible combinations of having performed as it did (C) and let $X = 1 - PC$, then

$$S = \begin{cases} X & \text{if } K \geqslant \sum_{i=1}^{K+J} \left(\frac{1}{N_i}\right) \\ -X & \text{otherwise} \end{cases}$$

is the score assigned to the forecaster. The score is positive if the number of years the forecaster was closest to actual is greater than or equal to the expected number of years the forecaster should have been closest to actual just by chance. If the model performs worst than expected then it is assigned a negative score. Then, a score close to 1 indicates a strong performance, while a score close to -1 indicates a weak performance.

4. Comparing forecast accuracy

Following Diebold and Mariano [3] we implement the sign test, which is an exact finite sample test. Let d_t denote the squared forecast error differential between any two models at time t and let T be the forecast horizon, then the statistic

$$S_1 = \sum_{t=1}^{T} I_+(d_t)$$

follows a binomial distribution under the null hypothesis of zero-median squared forecast error differential, where

$$I_+(d_t) = \begin{cases} 1 \text{ if } d_t > 0 \\ 0 \text{ otherwise} \end{cases}$$

Assuming normality we can compare the following sign-test statistic against a standard normal:

$$S_2 = \frac{S_1 - .5T}{\sqrt{.25T}}$$

Results from this test indicate that only in three cases (Colorado, Oregon and Texas) did the forecaster/model with the best RMSE and U-statistic perform significantly better than the next best forecaster at the five percent level. In the case of Arizona, where a considerably larger sample of forecasters is available, the forecaster with the best RMSE performed significantly better than six out of eight of the other forecasters.

5. Results

The results for forecasters that consistently participated over a period of years are shown in Table 1. The probability scores are based on the entire sample of forecasters for each year not just for the forecasters shown. For many of the states it is difficult to draw conclusions since there are only two or three forecasters that consistently participated over at least a 10 year period. Arizona is an exception to this and, to a lesser degree, so is Texas. In general the results show that some forecasters consistently perform better than others. The results also indicate that in 7 out of 10 cases the models that have been closest to actual the most years are also those with the best RMSE and U coefficients.

In an effort to learn about which models worked the best, we sent e-mails asking forecasters what type of models were used and if they remained constant over time. The survey got only a small response although in general it appeared that the models with better RMSE and U coefficients are formal econometric models (ARIMA or structural), rather than purely judgmental.

As mentioned, however, clear winners from this exercise were the Dallas Fed and, to a lesser degree, the University of Arizona. Marshall Vest with the University of Arizona reported that they rent a General Equilibrium Style Model from Global Insight that they have used over the time period of this evaluation. Vest commented that a key to the models accuracy is that they have the best estimate of what job growth has been over the past 12 months before they estimate their forecasts for the following year. To achieve this, they roll in early benchmark data from the state as it becomes available and for the estimates outside this benchmark, they estimate what the benchmark revisions will likely be.

While the forecasting model of the University of Arizona is proprietary, the model used at the Federal Reserve Bank of Dallas can be described in detail. One interesting thing is that the Dallas Fed also credits much of its forecasting accuracy to taking steps to improve the recent estimates of job growth through early benchmarking and proper seasonal adjustment. For the period outside the benchmark, the Dallas Fed uses a simple estimate of cyclical bias to estimate the benchmark revision.

6. Dallas fed model

The data series that is forecast in the Western Blue Chip is non-farm payroll employment from the Current Establishment Survey (CES) program, produced by the Texas Employment Commission in cooperation with the Bureau of Labor Statistics. One issue with the state employment data, first discovered by Berger and Phillips [1], is that the series are actually two different series spliced together and these two series have different seasonal patterns. The bulk of the data is based on reports filed by firms covered by unemployment insurance (UI), while the most recent ten to twenty-two months of data are based on a survey of business establishments. Running a standard

Table 1

State forecast performance measures

	(Sign-test critical value = 1.96)					
Arizona	RMSE	U-Stat	U-bias	U-var	U-cov	Score
Arizona Public Service	1.659004	0.225048	0.142079	0.466099	0.391823	−0.59025
ASU	1.665261	0.229974	0.207557	0.466881	0.325562	−0.59025
Department of Economic Security	1.547781	0.207529	0.139778	0.38941	0.470812	0.615858
Eggert Economic Enterprises, Inc.	1.484911	0.194287	0.048002	0.700022	0.251976	−0.59025
Joint Legislative Budget Committee	1.500631	0.200201	0.1046	0.630817	0.264583	0.658541
The Maguire Company	1.525307	0.202758	0.080215	0.761577	0.158208	0.638455
NAU-BBER	1.508164	0.195402	0.029416	0.621069	0.349515	0.831279
Salt River Project	1.621982	0.214611	0.06245	0.621713	0.315837	0.840652
U of A	1.130498	0.141601	0.007617	0.416039	0.576344	0.992384
Best RMSE	1.13	U of A				
Best U-Stat	0.14	U of A				
Most Unbiased	0.01	U of A				
Best Variance	0.39	Department of Economic Security				
Best Score	0.99	U of A				
Most times closest to actual	4 out of 15	U of A				
Sign-test statistic	−1.80	U of A and Eggert Economic Enterprises, Inc.				
	−1.29	U of A and NAU-BBER				
	−2.32	U of A and Joint Legislative Budget Committee				
California	RMSE	U-Stat	U-bias	U-var	U-cov	Score
L.A. County Econ. Dev. Corp.	1.043165	0.254015	0.002814	0.255399	0.741787	0.544271
UCLA – Busn. Forecasting Project	0.848956	0.198306	0.003656	0.282961	0.713383	0.801553
Best RMSE	0.85	UCLA – Busn. Forecasting Project				
Best U-Stat	0.20	UCLA – Busn. Forecasting Project				
Most Unbiased	0.00	L.A. County Economic Development Corp				
Best Variance	0.26	L.A. County Economic Development Corp				
Best Score	0.80	UCLA – Business Forecasting Project				
Most times closest to actual	3 out of 15	UCLA – Business Forecasting Project				
Sign-test statistic	−0.26					

Table 1, continued

	(Sign-test critical value = 1.96)					
Colorado	RMSE	U-Stat	U-bias	U-var	U-cov	Score
Colorado Legislative Council	1.586912	0.298429	0.075853	0.640183	0.283964	0.987314
Office of State Planning and Budgeting	1.714247	0.326132	0.083678	0.564031	0.352291	−0.7783
Best RMSE	1.59	Colorado Legislative Council				
Best U-Stat	0.30	Colorado Legislative Council				
Most Unbiased	0.08	Colorado Legislative Council				
Best Variance	0.56	Office of State Planning and Budgeting				
Best Score	0.99	Colorado Legislative Council				
Most times closest to actual	4 out of 17	Colorado Legislative Council				
Sign-test statistic	−3.15					
Idaho						
Idaho Division of Financial Mgmt.	1.618109	0.28504	0.271468	0.354277	0.374255	0.962478
Idaho State Univ.	1.789491	0.304883	0.123037	0.327118	0.549845	−0.7544
Best RMSE	1.62	Idaho Division of Financial Mgmt.				
Best U-Stat	0.24	Idaho Division of Financial Mgmt.				
Most Unbiased	0.01	Idaho State Univ.				
Best Variance	0.59	Idaho State Univ.				
Best Score	0.97	Idaho Division of Financial Mgmt.				
Most times closest to actual	4 out of 13	Idaho Division of Financial Mgmt.				
Sign-test statistic	−0.28					
Nevada	RMSE	U-Stat	U-bias	U-var	U-cov	Score
Legis. Council Bureau	2.901057	0.291349	0.16358	0.288389	0.548031	0.742611
Southwest Gas	2.814569	0.266427	0.050812	0.185842	0.763346	−0.91757
Univ. of Nevada at Las Vegas – CBER	2.542553	0.253638	0.217015	0.245588	0.537398	0.909172
Best RMSE	2.54	Univ. of Nevada at Las Vegas – CBER				
Best U-Stat	0.25	Univ. of Nevada at Las Vegas – CBER				
Most Unbiased	0.05	Southwest Gas				
Best Variance	0.19	Legis. Council Bureau				
Best Score	0.91	Legis. Council Bureau				
Most times closest to actual	5 out of 18	Univ. of Nevada at Las Vegas – CBER				
Sign-test statistics	−1.41	Legis. Council Bureau and UNLV-CBER				

Table 1, continued

	(Sign-test critical value = 1.96)					
New Mexico	RMSE	U-Stat	U-bias	U-var	U-cov	Score
NMSU – CEMAF	1.029166	0.2149	0.031597	0.264848	0.703555	−0.694
University of New Mexico – BBER	0.805792	0.169223	0.108395	0.22701	0.664595	0.87148
Best RMSE	0.81	University of New Mexico – BBER				
Best U-Stat	0.17	University of New Mexico – BBER				
Most Unbiased	0.03	NMSU – CEMAF				
Best Variance	0.23	University of New Mexico – BBER				
Best Score	0.87	University of New Mexico – BBER				
Most times closest to actual	4 out of 17	University of New Mexico – BBER				
Sign-test statistic	−1.41					
Oregon	RMSE	U-Stat	U-bias	U-var	U-cov	Score
Oregon Executive Dept.	1.287956	0.278715	0.00761	0.554828	0.437562	−0.72624
US Bancorp	1.154992	0.241027	0.007408	0.809935	0.182657	0.670472
Best RMSE	1.15	US Bancorp				
Best U-Stat	0.24	US Bancorp				
Most Unbiased	0.01	Oregon Executive Department				
Best Variance	0.55	Oregon Executive Department				
Best Score	0.67	US Bancorp				
Most times closest to actual	5 out of 17	US Bancorp				
Sign-test statistic	−2.35					
Texas	RMSE	U-Stat	U-bias	U-var	U-cov	Score
Econoclast	1.597366	0.346879	0.00196	0.785263	0.212777	−0.71738
Dallas Fed	0.96348	0.203512	0.00653	0.717136	0.276334	0.999989
Ed McClelland	1.582044	0.341794	0.001272	0.709461	0.289267	−0.71738
Perryman group	1.416084	0.294837	0.005801	0.779916	0.214282	0.642796
TX State Comptroller of Public Accounts	1.393969	0.307381	0.018999	0.646105	0.334896	0.939858
Best RMSE	0.96	Dallas Fed				
Best U-Stat	0.20	Dallas Fed				
Most Unbiased	0.00	Ed McClelland				

Table 1, continued

	(Sign-test critical value = 1.96)					
Best Variance	0.65	TX State Comptroller of Public Accounts				
Best Score	1.00	Dallas Fed				
Most times closest to actual	8 out of 14	Dallas Fed				
Sign-test statistic	−2.67	Dallas Fed and Texas State Comptroller				
	−2.67	Dallas Fed and Perryman Group				
Utah	RMSE	U-Stat	U-bias	U-var	U-cov	Score
Thredgold Economic Associates	1.472128	0.25262	0.01397	0.48023	0.5058	0.957523
Utah State Tax Commission	1.525503	0.257834	0.023529	0.357922	0.618548	−0.74716
Best RMSE	1.47	Thredgold Economic Associates				
Best U-Stat	0.25	Thredgold Economic Associates				
Most Unbiased	0.01	Thredgold Economic Associates				
Best Variance	0.36	Utah State Tax Commission				
Best Score	0.96	Thredgold Economic Associates				
Most times closest to actual	4 out of 10	Thredgold Economic Associates				
Sign-test statistic	−1.89					
Washington	RMSE	U-Stat	U-bias	U-var	U-cov	Score
Dick Conway & Associates	0.980383	0.242142	0.025459	0.172216	0.802325	0.94877
Doug Pedersen & Associates	0.926981	0.208117	0.000499	0.251262	0.748239	0.795081
Office of Forecast Council	0.979654	0.251763	0.093331	0.239252	0.667416	−0.93122
Best RMSE	0.93	Doug Pedersen & Associates				
Best U-Stat	0.21	Doug Pedersen & Associates				
Most Unbiased	0.00	Doug Pedersen & Associates				
Best Variance	0.17	Dick Conway & Associates				
Best Score	0.95	Dick Conway & Associates				
Most times closest to actual	5 out of 14	Dick Conway & Associates				
Sign-test statistic	0					

census X-11 or X-12 seasonal adjustment procedure on the combined CES data series results in seasonal factors which are essentially based on the UI data. When these seasonal factors are applied to the establishment survey data at the end of the series, it often results in a January jump and other irregularities that are revised away when the data are rebenchmarked to the UI data once a year with the release of the January data.

The data used by the Dallas Fed to forecast Texas job growth is seasonally adjusted using a two-step seasonal adjustment process described in Berger and Phillips [1, 2], that estimates and applies two separate seasonal adjustment factors for the two separate parts of the data. This process helps eliminate erroneous seasonal factors that might lead to poor forecasts for the following year.

Another adjustment that BP makes to the employment data is early benchmarking. Once a year, concurrent with the release of the January CES data, the BLS revises the previously estimated data based on a years worth of UI data, a process called benchmarking. The benchmark period covers from July two-years-prior to June of the previous year. Preliminary UI data for Texas at the three-digit North American Industrial Classification System (NAICS) are available with about a three-quarter lag after the reporting quarter. Berger and Phillips [1] show that this preliminary data is very close to the final data used for the annual benchmark and thus can accurately be used to estimate the benchmark revision.

Finally for the data that does not incorporate an official benchmark or an early benchmark, the Dallas Fed estimates a simple bias adjustment factor based on past observations. Past data for Texas employment shows that the benchmark revision is usually negative in recessions and positive in expansions. For example, the data from the start of the most recent expansion in Texas, mid-2003 through mid-2007, every quarter of employment data was upwardly revised due to the benchmark revision with an average revision of 1.4 percentage points on an annual basis. When the Dallas Fed forecast model was run in December of 2007, it incorporated a 1.4 percent bias adjustment to the annualized growth in the out-of -benchmark data from July to November of 2007.

Using this employment data, that incorporates both an early benchmark and the correct seasonal adjustment, sharply reduces later revisions and thus more accurately represents the data for the current year.

The structure of the Dallas Fed model is a transfer function model that utilizes the autoregressive, moving-average process in the employment data along with a Texas Leading Index derived by Phillips [4]. The Texas Leading index is similar in methodology and construction as the U.S. Leading Index produced by the Conference Board. Phillips [5] discusses how this methodology has been useful in predicting recessions in the U.S. economy.

The first step in the transfer function is to specify the ARIMA model that best fits the employment data. The best fit was determined to be a (1,1,1) model. The next step is to pre-filter both the log employment series and the log Texas Leading Index by this (1,1,1) process and to analyze the cross-correlation matrix to look for statistically

Table 2
Texas forecasting model, measures of fit and analysis of residuals

		Conditional Least Squares Estimation					
Parameter	Estimate	Std. Error	t	Pr > $\lvert t \rvert$	Lag	Variable	Shift
MU	0.0011558	0.0002351	4.92	<0.0001	0	lntxnag	0
MA1,1	0.85236	0.10149	8.40	<0.0001	1	lntxnag	0
AR1,1	0.92130	0.07406	12.44	<0.0001	1	lntxnag	0
NUM1	0.06117	0.0081604	7.50	<0.0001	0	lnleadi	3
DEN1,1	0.87980	0.01880	46.79	<0.0001	1	lnleadi	3
Constant Estimate		0.000091					
Variance Estimate		4.29E-6					
Std Error Estimate		0.002071					
AIC[2]		−2927.59					
SBC[3]		−2908.94					
Number of Residuals		308					

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	12.99	4	0.0113	−0.098	−0.035	0.131	0.041	−0.059	0.090
12	13.26	10	0.2098	−0.007	0.009	0.012	0.014	−0.007	0.018
18	23.05	16	0.1123	−0.006	−0.019	−0.029	−0.018	0.112	−0.125

Crosscorrelation Check of Residuals with Input lnleadi

To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	2.92	5	0.7122	−0.016	0.002	−0.011	−0.057	0.018	0.075
11	16.42	11	0.1261	0.049	0.144	−0.073	0.115	−0.048	0.022
17	20.36	17	0.2561	−0.011	−0.076	0.077	−0.019	0.028	0.006
23	23.56	23	0.4286	0.009	0.001	0.077	−0.065	0.015	0.006

[2] AIC does not include log determinant.
[3] SBC does not include log determinant.

significant relationships between changes in the leading index and changes in Texas employment. The prefiltering of the data ensures that the relationships shown in the cross-correlation matrix are not spurious due to the two data series following the same autoregressive process. In analyzing the cross-correlation matrix there was a significant coefficient at a 3-month lead. Thus there is a delay of three months between changes in the leading index and changes in Texas employment. There were also some significant coefficients after the three month lead that tended to die off. This relationship of changes in the leading index (X_t) to changes in employment can be shown as the following where B is the backshift operator:

$$e_t = \omega_0 \left(1 + \sum_{i=n}^{\infty} \delta_1^i B^i\right) X_t$$

Where e_t is the first difference of the natural log of Texas employment and X_t is the first difference of the natural log of the Texas Leading Index. Notice that in the transfer function that shocks to the leading index die out over time depending on the closeness of δ_1^i to 1. In the Texas model δ_1^i is estimated to be 0.88 so that changes in the leading index have an impact on employment that die out slowly and thus the forecast tends to be rather smooth. The delay in the model is three months so $n = 3$. The results from the most recent run of the model are shown in Table 2. As shown here, the T-statistics are quite high, particularly for the lagged impact of changes in the leading index on changes in employment. Also shown are the autocorrelation check of the models residuals and the cross-correlation of the residuals with changes in the leading index. In this particular run of the model, there is some evidence of some autocorrelation in the model that may indicate that the autoregressive part of the model may need some adjustment. The cross-correlation results show that the predictive content of the leading index is contained in the model so that the error terms have no statistical relationship with the leading index.

7. Conclusion

An analysis of the real-time job growth forecasts across 10 major states in the U.S. reveals that some forecasters consistently outperform others. Generally it seems that formal econometric models outperform judgment. One model that has performed well, the Dallas Fed Texas model, is a simple transfer function model which utilizes the autoregressive movements in employment and changes in the Texas Leading Index, an index which was designed to predict turning points in the economy. While the simplicity of the model likely contributes to its relatively strong performance, the model is also helped by using employment data that is early benchmarked and has appropriate seasonal factors. The University of Arizona, which also performed well, also gave much credit to using early benchmarking and estimating out-of-benchmark revisions to the existing data.

Acknowledgements

The two-step seasonal adjustment and early benchmarking of the Texas employment data is performed by Frank Berger at the Federal Reserve Bank of Dallas using SAS version 9.2. This data along with the Texas Leading Index is updated monthly and available on the Dallas Fed web page at http://dallasfed.org/data/index.html. The Dallas Fed forecasting model for Texas is run using SAS version 9.2. The data for the state forecasts were gathered mostly from hard-copy issues of the *Western Blue Chip Economic Forecast.* The statistical calculations in Table 1 were performed in Excel 2007 and E-views version 5.1.

References

[1] F.D. Berger and K. Phillips, *Reassessing Texas Employment Growth*, The Southwest Economy, Federal Reserve Bank of Dallas, July/August, 1993.

[2] F.D. Berger and K. Phillips, Solving the Mystery of the Disappearing January Blip in State Employment Data, *Economic Review*, Federal Reserve Bank of Dallas, Second Quarter, 1994.
[3] F.X. Diebold and S.M. Roberto, Comparing Predictive Accuracy, *Journal of Business and Economics Statistics* **13**(3) (July 1995), 253–263.
[4] K.R. Phillips, The Texas Index of Leading Economic Indicators: A Revision and Further Evaluation, *Economic Review*, Federal Reserve Bank of Dallas, July 1990).
[5] K.R. Phillips, The Composite Index of Leading Economic Indicators: A Comparison of Approaches, *Journal of Economic and Social Measurement* **25**(3,4), Special Issue on Regional Economic Indicator Models, (1998/1999).
[6] R.S. Pindyck and L.R. Daniel, *Econometric Models and Economic Forecasts*, (4th ed.), McGraw-Hill, 1998.