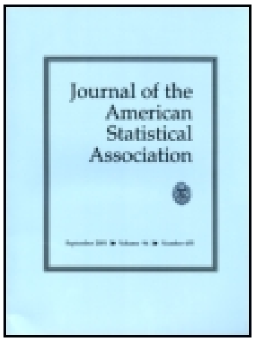

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Forecasting the U.S. Unemployment Rate

Alan L. Montgomery , Victor Zarnowitz , Ruey S. Tsay & George C. Tiao

Forecasting the U.S. Unemployment Rate

Alan L. Montgomery, Victor Zarnowitz, Ruey S. Tsay, and George C. Tiao

This article presents a comparison of forecasting performance for a variety of linear and nonlinear time series models using the U.S. unemployment rate. Our main emphases are on measuring forecasting performance during economic expansions and contractions by exploiting the asymmetric cyclical behavior of unemployment numbers, on building vector models that incorporate initial jobless claims as a leading indicator, and on utilizing additional information provided by the monthly rate for forecasting the quarterly rate. Comparisons are also made with the consensus forecasts from the Survey of Professional Forecasters. In addition, the forecasts of nonlinear models are combined with the consensus forecasts. The results show that significant improvements in forecasting accuracy can be obtained over existing methods.

KEY WORDS: Autoregressive integrated moving average models; Forecast comparisons; Markov switching autoregressive models; Nonlinearity; Threshold autoregressive models.

1. INTRODUCTION AND SUMMARY

A primary application of many econometric models is forecasting, although often a thorough study of the forecasting properties of these models is neglected. This article presents an in-depth study of the forecasts for the quarterly U.S. unemployment rate using various econometric and time series models. Our main emphases are on measuring forecasting performance during the rise and decline of unemployment, on building vector models that incorporate initial jobless claims as a leading indicator, and on utilizing additional information provided by the monthly rate for forecasting the quarterly rate. Comparisons between these forecasting methods are made to further our understanding of the strengths and deficiencies of these methods. These comparisons provide several insights about using the various methods in forecasting the unemployment rate that can be exploited by future researchers. In particular, we find that nonlinear models improve the forecastability of the unemployment series during economic contractions.

In this article we are concerned exclusively with the behavior over time of a single most important measure in the study of employment: the U.S. civilian unemployment rate, seasonally adjusted (u). The quarterly u series is the average of the reported monthly rates and is plotted in Figure 1 for 1948–1993. We base much of our analysis on the quarterly data but also demonstrate that the monthly series provides valuable additional information for purposes of short-term forecasting (see Sec. 3.3). Of course, the quarterly series is smoother than the monthly series, but the two show the same cyclical and trend features. The U.S. unemployment rate displays steep increases that end in sharp peaks and alternates with much more gradual and longer declines that end in mild troughs. Such cyclical asymmetries have long been noted and much debated (De Long and Summers 1986; Elwood 1996; Mitchell 1927; Neftçi 1984; Rothman 1996; Sichel 1989; Zarnowitz 1992, chap. 8).

The shaded areas in Figure 1 denote the business cycle contractions that run from peak to trough as dated by the National Bureau of Economic Research (NBER). The NBER uses many sources of information to determine business cycles, including total output, income, employment, and trade (Moore 1983). Evidently the unemployment rate has a strong tendency to move countercyclically, upward in general business slowdowns and contractions and downward in speedups and expansions. The cyclical troughs in u lead by short intervals at—or occasionally coincide with—the peaks P that mark the beginning of each business contraction (or recession, to use a less technical term). The cyclical peaks in u coincide with or lag behind the troughs T that mark the beginning of each business expansion (or recovery).

These asymmetries in the unemployment rate are important in our analysis for two reasons. First, univariate linear models are not able to accurately represent these asymmetric cycles. Thus we would expect that the greatest contribution of recently developed nonlinear models will be to help forecast during these contrasting cyclical phases. Second, economic contractions result in much economic hardship for the unemployed, leading to social and political problems. Thus general interest in forecasting unemployment will be greater during contractionary periods, which makes it particularly important to know the duration and turning points of these cycles. This suggests that any forecasting gains that can be made during contractionary periods will be valued more than those made during expansionary periods.

The data used cover the period 1948–1993. The forecasts are generated from prediction methods using various models with a rolling estimator, beginning with the third quarter of 1968 as the initial origin of forecast and ending with the third quarter of 1993 as the final origin. Models considered include linear univariate autoregressive integrated moving

Alan L. Montgomery is Assistant Professor of Marketing at the Wharton School, University of Pennsylvania, Philadelphia, PA 19104 (E-mail: alm@wharton.upenn.edu). Victor Zarnowitz is Professor Emeritus of Economics and Finance, Graduate School of Business, University of Chicago, Chicago, IL, 60637 and Director of Research at the Foundation for International Business and Economic Research (FIBER), New York, NY 10168 (E-mail: fiber@ziplink.net). Ruey S. Tsay is Professor of Econometrics and Statistics (E-mail: rst@gsbrst.uchicago.edu) and George C. Tiao is the W. Allen Wallis Professor of Econometrics and Statistics, Graduate School of Business, University of Chicago, Chicago, IL 60637 (E-mail: gct@gsbgct.uchicago.edu). The authors thank the editor and referees for helpful comments.

Journal of the American Statistical Association
June 1998, Vol. 93, No. 442, Applications and Case Studies

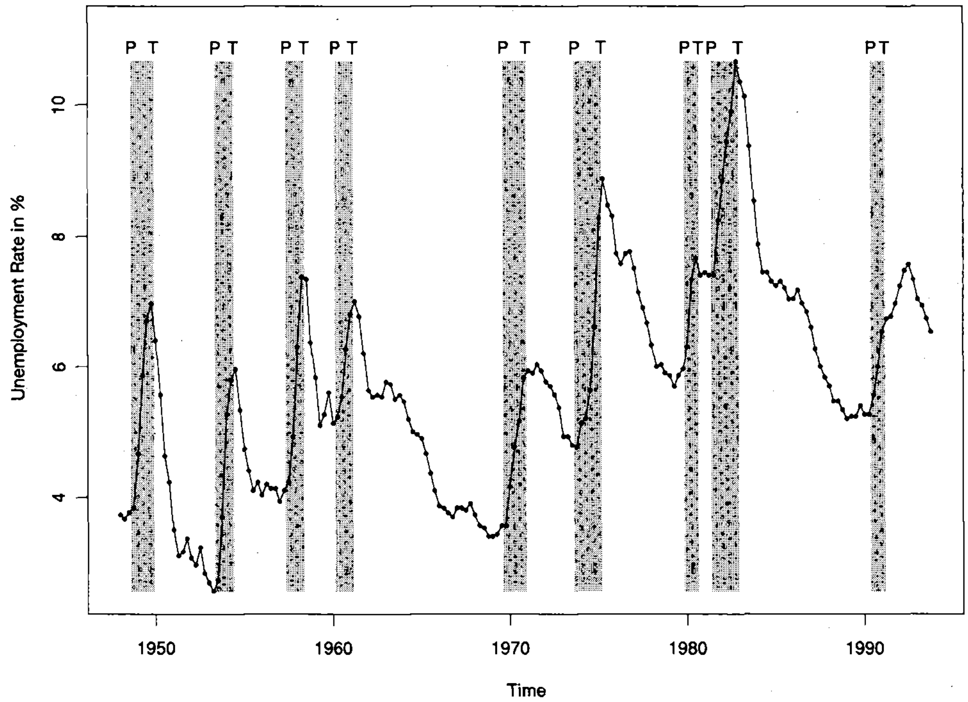

Figure 1. Quarterly U.S. Unemployment Rate. ▭, *Business contractions;* P, *U.S. business cycle peak;* T, *U.S. business cycle trough.*

average (ARIMA) models, bivariate vector autoregressive moving average (VARMA) models, threshold autoregressive (TAR) models, and Markov switching autoregressive (MSA) models. Comparisons are also made with the consensus group median forecasts of the Survey of Professional Forecasters (SPF), which are collected on a quarterly basis. This survey was originally conducted by the American Statistical Association (ASA) and NBER, although since 1990 it has been conducted by the Federal Reserve Bank of Philadelphia (Croushore 1993).

The outline of this article is as follows. Section 2 sketches the development of statistical models for forecasting the quarterly unemployment rate. Section 3 compares the out-of-sample performance of one- to five-quarter–ahead forecasts constructed from these models. All comparisons are made in terms of the ratio of mean squared errors (MSE) with respect to a seasonal univariate ARIMA model for quarterly unemployment rate as the benchmark. Additionally, a new combined forecast based on the consensus forecasts and those from the quarterly TAR model is presented. Sections 4 and 5 conclude the paper with a discussion of the implications of our findings for modeling the unemployment rate.

Our principal empirical findings can be summarized as follows:

a. Reductions in MSE are present when the forecasts occur during periods of economic contractions—rapidly rising unemployment; for example, a TAR model yields up to a 28% reduction in MSE for longer term forecasts (see Table 2, line 3). Thus nonlinear models can significantly improve the forecasting performance during certain periods. Although the TAR model yields an adequate nonlinear model, it remains to be determined whether this is the best nonlinear model.
b. For models based on monthly data, but using the same forecast origins as the quarterly models, the performance compares favorably (as expected) with the benchmark and with their quarterly counterparts for one- or two-quarter–ahead forecasts, whereas the gain becomes much smaller as the forecasting horizon is increased (see Table 4). The short-term performance can be improved considerably if the origin is advanced 1 month into the quarter after the forecasting origin. Assuming that the first month of the quarter after the forecast origin is known, the monthly models show a reduction in MSE of more than 75% for one-quarter–ahead forecasts and more than a 40% reduction for two-quarter–ahead forecasts.
c. Compared to the consensus SPF forecasts, the performance of all of the quarterly and monthly models is inferior when the forecasts are generated from the same origin. When the origin is advanced 1 month into the quarter, the monthly models and the consensus forecasts become more comparable, indicating that some of the improvement in the short-term consensus forecasts may be due to some forecasters knowing the first month.
d. When the consensus SPF forecasts are combined with simple univariate TAR models, significant gains in forecasting the unemployment result. These new combined forecasts show a 25% reduction in overall MSE for five-step–ahead forecasts, and during periods of rapidly increasing unemployment we find a 35% reduction in MSE. This demonstrates that nonlinearity is not fully exploited by forecasting methods currently

in use and suggests that future research in this area may be warranted.

e. All models are misspecified, as there exists no true model for the U.S. unemployment rate. Our study shows that different model approximations may be useful, depending on the conditioning information set used when constructing forecasts.

2. METHODS AND MODELS FOR FORECASTING

In this section we briefly discuss statistical methods and models used in our analysis of the U.S. unemployment rate data. For each model used, we report parameter estimates based on all available data; however, in our forecasting comparisons, we use out-of-sample forecasts and re-estimate each model at every forecast origin. We begin in Section 2.1 by building a univariate linear time series model for the quarterly unemployment rate u_t which serves as a benchmark for forecasting comparisons. In Section 2.2 we introduce multivariate linear models that provide a framework for incorporating leading indicators into our forecasts. In Sections 2.3 and 2.4 we consider nonlinear adaptations of the autoregressive model. The first is the threshold autoregressive model (TAR); the second, a Markov switching autoregression (MSA). Finally, in Section 2.5 we consider a final departure from the benchmark model by considering the use of monthly data to improve the forecasting accuracy of the quarterly series.

2.1 Univariate Linear Models

A commonly used statistical model in linear time series analysis is the ARIMA model (Box, Jenkins, and Reinsel 1994). For the U.S. quarterly unemployment rate series u_t, the model can be written as

$$(1-\phi_1 L-\cdots-\phi_p L^p)(1-L)^d u_t = c+(1-\theta_1 L-\cdots-\theta_q L^q)\varepsilon_t \quad (1)$$

where p, d, and q are nonnegative integers; c, ϕ_i, and θ_j are parameters; L is the lag operator such that $Lu_t = u_{t-1}$; and $\{\varepsilon_t\}$ is a sequence of iid random variables with mean 0 and variance σ_ε^2. We assume that the two polynomials $\phi(L) = 1-\phi_1 L-\cdots-\phi_p L^p$ and $\theta(L) = 1-\theta_1 L-\cdots-\theta_q L^q$ have no common factors and have all their 0s outside the unit circle. When $d = 0, u_t$ is weakly stationary, and when $d > 0$, it is unit-root nonstationary.

Empirical Result. To construct an ARIMA model, we follow the three-stage iterative procedure of Box, Jenkins, and Reinsel (1994) consisting of model specification (Tsay and Tiao 1984), estimation (Ansley 1979; Hillmer and Tiao 1979), and diagnostic checking (Ljung and Box 1978). Using the U.S. quarterly unemployment rate series u_t for 1948–1993, we obtain an ARIMA(1, 1, 0) model, which is commonly used to model the unemployment rate (Sims and Todd 1991):

$$(1-\underset{(.06)}{.64}L)(1-L)u_t = \varepsilon_t, \qquad \hat{\sigma}_\varepsilon^2 = .109. \quad (2)$$

The standard error of the parameter estimate is printed underneath the estimate in parentheses. The Ljung–Box statistic for this fitted model is $Q(12) = 33.2$. Compared to the chi-squared distribution with 11 df the statistic indicates that some residual serial correlations are present.

In consideration of the seasonal adjustment process used to filter the unemployment series, we modify model (2) by including a multiplicative seasonal ARMA(4, 4) factor to

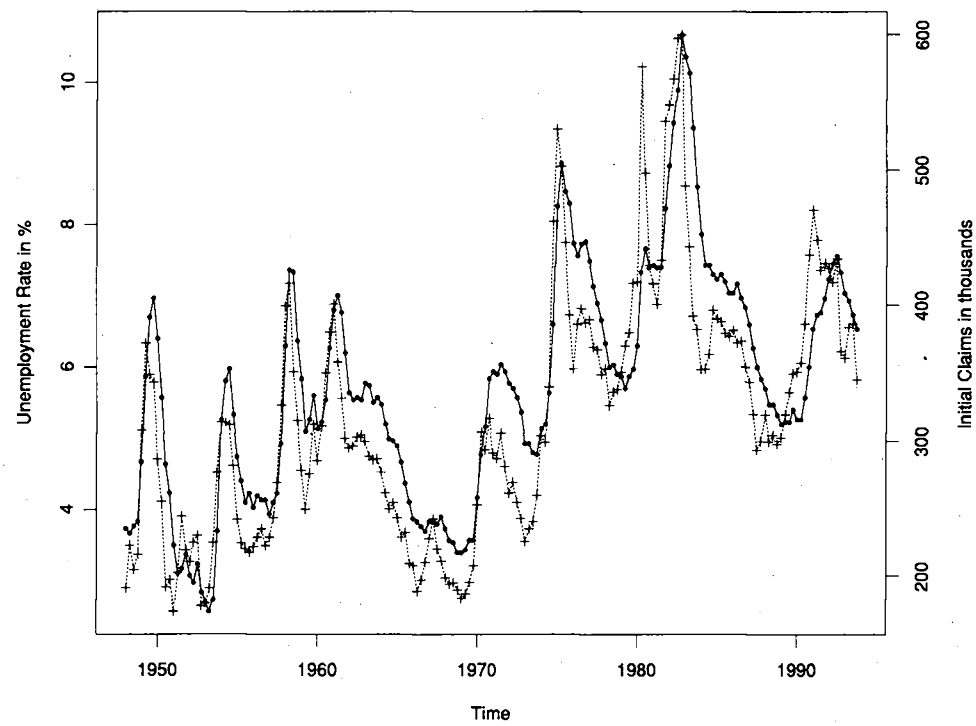

Figure 2. Quarterly U.S. Unemployment Rate and Initial Claims. ——, Unemployment rate; · · ·, initial claims.

account for any residual seasonal effects associated with the adjustment process:

$$\underset{(.11)}{(1-.31L^4)}\ \underset{(.06)}{(1-.65L)}\ (1-L)u_t = \underset{(.07)}{(1-.78L^4)}\varepsilon_t, \qquad \hat{\sigma}_\varepsilon^2 = .090. \quad (3)$$

The Ljung–Box statistic for this fitted model is $Q(12) = 9.8$, indicating that the model is adequate.

The unit-root nonstationarity of models (2) and (3) might be hard to justify for the U.S. unemployment rate, because existing historical data suggest that the rates are changing only within a limited range. However, we regard the ARIMA(1, 1, 0)(4, 0, 4) model in (3) as a statistical model for short-term forecasting, not a description of the long-term dynamic structure of the series.

2.2 Multivariate Linear Models

In many applications, related variables are available, and one would like to make use of all relevant information in forecasting. In predicting the U.S. unemployment rate, there exist data on initial jobless claims and other leading indicators. To utilize such information, we consider all the viable time series jointly and define $\mathbf{u}_t = (u_{1t}, \ldots, u_{kt})'$ to be a vector of k such series. The vector generalization of the ARIMA model, called the vector autoregressive moving average (VARMA) model, can be written as

$$(\mathbf{I} - \mathbf{\Phi}_1 L - \cdots - \mathbf{\Phi}_p L^p)\mathbf{u}_t = \mathbf{c} + (\mathbf{I} - \mathbf{\Theta}_1 L - \cdots - \mathbf{\Theta}_q L^q)\boldsymbol{\varepsilon}_t \quad (4)$$

where c is a constant vector, $\mathbf{\Phi}_i$ and $\mathbf{\Theta}_j$ and $k \times k$ matrices, I is the $k \times k$ identity matrix, and $\{\boldsymbol{\varepsilon}_t\} = \{(\varepsilon_{1t}, \ldots, \varepsilon_{kt})'\}$ is a sequence of iid random vectors with mean 0 and positive-definite covariance matrix $\mathbf{\Sigma}$. We assume that the two matrix polynomials $\mathbf{\Phi}(L)$ and $\mathbf{\Theta}(L)$, defined in a similar way as those of (1), are left-coprime and that all of the 0s of the determinants $|\mathbf{\Phi}(L)|$ and $|\mathbf{\Theta}(L)|$ are on or outside the unit circle. (For further properties of this model, see Hannan and Deistler 1988 and Tiao and Tsay 1989.)

The VARMA models in (4) are very flexible. They encompass many familiar models in the literature. Consider, for example, the case of $k = 2$. If all of the lower off-diagonal elements of $\mathbf{\Phi}_i$ and $\mathbf{\Theta}_j$ are 0, then the model reduces to a transfer function model (also referred to as a dynamic linear model) with u_{2t} and u_{1t} as its input and output variables. (See Tiao and Box 1981 for a further discussion of building VARMA models.)

Empirical Result. A possible leading indicator variable for the unemployment rate is the number of initial claims of unemployment. Let $\mathbf{u}_t = (u_{1t}, u_{2t})'$, where $u_{1t} = u_t$ is the quarterly unemployment rate and u_{2t} is the quarterly number of initial claims of unemployment (divided by 100). Figure 2 illustrates the dynamic relationship between initial claims and the unemployment rate. The initial claims provide an indication of whether unemployment is improving or worsening, and typically leads the unemployment rate by one or two quarters. Using the data for 1948–1993, we build the following VARMA model for $\mathbf{u}_t$:

$$(\mathbf{I} - \mathbf{\Phi}_4 L^4)(\mathbf{I} - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2)\mathbf{u}_t = \mathbf{c} + (\mathbf{I} - \mathbf{\Theta}_4 L^4)\boldsymbol{\varepsilon}_t, \qquad \operatorname{cov}(\boldsymbol{\varepsilon}_t) = \mathbf{\Sigma}, \quad (5)$$

where

$$\mathbf{\Phi}_1 = \begin{bmatrix} 1.29 & .61 \\ (.07) & (.09) \\ -.15 & 1.36 \\ (.08) & (.10) \end{bmatrix}, \quad \mathbf{\Phi}_2 = \begin{bmatrix} -.39 & -.51 \\ (.08) & (.10) \\ .13 & -.36 \\ (.07) & (.10) \end{bmatrix},$$

$$\mathbf{\Phi}_4 = \begin{bmatrix} .33 & .20 \\ (.10) & (.08) \\ -.27 & .60 \\ (.08) & (.09) \end{bmatrix}, \quad \mathbf{\Theta}_4 = \begin{bmatrix} .91 & -.14 \\ (.09) & (.11) \\ .24 & .43 \\ (.11) & (.14) \end{bmatrix},$$

$$\mathbf{c} = \begin{bmatrix} .05 \\ (.03) \\ .08 \\ (.05) \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} .069 & .054 \\ .054 & .081 \end{bmatrix}.$$

Note that the two lower triangular elements in the AR matrices $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are statistically insignificant at the 5% level based on their asymptotic distributions. This suggests that apart from seasonal relations, the initial claims series is indeed an input variable for the unemployment rate.

2.3 Threshold Autoregressive Models

We have seen in Figure 1 that the unemployment rate exhibits a pronounced asymmetric cyclical behavior; that is, it increases at a faster rate than it decreases. One nonlinear model that can be used to capture this behavior is the threshold autoregressive (TAR) model (Tong 1983). To introduce the TAR model, we consider a simple example. Define a time series x_t by

$$x_t = \begin{cases} -1.5x_{t-1} + \varepsilon_t & \text{if } x_{t-1} \le 0 \\ .6x_{t-1} + \varepsilon_t & \text{if } x_{t-1} > 0, \end{cases} \quad (6)$$

where ε_t is Gaussian noise with mean 0 and variance σ_ε^2. This is an ergodic and stationary series. For this model, the structural change is governed by the reference space of x_{t-1}, not by the time index t. If x_t denotes the growth rate of U.S. quarterly real gross national product, then model (6) simply says that the growth rate behaves differently between "expansions" (positive growth rate at $t-1$) and "contractions" (negative growth rate at $t-1$). Further, with the negative coefficient -1.5, the model is likely to produce a positive x_t when $x_{t-1} \le 0$. In contrast, the coefficient of .6 implies that a positive x_{t-1} tends to be followed by another positive x_t. The persistence of this autoregressive component tends to result in "expansions" that last for multiple periods. Thus this simple model captures the common knowledge that expansions tend to last longer than contractions.

More generally, the TAR model for a time series x_t is defined as

$$x_t = \phi_0^{(i)} + \phi_1^{(i)} x_{t-1} + \cdots + \phi_{p_i}^{(i)} x_{t-p_i} + \varepsilon_{it}, \quad \text{if } r_{i-1} < c_{t-d} \le r_i \quad \text{for} \quad i = 1, \ldots, R, \tag{7}$$

where d is a positive integer, r_i are real, p_i are nonnegative integers, R is the number of regimes, and $\{\varepsilon_{it}\}$ are independent Gaussian white noise series with mean 0 and variances σ_i^2. In (7), c_{t-d} is the threshold variable defining the reference space and in simple TAR models becomes $c_{t-d} = x_{t-d}$, d is a delay parameter, and r_i are the thresholds. Basically, the model in (7) describes changes in the dynamic structure of x_t at the thresholds r_i with respect to c_{t-d}.

Empirical Result. We estimate the following TAR model using quarterly data for 1948–1993:

$$\nabla u_t = \begin{cases} \underset{(.03)}{.01} + \underset{(.10)}{.73}\nabla u_{t-1} + \underset{(.12)}{.10}\nabla u_{t-2} + \varepsilon_{1t} & \text{if } \nabla u_{t-2} \le .1 \\ \underset{(.09)}{.18} + \underset{(.12)}{.80}\nabla u_{t-1} - \underset{(.16)}{.56}\nabla u_{t-2} + \varepsilon_{2t} & \text{otherwise,} \end{cases} \tag{8}$$

where $\nabla u_t = (1 - L)u_t = u_t - u_{t-1}$ and the sample variances of ε_{it} are .076 and .165 for $i = 1$ and 2. The two regimes of $c_{t-2} = u_{t-2} - u_{t-3}$ are denoted as $(c_{t-2} \le .1)$ and $(c_{t-2} > .1)$. In (8), the innovational series $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{2t}\}$ are assumed independent.

The specification of this model was done using the method proposed by Tsay (1989). The threshold of .1 is selected by the Akaike information criterion (AIC) and the Bayes information criterion (BIC). The coefficient estimates are posterior means obtained by the Gibbs sampler (Gelfand and Smith 1990; Geman and Geman 1984), and the associated posterior standard errors are given in parentheses. (For further discussion of implementating the Gibbs sampler for our problem, see McCulloch and Tsay 1994.) Estimates for this model may also be obtained using conditional least squares (Chan 1993). For this dataset, the estimates from the two methods are close.

For the model in (8), the threshold of .1 says that the change in the quarterly unemployment rate, ∇u_t, behaves as a piecewise linear model in the reference space of $u_{t-2} - u_{t-3}$. The change in the unemployment rate approximates the discriminant function for predicting whether an observation occurs during an economic contraction or an expansion. Intuitively, this model says that the dynamics of unemployment act differently depending on the magnitude of the recent change in the unemployment rate. In the first of our two regimes, the unemployment rate has had either a decrease or a minor increase. Here the economy should be stable, and essentially the change in rate follows a simple AR(1) model, because the lag 2 coefficient is insignificant. In the second regime, we could have a substantial jump in the unemployment rate. This typically corresponds to the contraction phase in the business cycle. It is also the period during which government interventions and industrial restructuring are likely to occur. Here the first difference of u_t follows an AR(2) model. The positive intercept indicates an upward trend. The AR(2) polynomial contains two complex characteristic roots, which indicate possible cyclical behavior in the change of u_t. Consequently, the chance of having a turning point in u_t increases, suggesting that the period of large increases in u_t should be short. This implies that the contraction phases in the U.S. economy tend to be shorter than the expansion phases. The behavior of our model shares some strong similarities with the one used by Tiao and Tsay (1994) in analyzing the series of quarterly real U.S. gross national product (GNP) in terms of both the structure of the model and the improvements in forecasting performance.

2.4 Markov Switching Autoregressive Models

An alternative model to capture the apparent asymmetrical behavior in the unemployment rate is a hidden Markov process that switches between two autoregressive models (Hamilton 1989, 1990). A time series x_t is said to follow a two-state Markov switching autoregressive (MSA) model when

$$x_t = \phi_0^{(s_t)} + \phi_1^{(s_t)} x_{t-1} + \cdots + \phi_p^{(s_t)} x_{t-p} + \varepsilon_{it}, \quad \text{if } s_t = i, \qquad i = 1, 2 \tag{9}$$

where p is a nonnegative integer, $\{\varepsilon_{it}\}$ are independent Gaussian white noise series with mean 0 and variances σ_i^2, and $\{s_t\}$ denotes the state process with transition probability

$$P\ (s_t = 2 | s_{t-1} = 1) = \alpha_1$$

and

$$P\ (s_t = 1 | s_{t-1} = 2) = \alpha_2. \tag{10}$$

Empirical Result. Following the approach proposed by McCulloch and Tsay (1994) and using data for 1948–1993, we obtain the MSA model

$$\nabla u_t = \begin{cases} \underset{(.03)}{-.07} + \underset{(.14)}{.38}\nabla u_{t-1} - \underset{(.11)}{.05}\nabla u_{t-2} + \varepsilon_{1t} & \text{if } s_t = 1 \\ \underset{(.04)}{.16} + \underset{(.13)}{.86}\nabla u_{t-1} - \underset{(.14)}{.38}\nabla u_{t-2} + \varepsilon_{2t} & \text{if } s_t = 2. \end{cases} \tag{11}$$

The conditional means of ∇u_t are $-.10$ for $s_t = 1$ and .31 for $s_t = 2$. Thus the first state represents the expansionary periods in the economy, and the second state represents the contractions. The sample variances of ε_{it} are .031 and .192 for $i = 1$ and 2. Additionally, the estimates of the transition probabilities are $\alpha_1 = .084$ (.060) and $\alpha_2 = .126$ (.053). The coefficient estimates are posterior means obtained by the Gibbs sampler.

The model in (11) implies that in the second state, the unemployment rate u_t has an upward trend with an AR(2) polynomial possessing complex characteristic roots. This feature of the model is similar to the second regime of the

TAR model. In the first state, the unemployment rate u_t has a slightly decreasing trend with a much weaker autoregressive relationship.

2.5 Disaggregation Over Time

Because monthly unemployment data are available, forecasts of quarterly unemployment rate can be made from either monthly data or quarterly averages. The choice is related to the general problem of temporal aggregation in time series forecasting. In principle, if a "true" model for the monthly series existed and was known, then monthly series should be used to produce more accurate quarterly forecasts. In practice, there is no known true model for the monthly series. Therefore, it is unclear that using monthly data alone can necessarily produce better forecasts. But an advantage of using monthly data is that, depending on the time the forecast was made within the quarter, some new monthly observations may already be available. In this article we compare forecasting accuracy using monthly and quarterly data.

Empirical Result. For monthly unemployment rates for 1948–1993, we have gone through procedures similar to those used to build linear models for the quarterly data. The monthly unemployment rate is denoted by v_t. The univariate ARIMA model for v_t is

$$(1 - \underset{(.06)}{.45}L^{12})(1 - \underset{(.05)}{1.88}L + \underset{(.05)}{.88}L^2)v_t = \underset{(.003)}{.01} + (1 - \underset{(.04)}{.77}L^{12})(1 - \underset{(.07)}{.74}L)e_t, \quad (12)$$

where the sample variance of the innovational series e_t is .042. Next, let $\mathbf{v}_t = (v_t, i_t)'$, where i_t is the monthly number of initial claims of unemployment. A vector ARIMA model for $\mathbf{v}_t$ is

$$(\mathbf{I} - \mathbf{\Phi}_{12}L^{12})(\mathbf{I} - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - \mathbf{\Phi}_3 L^3)\mathbf{v}_t = \mathbf{c} + (\mathbf{I} - \mathbf{\Theta}_{12}L^{12})(\mathbf{I} - \mathbf{\Theta}_1 L - \mathbf{\Theta}_2 L^2)\mathbf{e}_t \quad (13)$$

with $\operatorname{cov}(\mathbf{e}_t) = \mathbf{\Sigma}_e$, where the coefficient matrices are

$$\mathbf{\Phi}_1 = \begin{bmatrix} 1.62 & .29 \\ (.25) & (1.01) \\ .17 & .77 \\ (.14) & (.36) \end{bmatrix}, \quad \mathbf{\Phi}_2 = \begin{bmatrix} -.61 & .68 \\ (.38) & (.93) \\ -.26 & .01 \\ (.19) & (.33) \end{bmatrix},$$

$$\mathbf{\Phi}_3 = \begin{bmatrix} -.02 & -.93 \\ (.17) & (.52) \\ .08 & .19 \\ (.08) & (.22) \end{bmatrix}, \quad \mathbf{\Phi}_{12} = \begin{bmatrix} -.02 & 1.02 \\ (.09) & (.16) \\ -.16 & .41 \\ (.05) & (.08) \end{bmatrix},$$

$$\mathbf{c} = \begin{bmatrix} -.11 \\ (.12) \\ .11 \\ (.05) \end{bmatrix}, \quad \mathbf{\Theta}_1 = \begin{bmatrix} .83 & -.01 \\ (.26) & (1.01) \\ .13 & -.14 \\ (.13) & (.36) \end{bmatrix},$$

$$\mathbf{\Theta}_2 = \begin{bmatrix} -.09 & .86 \\ (.21) & (.47) \\ -.11 & -.17 \\ (.10) & (.21) \end{bmatrix}, \quad \mathbf{\Phi}_{12} = \begin{bmatrix} .26 & 1.03 \\ (.09) & (.16) \\ .02 & .43 \\ (.05) & (.08) \end{bmatrix},$$

$$\mathbf{\Sigma}_e = \begin{bmatrix} .030 & .014 \\ .014 & .050 \end{bmatrix}.$$

As in (5), except for seasonal relations, notice the upper triangular pattern in the parameter matrices. This again implies that initial claims are an input for the unemployment rate.

We do not consider nonlinear models for the monthly unemployment rate, because quarterly aggregates of these models no longer have the same model structure. This leads to model inconsistencies that complicate the modeling process and require further study.

3. COMPARATIVE PERFORMANCE OF ROLLING FORECASTS

Comparisons of models are usually based on their likelihoods. In linear Gaussian ARIMA time series models, a sufficient statistical reduction allows us to consider the MSE of the residuals. This in turn is equivalent to assessing the models according to their within-sample one-step–ahead forecasts. But it is not known whether the true univariate generating process of the U.S. unemployment belongs to the class of Gaussian linear models. This means that comparing models based upon within-sample one-step–ahead forecasts may not be adequate. Because a primary use of time series models is forecasting, an alternate and perhaps more relevant criterion in judging model performance is the MSE of out-of-sample multistep-ahead forecasts. Its use reflects the need to consider not only improvements in the fit of the model, but also improvements in longer-term forecasting performance. This criterion may also allow us to address inadequacies in the models by identifying any patterns and biases that exist.

In this section we consider out-of-sample forecasts, to better reflect the actual forecasting environment within which practitioners must operate. The forecasts are generated from a rolling forecast method. First, the model parameters are estimated using all observations through a given forecasting origin; next, the forecasts are generated for this origin. This procedure is then repeated for all forecasting origins in the period of interest. In this article we consider all forecasting origins beginning with the third quarter of 1968 and ending with the third quarter of 1993. The beginning quarter was chosen as the starting point because this is the period in which the SPF survey begins. The initial 83 quarters, 1948:1–1968:2, permit fairly stable estimates for the time series models. The third quarter of 1993 is the final quarter before the Labor Department changed the survey technique, which resulted in an increase in the unemployment rate of .5%.

Tables 1–4 summarize the forecasting performance of the various time series models that will be discussed throughout the rest of this section. Table 1 summarizes the mean of the forecast errors and the MSEs of one- to five-quarter–ahead

Table 1. Comparison of Quarterly Forecasts for Various Methods

Model description	Relative MSE of forecasts					Means of forecast errors				
	One-step-ahead	Two-step-ahead	Three-step-ahead	Four-step-ahead	Five-step-ahead	One-step-ahead	Two-step-ahead	Three-step-ahead	Four-step-ahead	Five-step-ahead
Seasonal ARIMA	1.00	1.00	1.00	1.00	1.00	.03	.09	.17	.25	.33
ARIMA(1, 1, 0)	1.04	1.12	1.12	1.14	1.20	.01	.03	.05	.08	.11
TAR	1.00	1.04	.99	.98	1.03	−.01	−.02	−.03	−.03	−.01
MSA	1.19	1.39	1.40	1.45	1.61	.00	−.02	−.04	−.07	−.12
Bivariate AR	1.20	1.21	1.08	1.04	1.03	.02	.07	.16	.25	.34
SPF median forecasts	.31	.53	.58	.60	.65	−.04	−.04	.02	.10	.20
New combined forecasts	.27	.44	.50	.49	.49	.00	.00	.00	.00	.00
MSE of the benchmark model	.08	.31	.67	1.13	1.54					

forecasts for the various models considered. Table 2 provides these statistics broken out for economic contractions and expansions. Table 3 considers the MSE of the forecasts made during rapidly increasing periods of unemployment, which serves as another proxy for economic contractions. Finally, Table 4 reports the MSEs using monthly model forecasts.

Each row in the tables refers to a particular time series model described in Section 2. For a given comparison in every table, each of five columns corresponds to a particular forecast horizon. All the MSEs of the forecasts are expressed in relative terms; that is, relative to the corresponding MSEs of the quarterly ARIMA model in (3), which serves as our primary benchmark. The actual values of the MSEs for the benchmark quarterly ARIMA model are listed in the last row. Note that the relative MSEs do not reflect the absolute increases in the MSEs. The latter indicates that it is much more difficult to create good long-term forecasts than short-term ones, because there is on average a twentyfold increase in MSE from one- to five-step–ahead forecasts.

3.1 Forecasts Using Univariate Time-Series Models

Table 1 shows that the overall biases of all models considered are small. Thus we focus our discussion on the MSE shown in the table. The first two rows report the relative MSE of ARIMA models for the unemployment series. The first row corresponds to the baseline model, which includes a multiplicative seasonal ARMA(4, 4) factor, whereas the second row is the commonly used ARIMA(1, 1, 0) model. The difference in forecasting performance between the two illustrates the importance of properly adjusting for any residual seasonal effects that may remain after the seasonal adjustment process. But it should be noted that if the model is misspecified (i.e., is truly nonlinear), then some of the decrease in the MSE from using the seasonal ARIMA model may be due to model misspecification and not due solely to residual seasonal effects. Additionally, the seasonal adjustment process itself may induce nonlinearity (Ghysels, Granger, and Siklos 1996).

When considering the nonlinear univariate models in lines 3 and 4, we do not observe any overall improvement in the forecasts. In fact, the MSEs of the TAR model are indistinguishable from those of the benchmark model. The MSA model shows a deterioration of forecasting performance, especially for the multistep-ahead forecasts. One explanation for this poor forecasting performance is that for the MSA model, the classification of the state at any point of time including the forecast origin is rarely completely certain, so that the forecasts are complex mixtures of the predictions of the individual states. Consequently, the MSE may not be the appropriate criterion for evaluating the performance of the model. For example, Engel (1994) also considered a MSA model and found that it does not forecast exchange rates better than a random walk, although the Markov model showed some improvement at predicting the direction of change.

Because we do not know the true model from which the unemployment series was generated, all models must be viewed as an approximation to the true generating process, and it is possible that one model may dominate in a specific regime/state compared with others. To understand the merits of the nonlinear model better, we study the forecasts during different periods in the economy. For purposes of this discussion, we report the forecasting performance of the models separately in Table 2 for contractionary and expansionary periods, as defined by the NBER and shown in Figure 1.

In contrast to the small overall biases in Table 1, Table 2 shows a difference in the biases during expansionary and contractionary periods. During economic contractions, we can notice a consistent positive bias (underpredicting the unemployment rate) for all the models considered, with the bias becoming larger for longer-term forecasts. Notice the MSA model yields the smallest bias of all models considered during economic contractions.

Table 2 also provides the MSEs of the forecasts in different economic regimes for the various models. The TAR model (line 3 of Table 2) shows a decrease in MSE of up to 28% for forecasting origins that occur during an economic contraction. Although the ARIMA(1, 1, 0) model (line 2) shows a similar improvement during periods of economic contractions, it also shows much worse performance in economic expansions, which leads to an overall decline in forecasting performance (line 2 of Table 1). The TAR model (line 3) does not exhibit such a marked trade-off.

Table 2. Comparison of Quarterly Forecasts for Various Methods Whose Forecasting Origin Occurs During an Economic Contraction/Expansion

	Relative MSE of forecasts										*Mean of forecast errors*									
	Economic contractions					*Economic expansions*					*Economic contractions*					*Economic expansions*				
Model description	*One-step-ahead*	*Two-step-ahead*	*Three-step-ahead*	*Four-step-ahead*	*Five-step-ahead*	*One-step-ahead*	*Two-step-ahead*	*Three-step-ahead*	*Four-step-ahead*	*Five-step-ahead*	*One-step-ahead*	*Two-step-ahead*	*Three-step-ahead*	*Four-step-ahead*	*Five-step-ahead*	*One-step-ahead*	*Two-step-ahead*	*Three-step-ahead*	*Four-step-ahead*	*Five-step-ahead*
Seasonal ARIMA	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.31	.68	1.08	1.41	1.38	−.01	.00	.03	.08	.17
ARIMA(1, 1, 0)	.85	.83	.78	.71	.76	1.13	1.31	1.37	1.43	1.39	.26	.52	.71	.73	.49	−.03	−.05	−.05	−.03	.05
TAR	.89	.91	.83	.72	.72	1.06	1.13	1.10	1.15	1.17	.24	.56	.87	1.01	.86	−.05	−.11	−.17	−.19	−.14
MSA	.97	1.03	.96	.86	1.02	1.31	1.64	1.73	1.84	1.87	.20	.41	.57	.52	.14	−.03	−.08	−.13	−.17	−.16
Bivariate AR	1.15	1.23	1.16	1.05	1.05	1.22	1.19	1.03	1.03	1.02	.03	.23	.55	.79	.86	.01	.05	.10	.17	.26
SPF forecasts	.16	.48	.60	.63	.67	.39	.56	.56	.57	.64	.08	.41	.75	.98	.94	−.06	−.10	−.10	−.03	.09
Combined forecasts	.15	.41	.51	.49	.43	.34	.45	.50	.50	.51	.07	.31	.58	.70	.54	−.01	−.05	−.09	−.11	−.08
MSE of benchmark	.22	.97	2.14	3.38	3.46	.06	.21	.45	.78	1.24										

Contractionary phases can also be captured by large increases in unemployment, as reported in Table 3. We identify large increases as those changes where $x_t - x_{t-1} > .17$, which is the upper quartile of the differenced unemployment series. The results in Table 3 are similar to those in Table 2, because the lagged difference of the unemployment rate is a fairly good linear discriminator of economic contractions and expansions.

These results suggest that the TAR model better approximates the long-term dynamic structure of the unemployment series during economic contractions, whereas the ARIMA model may better represent its short-term structure. For these results to be of value in practice, long-term forecasts during contractionary periods must be of greater interest than those of overall short-term forecasts. We argue that this is precisely the case. The most interesting forecasts are those when unemployment is rising rapidly (contraction), because it is during those times that the social impact of unemployment is greatest and the desire to forecast a change in unemployment will be most acute. Formally, these arguments can be expressed by saying that the loss attached to forecast errors during a contraction is greater than that during an expansion (cf. Christoffersen and Diebold 1996). Additionally, the long-term forecasts will be of more importance, because unemployment is much more difficult to forecast long-term than short-term.

3.2 Forecasts Using a Leading Indicator

The bivariate AR model using the initial claims does not result in any overall improvement in the predictive ability over the univariate benchmark model, as shown in line 5 of Table 1. The reason for the increase in the MSE of the short-term forecasts is probably due to the increased number of model parameters in the bivariate model. But the initial claims help improve the long-term forecastability of the series during rapidly increasing periods of unemployment, as shown in line 5 of Table 3. Intuitively, it is during periods of rapidly rising unemployment that the information provided about whether a turning point has occurred is crucial for improving the forecasts of the model.

The disadvantage of using a leading indicator is that it essentially transfers the problems of generating long-term forecasts from one series to another. If we were interested primarily in short-term forecasts, this would not be a serious deficiency, but here we are interested in longer-term characteristics of the model. An additional disadvantage of using a leading indicator is that it simply transfers the nonlinearity from the unemployment rate to initial claims. In contrast, the univariate TAR model can endogenously generate asymmetries in the unemployment rate.

3.3 Comparing Monthly and Quarterly Models

Our primary interest has been in forecasting the unemployment rate on a quarterly basis. Because monthly data are available, we also model the unemployment rate on a monthly basis and aggregate the forecasts to a quarterly level. Given a monthly linear time series model, it is straightforward to obtain the corresponding linear model

Table 3. Comparison of Quarterly Forecasts for Various Methods During Rapidly Increasing Periods of Unemployment ($u_{t-1} - u_{t-2} > .17$)

Model description	One-step-ahead	Two-step-ahead	Three-step-ahead	Four-step-ahead	Five-step-ahead
	Relative MSE of forecasts				
Seasonal ARIMA	1.00	1.00	1.00	1.00	1.00
ARIMA(1, 1, 0)	1.08	1.02	.92	.88	.95
TAR	.92	.85	.78	.68	.70
MSA	1.17	1.34	.98	.89	.94
Bivariate AR	.99	1.05	.82	.84	.98
SPF median forecasts	.19	.41	.54	.58	.65
New combined forecasts	.18	.34	.44	.40	.42
MSE of the benchmark model	.19	.83	1.66	2.99	3.92

for the quarterly time series (Telser 1967). But these aggregations assume that both series are linear. In constructing the monthly models, we have ignored these consistency issues and concentrated on finding the best model possible based on the monthly data. To discern the effect of using monthly data to do quarterly forecasts, we construct monthly ARIMA and bivariate models just as before. These disaggregated models are given in Section 2.5. Specifically, the monthly univariate ARIMA model is given in (12); the bivariate ARMA model, in (13).

The use of monthly data results in substantial improvements in short-term forecasts, as reported in lines 2 and 3 of the first part of Table 4. These improvements range from a 20% decrease in one-step–ahead MSE for univariate linear models to a more than 40% decrease for bivariate linear models. But when we consider longer-term forecasts the performance of the monthly ARIMA model is worse, and that of the monthly bivariate ARMA model is better. Some of this gain can be explained by the increased efficiency of the estimates, because monthly data have three times as many observations as the quarterly data. Also, forecasting temporal aggregates using disaggregated series can result in substantial gains in forecasting efficiency. (See Tiao 1972 for a discussion using an ARIMA$(p, 1, q)$ model.)

Another issue of interest when dealing with disaggregated data is whether the knowledge of an additional month in the first quarterly period to forecast will substantially alter the forecasting performance of the model. We have computed the rolling forecasts of the ARIMA and bivariate ARMA models using monthly data conditional on the first month or first two months of the initial quarterly period to be forecasted. Lines 4 and 5 of Table 4 show that, on inclusion of the first month, the MSE of the one-quarter–ahead forecast shows more than a 75% reduction from the MSE of the benchmark quarterly ARIMA model for both classes of models. Although there is some gain to using this additional month for longer-term forecasts for the linear models, these decreases diminish significantly. Inclusion of the first two months of the forecasting quarter results in an even more pronounced reduction in one-step–ahead MSE, of more than 96% (lines 6 and 7 of Table 4).

Theoretically, using the monthly ARMA model in (12) and knowledge of the first month of the quarter should lead to a 68% reduction in the variance of the one-step–ahead forecast. Even more dramatically, when advancing the forecasting origin two months into the quarter, the reduction should be 94%. A sketch of the derivation of these calculations is given in the Appendix. Note that these reductions are very close to the empirical findings shown in Table 4. Howrey, Hymans, and Donihue (1991) and Preston and Chin (1996) have also considered the problem of forecasting quarterly series using disaggregate quarterly data and found similarly dramatic results.

We recompute the MSE of forecasts during an economic contraction from the monthly models as with the quarterly model and provide the results in the second panel of Table

Table 4. Comparison of Quarterly Forecasts Using Monthly Models

Model description	Overall: One-step-ahead	Overall: Two-step-ahead	Overall: Three-step-ahead	Overall: Four-step-ahead	Overall: Five-step-ahead	During economic contractions: One-step-ahead	During economic contractions: Two-step-ahead	During economic contractions: Three-step-ahead	During economic contractions: Four-step-ahead	During economic contractions: Five-step-ahead
	Relative MSE of forecasts									
Quarterly ARIMA—benchmark	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Monthly ARIMA	.80	.92	1.03	1.12	1.21	.61	.85	.94	.97	1.04
Monthly Bivariate ARMA	.58	.73	.78	.83	.89	.28	.65	.80	.88	1.06
First Month Known:										
Monthly ARIMA	.24	.58	.77	.90	1.06	.20	.66	.83	.87	.96
Monthly bivariate ARMA	.23	.63	.75	.81	.87	.11	.51	.72	.83	.97
First 2 Months Known:										
Monthly ARIMA	.03	.39	.63	.79	.97	.04	.45	.66	.80	.92
Monthly bivariate ARMA	.04	.42	.58	.69	.79	.02	.22	.43	.63	.78
SPF median forecasts	.31	.53	.58	.60	.65	.16	.48	.60	.63	.67
MSE of the benchmark model	.08	.31	.67	1.13	1.54	.22	.97	2.14	3.38	3.46

4. Note that there is an additional 24% drop in MSE for one-step–ahead forecasts from the monthly ARIMA model when considering forecasts only during economic contractions. There is an even more dramatic reduction of 52% in MSE for one-step–ahead forecasts from the monthly bivariate model during this same period. Although some improvements occur in the longer-term forecasts, these gains are not as impressive as those of the TAR model (line 3 of Table 2). To the best of our knowledge, such dramatic differences in forecasting performance during different periods of economic activity have not been discussed or demonstrated in the literature. In summary, of the models we have considered, we find that during economic contractions, short-term forecasts from monthly bivariate ARMA models are best, and in the longer-term, TAR models are better.

3.4 Comparisons with the Consensus Forecasts

Obviously, practitioners are not limited to considering only the past quarterly information on unemployment itself when forming their expectations about the future unemployment rate. They make use of exogenous and monthly data and can combine this information using complex functional models rather than simply relying on univariate or bivariate time series models. To gauge the importance of these additional sources of information, we report the performance of the median forecasts of selected participants of the SPF survey, which we call the consensus forecasts. (We reduce the sample of SPF survey to include only those forecasters who have participated in at least 10 surveys; this helps improve the consensus forecasts by reducing the number of spurious forecasters.) By comparing the forecasts from the TAR component model and the consensus forecasts, we can determine whether nonlinearities in the unemployment rate are fully exploited. A deficiency in the consensus forecasts—just as in the univariate ARIMA model—is consistently underpredicting turning points during contractionary periods. (All five peaks in the unemployment rate were underpredicted based upon the one-step-ahead forecast errors.)

The MSEs of these consensus forecasts are reported in line 6 of Tables 1 through 3 and line 8 of Table 4. For the one-step-ahead forecasts, the MSE of the consensus forecasts is only about 30% of that of the benchmark ARIMA model (see line 7 of Table 1). Some of this improvement may be explained by the forecasters having access to the first month of the quarter in which the survey is made. (The correlations between the forecast errors from the consensus and monthly model in which the first month of the forecasting origin quarter is known are .41, .69, .75, .78, and .81 for the one- through five-step–ahead forecasts.) Because the date on which the surveys are returned can vary, some forecasters have no additional information from the first month, whereas others can have up to two additional months (Swanson and White 1997).

Thus a more unprejudiced comparison is between the monthly models that are conditioned on an additional month of data (compare line 8 with lines 4–5 of Table 4). When considering this fact and the more natural comparisons between the monthly ARIMA and bivariate models (see lines 4 and 5 of Table 4), the model-based forecasts outperform the consensus forecasts in the first step ahead. But by the third step ahead, the consensus forecasts tend to outperform the monthly ARIMA and bivariate AR models. This may reflect the more efficient use of external information by the consensus forecasts.

3.5 Illustrating the Forecasts

For a better demonstration of the forecasting properties of the models, in Figure 3 we plot the multistep-ahead forecasts for one business cycle (the third quarter of 1980 to the fourth quarter of 1982) from each origin for the ARIMA, TAR, MSA, and consensus forecasts. The integrated component of the ARIMA(1, 1, 0) model results in level long-term forecasts, although the short-term forecasts are modified by an autoregressive component reflecting the change in the unemployment rate at the forecasting origin. A shortcoming of this model is its lack of "knowledge" about the business cycle. The seasonal ARIMA(1, 1, 0) model overcomes this shortcoming by allowing for complex roots in the autoregressive polynomial, which allow for turning points in the forecasts. Unfortunately, the model underpredicts the unemployment rate during the rapid increase of 1982 and exhibits forecasts that fluctuate a great deal more during stable periods of unemployment.

The TAR model performs much better during the rapid increase and decline in unemployment during the early 1980s. As with the other models, the TAR model also has difficulty determining when the turning point will occur, although it is better able to forecast high unemployment or continued high unemployment during the rapid increase in unemployment during 1982. Unfortunately, the TAR model performs poorly during the slowly declining periods of unemployment. The self-exciting nature of the model results in larger-than-observed probabilities that unemployment will shift from the first regime (declining or small increasing) to the second regime (strongly increasing) during some point within a five-step–ahead forecast. This indicates that future extensions to the TAR model should incorporate some additional persistence in shifting from the first regime to the second regime as does the MSA model.

The SPF consensus forecasts are overly optimistic during increasing periods of unemployment, consistently underpredicting the peak values of unemployment during 1983. In contrast, they are overly pessimistic during the rapid decrease in unemployment during the early 1980s. In comparison, the consensus forecasts do a much better job of reflecting the stability of the unemployment forecasts during the slow decline of the later 1980s. These forecasts have an unusual property in that the longer-term forecasts are almost always less than those of the near term forecasts.

Although the TAR model performs the best during this period, some proposed extensions may yield even better forecasting results. Montgomery, Zarnowitz, Tsay, and Tiao (1996) suggested decomposing the series into a linear and nonlinear component. The linear component is meant to capture the overall movement in the series, and the nonlinear component is meant to better model the asymmetry in the business cycles. Another change to the TAR model that may yield additional forecasting improvements is better classification of observations to the regimes. This could be accomplished through the use of exogenous or stochastic thresholds.

3.6 Improving the Forecasting Performance of the Models

All of these forecasts share certain strengths and weaknesses, but clearly they are not perfectly correlated, and none dominates the others. Thus an important question is whether any additional improvements can be made to the consensus forecasts by using these linear and nonlinear models. To answer this question, we construct a new combined forecast, which works by exploiting the differences in correlation among the forecast errors (For a discussion of combined forecasts see Granger and Newbold 1986, chap. 9.2). This new combined forecast is constructed by taking a weighted average of the forecasts from the consensus and the quarterly TAR component model. The weights are determined through usual least squares regression techniques, with a constant included to account for any systematic biases in the estimates. Furthermore, we compute different weights for each forecast horizon and allow for four different sets of weights. We allow for a separate set of weights depending on whether the forecasting origin was occurred during a strongly decreasing ($u_{t-1} - u_{t-2} < -.2$), decreasing ($-.2 \leq u_{t-1} - u_{t-2} < -.033$), increasing ($-.033 \leq u_{t-1} - u_{t-2} < .167$), or strongly increasing ($.167 \leq u_{t-1} - u_{t-2}$) period in the unemployment rate. These values correspond to the quartiles of the change in the unemployment rate. This yields a total of 20 different sets of linear combinations (five forecasting horizons by four regimes; e.g., the first set of weights relate to the one-step–ahead forecasts whose origin experiences strong decreases in the unemployment rate).

The weights to create the new combined forecasts are listed in Table 5, and the relative MSE of these new combined forecasts are listed in line 7 of Tables 1–3. (The MSEs of the new combined forecasts have been adjusted for the number of estimated parameters to yield unbiased MSE estimates. The MSE of each regime is $\sum_{i=1}^{n_i} \hat{\varepsilon}_{(i)}^2/(n_i - 3)$, because there are three parameters in each regime, where n_i denotes the number of observations in the regime and (i) is the ith prediction. There are 25 one-step-ahead forecasts for three regimes and 26 forecasts in the remaining regime.) These weights allow us to evaluate the marginal

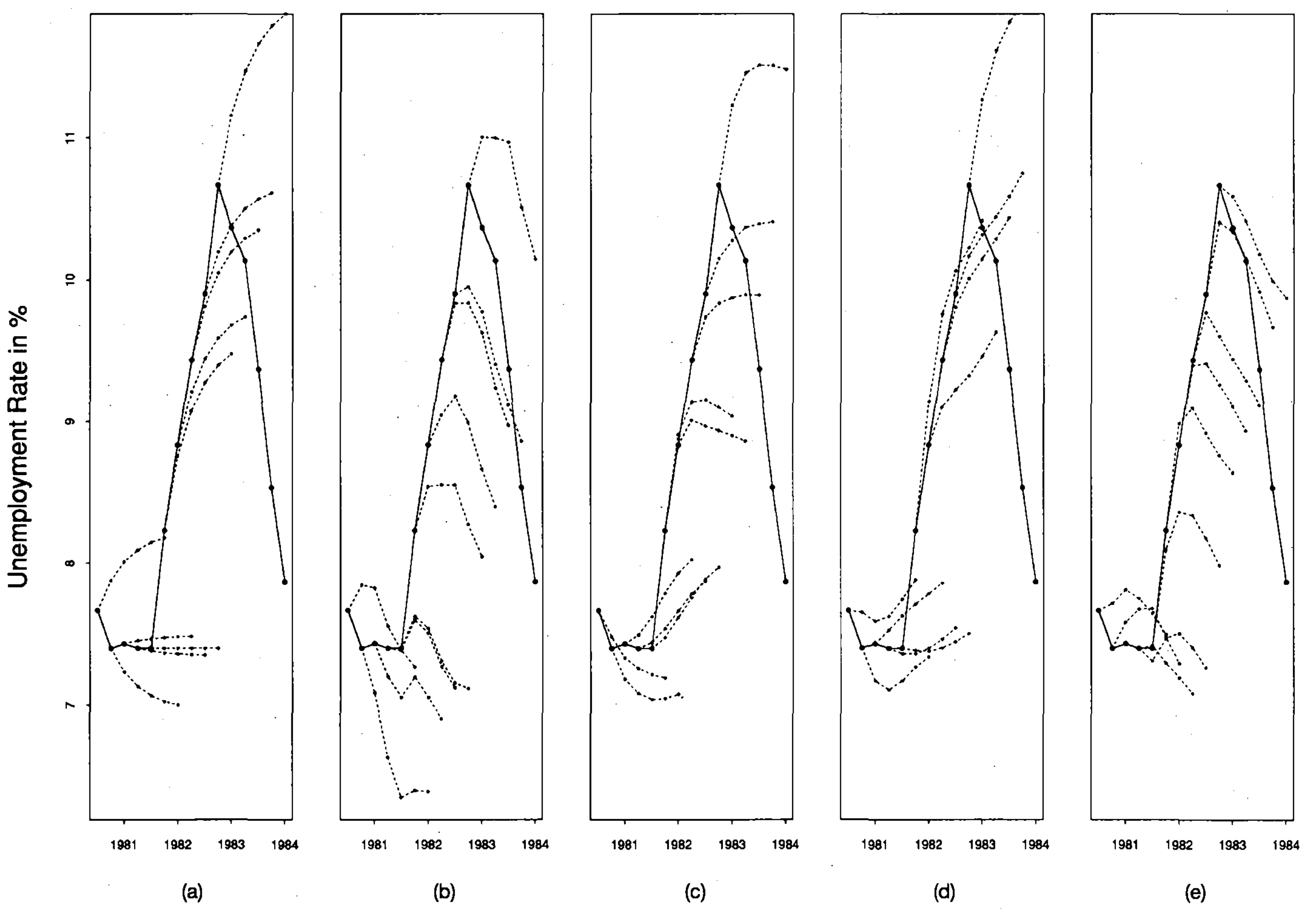

Figure 3. Comparison of Forecasts for 1980 Third Quarter–1982 Fourth Quarter Business Cycle. (a) ARIMA(1, 1, 0); (b) Seasonal ARIMA(1, 1, 0); (c) TAR model; (d) MSA model; (e) consensus. ——, Actual; · · · , forecast.

Table 5. Regression Weights of Forecasts from Various Models to Construct the New Combined Forecasts

Step ahead	Parameter	Strongly decreasing	Decreasing	Increasing	Strongly increasing
		Quartile of unemployment rate change at the forecasting origin			
One	Constant	.24[c]	−.25[b]	−.05	.13
	Consensus	.76[a]	.78[a]	1.27[a]	1.12[a]
	TAR	.19[c]	.26[a]	−.26[a]	−.14
Two	Constant	.50	−.33	−.37	.83[b]
	Consensus	.72[a]	.70[a]	1.52[a]	.95[a]
	TAR	.18	.34[a]	−.44[b]	−.05
Three	Constant	.77	−.57[b]	−.02	1.90[b]
	Consensus	.56[b]	.98[a]	1.35[a]	.80[a]
	TAR	.29	.10	−.30	−.04
Four	Constant	.87	−.49	.85	3.08[c]
	Consensus	.56[b]	1.30[a]	1.17[a]	.84[a]
	TAR	.26	−.22[a]	−.23	−.22
Five	Constant	.98	.44	1.52	3.87[c]
	Consensus	.54[c]	1.49[a]	1.25[a]	.92[a]
	TAR	.28	−.54[b]	−.38[b]	−.40[b]

NOTE: The superscript denotes the significance of the parameter using White's (1980) heteroscedasticity-consistent covariance estimator to reflect possible cross- and serial-correlations in the residuals:
[a](≤1%),
[b](≤5%),
[c](≤10%).

contributions of these model-based forecasts over the consensus forecasts. For example, if the TAR component model had no incremental forecasting benefit over the consensus forecasts, then the weights for the TAR component model would simply be 0. Clearly, this is not the case, although the importance of the TAR component model varies according to the rate of change of the unemployment rate and the horizon of the forecast. The most dramatic improvements over the consensus forecasts occur in the multistep forecasts of periods of rapidly increasing unemployment, where we find reductions in MSE of almost 40% (compare lines 6 and 7 of Table 3). Although some of this decrease can be attributed to the inclusion of the constant, the rest must be attributed to the additional information provided by the TAR component model, with this information weighted more strongly in longer-term forecasts.

4. DISCUSSION

Up to this point, we have contrasted the various models by comparing the MSEs of the forecasts. But because these errors may be correlated, it is important to take this into account when comparing the forecasts. A method for doing so was discussed by Granger and Newbold (1986, p. 279). Consistent with our previous models, we assume that the forecast errors generated by two methods are from a bivariate normal distribution, which we denote as (a_t, b_t), where a_t and b_t denote the forecasts errors of our new model and the benchmark model. Now consider the distribution of the transformed series $(v_t, w_t) = (a_t + b_t, a_t - b_t)$,

$$\begin{bmatrix} a_t \\ b_t \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} \right) \Rightarrow \begin{bmatrix} v_t \\ w_t \end{bmatrix} = \begin{bmatrix} a_t + b_t \\ a_t - b_t \end{bmatrix}$$

$$\sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_a^2 + 2\sigma_{ab} + \sigma_b^2 & \sigma_a^2 - \sigma_b^2 \\ \sigma_a^2 - \sigma_b^2 & \sigma_a^2 - 2\sigma_{ab} + \sigma_b^2 \end{bmatrix} \right). \quad (14)$$

We wish to test whether the variance of our new forecasts (σ_a^2) is significantly less than that of our benchmark model (σ_b^2). This can be done by testing whether the covariance of the transformed series $(E[v_t w_t] = \sigma_a^2 - \sigma_b^2)$ is 0. A simple test for this is the usual test of whether the correlation is 0.

An added complication is that the series of q-step-ahead forecast errors will follow an MA$(q-1)$ process, at least for a correctly specified ARIMA model. But the Granger–Newbold test is valid only for errors that are not serially correlated. Again, consider the same transformation as before and, relaxing the distributional assumptions, Meese and Rogoff (1988) showed that

$$\sqrt{T}\hat{\gamma}_{vw} \xrightarrow{d} \mathrm{N}(0, \Sigma),$$

where $\mathbf{v} = (v_1, \ldots, v_T)$, $\mathbf{w} = (w_1, \ldots, w_T)$, $\hat{\gamma}_{vw} = \mathbf{v}'\mathbf{w}/T$, $\Sigma = \sum_{\tau=-\infty}^{\infty} [\gamma_{vv}(\tau)\gamma_{ww}(\tau) + \gamma_{vw}(\tau)\gamma_{wv}(\tau)]$, $\gamma_{vw}(\tau) = \mathrm{cov}(v_t, w_{t-\tau})$, $\gamma_{wv}(\tau) = \mathrm{cov}(w_t, v_{t-\tau})$, $\gamma_{vv}(\tau) = \mathrm{cov}(v_t, v_{t-\tau})$, and $\gamma_{ww}(\tau) = \mathrm{cov}(w_t, w_{t-\tau})$. We replace Σ with a corresponding consistent estimate from the sample. (See Diebold and Mariano 1995 for a more general discussion of comparisons of predictive accuracy.)

We provide the p values of the Meese–Rogoff test between the quarterly ARIMA benchmark model and the other models in the first part of Table 6 (H_0: $\sigma_a^2 = \sigma_b^2$ versus H_a: $\sigma_a^2 < \sigma_b^2$, where the b subscript denotes the benchmark model). We use a one-sided test, because our benchmark model is a special case of all other models considered, and thus we are interested only in testing for improvements in forecasting performance. Consistent with our previous observations, we find little statistical support for the hypothesis that the MSE of the quarterly ARIMA benchmark model is different than the MSE of the other quarterly time series models (lines 1 through 4 of Table 6). But the one- to three-step-ahead forecasts for the monthly univariate ARIMA model when the first month of the quarter after the forecasting origin is known (line 5) show significant improvements over the quarterly model. The monthly bivariate ARIMA model (line 6) shows significant improvements up until the fourth-step–ahead. Finally, there is very strong evidence that the SPF forecasts are better than the benchmark forecasts. An additional test of the SPF forecasts against the new combined forecasts demonstrates that the new combined forecasts significantly outperform the SPF forecasts, with p values $<.01$ for all five steps.

These overall results are masking an important fact that the TAR model performs substantially better when the forecasts occur during economic contractions; i.e., when the unemployment rate is increasing quickly. To illustrate this fact, we recompute the first panel of Table 6 only for those forecasts whose origins occurred during economic contractions (as in Table 2) and list the results from the Granger–Newbold tests in the second panel of Table 6. Although it would be more proper to use the Meese–Rogoff test, this test cannot be directly applied because the forecast errors during economic contractions are noncontiguous, therefore we use the Granger–Newbold test to simplify the calculations. (But the discontinuity should lessen any serial cor-

Table 6. p Values of the Meese–Rogoff Tests for Significance of Improvements in Forecast MSE for One- to Five-Step-Ahead Forecasts by Various Methods, H_0: $\sigma_a^2 = \sigma_b^2$ versus H_a: $\sigma_a^2 < \sigma_b^2$

	Overall					Economic contractions				
Models	One-step-ahead	Two-step-ahead	Three-step-ahead	Four-step-ahead	Five-step-ahead	One-step-ahead	Two-step-ahead	Three-step-ahead	Four-step-ahead	Five-step-ahead
ARIMA(1, 1, 0)	.62	.82	.80	.79	.84	.21	.18	.14	.12	.22
TAR	.51	.64	.46	.45	.57	.35	.33	.16	.04	.08
MSA	.94	1.00	.99	.99	.99	.47	.53	.44	.34	.52
Bivariate AR	.93	.94	.75	.61	.59	.61	.68	.67	.56	.57
Monthly ARIMA (1st month known)	.00	.00	.07	.32	.58	.00	.08	.19	.23	.38
Monthly bivariate ARIMA (1st month known)	.00	.01	.02	.06	.15	.00	.09	.19	.27	.45
SPF median forecasts	.00	.00	.00	.00	.01	.00	.01	.02	.02	.05
New combined forecasts	.00	.00	.00	.00	.01	.00	.01	.01	.01	.01

relations present. There are five contractionary runs, for a total of 14 observations out of 101 forecasts.) Here we observe support for the proposition that the TAR model (see line 3 of Table 3) is performing better than the benchmark model for the four- and five-step-ahead forecasts (p values of .04 and .08). Additionally, if we recompute these statistics for those forecasts that were realized to occur during economic contractions, we find even stronger support showing that the three- to five-step-ahead forecasts from the TAR model are lower than those of the benchmark model (p values all <.01). Similar results are obtained for forecasts that fall during periods of rapidly increasing periods of unemployment (as in Table 3), with p values all <.01.

In Table 7 we compare our forecasts against several econometric models (see Zarnowitz and Braun 1993, pp. 11–84) in terms of square root of the mean squared error (RMSE). Throughout this article we have used the SPF survey to represent the richer set of information available to practicing economic forecasters. On comparison with the other econometric and model-based methods of models, we find that the SPF survey dominates these forecasts in terms of RMSE (except for long-term forecasts of the Michigan RSQE and BVAR model forecast models. This supports our use of these forecasts as a proxy for a full-information (although not an optimal) forecast.

Although our TAR model does not perform as well as the SPF survey, it does compare favorably with models that use additional information, such as the Sims model (Sims 1989). But when the forecasts of the TAR model and SPF survey are combined, there is substantial improvement in the resulting forecasts, as shown in the last column. These new combined forecasts dominate all other forecasts, with the exception of equal performance for one-step–ahead forecasts with the monthly ARIMA model when the first month of the quarter after the forecasting origin is known. In fact, it appears that much of the improvement in the short-term forecasts of the SPF survey and Michigan RSQE over the simple univariate ARIMA model is due to the incorporation of monthly information through forecasts of exogenous variables and constant adjustments of the forecasts (Donihue 1993).

5. CONCLUSIONS AND IMPLICATIONS

Theory and data agree on a number of analytical results that helped guide our empirical work:

1. Over the last century, the U.S. unemployment rate (u) had no consistent trend at all, and there are no good reasons why it should have either risen or fallen secularly. It has asymmetrical cyclical movements, particularly during the severe downward cycles, which dominate the behavior over time of the u rate. Short and steep rises in u, ending in sharp peaks, are characteristic of general business contractions; long and gradual declines, of business expansions. After World War II, u drifted slowly and irregularly upward, except during the long business expansions of the 1960s and 1980s.

Table 7. Comparisons for Various Econometric and Time Series Models of the U.S. Unemployment Rate in Terms of Root Mean Squared Errors

Target quarter	Quarterly ARIMA benchmark	Monthly ARIMA model	Quarterly TAR model	SPF survey	University of Michigan RSQE	BVAR model forecast	Sims model	New combined forecasts
1	.31	.15	.31	.17	.17	.28	.55	.15
2	.59	.44	.59	.42	.44	.50	.79	.37
3	.86	.75	.84	.64	.67	.66	1.03	.60
4	1.11	1.04	1.08	.84	.78	.78	1.23	.78
5	1.29	1.31	1.29	1.03	.93	.85	1.40	.91

NOTE: This table is based on forecast origins from the third quarter of 1968 through the fourth quarter of 1989. The monthly ARIMA model assumes the first month of quarter after the forecasting origin is known. The University of Michigan RSQE model is the Michigan Research Seminar in Quantitative Economics macroeconomic model forecasts. These forecasts begin in the fourth quarter 1970 and were not made in the first quarter of 1975 and 1976, and in the second quarter of 1971–1975 and 1977–1979. The RMSEs for the SPF survey forecasts in quarters strictly matching those covered by RSQE are (for target quarters 1, ..., 5, respectively): .17, .43, .68, .85, and .96. The BVAR model forecast model is a Bayesian vector autoregression model with six variables and six quarterly lags estimated sequentially with data available in 1993. The Sims model is a nine-variable, five-lag BVAR model that allows time variation in coefficients and forecast error variance and nonnormality in disturbances (Sims 1989).

2. For quarterly data, the TAR and MSA models outperform the linear benchmark model in terms of the MSE for multistep-ahead forecasts during contractions or periods of rapidly increasing unemployment. Specifically, the TAR model shows a 28% reduction in the MSE of four- and five-step-ahead forecasts whose forecasting origin occurs during an economic contraction (see Table 2). It should also be noted that forecasting unemployment is much more difficult during periods when it is rapidly increasing than during more stable periods.

3. Initial claims for unemployment insurance under the state programs, which are available weekly, are used as a leading indicator of u, because they contain information on whether unemployment is rising or falling. A bivariate linear (AR) model with quarterly initial claims, although not superior overall, outperforms the univariate benchmark linear model during periods of rapidly increasing unemployment (with reductions in the MSE of 18% and 16% for three- and four-step-ahead forecasts).

4. The forecasting gains from using monthly instead of quarterly data are very large due to the high autocorrelation in the U.S. unemployment rate. This is so especially for the shortest (one- and two-quarter–ahead) forecasts, as would be expected. As the horizon is increased, the gains become much less. If the origin is advanced 1 month into the quarter (to allow for the advantage of the real-time forecasters), then short-term performance is substantially improved. The reduction in the MSE for one-quarter–ahead forecasts is more than 75% for the two monthly models considered.

5. The best short-term forecasts using the models considered in this article were generated by a monthly bivariate ARMA model. There was a 42% reduction in the one-step–ahead MSE for this model compared to the quarterly ARIMA benchmark. There was an even greater (72%) reduction in the one-step–ahead MSE if we considered only forecasts generated during economic contractions.

6. Combining informed professional forecasts often results in substantial improvements: thus group mean/median forecasts from surveys of business and academic economists are generally much more accurate than most individual forecasts. The "consensus" forecasts from the SPF quarterly economic outlook survey performed better than selected econometric models and time-series forecasts for the one- to three-step–ahead forecasts (Zarnowitz and Braun 1993). These forecasts benefit from steady, up-to-date monitoring of weekly and monthly leading indicators and other timely data and also generally from pooling great amounts of complementary information from different markets and models. Hence it is not surprising that the SPF survey forecasts have much smaller errors than all of the different model forecasts that we have considered when the predictions are generated from the same periods of origin. But when the origin is advanced 1 month into the quarter, the monthly models and the consensus forecasts become more comparable.

7. The quarterly TAR model contains additional information to those forecasts generated by the SPF consensus forecast. A new combined forecast, constructed by taking a weighted average of the two, dominates—or at least equals—the RMSE of all other forecasts considered.

8. Future models of the unemployment rate can improve their forecasts by reflecting the asymmetry in the unemployment rate. Models of the unemployment rate should be able to generate rapid increases that will not consistently underpredict the turning points. At the same time, the model must be able to predict the stable, slowly declining periods of unemployment.

APPENDIX: AGGREGATION OF MONTHLY ARMA MODELS TO GENERATE QUARTERLY FORECASTS

Suppose that we have a monthly series (z_t) that follows an ARMA(p, q) model:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}, \qquad a_t \sim \mathrm{N}(0, \sigma^2). \quad \text{(A.1)}$$

We wish to compute the variance of the quarterly forecasts $(x_{t'})$ using the aggregated forecasts from our monthly model, where $x_{t'} = (z_{t+2} + z_{t+1} + z_t)/3$.

We denote the one-, two-, and three-step–ahead forecasts from time origin $t-1$ as

$$\hat{z}_{t|t-1} = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q},$$

$$\hat{z}_{t+1|t-1} = \phi_1 \hat{z}_{t|t-1} + \phi_2 z_{t-1} + \cdots + \phi_p z_{t-p+1} - \theta_2 a_{t-1} - \theta_3 a_{t-2} - \cdots - \theta_q a_{t-q+1},$$

and

$$\hat{z}_{t+2|t-1} = \phi_1 \hat{z}_{t+1|t-1} + \phi_2 \hat{z}_{t|t-1} + \phi_3 z_{t-1} + \cdots + \phi_p z_{t-p+2} - \theta_3 a_{t-1} - \theta_4 a_{t-2} - \cdots - \theta_q a_{t-q+2} \quad \text{(A.2)}$$

and the corresponding forecast errors as

$$\hat{e}_{t|t-1} = z_t - \hat{z}_{t|t-1} = a_t,$$

$$\hat{e}_{t+1|t-1} = z_{t+1} - \hat{z}_{t+1|t-1} = a_{t+1} + (\phi_1 - \theta_1) a_t,$$

and

$$\hat{e}_{t+2|t-1} = z_{t+2} - \hat{z}_{t+2|t-1} = a_{t+2} + (\phi_1 - \theta_1) a_{t+1} + [\phi_1(\phi_1 - \theta_1) + (\phi_2 - \theta_2)] a_t. \quad \text{(A.3)}$$

(See Box, Jenkins, and Reinsel 1994 for a further discussion.)

The variance of the quarterly forecast using the monthly model from time origin $t-1$ is

$$\begin{aligned}
&\mathrm{var}(x_{t'} | \mathbf{z}_{t-1}) \\
&= \mathrm{var}\left(\frac{1}{3}\,(z_{t+2} + z_{t+1} + z_t) | \mathbf{z}_{t-1}\right) \\
&= \frac{1}{9}\,\mathrm{var}(\hat{e}_{t+2|t-1} + \hat{e}_{t+1|t-1} + \hat{e}_{t|t-1}) \\
&= \frac{1}{9}\,\mathrm{var}(a_{t+2} + (1 + (\phi_1 - \theta_1)) a_{t+1} + (1 + (1 + \phi_1)(\phi_1 - \theta_1) + (\phi_2 - \theta_2)) a_t) \\
&= \frac{1}{9}\,[1 + (1 + (\phi_1 - \theta_1))^2 + (1 + (1 + \phi_1)(\phi_1 - \theta_1) + (\phi_2 - \theta_2))^2]\sigma^2, \quad \text{(A.4)}
\end{aligned}$$

where $\mathbf{z}_{t-1}$ denotes all information up to and including time period $t-1$.

The variance of the quarterly forecast when the first month of the quarter after the forecasting origin is known or from time origin t is

$$\begin{aligned}\operatorname{var}(x_{t'}|\mathbf{z}_t) &= \operatorname{var}\Big(\frac{1}{3}\,(z_{t+2}+z_{t+1}+z_t)|\mathbf{z}_t\Big)\\ &= \frac{1}{9}\,\operatorname{var}(\hat{e}_{t+2|t}+\hat{e}_{t+1|t})\\ &= \frac{1}{9}\,\operatorname{var}(a_{t+2}+(1+(\phi_1-\theta_1))a_{t+1})\\ &= \frac{1}{9}\,[1+(1+(\phi_1-\theta_1))^2]\sigma^2. \qquad \text{(A.5)}\end{aligned}$$

The variance of the quarterly forecast when the first 2 months of the quarter after the forecasting origin is known or from time origin $t+1$ is

$$\begin{aligned}\operatorname{var}(x_{t'}|\mathbf{z}_{t+1}) &= \operatorname{var}\Big(\frac{1}{3}\,(z_{t+2}+z_{t+1}+z_t)|\mathbf{z}_{t+1}\Big)\\ &= \frac{1}{9}\,\operatorname{var}(\hat{e}_{t+2|t+1})\\ &= \frac{1}{9}\,\operatorname{var}(a_{t+2}) = \frac{1}{9}\,\sigma^2 \qquad \text{(A.6)}\end{aligned}$$

The variance of the quarterly forecasts based on the quarterly data as given in (3) is .09. After evaluating (A.5) for the ARMA model of the monthly series given in (12), we would expect the variance of the quarterly forecasts to be .08, and if the first month or first 2 months of the quarter after the forecasting origin were known, then the variance of the quarterly forecasts as given in (A.6) and (A.7) would be .026 and .005. Thus, using the monthly model to generate the quarterly forecasts, we would expect the variance to decrease to 89% of the quarterly model. If the first month or first two months of the quarter after the forecasting origin were known, then we would expect to see even more dramatic declines, to 29% or 5.2% of the variance of the quarterly forecasts, respectively. These large improvements in forecasting performance of moving from monthly to quarterly data are the result of one of the roots of the ARMA polynomial in (12) being very close to the unit root. As this root moves away from the unit circle, the incremental forecasting performance will be much less.

[Received March 1997. Revised December 1997.]

REFERENCES

Ansley, C. F. (1979), "An Algorithm for the Exact Likelihood of a Mixed Autoregressive Moving Average Process," *Biometrika*, 66, 59–65.
Box, G. E. P., Jenkins, G. M., and Reinsel, (1994), *Time Series Analysis: Forecasting and Control*, (3rd ed.), Englewood Cliffs, NJ: Prentice-Hall.
Chan, K. S. (1993), "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21, 1, 520–533.
Christoffersen, P. F., and Diebold, F. X. (1996), "Further Results on Forecasting and Model Selection Under Asymmetric Loss," working paper, University of Pennsylvania, Dept. of Economics.
Croushore, D. (1993), "Introducing The Survey of Professional Forecasters," *The Federal Reserve Bank of Philadelphia Business Review*, Nov.–Dec., pp. 3–15.
Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.
DeLong, J. B., and Summers, L. H. (1986), "Are Business Cycles Symmetrical?," in *The American Business Cycle, Continuity and Changes*, ed. R. J. Gorton, Chicago: University of Chicago Press for NBER.
Donihue, M. R. (1993), "Evaluating the Role Judgement Plays in Forecasting Accuracy," *Journal of Forecasting*, 12, 2, 81–92.
Elwood, S. K. (1996), "Testing for the Symmetry in the Effects of Time Series Disturbances: An Unobserved Components-Threshold Approach," working paper, James Madison University, Department of Economics.
Engel, C. (1994), "Can the Markov Switching Model Forecast Exchange Rates?," *Journal of International Economics*, 36, 1–2, 151–165.
Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.
Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
Ghysels, E., Granger, C. W. J., and Siklos, P. L. (1996), "Is Seasonal Adjustment a Linear or Nonlinear Data-Filtering Process?" *Journal of Business & Economic Statistics*, 14, 374–386.
Granger, C. W. J., and Newbold, P. (1986), *Forecasting Economic Time Series*, New York: Academic Press.
Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.
——— (1990), "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45, 1, 2, 39–70.
Hannan, E. J., and Deistler, M. (1988), *The Statistical Theory of Linear Systems*, New York: Wiley.
Hillmer, S. C., and Tiao, G. C. (1979), "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models," *Journal of the American Statistical Association*, 74, 652–660.
Howrey, E. P., Hymans, S. H., and Donihue, M. R. (1991), "Merging Monthly and Quarterly Forecasts: Experience With MQEM," *Journal of Forecasting*, 10, 255–268.
Ljung, G. M., and Box, G. E. P. (1978), "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 65, 297–303.
McCulloch, R. E., and Tsay, R. S. (1994), "Bayesian Analysis of Threshold Autoregressive Processes with a Random Number of Regimes," *Journal of Time Series Analysis*, 15, 235–250.
Meese, R. A., and Rogoff, K. (1988), "Was it Real? The Exchange Rate-Interest Differential Relation Over the Modern Floating-Rate Period," *Journal of Finance*, 43, 933–948.
Mitchell, W. C. (1927), *Business Cycles: The Problem of Its Setting*. New York: NBER.
Montgomery, A. L., Zarnowitz, V., Tsay, R. S., and Tiao, G. C. (1996), "Modeling Nonlinearity in the U.S. Unemployment Rate with Unobserved Component Models," technical report, University of Chicago, Graduate School of Business.
Moore, G. H. (1983), *Business Cycles, Inflation, and Forecasting*. New York: Ballinger Pub. Co. for NBER.
Neftci, S. N. (1984), "Are Economic Time Series Asymmetric Over the Business Cycles?," *Journal of Political Economy*, 85, 281–291.
Preston, J. M., and Chin, D. M. (1996), "Using Monthly Data to Improve Quarterly Model Forecasts," *Federal Reserve Bank of Minneapolis Quarterly Review*, 20(2), 16–33.
Rothman, P. (1996), "Forecasting Asymmetric Unemployment Rates," working paper, East Carolina University, Dept. of Economics.
Sichel, D. E. (1989), "Business Cycle Asymmetry: A Deeper Look," Economic Working Paper 93, Federal Reserve Board of Governors.
Sims, C. A. (1989), "A Nine-Variable Probabilistic Macroeconomic Forecasting Model," Discussion Paper 14, Federal Reserve Bank of Minneapolis.
Sims, C. A., and Todd, R. M. (1991), "Evaluating Bayesian Vector Autoregressive Forecasting Procedures for Macroeconomic Data," paper presented at the NSF-NBER Seminar on Bayesian Inference in Econometrics and Statistics, St. Paul, April 1991.
Swanson, N., and White, H. (1997), "A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks," *Review of Economics and Statistics*, 79, 540–550.
Telser, L. (1967), "Discrete Samples and Moving Sums in Stationary Stochastic Processes," *Journal of the American Statistical Association*, 62, 484–499.
Tiao, G. C. (1972), "Asymptotic Behavior of Temporal Aggregates of Time Series," *Biometrika*, 59, 525–531.
Tiao, G. C., and Box, G. E. P. (1981), "Modeling Multiple Time Series With Applications," *Journal of the American Statistical Association*, 76, 802–816.
Tiao, G. C., and Tsay, R. S. (1989), "Model Specification in Multivariate Time Series" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 51, 157–213.

——— (1994), "Some Advances in Nonlinear and Adaptive Modelling in Time-Series," *Journal of Forecasting*, 13, 109–131.

Tong, H. (1983), *Threshold Models in Nonlinear Time Series Analysis*, Lecture Notes in Statistics 21, New York: Springer-Verlag.

Tsay, R. S. (1989), "Testing and Modeling Threshold Autoregressive Processes," *Journal of American Statistical Association*, 84, 231–240.

Tsay, R. S., and Tiao, G. C. (1984), "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models," *Journal of the American Statistical Association*, 79, 84–96.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

Zarnowitz, V. (1992), *Business Cycles: Theory, History, Indicators, and Forecasting*, Chicago: University of Chicago Press for NBER.

Zarnowitz, V., and Braun, P. (1993), "Twenty-Two Years of the SPF Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance," in *Business Cycles, Indicators, and Forecasting*, eds. J. H. Stock and M. W. Watson, Chicago: University of Chicago Press for NBER, pp. 11–84.