INTERNATIONAL ECONOMIC REVIEW
Vol. 46, No. 4, November 2005

OPTIMAL FORECAST COMBINATION UNDER REGIME SWITCHING*

BY GRAHAM ELLIOTT AND ALLAN TIMMERMANN[1]

University of California San Diego, U.S.A.

This article proposes a new forecast combination method that lets the combination weights be driven by regime switching in a latent state variable. An empirical application that combines forecasts from survey data and time series models finds that the proposed regime switching combination scheme performs well for a variety of macroeconomic variables. Monte Carlo simulations shed light on the type of data-generating processes for which the proposed combination method can be expected to perform better than a range of alternative combination schemes. Finally, we show how time variations in the combination weights arise when the target variable and the predictors share a common factor structure driven by a hidden Markov process.

1. INTRODUCTION

Forecast combinations have a proven track record.[2] There are good reasons why a combination of forecasts often outperforms the best individual forecasting model. Forecasting models provide simple approximations to data-generating processes that are likely to be far more complicated than assumed by the models. It is therefore unlikely that an individual model forecast encompasses all other models. Even if a single model always generates a lower expected loss than all other forecasts, so long as this model's forecast errors are not perfectly correlated with the forecast errors of competing models, diversification gains are possible from assigning some weight to other forecasts that may, individually, be dominated by the best model.

Consistent with the notion that all forecasting models are likely to be misspecified, the ranking of the forecasting performance of individual models has widely been found to change over time, cf. Stock and Watson (2003) and Aiolfi and Timmermann (2004). Some forecasting models may be able to rapidly adapt to events such as recessions and periods with high uncertainty about the economy's outlook, whereas others only adjust sluggishly. A rational decision maker would want to

* Manuscript received July 2002; revised June 2004.

[1] We thank three referees for many constructive comments on an earlier version of the article and also thank Carlos Capistran, Frank Diebold, Clive Granger, and Jim Hamilton for helpful discussions and comments. We are grateful for financial support from the NSF under grant number SES 0111238. Please address correspondence to: Allan Timmermann, Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508. E-mail: *atimmerm@ucsd.edu*.

[2] Numerous studies have found that forecast combinations tend to perform better than individual forecasts, cf. the comprehensive surveys by Clemen (1989) and Diebold and Lopez (1996) and the more recent work by Chen et al. (1999), Dunis et al. (2000), and Stock and Watson (1998, 1999).

put higher weight on the more adaptive forecasts during—or preferably prior to—such events but would also prefer at other times to use forecasts from more stable models with more precisely estimated parameters. Faced with such changes, it is natural for a decision maker to consider time-varying forecast combination schemes.[3]

Time-varying combination schemes were first proposed in the context of variance–covariance analysis by Granger and Newbold (1973) and in a regression framework by Diebold and Pauly (1987). Some papers have found empirical evidence that such schemes perform better than combinations that assume constant weights. For example, in an inflation forecasting experiment, Deutsch et al. (1994) consider switching regressions in which the regime is determined by some function of the lagged forecast error from a monetarist and a mark-up pricing model. Deutsch et al. use rolling regressions to estimate the parameters underlying forecast combination equations and find that using time-varying combination weights can lead to a substantial reduction in the mean squared forecast error.

This article proposes a new forecast combination method that lets the combination weights be driven by regime switching in a latent state variable. We show how time variations in the combination weights arise when the series that is being predicted and the predictors share a common factor structure driven by a hidden Markov process. This model is consistent with and potentially explains empirical findings of instability in the relative performance of different forecasting models. Although the underlying state is allowed to be unobserved, state probabilities can be filtered from the data and their updated values can be used to compute weighted out-of-sample forecasts. This framework provides a convenient and analytically tractable method for introducing dynamics in the combination weights.

Although there are often good reasons to let forecast combination weights be time varying, the specific form of time variation is often unclear. We therefore consider a variety of schemes, including regime switching driven by an observable state variable (as proposed by Deutsch et al., 1994), rolling window regression, and a time-varying parameter model estimated by the Kalman filter. In an empirical analysis of six popular macroeconomic variables, we study combination of forecasts from a simple autoregressive model with those from the survey of professional forecasters. We find that the proposed regime switching model generates the lowest out-of-sample mean squared forecast error for three of six series whereas none of the other combination schemes produces particularly promising results. In simulation experiments, we further document the types of underlying data-generating processes for which forecast combinations based on the regime switching estimation scheme provide more precise forecasts than alternative methods.

The plan of the article is as follows: Section 2 introduces the forecast combination methods studied in the article and considers their performance for a

[3] In a large study of structural instability, Stock and Watson (1996) report that a majority of macroeconomic time series models undergo structural change, suggesting another reason why relying on a single forecasting model is unlikely to be the best strategy.

range of macroeconomic time series. Section 3 proposes a common factor model with regime switching that gives rise to the regime switching model studied in Section 2. This section also conducts a Monte Carlo study of the performance of various forecasting schemes under a range of alternative data-generating processes. Section 4 concludes.

2. FORECAST COMBINATION METHODS

Assume that a decision maker is interested in predicting some univariate series, y_{t+1}, conditional on information at time t, $\mathcal{I}_t$, which comprises a set of individual forecasts $\hat{\mathbf{y}}_{t+1} = (\hat{y}_{1t+1}, \ldots, \hat{y}_{mt+1})'$ in addition to current and past values of y, i.e., $\mathcal{I}_t = \{\hat{\mathbf{y}}_{\tau+1}, y_\tau\}_{\tau=1}^t$. One option for the decision maker is to simply put a weight of unity on a particular model's forecast and zero weights on all other forecasts. This strategy could make sense if one of the forecasts always dominates the others by some margin in the sense that it is expected to lead to much smaller forecast errors.[4] In practice, forecasting models are likely to be misspecified, so the situation in which a single model always dominates its alternatives is unlikely to arise very often. Furthermore, even if a dominant model exists, it is far from sure that the forecaster will be able to identify this model in small samples. In this situation, an attractive option is to combine forecasts.

The vast majority of studies in the forecasting literature have considered linear forecast combinations of the form

$$y_{t+1} = \omega_0 + \boldsymbol{\omega}'\hat{\mathbf{y}}_{t+1} + \varepsilon_{t+1} \tag{1}$$

where ω_0 is an intercept term and $\boldsymbol{\omega}$ is an m-Vector of regression coefficients or "weights." An intercept term is included following the suggestion by Granger and Ramanathan (1984) to ensure that the bias of the forecast is optimally determined. Assuming that the regression coefficients in (1) are constant, the intercept and weights $(\omega_0, \boldsymbol{\omega})$ can be estimated by OLS using an expanding window of the data, $\mathbf{Y}_{1,t} = (y_1, \ldots, y_t)'$ and $\hat{\mathbf{Y}}_{1,t} = (\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_t)'$.

It is also common to use equal weights, i.e., $\boldsymbol{\omega} = \boldsymbol{\iota}/m$, where $\boldsymbol{\iota}$ is an m-Vector of ones, in the construction of forecasts. Many empirical studies have found that it is difficult to produce more precise forecasts than those generated by such equal weighted combinations, cf. Clemen (1989). Equal weights can be viewed as a way to deal with estimation error in the combination weights when only a relatively short data sample is available. As we will see, instability in the covariance between forecast errors from different models may be a further reason for using equal weights.

2.1. *Time-Varying Combination Weights.*

There are often good reasons to expect that the optimal forecast combination weights change over time. Individual

[4] If the best model only dominates other models by a small margin, diversification gains from combining its forecasts with those from other models could still be sufficiently large to justify forecast combination.

forecasting models can be viewed as local approximations to the true underlying data-generating process and their ability to approximate can be expected to change in the presence of structural breaks and as a function of the state of the economy. This suggests extending (1) to consider time-varying forecast combination weights

$$y_{t+1} = \omega_{0t} + \boldsymbol{\omega}_t' \hat{\mathbf{y}}_{t+1} + \varepsilon_{t+1} \tag{2}$$

where $(\omega_{0t}, \boldsymbol{\omega}_t)$ are adapted to the current information set, $\mathcal{I}_t$. In the absence of specific knowledge about the process giving rise to time variations in the combination weights, it is natural to consider several combination methods. We discuss such schemes in the following.

A commonly used method for dealing with slowly moving nonstationarities in a regression context is to use a rolling estimation window. To estimate the combination weights used to forecast y_{t+1}, rolling window regressions make use of the data $\mathbf{Y}_{t-c+1,t} = (y_{t-c+1}, \ldots, y_t)$ and $\hat{\mathbf{Y}}_{t-c+1,t} = (\hat{\mathbf{y}}_{t-c+1}, \ldots, \hat{\mathbf{y}}_t)'$, where c is the (fixed) window length. A problem with this approach is that it removes data in an arbitrary fashion without basing this decision on tests for structural breaks or other types of nonstationarity. Hence, it is easy to construct examples in which this scheme does poorly, despite its prominent use, cf., Pesaran and Timmermann (2003).

A second method that deals with instability in the combination weights assumes that (2) takes the form of a time-varying parameter (TVP) model

$$\begin{aligned} y_{t+1} &= \boldsymbol{\beta}_t' \mathbf{z}_{t+1} + \varepsilon_{t+1}, \varepsilon_{t+1} \sim \text{i.i.d.}\left(0, \boldsymbol{\sigma}_\varepsilon^2\right) \\ \boldsymbol{\beta}_t' &= \boldsymbol{\beta}_{t-1}' + \boldsymbol{\eta}_t, \ \boldsymbol{\eta}_t \sim \text{i.i.d.}\left(0, \boldsymbol{\sigma}_\eta^2\right), \ \text{Cov}\left(\varepsilon_{t+1}, \boldsymbol{\eta}_t\right) = \mathbf{0} \end{aligned} \tag{3}$$

Here $\mathbf{z}_{t+1} = (1\hat{\mathbf{y}}_{t+1}')'$ and $\boldsymbol{\beta}_t' = (\omega_{0t}\ \boldsymbol{\omega}_t)'$. This type of forecast combination scheme has been considered by, e.g., Zellner et al. (1991).

A third combination method proposed by Deutsch et al. (1994) (DGT in the sequel) lets the combination weights switch discretely through time, driven by an observable state variable such as the current forecast error

$$y_{t+1} = I_{\mathbf{e}_t \in A}(\omega_{01} + \boldsymbol{\omega}_1' \hat{\mathbf{y}}_{t+1}) + (1 - I_{\mathbf{e}_t \in A})(\omega_{02} + \boldsymbol{\omega}_2' \hat{\mathbf{y}}_{t+1}) + \varepsilon_{t+1} \tag{4}$$

Here $\mathbf{e}_t = \iota y_t - \hat{\mathbf{y}}_t$ is the vector of period-t forecast errors and $I_{\mathbf{e}_t \in A}$ is an indicator function taking the value unity when $\mathbf{e}_t \in A$ and zero otherwise. For example, the set A could be based on the sign of the first models's forecast error, so that $I_{\mathbf{e}_t \in A}$ is unity if the forecast error is positive, zero otherwise. Hence the forecast errors, $\mathbf{e}_t$, determine which of the two sets of combination weights—$\boldsymbol{\omega}_1$ or $\boldsymbol{\omega}_2$—is used to compute the forecast. As the set A is varied, different nonlinear forecasting models are obtained. Note that $I_{\mathbf{e}_t \in A}$ is known given $\mathcal{I}_t$, so it can be conditioned upon when forecasting y_{t+1}.

2.1.1. *Regime switching weights.* To be practically useful, time-varying optimal forecast combination weights must possess several properties. First, since the

relative ranking of different forecasting models has been found to change over time, the bias and variance of the forecast errors generated by the individual models should be allowed to be time varying. Furthermore, the correlation between the prediction errors should be allowed to vary through time in a persistent manner, thus giving rise to persistent shifts in the optimal combination weights. If changes in the relative performance and correlations between forecast errors are not at least mildly persistent it will be difficult to design a time-varying forecast combination method that can hope to outperform combination methods based on constant weights. Finally it is also important to account for the possibility of non-Gaussian forecast errors since this is again a standard feature of many time series that economists are interested in.

Here, we propose a simple regime switching combination scheme that captures these features. Assume that the joint distribution of the target variable and the vector of forecasts is driven by a latent state variable, S_{t+1}, which assumes one of k possible values, i.e., $S_{t+1} \in (1, \ldots, k)$. S_{t+1} is not assumed to be known, so $S_{t+1} \notin \mathcal{I}_t$. However, conditional on $\mathcal{I}_t$ and the underlying state, $S_{t+1} = s_{t+1}$, we assume that the joint distribution of y_{t+1} and $\hat{\mathbf{y}}_{t+1}$ is Gaussian

$$\begin{pmatrix} y_{t+1} \\ \hat{\mathbf{y}}_{t+1} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{ys_{t+1}} \\ \boldsymbol{\mu}_{\hat{\mathbf{y}}s_{t+1}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{ys_{t+1}} & \boldsymbol{\sigma}'_{y\hat{\mathbf{y}}s_{t+1}} \\ \boldsymbol{\sigma}_{y\hat{\mathbf{y}}s_{t+1}} & \boldsymbol{\Sigma}_{\hat{\mathbf{y}}s_{t+1}} \end{pmatrix} \right) \tag{5}$$

Following Hamilton (1989), we further assume that the states are generated by a first-order Markov chain with transition probability matrix

$$\mathbf{P}(s_{t+1} \mid s_t) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & \vdots \\ \vdots & \vdots & \cdots & p_{k-1k} \\ p_{k1} & \cdots & p_{kk-1} & p_{kk} \end{pmatrix} \tag{6}$$

Conditional on $S_{t+1} = s_{t+1}$, the expectation of y_{t+1} is linear in the prediction signals and thus takes the form of state-dependent intercept and combination weights

$$E[y_{t+1} \mid \mathcal{I}_t, s_{t+1}] = \mu_{ys_{t+1}} + \boldsymbol{\sigma}'_{y\hat{\mathbf{y}}s_{t+1}} \boldsymbol{\Sigma}^{-1}_{\hat{\mathbf{y}}s_{t+1}} (\hat{\mathbf{y}}_{t+1} - \boldsymbol{\mu}_{\hat{\mathbf{y}}s_{t+1}}) \tag{7}$$

where we recall that $\hat{\mathbf{y}}_{t+1} \in \mathcal{I}_t$. Hence, if the future value of the state variable, S_{t+1}, was known, the population value of the optimal combination weights under mean squared forecast error (MSFE) loss would be

$$\begin{aligned} \omega_{0s_{t+1}} &= \mu_{ys_{t+1}} - \boldsymbol{\sigma}'_{y\hat{\mathbf{y}}s_{t+1}} \boldsymbol{\Sigma}^{-1}_{\hat{\mathbf{y}}s_{t+1}} \boldsymbol{\mu}_{\hat{\mathbf{y}}s_{t+1}} \\ \boldsymbol{\omega}_{s_{t+1}} &= \boldsymbol{\Sigma}^{-1}_{\hat{\mathbf{y}}s_{t+1}} \boldsymbol{\sigma}_{y\hat{\mathbf{y}}s_{t+1}} \end{aligned} \tag{8}$$

In practice, the underlying state is likely to be unobservable, so the decision maker is interested in minimizing the MSFE conditional only on the current information, $\mathcal{I}_t$

$$E\left[e_{t+1}^2 \mid \mathcal{I}_t\right] = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t}\left\{\mu_{es_{t+1}}^2 + \sigma_{es_{t+1}}^2\right\} \tag{9}$$

where $e_{t+1} = y_{t+1} - \hat{y}_{t+1}$ is the scalar forecast error from the combination ($\hat{y}_{t+1} = \hat{\omega}_{0t} + \hat{\boldsymbol{\omega}}_t'\hat{\mathbf{y}}_{t+1}$), $\pi_{s_{t+1},t} = \Pr(S_{t+1} = s_{t+1} \mid \mathcal{I}_t)$ is the (filtered) probability of being in state s_{t+1} in period $t+1$ conditional on current information, $\mathcal{I}_t$, and, assuming a linear forecast combination in the general class (2),

$$\begin{aligned} \mu_{es_{t+1}} &= E[e_{t+1} \mid \mathcal{I}_t, s_{t+1}] = \mu_{ys_{t+1}} - \omega_{0t} - \boldsymbol{\omega}_t'\boldsymbol{\mu}_{\hat{\mathbf{y}}s_{t+1}}, \\ \sigma_{es_{t+1}}^2 &= Var(e_{t+1} \mid \mathcal{I}_t, s_{t+1}) = \sigma_{ys_{t+1}}^2 + \boldsymbol{\omega}_t'\boldsymbol{\Sigma}_{\hat{\mathbf{y}}s_{t+1}}\boldsymbol{\omega}_t - 2\boldsymbol{\omega}_t'\boldsymbol{\sigma}_{y\hat{\mathbf{y}}s_{t+1}} \end{aligned} \tag{10}$$

Differentiating (9) with respect to ω_{0t} and $\boldsymbol{\omega}_t$ and solving the first-order conditions, we get the following combination weights, ω_{0t}^*, $\boldsymbol{\omega}_t^*$

$$\begin{aligned} \omega_{ot}^* &= \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t}\mu_{ys_{t+1}} - \left(\sum_{s_{t+1}=1}^{k} \hat{\pi}_{s_{t+1},t}\boldsymbol{\mu}_{\hat{\mathbf{y}}s_{t+1}}'\right)\boldsymbol{\omega}_t \equiv \bar{\mu}_{yt} + \bar{\boldsymbol{\mu}}_{\hat{\mathbf{y}}t}'\boldsymbol{\omega}_t, \\ \boldsymbol{\omega}_t^* &= \left(\sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t}(\boldsymbol{\mu}_{\hat{\mathbf{y}}s_{t+1}}\boldsymbol{\mu}_{\hat{\mathbf{y}}s_{t+1}}' + \boldsymbol{\Sigma}_{\hat{\mathbf{y}}s_{t+1}}) - \bar{\boldsymbol{\mu}}_{\hat{\mathbf{y}}t}\bar{\boldsymbol{\mu}}_{\hat{\mathbf{y}}t}'\right)^{-1} \\ &\quad \times \left(\sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t}(\mu_{ys_{t+1}}\boldsymbol{\mu}_{\hat{\mathbf{y}}s_{t+1}} + \boldsymbol{\sigma}_{y\hat{\mathbf{y}}s_{t+1}}) - \bar{\mu}_{yt}\bar{\boldsymbol{\mu}}_{\hat{\mathbf{y}}t}\right) \end{aligned} \tag{11}$$

where $\bar{\mu}_{yt} = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t}\mu_{ys_{t+1}}$ and $\bar{\boldsymbol{\mu}}_{\hat{\mathbf{y}}t} = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t}\boldsymbol{\mu}_{\hat{\mathbf{y}}s_{t+1}}$. This generalizes the standard least squares formula for optimal forecast combination weights derived in a regression context by Granger and Ramanathan (1984), which of course is obtained when $k = 1$. It is clear from (11) that the optimal combination weights will, in general, be state dependent in this setting with weights that vary over time as the state probabilities are updated.[5] Consistent estimates of the combination weights for each state can readily be computed by estimating the model

$$y_\tau = \omega_{0s_\tau} + \boldsymbol{\omega}_{s_\tau}'\hat{\mathbf{y}}_\tau + \varepsilon_\tau,\ \varepsilon_\tau \sim N\left(0, \sigma_{\varepsilon s_\tau}^2\right),\ \tau = 1, \ldots, t \tag{12}$$

to get parameter estimates $(\hat{\omega}_{0s_\tau}, \hat{\boldsymbol{\omega}}_{s_\tau})$ and state probabilities, $\pi_{s_t,t} = \Pr(S_t = s_t \mid \mathcal{I}_t)$. One-step-ahead state probabilities $\pi_{t+1,t} = \Pr(S_{t+1} = s_{t+1} \mid \mathcal{I}_t)$ can be computed from the filtered state probabilities $\pi_{t,t} = \Pr(S_t = s_t \mid \mathcal{I}_t)$ using the transition probabilities $p_{s_t,s_{t+1}}$ (cf. Hamilton, 1994, pp. 694–95). The (conditionally) optimal forecast of y_{t+1} is then obtained as

[5] As pointed out by a referee, this assumes, however, that the rank of the transition probability matrix, **P**, exceeds 1. If this condition is not satisfied, the weights need not vary even if the number of states, k, is greater than 1.

$$E[y_{t+1} \mid \mathcal{I}_t] = \sum_{s_{t+1}=1}^{k} \pi_{t+1,t} E[y_{t+1} \mid \mathcal{I}_t, s_{t+1}] \tag{13}$$

2.2. *Empirical Evidence.* To find out how well the combination schemes described so far work in practice, we consider predictions of six popular macroeconomic variables, namely the nominal Gross Domestic Product (GNP before 1992), inflation as measured through the GDP chain-weighted price index (GDPPI), corporate profits after taxes (CPAT), the civilian unemployment rate (CUR) computed as a three-month average, the industrial production index (IPI), and private housing starts (HS) measured in millions and computed as a three-month average. For all series except for the unemployment rate we study the rate of change in next quarter's value over the current value, calculated as first differences in logs. All series are seasonally adjusted.

We combine forecasts from survey data with forecasts from simple time-series regressions. To the extent that rises and falls in economic activity are not equally well predicted by means of leading indicators available to the survey participants (but not used in the autoregressive models), one would expect that the relative weights on the two forecasts should vary over time. Time-series forecasts could also be affected more by structural breaks in the data-generating process than survey forecasts provided that the forecasters are aware of and account for the presence of such breaks.

Survey forecasts were obtained from the Survey of Professional Forecasters (SPF, formerly known as the ASA/NBER survey) and are available from the Philadelphia Fed's Web site. Forecasts were based on the mean across survey participants, computed again as growth rates (i.e., first differences in the logarithm of the one-quarter ahead forecast relative to the current value) except for the unemployment rate that was not transformed. By design, these are real-time forecasts. Forecasts from time-series regressions were computed from simple autoregressive (AR) models with lag length selected according to the Schwarz information criterion (SIC). Pseudo real-time one-step-ahead forecasts, $\hat{y}_{t+1}^{AR}$, were computed using parameters estimated from the AR model at time t. The relative squared forecast error values of the time-series and survey forecasts are positively correlated for most of the variables under consideration. Since both types of forecasts are affected by the same unpredictable shocks to the target variables, this is not surprising.

The sample covers the period from the fourth quarter of 1968 to the first quarter of 2003. To make sure that we have enough data points to estimate the parameters reasonably precisely, we reserve the last 40 data points for the out-of-sample experiment. This gives us about 100 initial observations to estimate the combination weights.

We compare the performance of nine different forecasting schemes, namely (i) forecasts from the time-series model; (ii) SPF forecasts; (iii) combined forecasts with weights estimated using an expanding window of the data; (iv) combined forecasts with weights estimated using a rolling window with 40 observations or 10 years of data; (v) combined forecasts based on Markov switching weights; (vi) combined forecasts that use equal weights; (vii) combined forecasts with

weights based on the DGT model driven by the sign of the lagged forecast errors from the time-series model; (viii) combined forecasts with weights based on the DGT model driven by the sign of the lagged forecast errors from the SPF data; (ix) combined forecasts with weights based on the TVP model estimated from the Kalman filter.

An important issue that arises when forecasting with the regime switching model is how many states (k) to use. To answer this question, we experimented with both two-state and three-state models. There are good a priori grounds for preferring the two-state model in our application. The single-state (linear) model is already included in the comparisons. Furthermore, given the short sample it is difficult to estimate the parameters of a three-state model sufficiently precisely to produce good out-of-sample forecasts. In fact, the two-state forecast combination model requires the estimation of 10 parameters whereas the three-state specification requires estimation of 18 parameters.[6]

To explore whether benefits can be gained in real time from identifying two economic states and dynamically combining the autoregressive and survey forecasts using the estimated state probabilities as weights, we conduct an out-of-sample forecasting exercise that estimates at each point in time the probability of the current state, s_t, as well as the parameters, $\hat{\boldsymbol{\theta}}_{s_t} = \{\hat{\omega}_{0s_t}, \hat{\omega}_{1s_t}, \hat{\omega}_{2s_t}, \hat{\sigma}_{s_t}, \hat{p}_{s_t s_t}\}$ and use these to compute one-step-ahead forecasts through (13).

Table 1 reports the out-of-sample forecasting performance produced by the different forecasting schemes. For three of the six series—the unemployment rate, inflation, and GDP growth—the two-state regime switching combination scheme produces the lowest MSFE values. The survey forecasts generate the lowest MSFE values for two series, namely housing starts and corporate profits, whereas the time-series forecasts are best for industrial production.

Compared to the other time-varying forecast combination methods the regime switching method produces a lower MSFE in four (against the rolling window and Kalman filter forecasts) or five (against the DGT forecasts) out of six cases. This provides fairly compelling evidence in support of the proposed two-state regime switching scheme. It further suggests that the best forecast combination method allows the combination weights to vary over time but in a mean-reverting manner. Unsurprisingly, allowing for three states leads to worse forecasting performance for four of the six variables under consideration.

Plots of the filtered state probabilities for the forecast combination regression that assumes two states are shown in Figure 1, ordered so that state 1 is the state in which the weight on the time-series forecast is high relative to its value in state 2. For housing starts, the unemployment rate, and growth in industrial production the time series indicate multiple switches back and forth between the two regimes, although there is evidence of a more permanent break toward the end of the sample for housing starts and industrial production. For the remaining three variables the evidence supports a single structural break interpretation. Due to

[6] We also compared the value of information criteria for the single-, two-, and three-state models. For all six series the Schwarz information criterion selected a two-state model. The Akaike information criterion also supported a two-state specification for the majority of the series.

TABLE 1

EMPIRICAL OUT-OF-SAMPLE FORECASTING PERFORMANCE

	Housing Starts	Industrial Production	Unemployment Rate	Corporate Profits	GDP Producer Price Index	Nominal GDP
A. Root mean squared forecast errors						
Expanding window weights	4.703	0.873	0.209	4.548	0.187	0.523
Markov switching (2 states)	4.815	0.864	0.201	4.572	0.172	0.493
Markov switching (3 states)	4.714	0.805	0.209	4.600	0.176	0.507
AR forecasts	4.860	0.732	0.232	4.607	0.222	0.635
SPF	4.163	0.955	0.226	4.452	0.198	0.515
Equal weights	4.385	0.825	0.221	4.482	0.196	0.518
Rolling window weights	4.391	0.811	0.226	4.872	0.180	0.520
DGT-1	4.968	0.802	0.211	4.769	0.186	0.501
DGT-2	4.701	0.803	0.204	4.638	0.195	0.538
TVP	4.351	0.780	0.205	4.591	0.186	0.502
B. Diebold–Mariano statistics: Two-state Markov switching versus alternative models						
vs. expanding window	−0.66	0.24	0.51	−1.14	1.94	1.31
vs. AR forecasts	0.12	−2.41	1.96	0.32	2.68	2.39
vs. SPF	−2.37	1.43	1.69	−0.82	2.16	1.00
vs. equal weights	−1.48	−0.80	1.41	−1.13	1.88	0.82
vs. rolling window	−1.42	−0.95	1.28	1.64	1.28	1.38
vs. DGT-1	0.89	−1.10	0.45	1.06	2.61	0.40
vs. DGT-2	−0.55	−1.12	0.22	0.42	2.59	1.48
vs. TVP	−1.65	−1.95	0.25	0.32	1.13	0.70
C. Giacomini–White statistics: Two-state Markov switching versus alternative models						
vs. expanding window	0.44	0.06	0.27	1.29	3.53	1.69
vs. AR forecasts	0.01	5.20	3.57	0.10	6.22	5.13
vs. SPF	5.04	2.00	2.72	0.68	4.28	1.01
vs. equal weights	2.14	0.64	1.95	1.28	3.34	0.68
vs. rolling window	1.96	0.90	1.62	2.58	1.62	1.86
vs. DGT-1	0.80	1.21	0.20	1.13	5.93	0.16
vs. DGT-2	0.30	1.25	0.05	0.18	5.86	2.14
vs. TVP	2.59	3.56	0.06	0.11	1.27	0.50

NOTES: Details of the forecasting methods are explained in the text. AR represent forecasts from autoregressive models, SPF are forecasts from the Survey of Professional Forecasters. The rolling window assumes a window length of 40 observations. DGT-1 and DGT-2 assume a nonlinear forecast combination model with weights based on the sign of the AR or SPF forecast errors. TVP weights are estimated using the Kalman filter. DM test statistics are asymptotically distributed as a standard normal random variable with positive values indicating that the two-state regime switching model performs better than the alternative model. The Giacomini–White test statistic is asymptotically distributed as a chi-squared random variable with one degree of freedom.

such structural breaks, it is unsurprising that the correlations between the NBER recession indicator and these state probabilities plotted in Figure 1 tend to be very weak and below 0.10. Many of the state probabilities are, however, strongly correlated: For example, for housing starts and industrial production or housing starts and producer price changes, pairwise correlations are close to 0.60.

The plots in Figure 1 do not reveal the extent of time-variation in combination weights, which depends, of course, on the difference between the coefficient estimates ω_{s_t} across the two states. To gain insights into such variations, Figure 2 plots the combination weights associated with the time-series and survey forecasts. Many interesting observations emerge from these plots. Note the strong negative

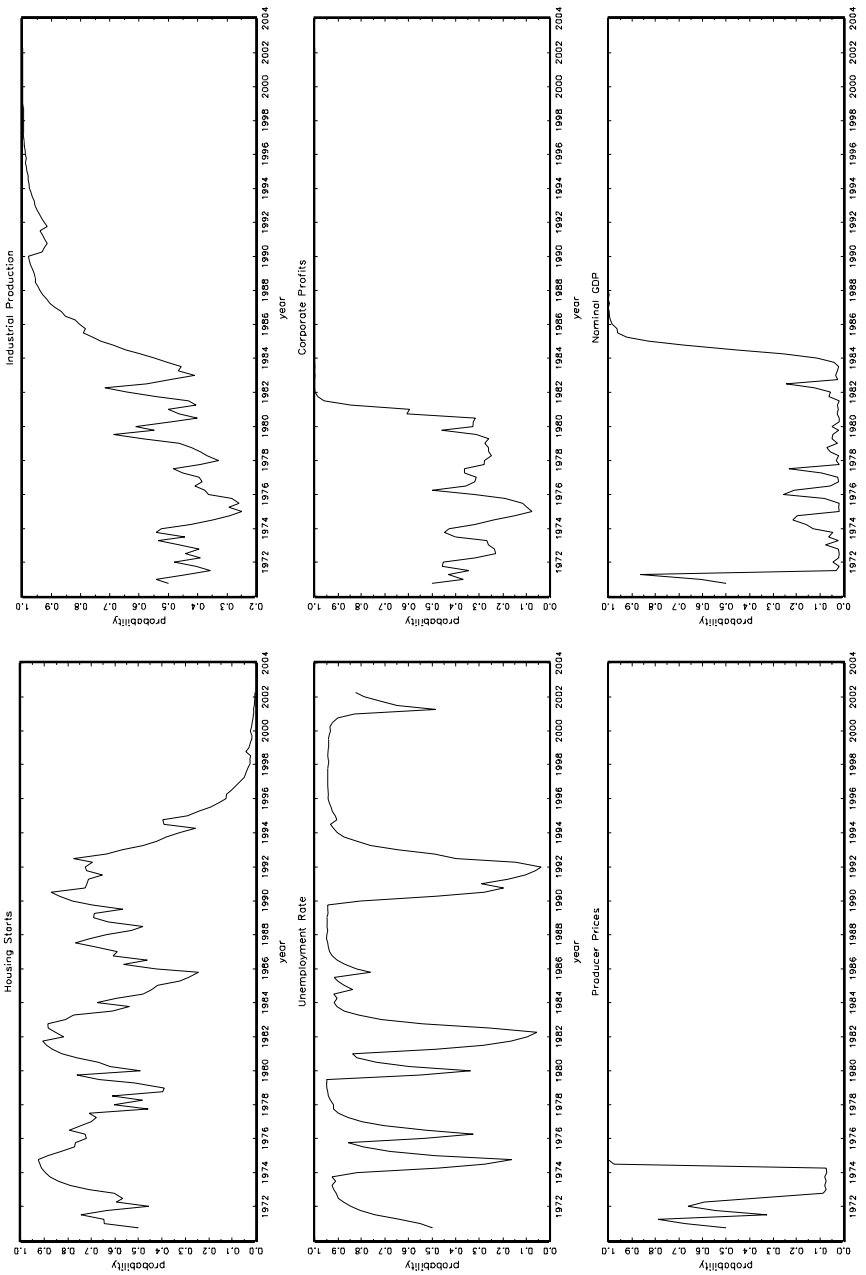


FIGURE 1

FILTERED STATE-1 PROBABILITIES

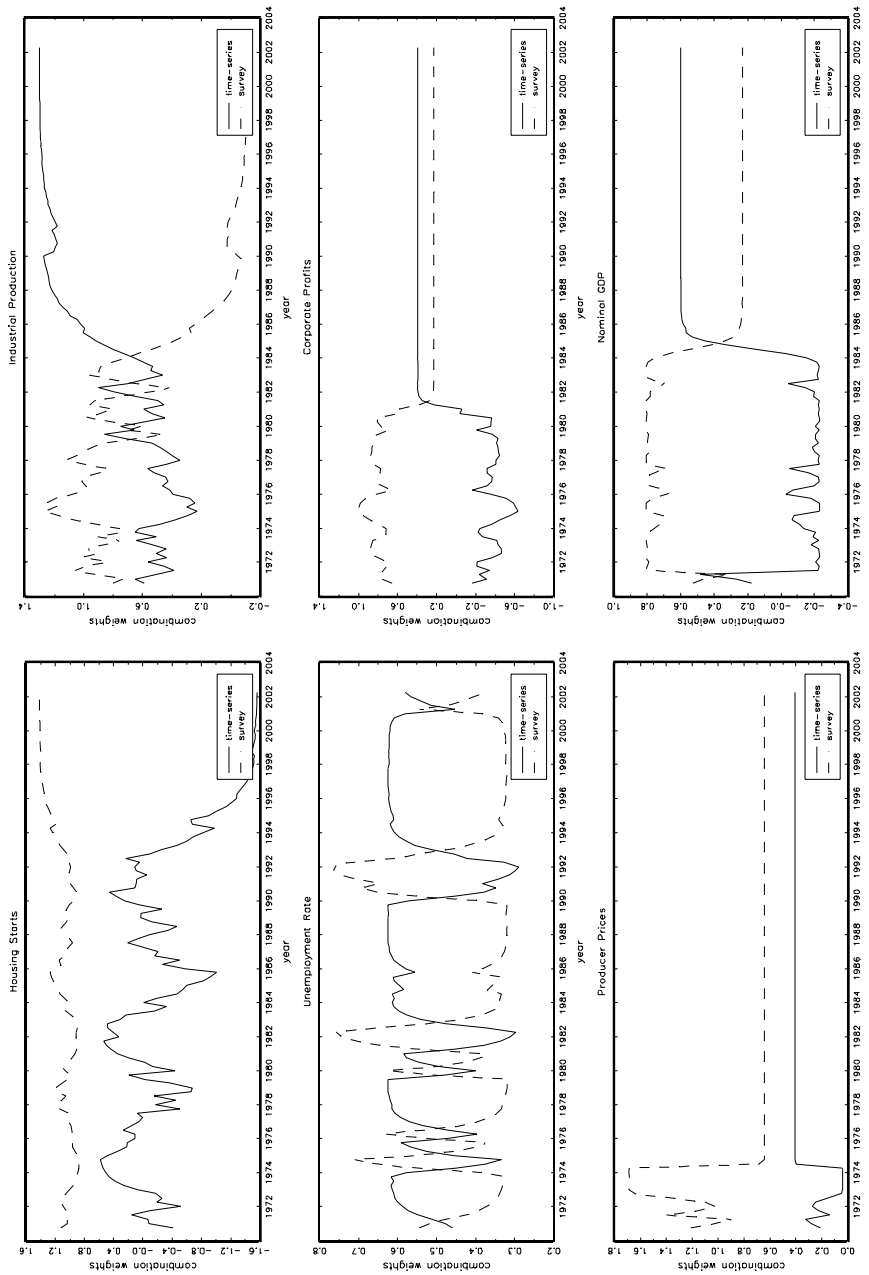


FIGURE 2

COMBINATION WEIGHTS

correlation between pairs of combination weights for all series except for housing starts. As the weight on one forecast goes up, the weight on the other forecast declines. This is to be expected from the strong positive correlation between the two sets of forecasts.

For housing starts the weight on the survey forecast is relatively stable and large (around one), whereas the weight on the time-series forecast is small and declining. A very different pattern is observed for the industrial production variable in which the weight on the time-series forecast increases over time, while conversely the weight on the survey forecast declines. For the three variables most strongly affected by a single structural break we observe similar patterns: In the case of corporate profits both combination weights stabilize at about 0.3 after the break around 1982; for the producer price series the survey weight stabilizes around 0.6 whereas the time-series weight stabilizes around 0.4; finally for the nominal GDP series the time-series weight stabilizes around 0.6, survey weight stabilizing around 0.3 after the break around 1985.

The finding of a structural break in the combination weights of many of our series could well be related to the structural decline in the volatility of real U.S. GDP growth identified in the first quarter of 1984 by McConnell and Perez-Quiros (2000). This is close to the time where the combination weights shift systematically for the industrial production, corporate profits, and nominal GDP series.

2.3. *Forecast Evaluation.* To assist in evaluating the forecasting performance of the Markov switching combination model against that of the alternative forecasting schemes, we compute Diebold–Mariano (1995) statistics based on the sequence of recursive one-step-ahead out-of-sample forecast errors. The values of this test statistic should be interpreted with caution and are best viewed mostly as a diagnostic since the asymptotic distribution of the test statistic depends on the sampling experiment that the researcher has in mind and the test also ignores parameter estimation uncertainty. Assuming nested models, Corradi and Swanson (2002) show that, provided the same loss function is used in the in-sample and out-of-sample estimation, parameter estimation error will vanish asymptotically. Here we assume that the benchmark or null model in the tests is the regime switching model. This means that the alternative models do not nest the null model, as is appropriate when using the Diebold–Mariano test.

Results are provided in panel B of Table 1. The majority of test statistics are not statistically significant even under conventional critical values, which there may be good reasons not to rely on. Such a finding is unsurprising in the light of the relative short out-of-sample period and the poor small-sample properties of out-of-sample tests for predictive ability documented in Chao et al. (2001). Even so, it is clear that the regime switching forecasts generally do quite well against the other forecasting methods, particularly for producer price inflation and nominal GDP growth.

To account for the effect of estimation uncertainty on forecasting performance and to have a method that applies both to nested and nonnested models, we next computed the test statistic proposed in corollary 2 in Giacomini and White (2004).

This is based on the uncentered squared multiple correlation coefficient from a regression of loss differentials on a vector of instruments. We used a constant as our only instrument, so this test statistic is asymptotically distributed as a chi-squared variable with one degree of freedom. Results are reported in panel C of Table 1 and are consistent with the Diebold–Mariano test statistics, indicating that the two-state model does well for the combined forecasts of the unemployment rate, inflation, and nominal GDP growth.

3. A MARKOV SWITCHING COMMON FACTOR MODEL

Although dynamic—or time-varying—forecast combination weights have been proposed in the literature as early as Bates and Granger (1969) and have further been advanced in textbooks on forecasting such as Diebold (2001), they have not proved to be of much practical use so far. One reason is that only a little work has been undertaken on understanding which mechanism gives rise to time variation in the optimal combination weights. In this section, we therefore propose a simple factor model that generates regime switching in the optimal combination weights. Doing so leads to a theoretically reasonable model that generates time variations in the forecast combination weights. The model also makes it easier to interpret and choose the parameters in simulation experiments, a feature we exploit in Section 3.2. Although, to our knowledge, this type of model has not previously been considered in the context of forecast combination, we will see that it is ideally suited for this purpose.[7]

3.1. *The Model.* Assume that y_{t+1} is a linear function of an $n \times 1$ vector of unobserved factors, $\mathbf{F}_{t+1}$, with loading coefficients, $\boldsymbol{\beta}'_{ys_{t+1}}$, that depend on some underlying state process, $S_{t+1} \in \{1, \ldots, k\}$

$$y_{t+1} = \mu_{ys_{t+1}} + \boldsymbol{\beta}'_{ys_{t+1}} \mathbf{F}_{t+1} + \varepsilon_{t+1}, \tag{14}$$
$$\varepsilon_{t+1} \sim N\left(0, \sigma^2_{\varepsilon s_{t+1}}\right)$$

The mean of y_{t+1} is controlled by the state-dependent scalar intercept, $\mu_{ys_{t+1}}$ and y_{t+1} is also affected by an unpredictable innovation term, ε_{t+1}, whose state-dependent variance is $\sigma^2_{\varepsilon s_{t+1}}$. The $n \times n$ variance–covariance matrix of $\mathbf{F}_{t+1}(\boldsymbol{\Sigma}_{Fs_{t+1}})$ and the $n \times 1$ vector of factor loadings $(\boldsymbol{\beta}_{ys_{t+1}})$ can also vary across states. $\mathbf{F}_{t+1}$ is a vector of "common factors" that are assumed to be correlated with the prediction signals and are Gaussian conditional on S_{t+1}.

Conditional on a given state $(S_{t+1} = s_{t+1})$, y_{t+1} is therefore Gaussian, but the states are assumed to be unobserved, so that y_{t+1} can be strongly non-Gaussian conditional only on $\mathcal{I}_t$. Although S_{t+1} and $\mathbf{F}_{t+1}$ are assumed to be unknown at the time when the prediction is computed, the forecaster is again assumed to observe

[7] Deutsch et al. (1994) consider a range of nonlinear models but do not study latent Markov models of the type analyzed here.

m prediction signals, $\{\hat{y}_{it+1}\}_{i=1}^{m}$ each of which reflects the true factors ($\mathbf{F}_{t+1}$) and noise, ς_{it+1}[8]

$$\hat{y}_{it+1} = \mu_{is_{t+1}} + \boldsymbol{\beta}'_{is_{t+1}}\mathbf{F}_{t+1} + \varsigma_{it+1}, \quad i = 1, \ldots, m, \tag{15}$$
$$\varsigma_{it+1} \sim N(0, \sigma^2_{is_{t+1}})$$

The noise terms, ς_{it+1}, are assumed to be mutually uncorrelated and uncorrelated both with ε_{t+1} and $\mathbf{F}_{t+1}$ and are thus uncorrelated with the innovation in y_{t+1}

$$E[\varepsilon_{t+1}\varsigma_{it+1}] = E[\varsigma_{jt+1}\varsigma_{it+1}] = 0 \quad \text{for all } t \text{ and all } i \neq j, \tag{16}$$
$$E[\mathbf{F}_{t+1}\varsigma_{jt+1}] = \mathbf{0}, \text{ for all } t \text{ and all } j$$

We also assume that the noise terms $\{\varepsilon_{t+1}, \varsigma_{1t+1}, \ldots, \varsigma_{mt+1}\}$ are serially uncorrelated. This condition can be relaxed at the cost of introducing extra terms in the conditional mean of $\hat{y}_{it+1}$ and y_{t+1}.

This model is quite general. The individual forecasts can be biased, conditionally as well as unconditionally, which happens when $\mu_{is_{t+1}} \neq \mu_{ys_{t+1}}$ for at least one value of $s_{t+1} \in (1, \ldots, k)$. The mean parameters control the centering of the distribution of the forecasts and of the target variable, y_{t+1}. Even the best forecast is never perfectly correlated with the realization provided that $\sigma_{\varepsilon s_{t+1}} \neq 0$. The variance of this "noisy," unpredictable part of y is allowed to vary across states. The forecasts are also contaminated by noise, possibly due to their inclusion of irrelevant variables reflected in ς_{it+1}. Again this source of noise is state dependent and can thus vary over time.

For a given state, a good individual forecasting model has factor loadings ($\boldsymbol{\beta}_{is_{t+1}}$) that are similar to those of the predicted variable ($\boldsymbol{\beta}_{ys_{t+1}}$) and also has a low variance of the noise component, $\sigma^2_{is_{t+1}}$. The ranking of different forecasts can thus change across different states in a way that gives rise to time variation in the optimal forecast combination weights. To see this, assume that there are two states ($k = 2$) and that the loading of y on factor 1 is high in state 1 and low in state 2, whereas the loading of y on factor 2 is high in state 2 and low in state 1. Further assume that the first prediction model has a high value of the first element of $\boldsymbol{\beta}_{1s_{t+1}}$ and a low value of $\sigma_{1s_{t+1}}$ when $s_{t+1} = 1$, whereas conversely the elements of $\boldsymbol{\beta}_{1s_{t+1}}$ are low and $\sigma_{1s_{t+1}}$ is high when $s_{t+1} = 2$. If the opposite conditions hold for the second prediction model, then a higher combination weight should be put on the first prediction model whenever the probability of state 1 is high, whereas the second prediction model should get a higher weight when state 2 has a high probability.

The dynamics of the optimal forecast combination weights are controlled by the state transition probability matrix, $\mathbf{P}$. This means our model can flexibly capture the idea that the identity of the best forecasting model can change over time in a

[8] Provided that the correlation between the factors and the target variable is time varying (which is ensured if $\boldsymbol{\beta}_{ys_{t+1}}$ is state dependent), the optimal combination weights will be state dependent irrespective of whether the factor loadings of the forecasts depend on the underlying state variable.

manner that is likely to be persistent but ultimately mean-reverts to weights that reflect the steady-state probabilities.

In this model, the conditional expectation of y_{t+1} given s_{t+1} is linear in the prediction signals and thus takes the form of a state-dependent intercept and vector of combination weights

$$E[y_{t+1} \mid \mathcal{I}_t, s_{t+1}] = \mu_{ys_{t+1}} + \boldsymbol{\beta}'_{ys_{t+1}} \boldsymbol{\Sigma}_{Fs_{t+1}} \mathbf{B}'_{s_{t+1}} \times (\mathbf{B}_{s_{t+1}} \boldsymbol{\Sigma}_{Fs_{t+1}} \mathbf{B}'_{s_{t+1}} + \boldsymbol{\Sigma}_{\varsigma s_{t+1}})^{-1} (\hat{\mathbf{y}}_{t+1} - \boldsymbol{\mu}_{\cdot s_{t+1}})$$

where $\boldsymbol{\mu}_{\cdot s_{t+1}} = (\mu_{1s_{t+1}}, \mu_{2s_{t+1}}, \ldots, \mu_{ms_{t+1}})'$, $\boldsymbol{\Sigma}_{\varsigma s_{t+1}} = E[\boldsymbol{\varsigma}_{t+1} \boldsymbol{\varsigma}'_{t+1} \mid s_{t+1}]$ is a diagonal covariance matrix of $\boldsymbol{\varsigma}_{t+1} = (\varsigma_{1t+1}, \ldots, \varsigma_{mt+1})'$, and $\mathbf{B}'_{s_{t+1}} = (\boldsymbol{\beta}_{1s_{t+1}}, \boldsymbol{\beta}_{2s_{t+1}}, \ldots, \boldsymbol{\beta}_{ms_{t+1}})$ is an $n \times m$ matrix of loading factors for the m forecasts on the n underlying factors in state s_{t+1}. This is similar to (7). In reality, the true underlying state is likely to be unobserved so the combined forecast has to be computed using state probability weights and $(\omega_{0t}, \boldsymbol{\omega}_t)$ gets based on the equivalent to (13)

(17)

$$\omega_{0t} = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} \left(\mu_{ys_{t+1}} - \boldsymbol{\beta}'_{ys_{t+1}} \boldsymbol{\Sigma}_{Fs_{t+1}} \mathbf{B}'_{s_{t+1}} (\mathbf{B}_{s_{t+1}} \boldsymbol{\Sigma}_{Fs_{t+1}} \mathbf{B}'_{s_{t+1}} + \boldsymbol{\Sigma}_{\varsigma s_{t+1}})^{-1} \boldsymbol{\mu}_{\cdot s_{t+1}} \right),$$

$$\boldsymbol{\omega}'_t = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} \left(\boldsymbol{\beta}'_{ys_{t+1}} \boldsymbol{\Sigma}_{Fs_{t+1}} \mathbf{B}'_{s_{t+1}} (\mathbf{B}_{s_{t+1}} \boldsymbol{\Sigma}_{Fs_{t+1}} \mathbf{B}'_{s_{t+1}} + \boldsymbol{\Sigma}_{\varsigma s_{t+1}})^{-1} \right)$$

where again $\pi_{s_{t+1},t} = \Pr(S_{t+1} = s_{t+1} \mid \mathcal{I}_t)$ is the future-state probability obtained from the transition probabilities and the (optimally) filtered-state probabilities. This means that the conditional distribution of y_{t+1} given $\hat{\mathbf{y}}_{t+1}$ becomes a time-varying mixture of normals that can approximate a large class of densities, cf. Marron and Wand (1992) and Timmermann (2000).

3.2. *Monte Carlo Simulations.* Armed with the factor model, we can study the comparative performance of the forecasting schemes introduced in Section 2 under a variety of data-generating processes. The data-generating processes that we study are special cases of the factor model with two factors ($n = 2$), two states ($k = 2$), and two prediction signals ($m = 2$). We simply assume that the factor loading coefficients are of identical magnitude but change their signs between the two states and choose our parameter values so that they do not appear to be too "extreme" compared to the differences in parameter values across regimes observed for the time series in the empirical section. Unconditionally, both predictions are unbiased observations of the respective factors but have an added noise component. The parameter values shared across the simulation experiment are as follows:

$$\boldsymbol{\beta}_{y1} = \begin{pmatrix} 3 \\ -3 \end{pmatrix}, \quad \boldsymbol{\beta}_{y2} = \begin{pmatrix} -3 \\ 3 \end{pmatrix}$$

$$\boldsymbol{\sigma}_{\varepsilon 1} = 3; \boldsymbol{\sigma}_{\varepsilon 2} = 5$$

$$\mathbf{B}_1 = \mathbf{B}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\boldsymbol{\mu}_y = \boldsymbol{\mu}_{.1} = \boldsymbol{\mu}_{.2} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_{F1} = \boldsymbol{\Sigma}_{F2} = \boldsymbol{\Sigma}_{\varsigma 1} = \boldsymbol{\Sigma}_{\varsigma 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Dynamics is already present in the model through the regime switching variable, S_{t+1}. However, we can also allow for serial correlation through the factor dynamics, e.g., by generating factor realizations from an AR(1) model

$$\mathbf{F}_t = \boldsymbol{\Gamma} \boldsymbol{F}_{t-1} + \mathbf{u}_t$$

where $\mathbf{u}_t$ is serially uncorrelated and independent of all other innovation terms.[9] In the main experiments, we assume that $\boldsymbol{\Gamma}$ is a diagonal matrix with 0.5 on the diagonal whereas $\mathbf{u}_t$ continues to be normally distributed with identity covariance matrix. We also considered scenarios with more or less persistence in the factor dynamics and found results very similar to those reported below. Forecast combinations conditional only on the concurrent forecast, $\hat{\mathbf{y}}_{t+1}$, are optimal either in the absence of such dynamics or—in the presence of first-order serial correlation in the factors—if the forecasts span the underlying factors. In the presence of more complicated factor dynamics, further lags of the forecasts can be included in the combination or serial correlation can be modeled explicitly in the regression residuals.

To study the performance of the various forecasting schemes, we consider six scenarios representing a variety of data-generating processes. The first experiment assumes that the process always stays in state 2, so there is no regime switching. Experiments two and three assume that there is regime switching in the combination weights with frequent shifts (experiment two, which sets $p_{11} = p_{22} = 0.7$) or more persistent states with infrequent shifts (experiment three, which sets $p_{11} = p_{22} = 0.95$). The fourth experiment maintains the persistent regimes from experiment three but considers the effect of fat tails on the results, assuming that the innovations $\{\varepsilon_{t+1}, \varsigma_{t+1}, \mathbf{u}_{t+1}\}$ are drawn from a fat-tailed Student-t distribution with five degrees of freedom.

[9] Alternatively, we could model regime switching in the dynamics of the factors, but assume that the factor loadings for the target variable were not regime dependent. This is the approach taken by Chauvet (1998).

To explore the possibility of nonrecurring structural breaks, the fifth experiment assumes a single structural break occurring halfway through the sample, leading to a permanent shift from the first to the second state. Such structural breaks are of interest for many economic time series and are particularly relevant given the evidence of breaks found in the empirical analysis in Section 2. Finally, the sixth experiment assumes that the true loading factors follow a time-varying parameter process, $\boldsymbol{\beta}_{yt} = \boldsymbol{\beta}_{yt-1} + \boldsymbol{\varepsilon}_{\beta t}$, with $Var(\boldsymbol{\varepsilon}_{\beta t}) = 0.1\mathbf{I}_2$, where $\mathbf{I}_2$ is a 2×2 identity matrix. The experimental setup is summarized below:

1. No regime switching (state 2 only)	$(p_{11}, p_{22}) = (0, 1)$
2. Frequent regime switching	$(p_{11}, p_{22}) = (0.7, 0.7)$
3. Highly persistent states	$(p_{11}, p_{22}) = (0.95, 0.95)$
4. Fat-tailed innovations	$(p_{11}, p_{22}) = (0.95, 0.95)$
5. Single break halfway through sample	
6. β_y follows TVP.	

We consider three sample sizes. First, we assume a short sample with 100 observations used to estimate the forecast combination weights. Parameter estimation errors are likely to be particularly important in such a small sample. We further study how fast the effect of estimation errors dies out in samples with 200 and 500 observations. To make sure that our results are comparable to those reported in the literature, we compute the one-step-ahead MSFE performance based on 5,000 simulations. MSFE results are reported as ratios relative to the MSFE produced by the forecast that uses combination weights estimated from an expanding window. This is a natural benchmark since it represents the standard way to obtain forecast combination weights. A value below one indicates superior performance of an alternative forecasting scheme relative to this benchmark.

Results from the simulations are reported in Table 2. In a stationary environment with no parameter instability (experiment 1), the expanding window OLS combination method produces the most accurate forecasts, as evidenced by the fact that the other forecasting schemes all produce relative performance values above one. The regime switching combination method does quite well with a relative performance ratio below 1.03 that shrinks further toward one as the sample size is increased and parameter estimation error gets reduced.

Next consider the experiment with frequent regime switching (experiment 2). In the smallest sample ($T = 100$) the best forecasts are produced by the simple average combination, followed closely by the first (individual) forecast. In the smallest samples, the worst forecasts in this setting are produced by the TVP model. This is likely a result of the heteroskedasticity and nonnormal shocks induced by the frequent regime shifts. The TVP algorithm is not designed to deal with such effects. For the smallest sample sizes, the regime switching combination scheme does not produce the best forecasts in this experiment despite the fact that it assumes the correct data-generating process. This can be explained by parameter estimation error, since the sample is so small that the combination weights are imprecisely estimated, which leads to a poor out-of-sample forecasting performance. Indeed, when the sample size is increased to 500 observations, the relative forecasting performance of the regime switching combination method improves and it dominates the other methods.

TABLE 2
MONTE CARLO SIMULATION RESULTS

Experiment	Regime Switching	First Forecast	Second Forecast	Simple Average	Rolling Window	DGT-I	DGT-II	TVP
Panel A: Sample size of 100 observations								
1. Stationary process	1.023	1.649	1.201	1.393	1.036	1.035	1.036	1.046
2. Frequent regime changes	1.007	1.019	1.026	0.993	1.035	1.030	1.030	1.067
3. Persistent regimes	0.915	1.014	1.015	0.986	1.013	0.994	0.985	0.972
4. Persistent regimes (*t*-distn.)	0.943	1.022	1.036	1.001	0.999	0.991	0.993	0.955
5. Single structural break	0.833	1.197	0.867	1.007	0.752	0.984	0.993	0.819
6. TVP process	1.025	1.294	1.267	1.172	1.002	1.029	1.030	1.025
Panel B: Sample size of 200 observations								
1. Stationary process	1.014	1.642	1.203	1.390	1.042	0.999	1.005	1.021
2. Frequent regime changes	0.994	1.042	1.040	1.013	1.066	1.000	1.000	1.050
3. Persistent regimes	0.859	1.051	1.049	1.022	1.020	0.979	0.978	0.938
4. Persistent regimes (*t*-distn.)	0.865	1.077	1.052	1.038	1.011	0.971	0.978	0.937
5. Single structural break	0.831	1.204	0.882	1.019	0.771	0.976	0.977	0.807
6. TVP process	0.991	1.347	1.319	1.241	0.969	1.007	1.007	0.961
Panel C: Sample size of 500 observations								
1. Stationary process	1.013	1.659	1.213	1.404	1.063	1.003	1.005	1.018
2. Frequent regime changes	0.980	1.036	1.038	1.010	1.082	1.000	1.008	1.040
3. Persistent regimes	0.879	1.050	1.030	1.012	1.019	0.968	0.971	0.953
4. Persistent regimes (*t*-distn.)	0.857	1.030	1.014	0.995	1.045	0.968	0.975	0.978
5. Single structural break	0.795	1.207	0.864	1.011	0.744	0.956	0.963	0.748
6. TVP process	0.969	1.392	1.369	1.300	0.955	1.001	0.993	0.935

NOTES: Performance is based on the MSFE of the one-step-ahead forecast measured relative to the MSFE performance of a combination method with estimated weights that uses an expanding data window. Values above one hence indicate relatively poor performance, whereas values below one indicate good performance. Results are averaged across 5000 Monte Carlo experiments.

When regimes are more persistent with an average duration of 20 periods (experiment 3), the regime switching combination scheme performs far better than the other forecasting methods irrespective of sample size. This reflects more precise estimates of the underlying parameters. The two nonlinear forecasting schemes and the TVP model also perform better—relative to the expanding window OLS method—in this setting than under more frequent regime switches. As expected, the performance of the time-varying (nonlinear) combination methods also improves when the sample size is increased.

Interestingly, in this experiment the MSFE performance of the regime switching combination model is even better around turning points than its average performance reveals. For example, when the sample size is 200, the relative MSFE value conditional on a sizable change in the estimated probability of the underlying state (defined as one that exceeds 0.25) declines from 0.86 to 0.76. A similar improvement did not occur in the model with very frequent regime shifts, since the turning points were too poorly identified in this model.

These findings are not overturned when the innovations are drawn from a fat-tailed *t*-distribution (experiment 4), although there is a slight deterioration in the performance of most of the more complicated forecast combination methods in the smallest samples. This is to be expected, since their parameter estimates can be more sensitive to deviations from normality—particularly the parameters of

the regime switching model that assumes Gaussian innovations conditional on the state.

Turning to the experiment with a single structural break occurring halfway through the sample (experiment 5), the best forecasts are produced by the rolling window method. This finding is unsurprising for the sample with 100 observations, since the break occurred 50 periods from the forecasting date. This is also the length of the rolling window, which is thus optimal under this scenario. Both the regime switching and the TVP forecast combination methods also perform very well under a structural break, suggesting that these methods adapt well even if they ultimately use misspecified models. Interestingly, the rolling window method continues to produce the best forecasts when the sample size is raised to 200 observations, followed closely again by the TVP and regime switching forecast combination schemes. The other forecasting methods perform significantly worse under this data-generating process.

Finally, when the factor exposures of the target variable follow a TVP process (experiment 6), the best forecasts are produced by the TVP combination scheme, particularly for the largest sample of 500 observations. The rolling window also produces good forecasts under this setting as does the regime switching combination method.

3.2.1. *Estimated versus equal weights.* A key question in the forecast combination literature is whether combination weights should be estimated—typically by using OLS estimation (cf. Granger and Ramanathan, 1984, and Diebold and Pauly, 1987)—or whether simple averages should be used. A number of studies have found that forecast combinations with estimated weights perform worse than combinations based on simple averages. For example, Kang (1986) concludes that "However genuine and sanguine the argument of the combination of forecasts using variance and covariance structures is in theory, it is not so in practice. A simple average should be used when underlying models are not known as in a survey . . . ." (p. 695). Kang attributes the poor performance of forecasts based on estimated combination weights to estimation uncertainty.

Ultimately the performance of the combined forecasts based on estimated versus equal-weighted averages depends on a number of factors—not least of which is the proximity of the population weights to equal weights: The closer the true population weights are to equal weights, the smaller the potential gains from using estimated combination weights.

Our simulations allow us to explore an alternative explanation, namely that the joint distribution of the predicted variable and the predictions is subject to instabilities induced by regime switching. This could make estimation of the combination weights more difficult than in the absence of regime switches. In the first Monte Carlo experiment, the combined forecast based on the equal-weighted average performs very poorly—generating MSFE values almost 40% higher than those based on the estimated forecast combination weights. Interestingly, however, once regime switching is introduced in the data-generating process, the simple average performs better than the forecasts based on estimated weights, irrespective

of whether the regime shifts occur very frequently (as in the second experiment) or infrequently (as in the third experiment).

In fact, if the main explanation of the poor performance of forecast combinations with estimated weights relative to the performance of the simple average forecast reflected parameter estimation error, then the regime switching combination model should perform even worse, since it introduces additional parameters and solves a more complicated, nonlinear estimation problem. On the other hand, if the poor performance of the least-squares combination weights was due to parameter instability, then regime switching could well improve on the forecasting performance of a combination scheme that assumes constant weights. Our empirical results and Monte Carlo simulations suggest that, for a number of plausible data-generating processes, parameter instability is the more likely explanation of the poor performance of forecast combinations based on least-squares weights.

4. CONCLUSION

This article proposed a new approach to forecast combination that lets the combination weights depend on a regime switching process driven by a latent variable. This approach is theoretically appealing in the presence of instability of unknown form in the forecasting performance of individual models. Indeed, several mechanisms could give rise to time variations in the combination weights, such as changes between recession and expansion periods, institutional shifts, or even differences in the learning speed of individual forecasting models representing varying degrees of complexity. Under any of these scenarios, a forecasting strategy of keeping the combination weights constant through time is unlikely to be the best available option to a decision maker.

Monte Carlo simulations suggested that forecasts based on regime switching combination weights perform quite well for a range of data-generating processes, including those with persistent regimes as well as settings in which the weights are subject to a single structural break or subject to a time-varying parameter process. Our simulations also provided an example in which using estimated combination weights leads to a lower MSFE than simply averaging the forecasts, provided that there is no regime switching. Once regime switching is present, however, the averaged combination outperformed the estimated weights in an MSFE sense. These results thus offer a new explanation—namely, parameter instability—for why simple averaging generally works so well in practice.

Earlier papers have handled potential instabilities in combination weights by using rolling windows of the data for parameter estimation (Deutsch et al., 1994) or by using time-varying parameter models (Zellner et al., 1991). Although such strategies are commonly used, it can be difficult to characterize the underlying data-generating process that gives rise to such estimation strategies, particularly in the case of the rolling window method.

Regime switching combination weights can be viewed as a natural intermediary between using fixed (estimated) weights and simple averaging. Forecasts from the regime switching model are formed as a weighted average of forecasts from each regime, with weights that reflect the individual state probabilities. Compared to

the type of parameter variation accounted for by the time-varying parameter model, regime switching models represent the opposite end of the spectrum. The time-varying parameter model has weights that change every period and move farther away on average from where they started, inducing unit-root like behavior, whereas the regime switching model assumes a stable, mean-reverting process for the weights and also can handle cases with structural breaks.

Our approach can be extended in several interesting directions. One possibility is to consider density forecasting based on regime switching combination models, using the methods for predictive density evaluation developed by Diebold et al. (1998) and Berkowitz (2001). Another interesting question is related to the performance of different combination methods in multistep forecasts.

REFERENCES

Aiolfi, M., and A. Timmermann, "Persistence of Forecasting Performance and Combination Strategies," Mimeo, UCSD, 2004.

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

Berkowitz, J., "Testing Density Forecasts with Applications to Risk Management," *Journal of Business and Economic Statistics* 19 (2001), 465–74.

Chao, J., V. Corradi, and N. R. Swanson, "An Out of Sample Test for Granger Causality," *Macroeconomic Dynamics* 5 (2001), 598–620.

Chauvet, M., "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switches," *International Economic Review* 39 (1998), 4, 969–96.

Chen, J., J. H. Stock, and M. W. Watson, "A Dynamic Factor Model Framework for Forecast Combination," *Spanish Economic Review* 1 (1999), 91–121.

Clemen, R. T., "Combining Forecasts: A Review and Annotated Bibliography," *International Journal of Forecasting* 5 (1989), 559–81.

Corradi, V., and N. R. Swanson, "A Consistent Test for Nonlinear Out of Sample Predictive Accuracy," *Journal of Econometrics* 110 (2002), 353–81.

Deutsch, M., C. W. J. Granger, and T. Terasvirta, "The Combination of Forecasts Using Changing Weights," *International Journal of Forecasting* 10 (1994), 47–57.

Diebold, F. X., *Elements of Forecasting*, 2nd edition (Mason, OH: South-Western Publishing, 2001).

——, and J. A. Lopez, "Forecast Evaluation and Combination," in G. S. Maddala and C. R. Rao eds., *Handbook of Statistics* (Amsterdam: Elsevier, 1996).

——, and R. Mariano, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13 (1995), 253–65.

——, and P. Pauly, "Structural Change and the Combination of Forecasts," *Journal of Forecasting* 6 (1987), 21–40.

——, T. Gunther, and A. Tay, "Evaluating Density Forecasts," *International Economic Review* 39 (1998), 863–83.

Dunis, C., J. Laws, and S. Chauvin, "The Use of Market Data and Model Combinations to Improve Forecast Accuracy," presented at CF2000 Conference in London, 2000.

Giacomini, R., and H. White, "Tests of Conditional Predictive Ability," UCSD manuscript, 2004.

Granger, C. W. J., and P. Newbold, "Some Comments on the Evaluation of Economic Forecasts," *Applied Economics* 5 (1973), 35–47.

——, and R. Ramanathan, "Improved Methods of Combining Forecasts," *Journal of Forecasting* 3 (1984), 197–204.

Hamilton, J. D., "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica* 57 (1989), 357–86.

——, *Time Series Analysis*, Princeton, NJ: Princeton University Press, 1994.
KANG, H., "Unstable Weights in the Combination of Forecasts," *Management Science* 32 (1986), 683–95.
MARRON, J. S., AND M. P. WAND, "Exact Mean Integrated Squared Error," *Annals of Statistics* 20 (1992), 712–36.
MCCONNELL, M., AND G. PEREZ-QUIROS, "Output Fluctuations in the United States: What Has Changed since the Early 80s?" *American Economic Review* 90 (2000), 1464–76.
PESARAN, M. H., AND A. TIMMERMANN, "Model Instability and Choice of Observation Window," Mimeo, University of Cambridge and UCSD, 2003.
STOCK, J. H., AND M.W. WATSON, "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics* 14 (1996), 11–30.
——, AND——, "A Comparison of Linear and Nonlinear Models for Forecasting Macroeconomic Time Series," in R. Engle and H. White eds., *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger* (Oxford: Oxford University Press, 1998).
——, AND ——, "Forecasting Inflation," *Journal of Monetary Economics* 44 (1999), 293–375.
——, AND ——, "Combination Forecasts of Output Growth in a Seven-Country Data Set," Mimeo, Harvard University, 2003.
TIMMERMANN, A., "Moments of Markov Switching Models," *Journal of Econometrics*, 96 (2000), 75–111.
ZELLNER, A., C. HONG, AND C-K MIN, "Forecasting Turning Points in International Output Growth Rates Using Bayesian Exponentially Weighted Autoregression, Time-Varying Parameter, and Pooling Techniques," *Journal of Econometrics* 49 (1991), 275–304.