Averaging Forecasts from VARs with Uncertain Instabilities *

Todd E. Clark
Federal Reserve Bank of Kansas City

Michael W. McCracken
Board of Governors of the Federal Reserve System

June 2006

Abstract

A body of recent work suggests commonly–used VAR models of output, inflation, and interest rates may be prone to instabilities. In the face of such instabilities, a variety of estimation or forecasting methods might be used to improve the accuracy of forecasts from a VAR. These methods include using different approaches to lag selection, observation windows for estimation, (over-) differencing, intercept correction, stochastically time–varying parameters, break dating, discounted least squares, Bayesian shrinkage, and detrending of inflation and interest rates. Although each individual method could be useful, the uncertainty inherent in any single representation of instability could mean that combining forecasts from the entire range of VAR estimates will further improve forecast accuracy. Focusing on simple models of U.S. output, prices, and interest rates, this paper compares the effectiveness of combination in improving VAR forecasts made with real–time data. The combinations include simple averages, medians, trimmed means, as well as a number of weighted combinations, based on: restricted (Bates-Granger) regressions, factor model estimates, regressions involving just forecast quartiles, Bayesian model averaging, information criteria–based averaging, and predictive least squares–based weighting. Our goal is to identify those approaches that, in real time, yield the most accurate forecasts of these variables. We use forecasts from simple univariate time series models and the Survey of Professional Forecasters as benchmarks.

JEL Nos.: C53, E37, C32
Keywords: Forecast combination, real-time data, prediction, structural change

* *Clark (corresponding author)*: Economic Research Dept.; Federal Reserve Bank of Kansas City; 925 Grand; Kansas City, MO 64198; todd.e.clark@kc.frb.org. *McCracken*: Board of Governors of the Federal Reserve System; 20th and Constitution N.W.; Mail Stop #61; Washington, D.C. 20551; michael.w.mccracken@frb.gov. This paper was written for a Reserve Bank of New Zealand conference Macroeconometrics and Model Uncertainty. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City, Board of Governors, Federal Reserve System, or any of its staff.

1 Introduction

In previous work (Clark and McCracken, 2006) we considered the performance of various methods for improving the forecast accuracy of VARs in the presence of structural change. For trivariate VARs in a range of measures of output, inflation, and a short–term interest rate, these methods include: sequential updating of lag orders, using various observation windows for estimation, working in levels or differences, intercept corrections (as in Clements and Hendry (1996)), stochastically time–varying parameters, break dating, discounted least squares estimation, Bayesian shrinkage, and detrending of inflation and interest rates. While some of these methods performed well at various times, various forecast horizons, and for some variables, simple averages (across the various methods just described) were consistently among the best performers.

One interpretation of this result is that it is crucial to have an understanding of the form of instability when constructing good forecasts. Another, and the one we prefer, is that in practice it is very difficult to know the form of structural instability, and model averaging provides an effective method for forecasting in the face of such uncertainty. As Timmermann (2005) indicates, structural breaks are commonly cited (in studies such as Bates and Granger (1969), Diebold and Pauly (1987), and Hendry and Clements (2004)) as motivation for combining forecasts from alternative models.

As summarized by Timmermann (2005), competing models will differ in their sensitivity to structural breaks. Depending on the size and nature of structural breaks, models that quickly pick up changes in parameters may or may not be more accurate than models that do not. For instance, in the case of a small, recent break, a model with constant parameters may forecast more accurately than a model that allows a break in coefficients, due to the additional noise introduced by the estimation of post–break coefficients (see, for example, Clark and McCracken (2005c) and Pesaran and Timmermann (2006)). However, in the case of a large break well in the past, a model that correctly picks up the associated change in coefficients will likely forecast more accurately than models with constant or slowly changing parameters. Accordingly, Timmermann (2005) and Pesaran and Timmermann (2006) suggest that combinations of forecasts from models with varying degrees of adaptability to uncertain (especially in real time) structural breaks will be more accurate than forecasts from individual models. Min and Zellner (1993) consider averaging of fixed and time–varying parameter models as a means of managing the impacts of structural change on forecasts. Along the same lines, combining forecasts from models that allow for different forms of structural breaks (non–stationarities) may improve accuracy.

Accordingly, in this paper we provide empirical evidence on the ability of various forms of model averaging to improve the real–time forecast accuracy of small-scale macroeconomic VARs in the presence of uncertain forms of model instabilities. Focusing on six distinct trivariate models incorporating different measures of output and inflation and a common interest rate measure, we consider a wide range of approaches to averaging forecasts (or forecast models) obtained with a variety of the aforementioned primitive methods for managing model instability. The average forecasts include: equally weighted averages with and without trimming, medians, common factor-based forecasts, combinations estimated with ridge regression, MSE–weighted averages, lowest MSE forecasts (predictive least squares–based forecasts), information theoretic averages, Bayesian model averages, and combinations based on quartile average forecasts (as suggested by Aiolfi and Timmermann (2006)).[1] For each of these forms of forecast or model averaging we construct real time forecasts of each variable using real–time data. We compare our results to those from simple baseline univariate models, selected baseline VAR models, and forecasts from the Survey of Professional Forecasters.

We consider this problem to be important for two reasons. First, small-scale VARs are widely used in macroeconomics and central bank forecasting. Examples of VARs used to forecast output, prices, and interest rates are numerous, including Sims (1980), Doan, et al. (1984), Litterman (1986), Brayton et al. (1997), Jacobson et al. (2001), Robertson and Tallman (2001), Del Negro and Schorfheide (2004), and Favero and Marcellino (2005). Second, there is an increasing body of evidence suggesting that these VARs may be prone to instabilities.[2] Examples include Webb (1995), Kozicki and Tinsley (2001b, 2002), Cogley and Sargent (2001, 2005), Boivin (2005), and Beyer and Farmer (2006). Still more studies have examined instabilities in smaller models, such as AR models of inflation or Phillips curve models of inflation. Examples include Stock and Watson (1996, 1999b, 2003, 2005),

[1]Recent examples of studies incorporating averages of forecasts from models with different variables include Koop and Potter (2003), Stock and Watson (2003), Clements and Hendry (2004), Maheu and Gordon (2004), and Pesaran, Pettenuzzo and Timmermann (2004).

[2]Admittedly, while the evidence of instabilities in the relationships incorporated in small macroeconomic VARs seems to be growing, the evidence is not necessarily conclusive. Rudebusch and Svensson (1999) apply stability tests to the full set of coefficients of an inflation–output gap model and are unable to reject stability. Rudebusch (2005) finds that historical shifts in the behavior of monetary policy haven't been enough to make reduced form macro VARs unstable. Estrella and Fuhrer (2003) find little evidence of instability in joint tests of a Phillips curve relating inflation to the output gap and an IS model of output. Similarly, detailed test results reported in Stock and Watson (2003) show inflation–output gap models to be largely stable.

Levin and Piger (2003), Roberts (2004), and Clark and McCracken (2005a). Although many different structural forces could lead to instabilities in macroeconomic VARs (e.g., Rogoff (2003) and others have suggested that globalization has altered inflation dynamics), much of the aforementioned literature has focused on shifts potentially attributable to changes in the behavior of monetary policy. Even if monetary policy is a leading cause of structural instability in common VAR models, the timing of changes in policy regimes is rarely clear. Moreover, even if policy regimes were easily identified, the type of instability these regimes changes induce in VARs would remain unclear.

Our results indicate that while some of the primitive forms of managing structural instability sometimes provide the largest gains in terms of forecast accuracy — notably those models with some form of Bayesian shrinkage — model averaging is a more consistent method for improving forecast accuracy. Not surprisingly, the best type of averaging often varies with the variable being forecast, but several patterns do emerge. After aggregating across all models, horizons and variables being forecasted, it is clear that the simplest forms of model averaging — such as those that use equal weights across all models or those that average a univariate model with a particular VAR, such as a VAR(4) with inflation detrending — consistently perform among the best methods. At the other extreme, forecasts based on OLS–type combination, predictive least squares, and factor model–based combination rank among the worst.

The remainder of the paper proceeds as follows. Section 2 describes the real-time data and samples. Section 3 provides a synopsis of the forms of model averaging used to forecast in the presence of uncertain forms of structural change. Section 4 presents our results on forecast accuracy, including root mean square errors of the methods used. Section 5 concludes.

2 Data and model details

We consider the real–time forecast performance of models with three different measures of output (y), two measures of inflation (π), and a short–term interest rate (i). The output measures are GDP or GNP (depending on data vintage) growth, an output gap computed with the method described in Hallman, Porter, and Small (1991), and an output gap estimated with the Hodrick and Prescott (1997) filter. The first output gap measure (hereafter, the HPS gap), based on a method the Federal Reserve Board once used to estimate potential output for the nonfarm business sector, is entirely one–sided but turns out to be highly correlated with an output gap based on the Congressional Budget Office's (CBO's) estimate of potential output. The HP filter of course has the advantage of being widely used and easy to implement. We follow Orphanides and van Norden (2005) in our real time application of the filter: for forecasting starting in period t, the gap is computed using the conventional filter and data available through period $t-1$. The inflation measures include the GDP or GNP deflator or price index (depending on data vintage) and CPI price index. The short–term interest rate is measured with the 3–month Treasury bill rate (for which we have SPF forecasts); using the federal funds rate (for which we do not have SPF forecasts) yields qualitatively similar results. Note, finally, that growth and inflation rates are measured as annualized log changes (from $t-1$ to t). Output gaps are measured in percentages (100 times the log of output relative to trend). Interest rates are expressed in annualized percentage points.

The raw quarterly data on output, prices, and interest rates are taken from the Federal Reserve Bank of Philadelphia's Real–Time Data Set for Macroeconomists (RTDSM), the Board of Governor's FAME database, and the website of the Bureau of Labor Statistics. Real–time data on GDP or GNP and the GDP or GNP price series are from the RTDSM. For simplicity, hereafter we simply use the notation "GDP" and "GDP price index" to refer to the output and price series, even though the measures are based on GNP and a fixed weight deflator for much of the sample. In the case of the CPI and the interest rates, for which real time revisions are small to essentially non–existent (see, for example, Kozicki (2004)), we simply abstract from real time aspects of the data. For the CPI, we follow the advice of Kozicki and Hoffman (2004) for avoiding choppiness in inflation rates for the 1960s and 1970s due to changes in index bases, and use a 1967 base year series taken from the BLS website in late 2005.[3] For the T-bill rate, we use a series obtained from FAME. Finally, we obtained Survey of Professional Forecasters' (SPF) projections of GDP/GNP growth, inflation, and the T-bill rate from the website of the Federal Reserve Bank of Philadelphia.[4]

The full forecast evaluation period runs from 1970:Q1 through 2005; as detailed in section 3, forecasts from 1965:Q4 through 1969:Q4 are used as initial values in the combination

[3]The BLS only provides the 1967 base year CPI on a not seasonally adjusted basis. We seasonally adjusted the series with the X-11 filter.

[4]The SPF data provide GDP/GNP and the GDP/GNP price index in levels, from which we computed log growth rates. We derived 1–year ahead forecasts of CPI inflation by compounding the reported quarterly inflation forecasts.

forecasts that require historical forecasts. Accordingly, we use real time data vintages from 1965:Q4 through 2005:Q4. As described in Croushore and Stark (2001), the vintages of the RTDSM are dated to reflect the information available around the middle of each quarter. Normally, in a given vintage t, the available NIPA data run through period $t-1$.[5] The start dates of the raw data available in each vintage vary over time, ranging from 1947:Q1 to 1959:Q3, reflecting changes in the published samples of the historical data. For each forecast origin t in 1970:Q1 through 2005:Q3, we use the real time data vintage t to estimate output gaps, estimate the forecast models, and then construct forecasts for periods t and beyond. The starting point of the model estimation sample is the maximum of (i) 1955:Q1 and (ii) the earliest quarter in which all of the data included in a given model are available, plus five quarters to allow for four lags and differencing or detrending.

We present forecast accuracy results for forecast horizons of the current quarter ($h = 0Q$), the next quarter ($h = 1Q$), and four quarters ahead ($h = 1Y$). In light of the time $t-1$ information actually incorporated in the VARs used for forecasting at t, the current quarter (t) forecast is really a 1–quarter ahead forecast, while the next quarter ($t+1$) forecast is really a 2–step ahead forecast. What is referred to as a 1–year ahead forecast is really a 5–step ahead forecast. In keeping with conventional practices and the interests of policymakers, the 1–year ahead forecasts for GDP/GNP growth and inflation are four–quarter rates of change (the percent change from period $t+1$ through $t+4$). The 1–year ahead forecasts for output gaps and interest rates are quarterly levels in period $t+4$.

As discussed in such sources as Romer and Romer (2000), Sims (2002), and Croushore (2005), evaluating the accuracy of real time forecasts requires a difficult decision on what to take as the actual data in calculating forecast errors. The GDP data available today for, say, 1970, represent the best available estimates of output in 1970. However, output as defined today is quite different from the definition of output in 1970. For example, today we have available chain weighted GDP; in the 1970s, output was measured with fixed weight GNP. Forecasters in 1970 could not have foreseen such changes and the potential impact on measured output. Accordingly, in our baseline results, we follow Romer and Romer (2000) and use the second available estimates of GDP/GNP and the GDP/GNP deflator as actuals in evaluating forecast accuracy. In the case of h–step ahead (for $h = 0, 1$,and 4) forecasts made for period $t+h$ with vintage t data ending in period $t-1$, the second

[5] In the case of the 1996:Q1 vintage, with which the BEA published a benchmark revision, the data run through 1995:Q3 instead of 1995:Q4.

available estimate is normally taken from the vintage $t + h + 2$ data set. In light of our abstraction from real time revisions in CPI inflation and interest rates, the real time data correspond to the final vintage data.

3 Methods used

The forecasts of interest in this paper are combinations of forecasts from a wide range of approaches to allowing for structural change in trivariate VARs: sequential updating of lag orders, using various observation windows for estimation, working in levels or differences, intercept corrections, stochastically time–varying parameters, break dating, discounted least squares estimation, Bayesian shrinkage, and detrending of inflation and interest rates. Table 1 lists the set of individual VAR forecast methods considered in this paper, along with some detail on forecast construction. To be precise, for each model — defined as being a baseline VAR in one measure of output (y), one measure of inflation (π), and one short–term interest rate (i) — we apply each of the estimation and forecasting methods listed in Table 1.

Note that, although we simply refer to all the underlying forecasts as VAR forecasts, in fact the list of individual models includes a univariate specification for each of output, inflation, and the interest rate. For output, widely modeled as following low-order AR processes, the univariate model is an AR(2). In the case of inflation, we follow Stock and Watson (2005) and use an MA(1) process for the change in inflation ($\Delta\pi$), estimated with a rolling window of 40 observations. Stock and Watson find that the IMA(1) generally outperforms random walk or AR model forecasts of inflation. For simplicity, in light of some general similarities in the time series properties of inflation and short–term interest rates and the IMA(1) rationale for inflation described by Stock and Watson, the univariate model for the short-term interest rate is also specified as an MA(1) in the first difference of the series (Δi).

Table 2 provides a comprehensive list, with some detail, of the approaches we use to combining forecasts from these underlying models. The remainder of this section explains the averaging methods.

3.1 Equally weighted averages

We begin with seven distinct, simple forms of model averaging, in each case using what could loosely be described as equal weights. The first is an equally weighted average of all the VAR forecasts in Table 1, for a given triplet of variables. More specifically, for a given combination of measures of output, inflation, and the interest rate (for example, for the combination GDP growth, GDP inflation, and the T-bill rate), we average forecasts from the 54 VARs listed in Table 1.

Three more equally weighted average forecasts are constructed in much the same way but with an eye towards making the model average robust to individual forecasts that might be considered outliers. Specifically, following Stock and Watson (1999a), we consider 10 percent and 20 percent trimmed means. To implement these, at each forecast origin t, we first order the 54 VAR forecasts. With 10 percent trimming, we remove the 3 highest and lowest forecasts, and take an equally weighted average of the remaining forecasts. With 20 percent trimming, we do the same but after removing the 5 highest and lowest forecasts. Finally, we also consider using the median rather than the equally weighted average of all the forecasts (a rule that might be thought of as a trimmed forecast as the trimming parameter approaches 50 percent).

We include a fifth average forecast approach motivated by the results of Clark and McCracken (2005c), who show that forecast accuracy can be improved by combining forecasts from models estimated with recursive (all available data) and rolling samples. For a given VAR(4), we form an equally weighted average of the model forecasts constructed using parameters estimated (i) recursively (with all of the available data) and (ii) with a rolling window of the past 60 observations. Three other averages are motivated by the Clark and McCracken (2005b) finding that combining forecasts from nested models can improve forecast accuracy. We consider an average of the univariate forecast with the VAR(4) forecast, an average of the univariate forecast with the DVAR(4) forecast, and an average of the univariate forecast with a forecast from a VAR(4) in output, detrended inflation, and the detrended interest rate (Table 1 and section 3.7 provide more information on the VAR(4) with inflation detrending).

3.2 Combinations based on Bates–Granger/ridge regression

We also consider a large number of average forecasts based on historical forecast performance — one such approach being forecast combination based on Bates and Granger (1969)–type regression. For these methods, we need an initial sample of forecasts preceding the sample to be used in our formal forecast evaluation. With the formal forecast evaluation sample beginning with 1970:Q1, we use an initial sample of forecasts from 1965:Q4 (the starting point of the RTDSM) through 1969:Q4. Therefore, in the case of current quarter forecasts constructed in 1970:Q1, we have an initial sample of 17 forecasts to use in estimating combination regressions, forming MSE weights, etc. Note also that these performance-based combinations are based on real time forecast accuracy. That is, in period t, in deciding how best to combine forecasts based on historical performance, we use the historical real time forecasts compared to real time data in determining the combinations.

To obtain combinations based on the Bates–Granger approach, for each of output, inflation, and the interest rate we use the actual data that would have been available to a forecaster in real time to estimate a generalized ridge regression of the actual data on the 54 VAR forecasts, shrinking the coefficients toward equal weights. Our implementation follows that of Stock and Watson (1999a): letting $Z_{t+h|t}$ denote the vector of 54 forecasts of variable z_{t+h} made in period t and β^{equal} denote a 54×1 vector filled with 1/54, the combination coefficient vector estimate is

$$\hat{\beta} = (cI_{54} + \sum_t Z_{t+h|t} Z'_{t+h|t})^{-1} (c\beta^{equal} + \sum_t Z_{t+h|t} z_{t+h}), \tag{1}$$

where $c = k \times \text{trace}(54^{-1} \sum_t Z_{t+h|t} Z'_{t+h|t})$. We consider three different forecasts, based on different values of the shrinkage coefficient k: .001, .25, and 1. A smaller (larger) value of k implies less (more) shrinkage. Following Stock and Watson (1999a), we use a value .001 to approximate the OLS combination of Bates and Granger (1969). For each value of k, we consider forecasts based on both a recursive estimate of the combination regression (using all available forecasts) and a 10–year rolling sample estimate (using just the most recent 10 years of forecasts, or all available if less than 10 years are available).

3.3 Common factor combinations

Stock and Watson (1999a) develop another approach to combining information from individual model forecasts: estimating a common factor from the forecasts, regressing actual data on the common factor, and then using the fitted regression to forecast into the future. Therefore, using the real time forecasts available through the forecast origin t, we estimate (by principal components) one common factor from the set of 54 VAR forecasts for each of output, inflation, and the interest rate (estimating one factor for output, another for inflation, etc.). We then regress the actual data available in real time as of t on a constant and the factor. The factor–based forecast is then obtained from the estimated regression, using the factor observation for period t. As in the case of the ridge regressions, we compute factor–combination forecasts on both a recursive (using all available forecasts) and 10–year rolling (using just the most recent 10 years of forecasts, or all available if less than 10 years are available) basis.

3.4 MSE–weighted and PLS forecasts

As noted in Bates and Granger (1969), if one ignores the covariances of the forecast errors across models, the regression-based method above (with k set to 0) is equivalent to weighting each of the models by its inverse forecast MSE relative to the sum of those for the other models. Accordingly, we consider several average forecasts based on inverse MSE weights. At each forecast origin t, historical MSEs of the 54 VAR forecasts of each of output, inflation, and the interest rate are calculated with the available forecasts and actual data, and each forecast i of the given variable is given a weight of $MSE_i^{-1}/\sum_i MSE_i^{-1}$. In addition, following Stock and Watson (1999a) and Rapach and Strauss (2005), we consider a forecast based on a discounted mean square forecast error (in which, from a forecast origin of t, the squared error in the earlier period s is discounted by a factor δ^{t-s}), using a discount rate of $\delta = .95$.

We also consider a forecast based on the model(s) with lowest historical MSE — i.e., based on predictive least squares (PLS). At each forecast origin t, we identify the model forecast with the lowest historical MSE, and then use that single model to forecast into the future. In the event two or more models are equally accurate in the historical period under consideration at t, we use an equally–weighted average of the forecasts from these models.

We compute alternative MSE–weighted and PLS forecasts with several different samples of historical forecasts: all available forecasts (recursive), a 10 year rolling window of forecasts, and a 5 year rolling window of forecasts.

3.5 Quartile forecasts

Aiolfi and Timmermann (2005) develop alternative approaches to forecast combination that take into account persistence in forecast performance — the possibility that some models may be consistently good while others may be consistently bad. Their simplest forecast is an equally weighted average of the forecasts in the top quartile of forecast accuracy (that is, the forecasts with historical MSEs in the lowest quartile of MSEs). More sophisticated forecasts involve measuring performance persistence as forecasting moves forward in time, sorting the forecasts into clusters based on past performance, and estimating combination regressions with a number of clusters determined by the degree of persistence. For tractability in our extensive real–time forecast evaluation, we consider simple versions of the Aiolfi–Timmermann methods, based on just the first and second quartiles. Specifically, we consider a simple average of the forecasts in the top quartile of historical forecast accuracy. We also consider a forecast based on an OLS–estimated combination regression including a constant, the average of the first quartile forecasts, and the average of the second quartile forecasts. For simplicity, we systematically exclude the third and fourth quartiles in light of the tendency for some VAR forecasts to be persistently poor (see Clark and McCracken (2006)). We compute these quartile–based forecasts with several different samples of historical forecasts: all available forecasts (recursive), a 10 year rolling window of forecasts, and a 5 year rolling window of forecasts.

3.6 Weighted averages based on in–sample fit

We consider several average forecasts weighted by measures of in–sample fit at each point in the forecast sample. Following Kapetanios, Labhard, and Price (2005), one such forecast is based on the AIC; another is based on the BIC. Specifically, at each forecast origin t, for each of the model estimates listed in Table 1, we compute the AIC and BIC for each equation of the model. Letting IC denote the information criterion (AIC or BIC) and $\Psi_i = IC_i - \min_j IC_j$, where $\min_j IC_j$ is the lowest IC value across the 54 equations for a given variable, the forecast from model i is given a weight of

$$w_i = \frac{e^{-.5\Psi_i}}{\sum_j e^{-.5\Psi_j}}. \quad (2)$$

In our application, calculating the information criteria requires some decisions on how to deal with some of the important differences in estimation approaches (e.g., rolling versus recursive estimation) for the 54 underlying model forecasts. In the case of models estimated with a rolling sample of data, we calculate the AIC and BIC based on a model that allows a discrete break in all the model coefficients at the point of the beginning of the rolling sample (equivalently, based on two models, one estimated with the rolling window and the other estimated with data up to the start of the rolling window). Our rationale is the Ng and Perron (2005) argument that model selection based on information criteria should use a common estimation sample for all models. For models estimated by discounted least squares (DLS), we calculate the information criteria using residuals defined as actual data less fitted values based on the DLS coefficient estimates.

In the case of BVAR models, for simplicity we abstract from the prior and calculate the information criteria based on the residual sums of squares and simple parameter count.[6] The prior is asymptotically irrelevant in the sense that, as the sample grows, sample information dominates the prior. Admittedly, to the extent the priors are important for improving forecasting in finite samples, our approach will give too little weight to a BVAR relative to the corresponding VAR. However, taking (proper Bayesian) account of the finite–sample role of the Bayesian prior in combining forecasts from models estimated with different priors would require Monte Carlo integration, which is intractable in our large–scale, real–time forecast evaluation.

Following Wright (2003), among others, we also consider forecasts obtained by Bayesian model averaging (BMA). At each forecast origin t, for each equation of the 54 models listed in Table 1, we calculate a posterior probability using

$$\begin{aligned}
\text{Prob}(M_i|\text{data}) &= \frac{\text{Prob}(\text{data}|M_i) \times \text{Prob}(M_i)}{\sum_i \text{Prob}(\text{data}|M_i) \times \text{Prob}(M_i)} \qquad (3)\\
\text{Prob}(M_i) &\equiv \text{prior probability on model } i = 1/54 \\
\text{Prob}(\text{data}|M_i) &\propto (1+\phi)^{-p_i/2} S_i^{-(t+1)} \\
\phi &= \text{parameter determining rate of shrinkage toward the restricted model} \\
p_i &= \text{the number of explanatory variables in model } i \\
S_i^2 &= (Y - X_i\hat{\Gamma}_i)'(Y - X_i\hat{\Gamma}_i) + \frac{1}{1+\phi}(\hat{\Gamma}_i - \Gamma_{prior})'X_i'X_i(\hat{\Gamma}_i - \Gamma_{prior}) \\
X_i &= \text{matrix of regressors in model } i \\
\hat{\Gamma}_i &= \text{vector of estimates of the coefficients of model } i.
\end{aligned}$$

We report results for two different settings of the shrinkage parameter ϕ, one relatively high ($\phi = 2$) and one low ($\phi = .2$). Lower values of ϕ are associated with greater shrinkage toward the restricted model. In Wright's (2003) application, the BMA weight prior on all coefficient estimates (Γ_{prior}) was taken to be diffuse, and set to 0. In our application, however, the use of levels of inflation and interest rates in some models and differences (or detrended series) in others requires a different approach: if we were to use a prior mean of zero in all cases, the quadratic term $(\hat{\Gamma}_i - \Gamma_{prior})'X_i'X_i(\hat{\Gamma}_i - \Gamma_{prior})$ would be very large for models in

[6] For BVARs with TVP, at each point in time t we calculated the model residuals as a function of the period t coefficients and used these residuals to compute the residual sums of squares. In counting parameters for the information criteria calculations, we abstracted from the time variation in the coefficients. Our loose rationale is that, at least for forecasting into the future, the parameters are held constant, and therefore the count can be approximated as the simple count abstracting from time variation.

levels relative to models in differences (or detrended data). To avoid such a problem, in calculating BMA weights we use a prior mean of 0 for all coefficients of equations in $\Delta\pi$ and Δi (or detrended π and i) but a prior mean with a 1 for the own first lag in equations in π and i.

As in the case of the AIC and BIC-weighted forecasts, calculating the model averaging weights requires some decisions on how to deal with some of the important differences in estimation approaches for the 54 underlying model forecasts. For models estimated with rolling samples of data, DLS, and BVARs, we follow the approach described above for the AIC and BIC calculations.[7] Accordingly, in computing BMA weights for BVARs, for simplicity we abstract from the role of the prior in the coefficient estimates; as Phillips (1996) notes, the prior can be viewed as asymptotically irrelevant. Proper Bayesian treatment would require Monte Carlo integration of posterior densities.[8] In the case of the univariate models for inflation and interest rates, the MA specifications introduce a complication to BMA. For simplicity, we approximate the MA fits with AR(2) models estimated for $\Delta\pi$ and Δi (estimating separate models for the rolling sample and the earlier sample), and calculate approximate BMA weights using these AR(2) approximations.

3.7 Benchmark forecasts

To evaluate the practical value of all the averaging methods described above, we compare the accuracy of the above combination or average forecasts against various benchmarks. In light of common practice in forecasting research, we use forecasts from the univariate time series models as one set of benchmarks.[9] We also use SPF forecasts of growth, inflation, and interest rates as benchmarks. Using forecasts from the Federal Reserve's Greenbook yields qualitatively similar conclusions.

We also include for comparison forecasts from selected VAR methods that are either of general interest in light of common usage or performed relatively well in our prior work: a

[7] For BVARs with TVP, in calculating the BMA penalty term for forecasting from origin t, we use the last available coefficient and $X'X$ estimates, again because these are the values relevant for forecasting ahead.

[8] As Koop (2006) notes, BMA allows for two types of shrinkage: (1) through priors on parameters imposed in parameter estimation and (2) through the model priors in the calculation of the BMA weights. Accordingly, in practice, there is some interchangeability between the two types of shrinkage. Asymptotically, the first form becomes irrelevant asymptotically. Our simple approach corresponds to focusing entirely on the second form of shrinkage.

[9] Of course, the choice of benchmarks today is influenced by the results of previous studies of forecasting methods. Although a forecaster today might be expected to know that an IMA(1) is a good univariate model for inflation, the same may not be said of a forecaster operating in 1970. For example, Nelson (1972) used as benchmarks AR(1) processes in the change in GNP and the change in the GNP deflator (both in levels rather than logs). Nelson and Schwert (1977) first proposed an IMA(1) for inflation.

VAR(4); DVAR(4) (a VAR with inflation and the interest rate differenced); BVAR(4) with conventional Minnesota priors; BVAR(4) with stochastically time–varying (random walk) parameters; and a BVAR(4) in output, detrended inflation, and the interest rate less the inflation trend. The BVAR(4) with inflation detrending draws on the work of Kozicki and Tinsley (2001a,b, 2002) on models with learning about an unobserved time–varying inflation target of the central bank. For tractability in real time forecasting, we follow Cogley (2002) in estimating the inflation target or trend with exponential smoothing.[10] Table 1 provides additional detail on all of these model specifications.

4 Results

In evaluating the performance of the forecasting methods described above, we follow Stock and Watson (1996, 2003, 2005), among others, in using squared error to evaluate accuracy and considering forecast performance over multiple samples. Specifically, we measure accuracy with root mean square error (RMSE). The forecast samples are generally specified as 1970-84 and 1985-2005.[11] We split the full sample in this way to ensure our general findings are robust across sample periods, in light of the evidence in Stock and Watson (2003) and others of instabilities in forecast performance over time.

To be able to provide broad, robust results, in total we consider a large number of models and methods — too many to be able to present all details of the results. In the interest of brevity, we present more detailed results on forecasts of output and inflation than forecasts of interest rates, in light of generally greater interest in the former.

Tables 3 through 8 report forecast accuracy (RMSE) results for six combinations of output (GDP growth, HPS gap and HP gap) and inflation (GDP price index and CPI) and 39 forecast methods. In each case we use the 3-month T–bill as the interest rate. In Table 9 we report forecast accuracy results for the T-bill rate, from models using GDP growth and GDP inflation and models using the HPS gap and GDP inflation. In every case, the first row of the table provides the RMSE associated with the baseline univariate model, while the others report ratios of the corresponding RMSE to that for the benchmark univariate model. Hence numbers less than one denote an improvement over the univariate baseline while numbers greater than one denote otherwise.

[10] As noted in Clark and McCracken (2006), the resulting trend estimate is quite similar to measures of long–run inflation expectations.

[11] With forecasts dated by the end period of the forecast horizon $h = 0, 1, 4$, the VAR forecast samples are, respectively, 1970:Q1+h to 1984:Q4 and 1985:Q1 to 2005:Q3-h.

In Table 10 we take another approach to broadly determining which methods tend to perform better than the benchmark. Across each variable, model and horizon, we compute the average rank of the methods included in Tables 3-9. We present average rankings for every method we consider across every variable, forecast horizon, and the 1970-84 and 1985-05 samples (spanning all columns of Tables 3-9 plus unreported results for forecasts of the T-bill rate from models using the HP gap and GDP inflation, GDP growth and CPI inflation, HPS gap and CPI inflation, and HP gap and CPI inflation).

To determine the statistical significance of differences in forecast accuracy, we use White's (2000) bootstrap method to calculate p–values for each RMSE ratio in Tables 3-9, as well as for the best RMSE ratio in each column of the tables. The individual p–values represent a pairwise comparison of each VAR or average forecast to the univariate forecast. RMSE ratios that are significantly less than 1 at a 10 percent confidence interval are indicated with a *slanted* font. The best RMSE ratio p–values take into account the data snooping or search involved in selecting, for a given variable, sample, and forecast horizon, the best forecast among those included in the given column. The best RMSE ratio p–value indicates whether, after taking the search into account, the best forecast's improvement in accuracy (relative to the univariate benchmark) is statistically significant. For each column, a best RMSE ratio that is significantly less than 1 at a 10 percent significance level is indicated with a **bold** font. As explained in White (2000), the p–value calculated to take data snooping into account is higher than the pairwise p–value.[12] As a result, the best RMSE ratio in each column may be significantly less than 1 on a pairwise basis (and appear in a slanted font) but not once snooping is taken into account (not appear in bold font).

We implement White's (2000) procedure by sampling from the time series of forecast errors underlying the entries in Tables 3-9. For simplicity, we use the moving block method of Kunsch (1989) and Liu and Singh (1992) rather than the stationary bootstrap actually used by White; White notes that the moving block is also asymptotically valid. The bootstrap is applied separately for each forecast horizon, using a block size of 1 for the $h = 0Q$ forecasts, 2 for $h = 1Q$, and 5 for $h = 1Y$.[13] In addition, in light of the potential for changes over time

[12]To see why, suppose that, in truth, all of the VAR forecasts and the univariate forecast of GDP growth in 1970-84 are equally accurate. In the pairwise comparison of a VAR forecast to the univariate, there is a 10 percent probability of Type I error. However, if we were to select the lowest RMSE ratio among the 38 ratios considered and continue to use the pairwise–appropriate critical value, the probability of Type I error would be higher. Compared to the pairwise evaluation, the search for the best RMSE ratio increases the likelihood of finding a RMSE ratio that is less than one, even though, in population, all ratios have a value of 1. The bootstrap properly accounts for these effects of the search for a best RMSE ratio.

[13]For a forecast horizon of τ periods, forecast errors from a properly specified model will follow an MA($\tau-1$)

in forecast error variances, the bootstrap is applied separately for each subperiod. Note, however, that the bootstrap sampling preserves the correlations of forecast errors across forecast methods.

4.1 Declining volatility

While there are many nuances in the detailed results, some clear patterns emerge. The univariate RMSEs clearly show the reduced volatility of the economy since the early 1980s, particularly for output. For each horizon, the benchmark univariate RMSEs of GDP growth and HP gap forecasts declined by roughly two-thirds across the 1970-84 and 1985-05 samples; the benchmark RMSE for HPS gap forecasts declined by about half. The reduced volatility is less extreme for the inflation measures but still evident. For each horizon, the benchmark RMSEs fell by roughly half across the two periods, with the exception that at the $h = 1Y$ horizon the variability in GDP inflation declined nearly two-thirds. The reverse is true for the interest rate forecasts. For each horizon, the benchmark RMSEs fell by roughly two-thirds across the two periods, with the exception that at the $h = 1Y$ horizon the variability declined only by half.

4.2 Declining predictability

Consistent with the results in Campbell (2005), D'Agostino, et al. (2005), Stock and Watson (2005), and Tulip (2005), there are some clear signs of a decline in the predictability of both output and inflation: it has become harder to beat the accuracy of a univariate forecast. For example, for each forecast horizon, most methods or models beat the accuracy of the univariate forecast of GDP growth during the 1970-84 period (Tables 3 and 6). In fact, many do so at a level that is statistically significant; at each horizon the best performer is significant even after accounting for multiple testing. But over the 1985-2005 period, only the BVAR(4)-TVP models are more accurate at short horizons, and that improvement fails to be statistically significant. At the $h = 1Y$ horizon a handful of the methods continue to outperform the benchmark univariate, but very few are statistically significant.

The reduction in predictability is a bit more mixed when output is measured by an output gap. For example, when real activity is measured using the HPS output gap (Tables 4 and 7), several methods perform significantly better than the benchmark in the 1970-84 period — especially at longer horizons — while only the BVAR(4)-TVP significantly

process. Accordingly, we use a moving block size of τ for a forecast horizon of τ.

outperforms the benchmark in the 1985-05 period. However, when the HP measure of the output gap is used, a higher percentage of the methods outperform the univariate benchmark at all horizons, and the magnitude of the accuracy gains can be impressive. In fact, in Table 5 we see that the inflation detrended BVAR(4) model improves upon the benchmark in every horizon during the 1985-05 period and does so with statistically significant maximal gains of 20 and 30 percent at the $h = 1Q$ and $h = 1Y$ horizons.

The predictability of inflation has also declined, although less dramatically than for output. For example, in models with GDP growth and GDP inflation (Table 3), the best 1–year ahead forecasts of inflation improve upon the univariate benchmark RMSE by more than 10 percent in the 1970-84 period but only about 5 percent in 1985-05. The evidence of a decline in inflation predictability is perhaps most striking for CPI forecasts at the $h = 0Q$ horizon. In Tables 6-8, most of the models convincingly outperform the univariate benchmark during the 1970-84 period, with statistically significant maximal gains of roughly 20 percent. But in the following period, fewer methods outperform the benchmark, with gains typically about 4 percent.

Predictability of the T-bill rate has not so much declined as it has shifted to a longer horizon. In Table 9 we see that at the $h = 0Q$ horizon far fewer methods outperform the univariate benchmark as we move from the 1970-84 period to the 1985-05 period. However, the decline in relative predictability starts to weaken as the forecast horizon increases. At the $h = 1Q$ horizon some methods continue to beat the benchmark, although with maximal gains of only 5 percent. But at the $h = 1Y$ horizon, not only do a larger number of methods improve upon the benchmark, they do so with maximal gains that are a substantial, and statistically significant, 10 percent.

4.3 Averaging methods that typically outperform the benchmark

The sharp decline of predictability makes it difficult to identify models or averaging methods that consistently beat the accuracy of the univariate benchmarks. Note also that the considerable sampling error inherent in small sample forecast comparisons further compounds the difficulty of finding a method that always or nearly always beats the univariate benchmark. Suppose, for example, that there exists a model or average forecast that, in population, is somewhat more accurate (by 10 percent, say) than the univariate benchmark. For forecast samples of roughly 15 years, there is a good chance that, in a given sample, the univariate forecast will actually be more accurate (see, e.g., Clark and McCracken's

(2005a) results for Phillips curve forecasts of inflation). The sampling uncertainty grows with the forecast horizon. As a result, we probably shouldn't expect to be able to identify a particular forecast model or method that beats the univariate benchmark for every variable, horizon, and sample period. Instead, we might judge a model or method a success if it beats the univariate benchmark most of the time (with some consistency across the 1970-84 and 1985-05 samples) and, when it fails to do so, is not dramatically worse than the univariate benchmark.

With these considerations in mind, the best forecast would appear to come from the pairwise averaging class: the single best forecast is an average of the univariate forecast with the forecast from a VAR(4) with inflation detrending (a VAR(4) in y, $\pi - \pi^*_{-1}$, and $i - \pi^*_{-1}$, motivated by the work of Kozicki and Tinsley (2001a,b, 2002)). More so than any other forecast, the forecast based on an average of the univariate and inflation detrended VAR(4) projections beats the univariate benchmark a very high percentage of the time and, when it fails to do so, is generally comparable to the univariate forecast. For example, in the case of forecasts of GDP growth and GDP inflation from models in these variables and the T-bill rate (Table 3), this pairwise average's RMSE ratio is less than 1 for all samples and horizons, with the exception of $h = 0Q$ and $h = 1Q$ forecasts of GDP growth for 1985-05, in which cases the RMSE ratio is only slightly above 1. For instance, in the case of 1–year ahead forecasts of GDP growth, the RMSE of this average forecast is about 15 percent below the univariate benchmark for 1970-84 and 9 percent below for 1985-05; the corresponding figures for GDP inflation are each roughly 3 percent. Similarly, for forecasts of the HPS output gap and CPI inflation (Table 7), the RMSE ratio of the univariate–inflation detrended VAR(4) average forecast is less than 1 in all columns but one, in which the ratio is 1.010 ($h = 0Q$ forecasts of the HPS gap).

While not quite as good as the average of the univariate and inflation detrended VAR forecasts, some other averages also seem to perform well, beating the accuracy of the univariate benchmark with sufficient consistency as to be considered superior. In particular, two of the other pairwise forecasts — the VAR(4) with univariate and DVAR(4) with univariate averages — are often, although not always, more accurate than the univariate benchmarks. For instance, in forecasts of the HPS output gap and CPI inflation from models in these variables and the T-bill rate (Table 7), these pairwise averages' RMSE ratios are less than 1 in 8 of 12 columns, and only slightly to modestly above 1 (1.5 to 3.3 percent) in the exceptions. The VAR(4) with univariate average tends to have a more consistent advantage in 1985-05 forecasts. In addition, among the inflation forecasts, the three pairwise combinations (univariate with inflation detrended VAR(4), VAR(4) and DVAR(4)) are the most consistent out-performers of the univariate benchmark across both the 1970-84 and 1985-05 subsamples.

The rankings in Table 10 confirm that, from a broad perspective, the best forecasts are averages (averages of RMSE ratios instead of rankings confirm the same). In these rankings, the single best forecast is the average of the forecasts from the univariate and inflation detrended VAR(4). Across all variables, horizons, and samples, this forecast has an average ranking of 7.1; the next–best forecast, the average of the univariate and VAR(4) forecasts, has an average ranking of 12.5. While the univariate–inflation detrended VAR(4) average is, in relative terms, especially good for forecasting the T-bill rate (see column 5), this forecast retains its top rank even when interest rate forecasts are dropped from the calculations (column 2). This average forecast also performs relatively well for forecasting both output (column 3 shows it ranks a close second to the BVAR(4) with inflation detrending — column 3) and inflation (column 3 shows it ranks first). As to sample stability, the univariate–inflation detrended VAR(4) average is not best across both samples, but consistent. In the 1970-84 sample, this average is outranked by several others (column 6), but in the later sample, it is the top-ranked forecast.[14]

4.4 Averaging methods that only sometimes outperform the benchmark

Among other forecasts, it is difficult to identify any methods that might be seen as consistently equaling or materially beating the univariate benchmark. Take, for instance, the simple equally weighted average of all forecasts, applied to a model in GDP growth, GDP inflation, and the T-bill rate (Table 3). This averaging approach is consistent in beating the univariate benchmark in the 1970-84 sample, but typically fails to beat the benchmark in the 1985-05 sample. Similarly, in the case of T-bill forecasts from the same model (Table 9, left half), the all–model average loses out to the univariate benchmark for two of the six combinations of horizon and sample, while the generally best–performing method of averaging the univariate and inflation detrended VAR(4) forecasts beats the univariate

[14] Note that, although forecasts such as the best quartile (recursive) have a higher rank in the first sample, the average accuracy gain offered by the forecasts (for output and inflation) is small. For 1970-84, the average RMSE ratio of the best quartile (recursive) forecast is .918, the lowest of all methods; the average ratio for the univariate–inflation detrended VAR(4) forecast is .929.

benchmark in all cases.

A number of the other averaging methods perform quite comparably to the simple average — and thus, by extension, fail to consistently equal or beat (materially) the univariate benchmark. Among the broad average forecasts, from the results in Tables 3-9 there seems to be no advantage of a median forecast or trimmed means over the simple average. The accuracy of these forecasts tends to be quite similar. For example, in the case of 1-year ahead forecasts of GDP growth and GDP inflation for 1985-05, the 20 percent trimmed mean forecast's RMSE ratios are .991 (growth) and 1.033 (inflation), compared to the simple average's ratios of, respectively, .990 and 1.044 (Table 3).

Similarly, MSE–weighted forecasts are quite comparable to simple average forecasts, in terms of RMSE accuracy.[15] To use the same example of 1-year ahead forecasts of GDP growth and GDP inflation for 1985-05, the recursively MSE–weighted forecast's RMSE ratios are .998 (growth) and 1.037 (inflation), compared to the simple average's ratios of, respectively, .990 and 1.044 (Table 3). In 1-year ahead forecasts of CPI inflation (Table 6), the RMSE ratio of the recursively MSE–weighted forecast is .797 for 1970-84 and 1.034 for 1985-05, compared to the simple average forecast's RMSE ratios of .816 and 1.024, respectively.

Forecasts based on AIC and BIC weighting of models are even more similar to the simple average forecasts — so similar as to be hard to distinguish. Among all the results in Tables 3-9, the RMSE ratios of the AIC and BIC–based forecasts typically differ from the RMSE ratio of the simple average by no more than .003. For instance, in the case of 1-year ahead forecasts of GDP growth and GDP inflation for 1985-05, the BIC–weighted forecast's RMSE ratios are .991 (growth) and 1.047 (inflation), compared to the simple average's ratios of, respectively, .990 and 1.044 (Table 3). The similarities of the AIC and BIC–weighted forecasts to the simple average reflects similarities of the AIC and BIC weights to equal weights. For example, in the case of the Table 3 models estimated with 1988:Q1 vintage data, the BIC weights for GDP growth forecasts range from 1.34 percent to 2.12 percent; the AIC weights for GDP growth range from 1.59 percent to 1.97 percent (in all cases, the inflation weight ranges are comparable).

Using generalized ridge regression (with some shrinkage) to combine forecasts yields results that are more mixed when compared to the simple average, although neither con-

[15]However, in the case of forecasts of the HP output gap, the MSE–weighted averages are consistently slightly better than the simple averages.

sistently much better nor consistently much worse. For instance, in Table 6's results for forecasts of GDP growth, the RMSE ratios of the $k = 1$ recursive ridge regression forecast are consistently below those of the simple average forecast; in the case of Table 5's 1970-84 forecasts of CPI inflation, the strong shrinkage ridge forecast is less accurate (has higher RMSE ratios) than the simple average. Note that, within the set of ridge–based forecasts, using more shrinkage toward equal weights generally improves forecast accuracy.

Similarly, using the best–quartile forecast yields mixed results. The best quartile forecasts are sometimes more accurate and other times less accurate than the simple average and univariate forecasts. For example, in Table 6's results for 1-year ahead forecasts of GDP growth, the best quartile forecast based on a 10 year rolling sample has a RMSE ratio of .712 for 1970-84 and 1.107 for 1985-05, compared to the simple average forecast's RMSE ratios of, respectively, .816 and 1.024. Similarly, for Table 6's CPI inflation forecasts, the 10 year rolling best quartile approach yields a forecast that is more accurate than the simple average for 1970-84 and less accurate for 1985-05. Where the best quartile forecast seems to have a consistent advantage over a simple average is in output forecasts (especially for the HP output gap) for 1970-84.

The rankings in Table 10 confirm the broad similarity of the above methods — the simple average, MSE–weighted averages, AIC and BIC–weighted averages, ridge regression combinations, and best quartile forecasts. For example, the simple average forecast has an overall average ranking of 14.5, compared to rankings of 12.5 for the recursive MSE–weighted forecast and 16.1 for the AIC-weighted forecast (unreported average RMSE ratios are even more similar). By comparison, the best forecast, the univariate–inflation detrended VAR(4) average, has an overall ranking of 7.1. In a very broad sense, most of the aforementioned average forecasts are better than the univariate benchmarks in that they all have higher rankings than the univariate's average ranking of 18.7 (column 1). Note, however, that most of their advantage comes in the 1970-84 sample; in the later sample, the univariate forecast generally ranks higher. For instance, for 1970-84 output and inflation forecasts, the all–model average has an average accuracy rank of 13.2, compared to the univariate ranking of 26.0. But for 1985-05 forecasts, the all–model average has an average accuracy rank of 18.6, compared to the univariate ranking of 13.4.

4.5 Averaging methods that rarely outperform the benchmark

A number of the other averaging or combination methods are clearly dominated by univariate benchmarks (and, in turn, other average forecasts). OLS combinations or ridge combinations that approximate OLS often fare especially poorly. The OLS–approximating ridge regression combination (the one with $k = .001$) consistently yields poor forecasts. For example, in the case of 1985-05 1–year ahead forecasts of CPI inflation from models with GDP growth (Table 6), the RMSE ratio of the recursively estimated ridge regression with shrinkage parameter of .001 is 1.496. In other instances, the RMSE of the OLS–approximating ridge combination is about twice as large as that of the univariate benchmark. Similarly, the forecasts based on OLS combination regression using the first and second quartile average forecasts — especially those using rolling regressions — are generally (although not always, to be sure) dominated by other average forecasts. In the same example, the RMSE ratios of the forecasts based on rolling OLS combinations of the top two quartile forecasts are 1.129 (10 year rolling) and 1.701 (5 year rolling), respectively, compared to the all–average forecast's RMSE ratio of 1.024.

Forecasts based on using factor model methods to obtain a combination are also generally less accurate than alternatives such as the univariate and simple average forecasts. For example, in the case of 1-year ahead forecasts of GDP growth and GDP inflation for 1985-05, the recursively estimated factor combination forecast's RMSE ratios are 1.054 (growth) and 1.545 (inflation), compared to the simple average's ratios of, respectively, .990 and 1.044 (Table 3). The same is true for the PLS forecasts: although PLS forecasts are sometimes more accurate than the simple average, they are often much worse. In the same example, the recursive PLS forecast's RMSE ratios are 1.480 and 1.011, respectively. The BMA forecasts are also generally dominated by the simple average (dramatically in some instances, as with HP output gap forecasts), although BMA fares well in cases such as 1985-05 GDP growth forecasts. Continuing with the sample example, the BMA, $\phi = .2$ forecast has RMSE ratios of .965 (output) and 1.277 (inflation); the BMA, $\phi = 2$ forecast has RMSE ratios of 1.085 (output) and 1.184 (inflation). Finally, the forecast obtained by averaging recursive and rolling VAR(4) estimates is clearly dominated by the simple average forecast. To use the same application, this average forecast has RMSE ratios of 1.134 (output) and 1.123 (inflation).

The rankings in Table 10 provide a clear and convenient listing of the forecast methods that are consistently worse than the univariate benchmark and alternatives such as the best–performing pairwise average forecast and the all–model simple average. As previously mentioned, generalized ridge forecasts with little shrinkage ($k = .001$, so as to approximate OLS–based combination) typically perform among the worst forecasts for all horizons, variables and periods, with average ranks for all columns above 30. OLS combinations of quartiles using a 5 year rolling sample are similarly poor (low–ranked). Among the others mentioned as being poor performers (such as PLS, BMA, factor-based combination, and the simple average of recursive and rolling VAR(4)'s) a few of the average rankings drop as low as the high teens but are more frequently in the 20s.

4.6 Single VAR methods

Among the single VAR forecasts included for comparison, the BVAR(4) with inflation detrending is generally best. While shrinkage in the form of averaging forecasts from an inflation detrended VAR(4) with univariate forecasts is better than estimating the inflation detrended VAR(4) by Bayesian methods, the latter at least performs comparably to the simple average forecast. For example, as shown in Table 3, forecasts of GDP growth from the BVAR(4) with inflation detrending are often at least as accurate as the simple average forecasts (as, for example, with 1-year ahead forecasts for 1985-05). However, forecasts of GDP inflation from the same model are generally less accurate than the simple average (see, for example, the 1-year ahead forecasts for 1985-05). These examples reflect a pattern evident throughout Tables 3-8: while inflation detrending might be expected to most improve inflation forecasts, it instead most improves output forecasts. Although the accuracy of the other individual VAR models is more variable, overall these models are more clearly dominated by the univariate benchmark and others such as the simple average forecast. For example, in the case of the BVAR(4) using GDP growth and GDP inflation (and the T-bill rate), the simple average forecasts are generally more accurate than the BVAR(4) forecasts of growth over 1970-84, inflation over 1970-84, and inflation over 1985-05 (Table 3).

Consistent with these examples, in general, forecasts from single models are dominated by average forecasts. The pattern is clearly evident in the average rankings of Table 10. Across all variables, horizons, and samples, the best–ranked single model is the BVAR(4) with inflation detrending, which is out–ranked by 12 different average forecasts (however, in terms of average RMSE ratios, for a number of the averages with somewhat higher ranks there is little to distinguish this single model from the averages). The other single models rank well below the BVAR(4) with inflation detrending.

While averages are broadly more accurate than single model forecasts, it is less clear that they are consistently more accurate across sample periods. To check consistency, we calculated the correlation of the ranks of all 33 average forecasts and all 54 single model forecasts across the 1970-84 and 1985-05 periods, once averaging across the output and inflation measures as in the final two columns of Table 10 and then again including the T-bill rate. In both cases the correlation was roughly 50 percent for the single model forecasts. In contrast, the average forecasts had correlations of 70 and 80 percent, respectively. The implication is that not only is the typical average forecast more accurate than the typical single model forecast, it is also consistently so across the two periods.

4.7 Interpretation

Why might simple averages in general and the pairwise average of univariate and inflation–detrended VAR(4) forecasts be more accurate than any single model? As noted in the introduction, in practice it is very difficult to know the form of structural instability, and competing models will differ in their sensitivity to structural change. In such an environment, averages across models are likely to be superior to any single forecast. In line with prior research on combining a range of forecasts that incorporate information from different variables (such as Stock and Watson (1999a, 2004), Smith and Wallis (2005)), simple equally weighted averages are typically at least as good as averages based on weights tied to historical forecast accuracy. The limitations of weighted averages relative to simple averages are commonly attributed to difficulties in estimating potentially optimal weights in finite samples, especially the when cross–section dimension is large relative to the time dimension.

As to the particular success of forecasts using inflation detrending, one interpretation is that removing a smooth inflation trend — a trend that matches up well with long–term inflation expectations — from both inflation and the interest rate does a reasonable job of capturing non–stationarities in inflation and interest rates. Kozicki and Tinsley (2001a,b, 2002) have developed such VARs from models with learning about an unobserved, time–varying inflation target of the central bank.

However, such a single representation is surely not the true model, and noise in estimating the many parameters of the model likely have an adverse effect on forecast accuracy. Therefore, a better forecast can be obtained by applying some form of shrinkage. One approach, which primarily addresses parameter estimation noise, is to estimate the VAR with inflation detrending using Bayesian shrinkage. Another approach is to combine forecasts from the inflation detrended VAR with forecasts from an alternative model — in our case, the univariate benchmark (note that the IMA(1) benchmarks for inflation and the T-bill rate imply random walk trends).[16] Koop (2006) notes that such model averaging can be viewed as a form of shrinkage for addressing both parameter estimation noise and model uncertainty. The superiority of this average forecast can be interpreted as highlighting the value of inflation detrending, shrinkage of parameter noise, and shrinkage to deal with model uncertainty.[17]

4.8 Comparisons with the Survey of Professional Forecasters

Table 11 compares the accuracy of selected model and average forecasts (generally the better performing forecasts above) with the accuracy of SPF projections. The variables we report are those for which SPF forecasts exist: GDP growth, GDP inflation, CPI inflation (in the case of CPI inflation, the SPF forecasts don't begin until 1981, so we only report CPI results for the 1985-05 period), and the T-bill rate. The first row of the table provides the raw RMSEs of SPF forecasts; the remaining rows provide ratios of the RMSE of model or average forecasts relative to the RMSE of the corresponding SPF forecast.

Perhaps not surprisingly, in light of the results in studies such as Ang, Bekaert, and Wei (2006), the SPF forecasts generally dominate the time series model and average forecasts. For example, in $h = 0Q$ forecasts of GDP growth and GDP inflation for 1970-84, the (10 year rolling) MSE–weighted average forecast's RMSEs exceed the SPF RMSEs by about 44 and 17 percent, respectively. At such short horizons, of course, the SPF has a considerable information advantage over simple time series methods. As described in Croushore (1993), the SPF forecast is based on a survey taken in the second month of each quarter. Survey respondents then have considerably more information, on variables such as interest rates and stock prices, than is reflected in time series forecasts that don't include the same information (as reflected in the bottom panel of Table 11, that information gives the SPF its biggest advantage in near-term interest rates). However, the SPF's advantage over time series methods generally declines as the forecast horizon rises. For instance, in $h = 1Y$ forecasts

[16] As discussed in Stock and Watson (2005), suppose inflation is equal to the sum of a trend component and a cycle component. Moreover, suppose the trend is a random walk and the cycle is just white noise. The change in inflation is then equal to the sum of the trend innovation and the change in the cycle component, which is an MA(1) process.

[17] The results of Clark and McCracken (2005c) can be used to make a frequentist case for averaging the inflation detrended VAR with the univariate benchmark, based entirely on parameter estimation error.

of GDP growth and GDP inflation for 1970-84, the RMSE ratios of the MSE–weighted average forecasts are, respectively, .991 and 1.070.

Moreover, the SPF's advantage is typically much greater in the 1970-84 sample than the 1985-05 sample. Campbell (2005) notes the same for SPF growth forecasts compared to AR(1) forecasts of GDP growth, attributing the pattern to declining predictability. For example, in this later period, the RMSE ratio of $h = 0Q$ forecasts of GDP growth from the MSE–weighted average method is 1.120, compared to 1.437 in the earlier period. Reflecting the declining predictability of output and inflation and the reduced advantage of the SPF at longer horizons, for 1–year ahead forecasts in the 1985-05 period, the advantage of the SPF over the averages of VAR forecasts is often quite small (or nonexistent). For instance, in 1–year ahead forecasts, the RMSE ratios of MSE–weighted average forecasts of GDP growth, GDP inflation, and the T-bill rate are, respectively, 1.011, .987, and .998. However, in forecasts of CPI inflation, the SPF retains a sizable (at least 22 percent, in RMSE) advantage over the accuracy of all VAR–based forecasts.

5 Conclusion

In this paper we provide empirical evidence on the ability of several forms of model averaging to improve the real–time forecast accuracy of small-scale macroeconomic VARs in the presence of uncertain forms of model instability. Focusing on six distinct trivariate models incorporating different measures of output and inflation (but a common interest rate measure), we consider a wide range of approaches to averaging forecasts (or forecast models) obtained with a variety of primitive methods for managing model instability. These primitive methods include incorporating different choices of lag selection, observation windows used for estimation, levels or differences, intercept corrections, stochastically time–varying parameters, break dating, discounted least squares, Bayesian shrinkage, detrending of inflation and interest rates. The forms of forecast averaging include: equally weighted averages with and without trimming, medians, common factor-based factors, combinations estimated with ridge regression, MSE–weighted averages, lowest MSE forecasts (predictive least squares–based forecasts), information theoretic averages, Bayesian model averages, and combinations based on quartile average forecasts.

Our results indicate that some forms of model averaging do consistently improve forecast accuracy in terms of root mean square errors (RMSE). Not surprisingly, the best method often varies with the variable being forecasted, but several patterns do emerge. After aggregating across all models, horizons and variables being forecasted, it is clear that the simplest forms of model averaging — such as those that use equal weights across all models or those that average a univariate model with a particular VAR, such as a VAR(4) with inflation detrending — consistently perform among the best methods. At the other extreme, forecasts based on OLS–type combination, predictive least squares, and factor model–based combination rank among the worst.

References

Aiolfi, Marco, and Allan Timmermann (2006), "Persistence in Forecasting Performance and Conditional Combination Strategies," *Journal of Econometrics*, forthcoming.

Ang, Andrew, Geert Bekaert and Min Wei (2006), "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?" *Journal of Monetary Economics*, forthcoming.

Bates, J.M., and Clive W.J. Granger (1969), "The Combination of Forecasts," *Operations Research Quarterly* 20, 451-468.

Beyer, Andreas, and Roger E.A. Farmer (2006), "Natural Rate Doubts," *Journal of Economic Dynamics and Control*, forthcoming.

Boivin, Jean (2005), "Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data," *Journal of Money, Credit and Banking*, forthcoming.

Campbell, Sean D. (2005), "Stock Market Volatility and the Great Moderation," FEDs Working Paper No. 2005-47.

Clark, Todd E. and Michael W. McCracken (2005a), "The Predictive Content of the Output Gap for Inflation: Resolving In–Sample and Out–of–Sample Evidence," *Journal of Money, Credit, and Banking*, forthcoming.

Clark, Todd E. and Michael W. McCracken (2005b), "Combining Forecasts from Nested Models," manuscript, Federal Reserve Bank of Kansas City.

Clark, Todd E. and Michael W. McCracken (2005c), "Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts," manuscript, Federal Reserve Bank of Kansas City.

Clark, Todd E. and Michael W. McCracken (2006), "Forecasting with Small Macroeconomic VARs in the Presence of Instability'," manuscript, Federal Reserve Bank of Kansas City.

Clements, Michael P. and David F. Hendry (1996), "Intercept corrections and structural change," *Journal of Applied Econometrics* 11, 475-94.

Cogley, Timothy (2002), "A Simple Adaptive Measure of Core Inflation," *Journal of Money, Credit, and Banking* 34, 94-113.

Cogley, Timothy and Thomas J. Sargent (2001), "Evolving Post World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual* 16, 331-73.

Cogley, Timothy and Thomas J. Sargent (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the Post World War II U.S.," *Review of Economic Dynamics* 8, 262-302.

Croushore, Dean (1993), "Introducing: the Survey of Professional Forecasters," Federal Reserve Bank of Philadelphia *Business Review*, Nov./Dec., 3-13.

Croushore, Dean (2005), "Forecasting with Real–Time Macroeconomic Data," *Handbook of Forecasting*, forthcoming.

Croushore, Dean and Tom Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105, 111-30.

D'Agostino, Antonello, Domenico Giannone and Paolo Surico (2005), "(Un)Predictability and Macroeconomic Stability," manuscript, ECARES.

Del Negro, Marco and Frank Schorfheide (2004), "Priors from General Equilibrium Models for VARs," *International Economic Review* 45, 643-73.

Diebold, Frank and Peter Pauly (1987),"The Use of Prior Information in Forecast Combination," *Journal of Forecasting*, 6 , 503-08.

Doan, Thomas, Robert Litterman and Christopher Sims (1984), "Forecasting and Conditional Prediction Using Realistic Prior Distributions," *Econometric Reviews* 3, 1-100.

Estrella, Arturo and Jeffrey C. Fuhrer (2003), "Monetary Policy Shifts and the Stability of Monetary Policy Models," *Review of Economics and Statistics* 85, 94-104.

Favero, Carlo and Massimiliano Marcellino (2005), "Modelling and Forecasting Fiscal Variables for the Euro Area," *Oxford Bulletin of Economics and Statistics*, forthcoming.

Hallman, Jeffrey J., Richard D. Porter and David H. Small (1991), "Is the Price Level Tied to the M2 Monetary Aggregate in the Long Run?" *American Economic Review* 81, 841-58.

Hendry, David F., and Michael P. Clements (2004), "Pooling of Forecasts," *Econometrics Journal* 7, 1-31.

Hodrick, Robert and Edward C. Prescott (1997), "Post-War U.S. Business Cycles: A Descriptive Empirical Investigation," *Journal of Money, Credit, and Banking* 29, 1-16.

Jacobson, Tor, Per Jansson, Anders Vredin and Anders Warne (2001), "Monetary Policy Analysis and Inflation Targeting in a Small Open Economy: a VAR Approach," *Journal of Applied Econometrics* 16, 487-520.

Kapetanios, George, Vincent Labhard, and Simon Price (2005), "Forecasting Using Bayesian and Information Theoretic Model Averaging: An Application to UK Inflation," Bank of England Working Paper no. 268.

Koop, Gary and Simon Potter (2003), "Forecasting in Large Macroeconomic Panels Using Bayesian Model Averaging," manuscript, Federal Reserve Bank of New York.

Koop, Gary (2006), "Discussion of (Stock and Watson's) 'An Empirical Comparison of Methods for Forecasting Using Many Predictors'," Deutsche Bundesbank conference on New Developments in Economic Forecasting, http://www.bundesbank.de/download/vfz/konferenzen/20060502_06_eltville/paper_koop.pdf.

Kozicki, Sharon, and Barak Hoffman (2004), "Rounding Error: A Distorting Influence on Index Data," *Journal of Money, Credit, and Banking* 36, 319-38.

Kozicki, Sharon and Peter A. Tinsley (2001a), "Shifting endpoints in the term structure of interest rates," *Journal of Monetary Economics* 47, 613-52.

Kozicki, Sharon and Peter A. Tinsley (2001b), "Term Structure Views of Monetary Policy Under Alternative Models of Agent Expectations," *Journal of Economic Dynamics and*

Control 25, 149-84.

Kozicki, Sharon and Peter A. Tinsley (2002), "Alternative Sources of the Lag Dynamics of Inflation," in *Price Adjustment and Monetary Policy*, Bank of Canada Conference Proceedings, 3-47.

Kunsch, Hans R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics* 17, 1217-41.

Levin, Andrew T. and Jeremy Piger (2003), "Is Inflation Persistence Intrinsic in Industrial Economies?" Working Paper 2002-023B, Federal Reserve Bank of St. Louis.

Litterman, Robert B. (1986), "Forecasting with Bayesian Vector Autoregressions — Five Years of Experience," *Journal of Business and Economic Statistics* 4, 25-38.

Liu, Regina Y. and Kesar Singh (1992), "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in R. Lepage and L. Billiard, eds., *Exploring the Limits of Bootstrap*, New York: Wiley, 22-148.

Maheu, John M. and Stephen Gordon (2004), "Learning, Forecasting and Structural Breaks," manuscript, University of Toronto.

Min, Chung-ki and Arnold Zellner (1993), "Bayesian and non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates," *Journal of Econometrics* 56, 89-118.

Nelson, Charles R. (1972), "The Predictive Performance of the FRB-MIT-PENN Model of the U.S. Economy," *American Economic Review* 62, 902-17.

Nelson, Charles R. and G. William Schwert (1977), "Short-Term Interest Rates as Predictors of Inflation: On Testing the Hypothesis that the Real Rate of Interest is Constant," *American Economic Review* 67, 478-86.

Ng, Serena and Pierre Perron (2005), "A Note on the Selection of Time Series Models," *Oxford Bulletin of Economics and Statistics* 67, 115-34.

Orphanides, Athanasios and Simon van Norden (2005), "The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time," *Journal of Money, Credit, and Banking* 37, 583-601.

Pesaran, M. Hashem, D. Pettenuzzo and Allan Timmermann (2004), "Bayesian Regime Averaging for Time Series subject to Structural Breaks," manuscript, UCSD.

Pesaran, M. Hashem, and Allan Timmermann (2006), "Selection of Estimation Window in the Presence of Breaks," *Journal of Econometrics*, forthcoming.

Phillips, Peter C.B. (1996), "Econometric Model Determination," *Econometrica* 64, 763-812.

Rapach, David E. and Jack K. Strauss (2005), "Forecasting U.S. Employment Growth Using Forecast Combining Methods," manuscript, St. Louis University.

Roberts, John M. (2004), "Monetary Policy and Inflation Dynamics," FEDs Working Paper No. 2004-62.

Robertson, John and Ellis Tallman (2001), "Improving Federal-Funds Rate Forecasts in VAR Models Used for Policy Analysis," *Journal of Business and Economic Statistics* 19, 324-30.

Rogoff, Kenneth (2003), "Globalization and Global Disinflation," in *Monetary Policy and Uncertainty: Adapting to a Changing Economy*, Federal Reserve Bank of Kansas City.

Romer, Christina D. and David H. Romer (2000), "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review* 90, 429-57.

Rudebusch, Glenn D. (2005), "Assessing the Lucas Critique in Monetary Policy Models," *Journal of Money, Credit, and Banking* 37, 245-72.

Rudebusch, Glenn D. and Svensson, Lars E.O. (1999), "Policy Rules for Inflation Targeting." in J. Taylor, ed., *Monetary Policy Rules*, University of Chicago Press: Chicago, 203-46.

Sims, Christopher A. (1980), "Macroeconomics and Reality," *Econometrica* 48, 1-48.

Sims, Christopher A. (2002), "The Role of Models and Probabilities in the Monetary Policy Process," *Brookings Papers on Economic Activity* 2, 1-40.

Smith, Jeremy, and Kenneth F. Wallis (2005), "Combining Point Forecasts: The Simple Average Rules, OK?" manuscript, University of Warwick.

Stock, James H. and Mark W. Watson (1996), "Evidence on Structural Stability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics* 14, 11-30.

Stock, James H. and Mark W. Watson (1999a), "A Dynamic Factor Model Framework for Forecast Combination," *Spanish Economic Review* 1, 91-121.

Stock, James H. and Mark W. Watson (1999b), "Forecasting Inflation," *Journal of Monetary Economics* 44, 293-335.

Stock, James H. and Mark W. Watson (2003), "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature* 41, 788-829.

Stock, James H. and Mark W. Watson (2004), "Combination Forecasts of Output Growth in a Seven–Country Data Set,"*Journal of Forecasting* 23, 405-30.

Stock, James H. and Mark W. Watson (2005), "Has Inflation Become Harder to Forecast?" manuscript, Princeton University.

Timmermann, Allan (2005), "Forecast Combinations," *Handbook of Forecasting*, forthcoming.

Tulip, Peter (2005), "Has Output Become More Predictable? Changes in Greenbook Forecast Accuracy," FEDs Working Paper No. 2005-31.

Webb, Roy H. (1995), "Forecasts of Inflation from VAR Models," *Journal of Forecasting* 14, 267-85.

White, Halbert (2000), "A Reality Check for Data Snooping," *Econometrica* 68, 1097-1126.

Table 1: VAR forecasting methods

method	details
VAR(4)	VAR in y, π, i with fixed lag of 4
VAR(2)	same as above with fixed lag of 2
VAR(AIC)	VAR with system lag determined at each t by AIC
VAR(BIC)	VAR with system lag determined at each t by BIC
VAR(AIC, by eq.&var.)	VAR in y, π, i allowing different, AIC-chosen lags for each variable in each equation
VAR(BIC, by eq.&var.)	same as above, with BIC-determined lags
DVAR(4)	VAR in y, $\Delta\pi$, Δi with fixed lag of 4
DVAR(2)	same as above with fixed lag of 2
DVAR(AIC)	VAR in y, $\Delta\pi$, Δi with system lag determined at each t by AIC
DVAR(BIC)	VAR in y, $\Delta\pi$, Δi with system lag determined at each t by BIC
DVAR(AIC, by eq.&var.)	VAR in y, $\Delta\pi$, Δi allowing different, AIC-chosen lags for each variable in each equation
DVAR(BIC, by eq.&var.)	same as above, with BIC-determined lags
BVAR(4)	VAR(4) in y, π, i, est. with Minnesota priors, using $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$
BDVAR(4)	VAR(4) in y, $\Delta\pi$, Δi, est. with Minnesota priors, using $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$
VAR(4), rolling	VAR in y, π, i with fixed lag of 4, estimated with a rolling sample
VAR(2), rolling	same as above with fixed lag of 2
VAR(AIC), rolling	same as above with AIC–determined lag
VAR(BIC), rolling	same as above with BIC–determined lag
VAR(AIC, by eq.&var.), rolling	same as above with AIC-determined lags for each var. in each eq.
VAR(BIC, by eq.&var.), rolling	same as above with BIC-determined lags for each var. in each eq.
DVAR(4), rolling	VAR in y, $\Delta\pi$, Δi with fixed lag of 4, estimated with a rolling sample
DVAR(2), rolling	same as above with fixed lag of 2
DVAR(AIC), rolling	same as above with AIC–determined lag
DVAR(BIC), rolling	same as above with BIC–determined lag
DVAR(AIC, by eq.&var.), rolling	same as above with AIC-determined lags for each var. in each eq.
DVAR(BIC, by eq.&var.), rolling	same as above with BIC-determined lags for each var. in each eq.
BVAR(4), rolling	BVAR(4) in y, π, i with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$, est. with a rolling sample
BDVAR(4), rolling	BVAR(4) in y, $\Delta\pi$, Δi with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$, est. with a rolling sample
DLS, VAR(4)	VAR(4) in y, π, i, est. by DLS, using dis. rates of .01 for y eq. and .05 for π and i eq.
DLS, VAR(2)	same as above with fixed lag of 2
DLS, VAR(AIC)	same as above with lag determined from AIC applied to OLS estimates of system
DLS, DVAR(4)	VAR(4) in y, $\Delta\pi$, Δi, est. by DLS using dis. rates of .01 for y eq. and .05 for $\Delta\pi$ and Δi eq.
DLS, DVAR(2)	same as above with fixed lag of 2
DLS, DVAR(AIC)	same as above with lag determined from AIC applied to OLS estimates of system
VAR(AIC), AIC intercept breaks	VAR(AIC lags) in y, π, i, with intercept breaks (up to 2) chosen to minimize the AIC
VAR(AIC), BIC intercept breaks	same as above, using the BIC to determine the number of intercept breaks
VAR(4), intercept correction	VAR(4) forecasts adjusted by the average value of the last four OLS residuals
VAR(AIC), intercept correction	VAR(AIC lag) forecasts adjusted by the average value of the last four OLS residuals
VAR(4), inflation detrending	VAR(4) in y, $\pi - \pi^*_{-1}$, and $i - \pi^*_{-1}$, where $\pi^* = \pi^*_{-1} + .05(\pi - \pi^*_{-1})$
VAR(2), inflation detrending	same as above with fixed lag of 2
VAR(AIC), inflation detrending	same as above with AIC–determined lag for the system in y, $\pi - \pi^*_{-1}$, and $i - \pi^*_{-1}$
VAR(BIC), inflation detrending	same as above with BIC–determined lag for the system in y, $\pi - \pi^*_{-1}$, and $i - \pi^*_{-1}$
BVAR(4), inflation detrending	BVAR(4) in y, $\pi - \pi^*_{-1}$, and $i - \pi^*_{-1}$, using $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$
VAR(4), full exp. sm. detrending	VAR(4) in y, $\pi - \pi^*_{-1}$, and $i - i^*_{-1}$, where $\pi^* = \pi^*_{-1} + .05(\pi - \pi^*_{-1})$, $i^* = i^*_{-1} + .07(i - i^*_{-1})$
VAR(2), full exp. sm. detrending	same as above with fixed lag of 2
VAR(AIC), full exp. sm. detrending	same as above with AIC–determined lag for the system in y, $\pi - \pi^*_{-1}$, and $i - i^*_{-1}$
VAR(BIC), full exp. sm. detrending	same as above with BIC–determined lag for the system in y, $\pi - \pi^*_{-1}$, and $i - i^*_{-1}$
BVAR(4) with TVP	TVP BVAR(4) in y, π, i with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = .1$, $\lambda = .0005$
BVAR(4) with TVP, $\lambda_4 = .5, \lambda = .0025$	TVP BVAR(4) in y, π, i with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = .5$, $\lambda = .0025$
BVAR(4) with TVP, $\lambda_4 = 1000, \lambda = .005$	TVP BVAR(4) in y, π, i with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$, $\lambda = .005$
BVAR(4) with TVP, $\lambda_4 = 1000, \lambda = .0001$	TVP BVAR(4) in y, π, i with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$, $\lambda = .0001$
BVAR(4) with intercept TVP	BVAR(4) in y, π, i, TVP in intercepts, $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = .1$, $\lambda = .0005$
BVAR(4) with intercept TVP, $\lambda_4 = .5, \lambda = .0025$	BVAR(4) in y, π, i, TVP in intercepts, $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = .5$, $\lambda = .0025$
univariate	AR(2) for y, rolling MA(1) for $\Delta\pi$, rolling MA(1) for Δi

Notes:
1. The variables y, π, and i refer to, respectively, output (GDP growth, the HPS gap, or the HP gap), inflation (GDP or CPI inflation), and the 3-month T-bill rate.
2. Unless otherwise noted, all models are estimated recursively, using all data (starting in 1955 or later) available up to the forecasting date. The rolling estimates of the univariate models for $\Delta\pi$ and Δi use 40 observations. The rolling estimates of the VAR models use 60 observatinos.
3. The AIC and BIC lag orders range from 0 (the minimum allowed) to 4 (the maximum allowed).
4. The intercept correction approach takes the form of equation (40) in Clements and Hendry (1996).
5. In BVAR estimates, prior variances take the "Minnesota" style described in Litterman (1986). The prior variances are determined by hyperparameters λ_1 (general tightness), λ_2 (tightness of lags of other variables compared to lags of the dependent variable), λ_3 (tightness of longer lags compared to shorter lags), and λ_4 (tightness of intercept). The prior standard deviation of the coefficient on lag k of variable j in equation j is set to $\frac{\lambda_1}{k^{\lambda_3}}$. The prior standard deviation of the coefficient on lag k of variable m in equation j is $\frac{\lambda_1 \lambda_2}{k^{\lambda_3}} \frac{\sigma_j}{\sigma_m}$, where σ_j and σ_m denote the residual standard deviations of univariate autoregressions estimated for variables j and m. The prior standard deviation of the intercept in equation j is set to $\lambda_4 \sigma_j$. In fixed parameter BVARs, we use generally conventional hyperparameter settings of $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, and $\lambda_4 = 1000$. The prior means for all coefficients are generally set at 0, with the following exceptions: (a) prior means for own first lags of π and i are set at 1 (in models with levels of inflation and interest rates); (b) prior means for own first lags of y are set at 0.8 in models with an output gap; and (c) prior means for the intercept of GDP growth equations are set to the historical average of growth in BVAR estimates that impose informative priors ($\lambda_4 = .1$ or .5) on the constant term.
6. The exponential smoothing used in the models with detrending is initialized with the average value of inflation over the first five years of each sample.
7. The time variation in the coefficients of the TVP BVARs takes a random walk form. The variance matrix of the coefficient innovations is set to λ times the Minnesota prior variance matrix. In time–varying BVARs with flat priors on the intercepts ($\lambda_4 = 1000$), the variation of the innovation in the intercept is set at λ times the prior variance of the coefficient on the own first lag instead of the prior variance of the constant.

Table 2: Forecast averaging methods

method	details
avg. of VAR(4), univariate	average of forecasts from univariate model and VAR(4) in y, π, and i
avg. of infl. detr. VAR(4), univar.	average of forecasts from univariate model and VAR(4) in y, $\pi - \pi^*_{-1}$, and $i - \pi^*_{-1}$
avg. of DVAR(4), univariate	average of forecasts from univariate model and VAR(4) in Δy, $\Delta\pi$, and i
avg. of VAR(4), rolling VAR(4)	average of forecasts from recursive and rolling estimates of VAR(4) in y, π, and i
average of all forecasts	simple average of forecasts from models listed in Table 1
median	median of model forecasts
trimmed mean, 10%	average of model forecasts, excluding 3 highest and 3 lowest
trimmed mean, 20%	average of model forecasts, excluding 5 highest and 5 lowest
ridge: recursive, .001	combination of model forecasts, est. with ridge regression (), $k = .001$
ridge: recursive, .25	same as above, using $k = .25$
ridge: recursive, 1.	same as above, using $k = 1$
ridge: 10y rolling, .001	same as above, using $k = .001$ and a rolling window of 40 forecasts
ridge: 10y rolling, .25	same as above, using $k = .25$ and a rolling window of 40 forecasts
ridge: 10y rolling, 1.	same as above, using $k = 1$ and a rolling window of 40 forecasts
factor, recursive	forecast from regression on common factor in model forecasts
factor, 10y rolling	same as above, using rolling window of 40 forecasts
MSE weighting, recursive	inverse MSE–weighted average of model forecasts
MSE weighting, 10y rolling	same as above, using a rolling window of 40 forecasts
MSE weighting, 5y rolling	same as above, using a rolling window of 40 forecasts
MSE weighting, discounted	inverse discounted MSE–weighted average of model forecasts, with discount rate of .95
PLS, recursive	forecast from model with lowest historical MSE
PLS, 10y rolling	same as above, using a rolling window of 40 forecasts
PLS, 5y rolling	same as above, using a rolling window of 20 forecasts
best quartile, recursive	simple average of model forecasts in the top quintile of historical (MSE) accuracy
best quartile, 10y rolling	same as above, using a rolling window of 40 forecasts
best quartile, 5y rolling	same as above, using a rolling window of 20 forecasts
OLS comb. of quartiles, recursive	forecast from (OLS) regression on the avg. forecasts from the 1st and 2nd quartiles
OLS comb. of quartiles, 10y rolling	same as above, using a rolling window of 40 forecasts
OLS comb. of quartiles, 5y rolling	same as above, using a rolling window of 20 forecasts
AIC weighting	weighted–average of model forecasts, with weights based on AIC
BIC weighting	weighted–average of model forecasts, with weights based on BIC
BMA, $\phi = .2$	weighted–average of model forecasts, with BMA weights using $\phi = .2$
BMA, $\phi = 2$	same as above, using $\phi = 2$

Notes:
1. All averages are based on the 54 forecast models listed in Table 1, for a given combination of measures of output, inflation, and the short-term interest rate.
2. See the notes to Table 1.

Table 3: Real-time RMSE results for GDP growth and GDP inflation

(RMSEs in first row, RMSE ratios in all others)

	GDP growth forecasts						GDP inflation forecasts					
	1970-84			1985-2005			1970-84			1985-2005		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univar	4.550	5.023	3.633	1.755	1.826	1.367	1.911	2.242	2.466	.989	1.052	.743
VAR(4)	1.063	.949	.940	1.115	1.128	1.051	.994	1.014	1.066	1.000	*.944*	.936
DVAR(4)	1.035	.928	*.761*	1.219	1.252	1.086	.998	**.941**	**.901**	.996	*.956*	1.011
BVAR(4)	.958	*.921*	.975	1.030	1.034	.979	.956	1.049	1.096	1.016	1.012	1.146
BVAR(4) with TVP	.957	*.930*	.970	.995	.987	*.919*	.963	1.053	1.106	1.003	.977	1.003
BVAR(4), inflation detrending	**.883**	*.830*	*.811*	1.067	1.060	.937	1.001	1.063	1.073	1.010	1.027	1.287
avg. of VAR(4), univariate	.986	*.922*	.893	1.024	1.021	.956	*.956*	.976	.999	.981	*.957*	*.931*
avg. of DVAR(4), univariate	.955	*.889*	*.795*	1.071	1.080	.996	.965	*.952*	*.934*	*.979*	*.961*	.969
avg. of infl. detr. VAR(4), univar.	.958	*.894*	*.856*	1.017	1.016	*.918*	*.961*	.972	.975	.986	*.967*	.970
avg. of VAR(4), rolling VAR(4)	1.111	1.005	.979	1.110	1.155	1.134	.983	1.029	1.064	1.040	.997	1.123
average of all forecasts	*.932*	*.861*	*.812*	1.066	1.079	.990	*.934*	.991	.977	1.022	.997	1.044
median	.935	*.872*	*.835*	1.051	1.087	1.008	.952	1.008	1.000	1.014	.988	1.016
trimmed mean, 10%	.935	*.865*	*.821*	1.065	1.078	.990	*.940*	.993	.983	1.021	.993	1.036
trimmed mean, 20%	.938	*.868*	*.825*	1.064	1.077	.991	*.943*	.996	.987	1.020	.992	1.033
ridge: recursive, .001	2.143	1.739	1.347	1.195	1.316	1.474	1.238	1.531	1.834	1.162	1.167	1.627
ridge: recursive, .25	.938	*.880*	*.740*	1.088	1.125	1.078	.968	1.045	1.098	1.019	.986	1.037
ridge: recursive, 1.	*.917*	*.847*	*.731*	1.092	1.132	1.037	.964	1.040	1.090	1.017	.983	.980
ridge: 10y rolling, .001	2.340	1.953	1.409	1.502	1.724	1.782	1.206	1.504	1.811	1.187	1.204	1.623
ridge: 10y rolling, .25	.958	*.888*	*.742*	1.070	1.096	1.028	.967	1.037	1.084	1.026	.976	1.043
ridge: 10y rolling, 1.	.925	*.850*	*.734*	1.081	1.112	1.004	.962	1.032	1.077	1.022	.981	.993
factor, recursive	.970	.894	.871	1.109	1.136	1.054	1.005	1.088	.995	1.031	1.075	1.545
factor, 10y rolling	.976	.900	.880	1.124	1.130	1.120	1.021	1.125	1.042	1.017	1.011	1.425
MSE weighting, recursive	*.930*	*.857*	*.791*	1.064	1.082	.998	*.935*	.992	.982	1.021	.996	1.037
MSE weighting, 10y rolling	*.930*	*.857*	*.792*	1.062	1.078	.985	*.935*	.992	.983	1.021	.994	1.037
MSE weighting, 5y rolling	*.928*	*.856*	*.797*	1.056	1.073	.983	*.935*	.988	.971	1.022	.998	1.034
MSE weighting, discounted	*.930*	*.858*	*.794*	1.061	1.078	.992	*.934*	.990	.979	1.021	.995	1.031
PLS, recursive	.937	**.800**	*.752*	1.149	1.335	1.480	.993	1.071	1.203	.980	.978	1.011
PLS, 10y rolling	.917	*.800*	*.737*	1.092	1.237	1.273	1.043	1.048	1.253	1.045	.963	1.092
PLS, 5y rolling	*.915*	*.901*	*.848*	1.120	1.197	1.091	1.053	1.080	1.307	1.110	1.006	1.129
best quartile, recursive	*.922*	*.854*	**.717**	1.060	1.145	1.067	.946	.993	1.000	1.014	.980	1.060
best quartile, 10y rolling	*.925*	*.855*	*.725*	1.057	1.101	1.051	.948	.990	1.010	1.011	.977	1.034
best quartile, 5y rolling	*.917*	*.860*	*.766*	1.053	1.118	1.090	*.943*	.969	.969	1.023	.979	1.026
OLS comb. of quartiles, recursive	.951	*.895*	*.747*	1.229	1.193	.921	.997	1.097	1.149	1.030	1.000	1.403
OLS comb. of quartiles, 10 year rolling	.953	.919	*.766*	1.114	1.097	1.114	1.018	1.143	1.204	1.011	.955	1.349
OLS comb. of quartiles, 5 year rolling	1.019	.948	*.841*	1.269	1.375	1.703	1.008	1.065	1.353	1.088	1.062	1.052
AIC weighting	*.931*	*.860*	*.811*	1.068	1.081	.992	*.937*	.992	.982	1.021	.998	1.047
BIC weighting	*.929*	*.859*	*.812*	1.068	1.081	.991	*.938*	.994	.983	1.021	.999	1.047
BMA, $\phi = .2$	.991	.945	.957	1.028	1.054	.965	.960	1.031	1.049	1.015	1.031	1.277
BMA, $\phi = 2$	.966	*.905*	.891	1.145	1.166	1.085	.999	1.041	1.040	1.040	1.024	1.184

Notes:

1. The entries in the first row are RMSEs, for variables defined in annualized percentage points. All other entries are RMSE ratios, for the indicated specification relative to the corresponding univariate specification.
2. Individual RMSE ratios that are significantly below 1 according to bootstrap p–values are indicated by a *slanted* font. In each column, if the best RMSE ratio is significantly less than 1 according to data snooping–robust p–values (bootstrapped as in White (2000)), the RMSE ratio appears in a **bold** font.
3. In each quarter t from 1970:Q1 through 2005:Q4, vintage t data (which generally end in $t-1$) are used to form forecasts for periods t ($h = 0Q$), $t+1$ ($h = 1Q$), and $t+4$ ($h = 1Y$). The forecasts of GDP growth and inflation for the $h = 1Y$ horizon correspond to annual percent changes: average growth and average inflation from $t+1$ through $t+4$.
4. Tables 1 and 2 provide further detail on the forecast methods.

Table 4: Real-time RMSE results for the HPS output gap and GDP inflation

(RMSEs in first row, RMSE ratios in all others)

	HPS output gap forecasts						GDP inflation forecasts					
	1970-84			1985-2005			1970-84			1985-2005		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univar	1.174	1.994	3.814	.824	1.166	2.148	1.911	2.242	2.466	.989	1.052	.743
VAR(4)	1.134	1.064	1.043	1.065	1.117	1.177	.997	1.021	1.074	.992	*.943*	.934
DVAR(4)	1.074	.961	*.751*	1.049	1.074	1.053	1.007	.951	*.908*	1.012	.993	1.165
BVAR(4)	1.075	1.005	1.003	1.057	1.113	1.152	.961	1.030	1.047	1.003	.987	1.041
BVAR(4) with TVP	1.046	.993	.983	.983	*.970*	*.920*	.951	1.014	1.046	1.019	1.024	1.193
BVAR(4), inflation detrending	1.020	.918	*.864*	1.002	1.003	.934	.977	.995	.918	.968	.957	1.051
avg. of VAR(4), univariate	1.022	.951	.887	1.014	1.024	1.013	*.960*	.977	.993	*.978*	*.955*	*.929*
avg. of DVAR(4), univariate	.990	*.915*	*.804*	1.014	1.021	1.010	.968	*.953*	*.930*	.980	.965	.996
avg. of infl. detr. VAR(4), univar.	1.002	.918	*.855*	.991	.983	*.934*	*.958*	*.961*	*.928*	*.970*	*.943*	*.889*
avg. of VAR(4), rolling VAR(4)	1.162	1.103	1.032	1.043	1.106	1.180	.994	1.045	1.093	1.022	.981	1.078
average of all forecasts	1.004	.922	*.824*	1.006	1.019	.997	**.933**	.967	*.923*	1.012	.987	1.019
median	1.017	.924	*.838*	1.011	1.026	.998	*.950*	.984	*.931*	1.013	.980	1.000
trimmed mean, 10%	1.009	.929	*.830*	1.007	1.020	.998	*.936*	.968	*.925*	1.012	.985	1.020
trimmed mean, 20%	1.012	.932	*.834*	1.007	1.020	.998	*.938*	.969	*.926*	1.011	.984	1.019
ridge: recursive, .001	1.209	1.437	1.798	1.130	1.416	1.330	1.262	1.662	1.256	1.158	1.227	2.069
ridge: recursive, .25	1.029	.920	.898	1.053	1.113	1.178	.975	1.053	1.096	1.018	.996	1.095
ridge: recursive, 1.	1.038	.950	**.716**	1.055	1.100	1.132	.969	1.033	1.067	1.014	.989	1.041
ridge: 10y rolling, .001	1.218	1.386	1.796	1.300	1.598	1.765	1.289	1.698	1.168	1.277	1.098	1.295
ridge: 10y rolling, .25	1.040	.942	.928	1.043	1.095	1.280	.976	1.049	1.093	1.011	.968	1.013
ridge: 10y rolling, 1.	1.046	.964	*.731*	1.038	1.071	1.156	.968	1.028	1.062	1.013	.975	.996
factor, recursive	1.064	1.078	.942	1.089	1.129	1.031	1.010	1.088	.979	1.032	1.089	1.584
factor, 10y rolling	1.082	1.096	.958	1.070	1.102	1.047	1.024	1.121	1.022	1.008	1.012	1.423
MSE weighting, recursive	1.003	.921	*.810*	1.006	1.017	.996	*.934*	.968	*.925*	1.011	.986	1.016
MSE weighting, 10y rolling	1.004	.922	*.812*	1.009	1.024	1.006	*.934*	.968	*.926*	1.010	.982	1.004
MSE weighting, 5y rolling	1.000	.922	*.832*	1.007	1.021	1.005	*.934*	.966	*.921*	1.011	.987	.998
MSE weighting, discounted	1.004	.924	*.819*	1.007	1.021	1.002	*.933*	.968	*.926*	1.010	.984	1.009
PLS, recursive	1.087	1.010	*.843*	1.045	1.115	1.053	.988	1.058	.933	.978	1.027	1.247
PLS, 10y rolling	1.001	.991	.876	1.050	1.039	1.099	1.066	1.047	.972	1.049	.969	1.080
PLS, 5y rolling	.962	.967	.861	1.051	1.069	1.054	1.078	1.079	1.241	1.043	1.000	1.104
best quartile, recursive	1.012	.917	*.781*	1.002	1.014	1.009	.950	.976	*.917*	1.012	1.005	1.111
best quartile, 10y rolling	1.011	.936	*.783*	1.025	1.041	1.013	.952	.976	*.910*	1.003	.970	.987
best quartile, 5y rolling	.994	.927	*.870*	1.023	1.066	1.071	*.949*	.959	*.891*	1.008	.973	.997
OLS comb. of quartiles, recursive	1.083	1.098	1.133	1.050	1.170	1.002	1.006	1.115	1.407	1.017	1.017	1.284
OLS comb. of quartiles, 10 year rolling	1.091	1.142	1.138	1.035	1.164	.964	1.016	1.134	1.392	.985	.978	1.245
OLS comb. of quartiles, 5 year rolling	1.118	1.117	1.216	1.201	1.455	1.599	1.107	1.140	1.750	1.067	1.087	1.089
AIC weighting	1.005	.922	*.823*	1.007	1.021	1.000	*.936*	.967	*.925*	1.012	.987	1.019
BIC weighting	1.004	.921	*.825*	1.007	1.021	1.000	*.936*	.968	*.924*	1.011	.987	1.017
BMA, $\phi = .2$	1.029	1.036	1.024	1.010	1.015	1.000	*.939*	.975	.943	.989	.974	1.067
BMA, $\phi = 2$	1.052	1.075	1.042	1.025	1.045	1.035	.974	.984	.934	1.027	1.008	1.093

Notes:
1. The variables in each multivariate model are the HPS output gap, GDP inflation, and the T-bill rate.
2. See the notes to Table 3.

Table 5: Real-time RMSE results for the HP output gap and GDP inflation

(RMSEs in first row, RMSE ratios in all others)

	HP output gap forecasts						GDP inflation forecasts					
	1970-84			1985-2005			1970-84			1985-2005		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univar	.868	1.254	1.885	.375	.490	.813	1.911	2.242	2.466	.989	1.052	.743
VAR(4)	1.362	1.387	1.094	1.129	.985	.966	.948	.941	*.898*	1.010	.975	1.070
DVAR(4)	1.215	1.145	*.820*	1.121	.971	.981	1.078	1.018	*.837*	1.013	1.009	1.195
BVAR(4)	1.175	1.167	1.102	*.953*	*.876*	*.851*	.940	.984	.920	1.018	1.035	1.294
BVAR(4) with TVP	1.113	1.128	1.068	.983	*.941*	*.934*	.941	.986	.956	1.005	1.000	1.159
BVAR(4), inflation detrending	1.052	1.028	*.911*	*.907*	**.807**	**.692**	.980	.993	.883	1.046	1.108	1.607
avg. of VAR(4), univariate	1.151	1.137	.962	1.027	*.940*	*.933*	*.940*	*.942*	*.904*	.983	*.960*	.960
avg. of DVAR(4), univariate	1.079	1.041	*.895*	1.031	*.961*	.979	.993	.967	*.873*	.985	.975	1.031
avg. of infl. detr. VAR(4), univar.	1.128	1.111	*.925*	1.012	*.869*	*.747*	*.947*	*.941*	*.868*	.989	.977	1.047
avg. of VAR(4), rolling VAR(4)	1.469	1.532	1.267	1.087	1.015	.974	.940	.946	*.871*	1.047	1.051	1.357
average of all forecasts	1.203	1.192	.993	1.026	*.918*	*.832*	*.928*	.933	*.797*	1.031	1.028	1.226
median	1.200	1.186	1.000	1.029	*.926*	*.851*	.943	.938	*.821*	1.020	1.018	1.199
trimmed mean, 10%	1.199	1.188	.991	1.022	*.914*	*.837*	*.931*	.933	*.802*	1.030	1.026	1.227
trimmed mean, 20%	1.197	1.187	.990	1.023	*.915*	*.841*	*.932*	.933	*.803*	1.029	1.024	1.224
ridge: recursive, .001	1.106	1.410	1.944	.949	1.180	1.350	1.323	1.444	1.616	1.104	1.039	2.081
ridge: recursive, .25	**.907**	.909	.941	*.877*	1.014	1.100	.973	.998	.929	1.031	1.017	1.238
ridge: recursive, 1.	*.924*	.965	*.807*	*.886*	.974	1.050	.967	.990	.915	1.030	1.018	1.207
ridge: 10y rolling, .001	1.141	1.472	2.195	1.151	1.327	1.626	1.362	1.533	1.612	1.267	1.073	1.858
ridge: 10y rolling, .25	*.912*	.917	.924	*.867*	*.928*	1.039	.970	.990	.918	1.027	1.001	1.197
ridge: 10y rolling, 1.	.928	.971	**.801**	*.887*	*.916*	.972	.962	.981	*.905*	1.024	1.001	1.166
factor, recursive	.939	1.143	1.209	.921	.996	.980	1.021	1.056	.935	1.049	1.093	1.544
factor, 10y rolling	.942	1.149	1.195	.972	1.079	1.213	1.025	1.080	.973	1.045	1.063	1.506
MSE weighting, recursive	1.176	1.165	.977	1.013	*.903*	*.831*	*.926*	.933	*.798*	1.029	1.027	1.222
MSE weighting, 10y rolling	1.175	1.164	.974	1.018	*.904*	*.815*	*.927*	.933	*.798*	1.029	1.027	1.230
MSE weighting, 5y rolling	1.173	1.161	.974	1.019	*.906*	*.819*	*.927*	.933	*.799*	1.029	1.030	1.241
MSE weighting, discounted	1.175	1.167	.972	1.015	*.904*	*.822*	*.926*	.933	*.800*	1.029	1.025	1.224
PLS, recursive	1.016	1.098	1.029	1.017	.979	.981	.972	.976	*.813*	.984	1.154	1.478
PLS, 10y rolling	1.051	1.130	1.029	.983	.946	.878	.962	.971	*.813*	1.024	1.094	1.358
PLS, 5y rolling	1.066	1.111	1.121	1.016	*.892*	*.866*	.992	1.045	.977	1.116	1.073	1.379
best quartile, recursive	1.089	1.105	.964	.982	*.884*	*.891*	*.929*	*.932*	**.794**	1.024	1.028	1.230
best quartile, 10y rolling	1.083	1.096	.960	.990	*.911*	*.857*	.938	*.937*	*.795*	1.024	1.031	1.213
best quartile, 5y rolling	1.069	1.107	.980	1.013	*.893*	*.830*	*.940*	.945	*.822*	1.026	1.015	1.243
OLS comb. of quartiles, recursive	.954	1.115	.975	**.866**	*.930*	.885	.979	1.093	.962	1.041	1.056	1.230
OLS comb. of quartiles, 10 year rolling	.957	1.122	1.016	*.875*	1.032	1.008	.986	1.111	.984	1.049	1.062	1.377
OLS comb. of quartiles, 5 year rolling	.996	.962	1.015	1.049	1.303	1.206	1.027	1.119	1.047	1.093	1.210	1.169
AIC weighting	1.205	1.193	.993	1.026	*.918*	*.832*	*.928*	.932	*.799*	1.030	1.030	1.231
BIC weighting	1.200	1.187	.991	1.026	*.916*	*.831*	*.929*	.933	*.800*	1.030	1.028	1.226
BMA, $\phi = .2$	1.255	1.306	1.023	1.070	1.055	.970	.957	.969	*.856*	1.052	1.108	1.609
BMA, $\phi = 2$	1.271	1.322	1.047	1.061	1.028	.925	.978	.987	*.859*	1.063	1.111	1.635

Notes:
1. The variables in each multivariate model are the HP output gap, GDP inflation, and the T-bill rate.
2. See the notes to Table 3.

Table 6: Real-time RMSE results for GDP growth and CPI inflation

(RMSEs in first row, RMSE ratios in all others)

	GDP growth forecasts						CPI inflation forecasts					
	1970-84			1985-2005			1970-84			1985-2005		
forecast method	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
univar	4.550	5.023	3.633	1.755	1.826	1.367	2.117	2.733	2.970	1.340	1.460	1.254
VAR(4)	1.072	.976	.934	1.110	1.119	1.064	*.857*	.949	.993	.986	1.037	1.077
DVAR(4)	1.079	.957	*.768*	1.211	1.232	1.099	*.847*	.888	**.854**	.963	1.012	1.095
BVAR(4)	.955	*.914*	.937	1.027	1.027	.965	*.925*	1.033	1.106	.997	.987	1.001
BVAR(4) with TVP	.953	*.925*	.943	.993	.991	*.935*	*.914*	1.015	1.086	.986	.966	*.936*
BVAR(4), inflation detrending	**.869**	*.810*	*.777*	1.152	1.151	1.087	.944	1.013	1.033	.975	.986	1.061
avg. of VAR(4), univariate	.993	.934	*.875*	1.022	1.016	.964	*.863*	*.912*	*.916*	*.965*	.993	.996
avg. of DVAR(4), univariate	.980	*.915*	*.806*	1.068	1.071	1.003	*.862*	*.898*	*.894*	*.951*	.983	1.013
avg. of infl. detr. VAR(4), univar.	.960	*.895*	*.818*	1.038	1.038	.983	*.857*	*.895*	*.863*	.968	.999	1.019
avg. of VAR(4), rolling VAR(4)	1.106	1.025	.972	1.162	1.174	1.137	*.859*	.979	1.035	1.021	1.083	1.143
average of all forecasts	.937	*.886*	*.816*	1.087	1.096	1.024	*.827*	.926	.943	.983	1.013	1.071
median	.937	*.882*	*.859*	1.069	1.084	1.029	*.863*	.916	*.937*	.978	1.004	1.041
trimmed mean, 10%	.938	*.888*	*.826*	1.085	1.093	1.021	*.834*	.924	.939	.980	1.011	1.060
trimmed mean, 20%	.938	*.889*	*.830*	1.084	1.091	1.021	*.840*	.923	.937	.979	1.008	1.054
ridge: recursive, .001	1.993	1.426	1.145	1.170	1.475	1.020	1.077	1.361	2.155	1.112	1.364	1.496
ridge: recursive, .25	.940	*.830*	*.709*	1.053	1.070	.998	*.847*	.971	1.077	.985	.996	1.070
ridge: recursive, 1.	*.925*	*.840*	*.741*	1.073	1.091	.999	*.847*	.966	1.071	.977	.991	1.013
ridge: 10y rolling, .001	2.026	1.358	1.096	1.292	1.643	1.469	1.044	1.281	2.085	1.195	1.619	1.597
ridge: 10y rolling, .25	.963	*.843*	*.718*	1.025	1.074	1.017	*.842*	.965	1.062	.989	.992	.997
ridge: 10y rolling, 1.	.932	*.847*	*.745*	1.057	1.090	1.003	*.841*	.959	1.059	.981	.989	.955
factor, recursive	.986	.940	.865	1.100	1.116	1.049	*.834*	.945	.952	.992	1.064	1.256
factor, 10y rolling	.994	.948	.878	1.112	1.115	1.106	*.852*	.977	.990	.956	1.005	1.137
MSE weighting, recursive	.935	*.881*	*.797*	1.084	1.099	1.034	*.831*	.928	.943	.982	1.008	1.062
MSE weighting, 10y rolling	.936	*.881*	*.799*	1.081	1.095	1.024	*.830*	.928	.942	.981	1.008	1.062
MSE weighting, 5y rolling	.935	*.881*	*.812*	1.074	1.091	1.014	*.828*	.926	.941	.980	1.006	1.058
MSE weighting, discounted	.936	*.883*	*.803*	1.079	1.094	1.029	*.829*	.928	.945	.982	1.008	1.060
PLS, recursive	.951	**.780**	*.724*	1.191	1.400	1.421	.893	1.081	*.910*	1.005	1.016	1.262
PLS, 10y rolling	.950	*.780*	*.787*	1.140	1.251	1.151	*.869*	1.083	*.906*	1.005	1.043	1.201
PLS, 5y rolling	*.901*	.894	.888	1.128	1.194	1.110	.933	1.106	*.882*	1.020	1.049	1.088
best quartile, recursive	*.913*	*.854*	*.711*	1.071	1.136	1.132	*.872*	.930	*.907*	.979	1.014	1.085
best quartile, 10y rolling	*.919*	*.853*	*.712*	1.054	1.107	1.107	*.864*	.935	*.911*	.973	1.004	1.087
best quartile, 5y rolling	.944	*.872*	*.784*	1.065	1.111	1.053	*.838*	.937	.954	.979	1.018	1.037
OLS comb. of quartiles, recursive	.923	*.842*	*.698*	1.229	1.294	**.881**	*.894*	.959	.942	.984	1.033	1.161
OLS comb. of quartiles, 10 year rolling	.918	*.853*	**.684**	1.057	1.106	1.129	*.880*	.999	.973	.963	1.036	1.231
OLS comb. of quartiles, 5 year rolling	1.044	.945	*.806*	1.256	1.350	1.701	**.785**	1.072	2.182	.997	1.031	1.192
AIC weighting	.936	*.885*	*.814*	1.089	1.098	1.027	*.833*	.927	.940	.982	1.012	1.072
BIC weighting	.934	*.883*	*.815*	1.088	1.097	1.025	*.835*	.927	.941	.982	1.012	1.070
BMA, $\phi = .2$	1.024	1.090	1.044	1.061	1.077	.984	*.865*	.958	1.008	.979	1.009	1.038
BMA, $\phi = 2$	.982	.984	.961	1.182	1.202	1.166	*.869*	.961	.964	.983	1.037	1.106

Notes:
1. The variables in each multivariate model are GDP growth, CPI inflation, and the T-bill rate.
2. See the notes to Table 3.

Table 7: Real-time RMSE results for the HPS output gap and CPI inflation

(RMSEs in first row, RMSE ratios in all others)

	HPS output gap forecasts						CPI inflation forecasts					
	1970-84			1985-2005			1970-84			1985-2005		
forecast method	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
univar	1.174	1.994	3.814	.824	1.166	2.148	2.117	2.733	2.970	1.340	1.460	1.254
VAR(4)	1.135	1.062	.993	1.062	1.105	1.143	.909	1.016	1.111	.975	1.016	1.015
DVAR(4)	1.140	1.035	**.799**	1.051	1.079	1.062	.902	.959	.946	.975	1.010	1.055
BVAR(4)	1.079	1.008	.965	1.048	1.093	1.108	*.927*	1.021	1.083	.989	.973	.946
BVAR(4) with TVP	1.058	1.010	.984	*.981*	**.968**	**.917**	*.916*	1.011	1.103	1.003	1.014	1.068
BVAR(4), inflation detrending	1.007	*.893*	*.826*	1.013	1.032	.997	*.917*	.937	.879	*.945*	**.923**	**.838**
avg. of VAR(4), univariate	1.033	.968	*.887*	1.015	1.021	1.005	*.882*	.943	.957	*.962*	.987	.980
avg. of DVAR(4), univariate	1.023	.956	*.829*	1.016	1.024	1.015	*.886*	*.932*	*.931*	*.951*	.973	.976
avg. of infl. detr. VAR(4), univar.	1.010	.922	*.832*	.992	.989	.970	*.861*	**.894**	**.828**	*.946*	*.955*	*.877*
avg. of VAR(4), rolling VAR(4)	1.166	1.129	1.071	1.049	1.107	1.157	.900	1.054	1.168	1.024	1.070	1.052
average of all forecasts	1.014	.952	*.860*	1.009	1.024	1.000	*.837*	.929	.951	.979	.994	.979
median	1.032	.965	*.878*	1.006	1.022	.997	*.841*	.927	.950	.972	.995	.980
trimmed mean, 10%	1.016	.956	*.867*	1.008	1.023	1.000	*.843*	.928	.947	.976	.993	.970
trimmed mean, 20%	1.017	.958	*.871*	1.007	1.022	.999	*.845*	.928	.949	.975	.991	.969
ridge: recursive, .001	1.297	1.291	1.199	1.083	1.358	1.789	1.105	1.467	1.404	1.096	1.193	1.204
ridge: recursive, .25	1.051	.956	*.876*	1.052	1.105	1.168	*.872*	.992	1.058	.994	.989	.990
ridge: recursive, 1.	1.056	.989	*.804*	1.054	1.100	1.116	*.866*	.979	1.067	.983	.989	.966
ridge: 10y rolling, .001	1.330	1.334	1.129	1.227	1.463	1.667	1.154	1.487	1.215	1.128	1.220	1.129
ridge: 10y rolling, .25	1.064	.987	.932	1.037	1.081	1.221	*.869*	.990	1.052	.991	.975	*.865*
ridge: 10y rolling, 1.	1.068	1.007	*.832*	1.034	1.062	1.118	*.863*	.976	1.063	.983	.985	*.881*
factor, recursive	1.096	1.139	1.122	1.074	1.110	1.005	**.829**	.926	*.918*	1.002	1.088	1.302
factor, 10y rolling	1.119	1.160	1.127	1.066	1.094	1.032	*.841*	.945	.932	.963	1.022	1.178
MSE weighting, recursive	1.014	.950	*.847*	1.008	1.023	1.001	*.839*	.930	.948	.977	.988	.959
MSE weighting, 10y rolling	1.014	.951	*.848*	1.011	1.029	1.008	*.839*	.931	.948	.976	.986	.948
MSE weighting, 5y rolling	1.012	.952	*.862*	1.009	1.027	1.010	*.839*	.928	.943	.976	.989	.948
MSE weighting, discounted	1.014	.952	*.853*	1.009	1.026	1.007	*.838*	.931	.952	.977	.988	.951
PLS, recursive	1.046	.953	.858	1.011	1.052	1.076	*.901*	1.015	1.040	.993	1.010	*.850*
PLS, 10y rolling	.984	.946	.875	1.028	1.042	1.068	.931	1.003	1.084	1.023	1.010	*.858*
PLS, 5y rolling	1.031	1.048	.882	1.029	1.078	1.088	.965	1.079	1.018	1.039	1.087	*.893*
best quartile, recursive	1.011	.939	*.803*	1.008	1.025	1.050	*.867*	.934	.940	.976	.982	*.919*
best quartile, 10y rolling	1.011	.940	*.809*	1.030	1.054	1.021	*.869*	.934	.948	.974	.974	*.892*
best quartile, 5y rolling	.993	.933	*.828*	1.017	1.063	1.065	*.868*	.933	.928	.989	.979	*.900*
OLS comb. of quartiles, recursive	1.076	1.117	1.187	1.047	1.153	1.061	*.880*	.922	1.026	.967	1.021	1.069
OLS comb. of quartiles, 10 year rolling	1.108	1.141	1.198	1.239	1.260	1.081	*.860*	.908	.973	.967	1.012	1.198
OLS comb. of quartiles, 5 year rolling	1.104	1.212	1.221	1.256	1.386	1.359	1.064	1.085	1.115	.961	.998	1.308
AIC weighting	1.014	.951	*.858*	1.010	1.026	1.003	*.844*	.931	.949	.976	.991	.970
BIC weighting	1.013	.950	*.857*	1.010	1.026	1.003	*.844*	.930	.948	.976	.991	.970
BMA, $\phi = .2$	1.053	1.082	1.015	1.014	1.025	.996	*.873*	.955	1.006	.995	1.035	1.124
BMA, $\phi = 2$	1.098	1.120	1.065	1.030	1.052	1.028	*.856*	.958	.947	.994	1.077	1.116

Notes:
1. The variables in each multivariate model are the HPS output gap, CPI inflation, and the T-bill rate.
2. See the notes to Table 3.

Table 8: Real-time RMSE results for the HP output gap and CPI inflation
(RMSEs in first row, RMSE ratios in all others)

	HP output gap forecasts						CPI inflation forecasts					
	1970-84			1985-2005			1970-84			1985-2005		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univar	.868	1.254	1.885	.375	.490	.813	2.117	2.733	2.970	1.340	1.460	1.254
VAR(4)	1.433	1.498	1.168	1.112	.980	.951	*.849*	*.899*	.921	.993	1.067	1.184
DVAR(4)	1.346	1.343	*.834*	1.115	1.008	.964	*.856*	*.861*	*.739*	.979	1.080	1.288
BVAR(4)	1.173	1.162	1.085	*.954*	*.891*	*.846*	*.880*	.933	*.891*	1.008	1.041	1.182
BVAR(4) with TVP	1.118	1.133	1.067	.983	*.957*	*.917*	*.876*	.934	.932	.994	1.012	1.116
BVAR(4), inflation detrending	1.028	.976	*.852*	.959	*.909*	**.747**	*.914*	.917	*.796*	1.038	1.143	1.471
avg. of VAR(4), univariate	1.188	1.187	.983	1.023	*.945*	*.932*	*.862*	*.883*	*.864*	.964	.995	1.022
avg. of DVAR(4), univariate	1.141	1.128	*.902*	1.028	.975	.970	*.861*	*.870*	*.816*	*.951*	.996	1.076
avg. of infl. detr. VAR(4), univar.	1.162	1.134	*.880*	1.020	**.891**	*.766*	*.852*	*.850*	*.761*	.977	1.032	1.147
avg. of VAR(4), rolling VAR(4)	1.537	1.639	1.359	1.117	1.068	.962	*.845*	*.917*	.932	1.045	1.144	1.316
average of all forecasts	1.245	1.235	.998	1.049	.973	*.847*	**.797**	*.839*	*.738*	.996	1.054	1.224
median	1.242	1.224	1.000	1.049	.963	*.861*	*.802*	*.841*	*.757*	.980	1.043	1.180
trimmed mean, 10%	1.242	1.231	.997	1.043	.966	*.849*	*.803*	*.842*	*.736*	.994	1.051	1.218
trimmed mean, 20%	1.242	1.231	.994	1.043	.965	*.850*	*.804*	*.843*	*.738*	.992	1.047	1.212
ridge: recursive, .001	1.153	1.545	1.910	.912	1.168	1.487	1.057	1.501	1.679	1.163	1.408	1.500
ridge: recursive, .25	.915	.880	*.898*	*.877*	1.031	1.047	*.838*	.906	.917	1.001	1.040	1.283
ridge: recursive, 1.	.929	.943	**.793**	*.892*	.997	.986	*.830*	*.886*	*.848*	.994	1.040	1.239
ridge: 10y rolling, .001	1.219	1.630	2.024	1.063	1.640	1.763	1.093	1.636	1.773	1.278	1.506	1.401
ridge: 10y rolling, .25	.920	.885	.902	*.875*	.959	1.071	*.833*	*.901*	.902	.996	1.028	1.204
ridge: 10y rolling, 1.	.934	.948	*.795*	*.894*	.952	.965	*.823*	*.879*	*.837*	.985	1.018	1.154
factor, recursive	.950	1.154	1.252	.926	1.016	.979	*.839*	*.893*	*.839*	1.015	1.116	1.385
factor, 10y rolling	.957	1.167	1.250	.966	1.064	1.170	*.847*	.912	*.867*	.978	1.066	1.256
MSE weighting, recursive	1.211	1.199	.987	1.032	.948	*.834*	*.800*	*.839*	*.728*	.993	1.050	1.217
MSE weighting, 10y rolling	1.211	1.198	.985	1.037	.950	*.822*	*.801*	*.839*	*.726*	.993	1.052	1.223
MSE weighting, 5y rolling	1.215	1.203	.986	1.035	.955	*.841*	*.801*	*.839*	*.726*	.992	1.052	1.232
MSE weighting, discounted	1.212	1.204	.981	1.034	.949	*.831*	*.801*	*.841*	*.729*	.994	1.052	1.226
PLS, recursive	1.102	1.113	1.163	1.000	.941	.933	*.858*	**.825**	**.708**	.989	1.043	1.206
PLS, 10y rolling	1.118	1.111	1.157	1.034	.928	*.819*	.924	*.835*	*.708*	1.030	1.068	1.330
PLS, 5y rolling	1.159	1.262	1.125	1.002	1.016	.994	.990	.952	*.816*	1.019	1.181	1.574
best quartile, recursive	1.120	1.102	.944	.976	*.895*	*.803*	*.837*	*.836*	*.726*	.971	1.037	1.206
best quartile, 10y rolling	1.120	1.100	.946	.996	*.924*	*.794*	*.836*	*.847*	*.718*	.991	1.059	1.246
best quartile, 5y rolling	1.120	1.112	.999	1.003	.949	*.871*	*.827*	*.849*	*.725*	.992	1.048	1.296
OLS comb. of quartiles, recursive	.959	.881	.942	**.841**	*.908*	*.816*	*.896*	*.892*	*.857*	*.944*	1.051	1.281
OLS comb. of quartiles, 10 year rolling	.964	.883	.977	.937	.941	1.013	*.897*	*.906*	*.868*	1.030	1.134	1.311
OLS comb. of quartiles, 5 year rolling	1.017	.988	1.393	1.065	1.540	1.523	.970	.961	*.864*	1.079	1.197	1.794
AIC weighting	1.249	1.238	.999	1.050	.973	*.846*	*.800*	*.840*	*.733*	.995	1.056	1.231
BIC weighting	1.244	1.231	.995	1.048	.970	*.845*	*.801*	*.840*	*.733*	.994	1.054	1.227
BMA, $\phi = .2$	1.260	1.309	1.007	1.114	1.078	.957	*.833*	*.862*	*.775*	.992	1.022	1.181
BMA, $\phi = 2$	1.280	1.327	1.028	1.097	1.042	*.909*	*.852*	*.865*	*.749*	.986	1.095	1.387

Notes:
1. The variables in each multivariate model are the HP output gap, CPI inflation, and the T-bill rate.
2. See the notes to Table 3.

Table 9: Real-time RMSE results for the T-bill rate
(RMSEs in first row, RMSE ratios in all others)

	Models with GDP growth and GDP inflation						Models with HPS gap and GDP inflation					
	1970-84			1985-2005			1970-84			1985-2005		
forecast method	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
univar	1.305	2.098	2.821	.378	.778	1.625	1.305	2.098	2.821	.378	.778	1.625
VAR(4)	.940	.954	1.108	1.084	1.027	**.892**	.948	.982	1.134	1.087	1.021	*.895*
DVAR(4)	.933	.917	.981	1.137	1.131	1.083	.934	.917	.986	1.130	1.122	1.081
BVAR(4)	.949	*.926*	1.027	1.067	.986	*.907*	.961	.939	1.066	1.072	.987	*.915*
BVAR(4) with TVP	.949	*.933*	1.054	1.078	1.006	*.944*	.954	.941	1.110	1.108	1.045	1.029
BVAR(4), inflation detrending	.930	*.860*	**.908**	1.150	1.022	*.892*	.973	.909	.970	1.142	1.005	**.873**
avg. of VAR(4), univariate	*.934*	*.930*	1.030	.988	.966	*.925*	*.938*	*.945*	1.035	.987	.960	*.925*
avg. of DVAR(4), univariate	*.931*	*.909*	*.966*	1.027	1.036	1.032	*.932*	*.912*	*.974*	1.025	1.030	1.029
avg. of infl. detr. VAR(4), univar.	*.924*	*.910*	.964	.982	*.957*	*.908*	*.938*	.941	.984	.980	*.951*	*.904*
avg. of VAR(4), rolling VAR(4)	.967	.995	1.198	1.182	1.098	*.934*	.954	1.011	1.220	1.226	1.120	.996
average of all forecasts	.910	*.924*	1.007	1.035	.992	*.952*	.920	.949	1.050	1.034	.982	*.960*
median	.929	.939	.989	1.037	.999	*.957*	.966	.948	1.048	1.044	.997	.974
trimmed mean, 10%	*.910*	*.921*	.998	1.032	.991	*.956*	.925	.947	1.043	1.031	.981	*.961*
trimmed mean, 20%	.914	*.921*	.997	1.032	.991	*.957*	.925	.945	1.043	1.030	.982	*.963*
ridge: recursive, .001	.979	**.807**	1.884	1.101	1.208	1.136	.977	1.025	1.227	1.113	1.143	1.059
ridge: recursive, .25	.954	.962	1.122	1.032	.989	*.950*	.961	.984	1.109	1.051	1.016	1.010
ridge: recursive, 1.	.941	.960	1.124	1.037	.995	*.943*	.949	.981	1.129	1.054	1.018	1.007
ridge: 10y rolling, .001	1.127	.950	1.797	1.029	1.111	1.091	1.086	1.024	1.470	1.143	1.070	1.165
ridge: 10y rolling, .25	.969	.986	1.166	1.029	.981	*.915*	.975	1.009	1.150	1.054	1.009	.989
ridge: 10y rolling, 1.	.951	.976	1.155	1.034	.982	*.899*	.960	.998	1.165	1.058	1.009	*.964*
factor, recursive	.960	.957	1.189	1.133	1.124	1.105	.966	.966	1.174	1.172	1.176	1.179
factor, 10y rolling	.965	.958	1.261	1.211	1.208	1.314	.968	.962	1.210	1.266	1.275	1.347
MSE weighting, recursive	.916	*.923*	.996	1.037	.994	*.952*	.927	.948	1.038	1.037	.985	*.958*
MSE weighting, 10y rolling	.917	*.924*	.997	1.032	.993	*.953*	.928	.949	1.041	1.032	.983	*.958*
MSE weighting, 5y rolling	.920	.928	1.001	1.029	.991	*.957*	.931	.955	1.052	1.028	.983	.972
MSE weighting, discounted	.920	.926	1.002	1.034	.993	*.954*	.930	.952	1.049	1.035	.985	*.963*
PLS, recursive	.975	.997	1.033	1.104	.984	*.911*	.971	1.001	1.133	1.115	1.156	.984
PLS, 10y rolling	1.022	.967	1.075	1.125	.976	*.925*	.959	1.004	1.174	1.051	1.067	.996
PLS, 5y rolling	1.006	1.070	1.327	1.280	1.106	1.058	.976	1.187	1.301	1.230	1.176	1.196
best quartile, recursive	.960	.940	.977	1.054	1.010	.975	.975	.950	1.037	1.071	1.012	.977
best quartile, 10y rolling	.959	.948	.994	1.023	.984	*.959*	.969	.968	1.070	1.043	.992	*.956*
best quartile, 5y rolling	.964	.975	1.006	1.061	1.018	.997	.980	.991	1.083	1.086	1.026	1.014
OLS comb. of quartiles, recursive	1.033	1.029	1.298	1.160	1.159	1.130	1.020	1.036	1.183	1.157	1.205	1.110
OLS comb. of quartiles, 10 year rolling	1.070	1.092	1.426	1.219	1.198	1.455	1.016	1.112	1.203	1.314	1.222	1.373
OLS comb. of quartiles, 5 year rolling	1.097	1.094	1.311	1.386	1.569	1.862	1.187	1.304	1.194	1.464	1.485	1.910
AIC weighting	.914	.927	1.000	1.043	.999	*.953*	.925	.951	1.040	1.043	.990	*.959*
BIC weighting	.914	.926	.997	1.040	.997	*.953*	.925	.951	1.038	1.040	.988	*.958*
BMA, $\phi = .2$	.998	.993	1.002	1.157	1.072	1.020	1.009	1.019	1.046	1.150	1.078	1.039
BMA, $\phi = 2$	.968	.987	1.019	1.365	1.280	1.113	.972	.999	1.048	1.480	1.405	1.237

Notes:
1. In the results in the left half of the table, the variables are GDP growth, GDP inflation, and the T-bill rate. In the results in the right half of the table, the variables are the HPS output gap, GDP inflation, and the T-bill rate.
2. See the notes to Table 3.

Table 10: Average RMSE rankings in real-time forecasts

	all	all y, π	all y	all π	all i	all y, π 70-84	all y, π 85-05
avg. of infl. detr. VAR(4), univar.	7.1	9.0	9.9	8.1	3.4	13.5	4.5
avg. of VAR(4), univariate	12.5	14.7	18.1	11.2	8.1	20.5	8.8
MSE weighting, recursive	12.5	13.2	13.6	12.8	11.0	11.1	15.3
MSE weighting, 10y rolling	13.0	13.7	14.6	12.7	11.7	11.5	15.9
avg. of DVAR(4), univariate	13.1	13.0	16.2	9.9	13.3	14.6	11.5
MSE weighting, 5y rolling	13.4	13.1	14.1	12.1	13.8	10.5	15.8
MSE weighting, discounted	13.7	13.8	14.7	13.0	13.4	12.1	15.5
trimmed mean, 20%	14.1	15.4	17.2	13.5	11.5	16.1	14.7
trimmed mean, 10%	14.1	15.7	17.2	14.2	11.0	15.0	16.5
best quartile, recursive	14.2	12.8	11.3	14.3	16.9	9.6	16.0
best quartile, 10y rolling	14.3	12.9	12.9	13.0	17.1	10.8	15.1
average of all forecasts	14.5	15.9	17.2	14.7	11.6	13.2	18.6
BVAR(4), inflation detrending	14.7	15.5	9.1	21.9	13.2	16.3	14.7
BIC weighting	15.0	16.0	16.7	15.2	13.0	13.4	18.5
median	16.0	15.9	18.4	13.5	16.2	17.5	14.3
AIC weighting	16.1	16.7	18.1	15.2	14.9	13.8	19.5
ridge: 10y rolling, 1.	16.6	15.5	14.4	16.6	18.9	15.8	15.2
best quartile, 5y rolling	17.1	14.4	15.5	13.3	22.5	11.5	17.3
BVAR(4) with TVP	18.0	19.1	16.0	22.2	15.7	26.7	11.5
ridge: recursive, 1.	18.5	17.8	15.6	20.1	19.8	15.9	19.7
BVAR(4)	18.6	21.9	21.0	22.9	12.0	27.7	16.2
univariate	18.7	19.7	17.3	22.1	16.7	26.0	13.4
ridge: 10y rolling, .25	18.9	18.5	17.4	19.6	19.7	19.7	17.3
ridge: recursive, .25	20.2	20.8	17.4	24.3	19.0	18.9	22.7
PLS, recursive	21.8	21.5	21.7	21.4	22.4	19.6	23.5
DVAR(4)	22.4	21.4	26.2	16.5	24.4	19.2	23.6
BMA, $\phi = .2$	22.4	22.3	23.8	20.8	22.6	25.4	19.3
PLS, 10y rolling	22.8	21.9	18.5	25.3	24.6	18.9	25.0
VAR(4)	24.3	25.5	31.6	19.3	22.0	29.4	21.5
OLS comb. of quartiles, recursive	26.0	22.8	18.8	26.8	32.5	22.6	23.0
factor, recursive	27.6	26.9	25.8	28.0	28.9	23.7	30.1
OLS comb. of quartiles, 10 year rolling	27.8	24.1	21.9	26.4	35.2	24.6	23.7
factor, 10y rolling	28.2	26.5	27.8	25.2	31.5	26.8	26.2
BMA, $\phi = 2$	28.9	27.7	29.5	26.0	31.2	26.3	29.2
PLS, 5y rolling	29.0	27.4	23.0	31.8	32.3	25.9	28.8
avg. of VAR(4), rolling VAR(4)	30.6	31.0	33.8	28.2	29.9	31.1	30.9
ridge: recursive, .001	34.0	36.0	34.1	37.8	30.1	37.1	34.8
OLS comb. of quartiles, 5 year rolling	34.0	32.1	32.1	32.2	37.8	30.2	34.0
ridge: 10y rolling, .001	35.4	37.8	37.7	38.0	30.5	37.6	38.1
# of ratio observations	108	72	36	36	36	36	36

Notes:
1. The table reports average RMSE rankings of the full set of forecast methods or models included in Tables 3-9. The average rankings in the first column of figures are calculated, for each forecast method, across a total of 108 forecasts of output (3 measures: GDP growth, HPS gap, HP gap), inflation (2 measures: GDP inflation, CPI inflation), and interest rates (1 measure: T-bill rate) at horizons (3) of $h = 0Q$, $h = 1Q$, and $h = 1Y$ and sample periods (2) of 1970-84 and 1985-05. The average rankings in remaining columns are based on forecasts with models that include particular variables or forecasts of a particular variable, etc. For example, the average rankings in the second column are based on 72 forecasts of just output and inflation, with forecasts of interest rates omitted from the average ranking calculation.
2. See the notes to Table 3.

Table 11: Accuracy of select VAR and average forecasts compared to SPF forecasts, in real time data

(RMSEs in first row, RMSE ratios in all others)

	GDP growth forecasts					
	1970-84			1985-2005		
	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
SPF	2.945	3.943	2.903	1.664	1.798	1.331
univariate	1.545	1.274	1.251	1.055	1.015	1.027
BVAR(4) with TVP	1.478	1.185	1.214	1.049	1.002	.944
BVAR(4), inflation detrending	1.364	1.058	1.015	1.126	1.077	.962
avg. of VAR(4), univariate	1.523	1.174	1.118	1.080	1.037	.981
avg. of DVAR(4), univariate	1.474	1.133	.995	1.129	1.096	1.022
avg. of infl. detr. VAR(4), univar.	1.479	1.139	1.071	1.073	1.031	.942
average of all forecasts	1.440	1.096	1.016	1.124	1.096	1.017
MSE weighting, 10y rolling	1.437	1.092	.991	1.120	1.095	1.011
best quartile, 10y rolling	1.428	1.089	.907	1.114	1.118	1.079
	GDP inflation forecasts					
	1970-84			**1985-2005**		
	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
SPF	1.524	2.087	2.265	.826	.911	.780
univariate	1.254	1.074	1.089	1.198	1.155	.953
BVAR(4) with TVP	1.208	1.131	1.205	1.201	1.129	.955
BVAR(4), inflation detrending	1.256	1.141	1.168	1.210	1.186	1.226
avg. of VAR(4), univariate	1.199	1.049	1.088	1.176	1.106	.887
avg. of DVAR(4), univariate	1.210	1.022	1.018	1.172	1.110	.923
avg. of infl. detr. VAR(4), univar.	1.205	1.044	1.062	1.181	1.117	.924
average of all forecasts	1.171	1.065	1.064	1.224	1.152	.995
MSE weighting, 10y rolling	1.173	1.065	1.070	1.223	1.148	.987
best quartile, 10y rolling	1.189	1.064	1.100	1.212	1.129	.985
	CPI inflation forecasts					
	1970-84			**1985-2005**		
	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
SPF	NA	NA	NA	.793	1.252	.962
univariate	NA	NA	NA	1.690	1.167	1.304
BVAR(4) with TVP	NA	NA	NA	1.667	1.127	1.220
BVAR(4), inflation detrending	NA	NA	NA	1.648	1.150	1.383
avg. of VAR(4), univariate	NA	NA	NA	1.631	1.158	1.298
avg. of DVAR(4), univariate	NA	NA	NA	1.608	1.147	1.320
avg. of infl. detr. VAR(4), univar.	NA	NA	NA	1.636	1.165	1.329
average of all forecasts	NA	NA	NA	1.661	1.182	1.397
MSE weighting, 10y rolling	NA	NA	NA	1.658	1.176	1.385
best quartile, 10y rolling	NA	NA	NA	1.643	1.171	1.417
	T-bill rate forecasts					
	1970-84			**1985-2005**		
	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
SPF	.310	1.436	2.589	.105	.462	1.552
univariate	4.207	1.461	1.090	3.612	1.683	1.047
BVAR(4) with TVP	3.992	1.364	1.148	3.893	1.694	.989
BVAR(4), inflation detrending	3.913	1.257	.989	4.154	1.720	.934
avg. of VAR(4), univariate	3.931	1.359	1.122	3.568	1.626	.968
avg. of DVAR(4), univariate	3.916	1.329	1.052	3.710	1.743	1.080
avg. of infl. detr. VAR(4), univar.	3.886	1.330	1.050	3.547	1.610	.951
average of all forecasts	3.827	1.350	1.097	3.737	1.670	.997
MSE weighting, 10y rolling	3.859	1.351	1.087	3.726	1.671	.998
best quartile, 10y rolling	4.034	1.385	1.083	3.694	1.656	1.003

Notes:
1. The forecast errors are calculated using the second–available (real–time) estimates of output and inflation as the actual data on output and inflation.
2. All of the GDP growth, GDP inflation, and T-bill results correspond to those reported in Table 3, based on models in GDP growth, GDP inflation, and the T-bill rate.
3. The CPI results correspond to those reported in Table 5, based on models in GDP growth, CPI inflation, and the T-bill rate.
4. RMSEs for SPF forecasts of CPI inflation are not reported for the 1970-84 sample because the SPF data don't begin until 1981.
5. See the notes to Table 3.