# Averaging Forecasts from VARs with Uncertain Instabilities

## Todd E. Clark and Michael W. McCracken

# 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

October 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, different 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 models of U.S. output, prices, and interest rates, this paper examines the effectiveness of combination in improving VAR forecasts made with real–time data. The combinations include simple averages, medians, trimmed means, and a number of weighted combinations, based on: Bates-Granger regressions, factor model estimates, regressions involving forecast quartiles, Bayesian model 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 as benchmarks.

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

---

# 1 Introduction

Small–scale VARs are now widely used in macroeconomics and central bank forecasting. Examples of VARs used to forecast output, prices, and interest rates include 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). However, there is an increasing body of evidence suggesting that these VARs may be prone to instabilities.[1] Examples include Webb (1995), Kozicki and Tinsley (2001b, 2002), Cogley and Sargent (2001, 2005), Boivin (2006), and Beyer and Farmer (2006). 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.

Accordingly, in previous work (Clark and McCracken, 2006a) 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: sequentially updating lag orders, using various observation windows for estimation, working in differences rather than levels, making intercept corrections (as in Clements and Hendry (1996)), allowing stochastic time variation in model parameters, allowing discrete breaks in parameters, 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 summarized by Timmermann (2006), competing models will differ in their sensitivity to

---

[1] 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.

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 (2005b) 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 (2006) 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.

In this paper we provide empirical evidence on the ability of various forms of forecast 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 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, Bates–Granger combinations estimated with ridge regression, MSE–weighted averages, lowest MSE forecasts (predictive least squares forecasts), Bayesian model averages, and combinations based on quartile average forecasts (as suggested by Aiolfi and Timmermann (2006)). 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 and selected baseline VAR models.

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) using detrended inflation and interest rates — consistently perform among the best methods. At the other extreme, forecasts based on OLS–type combination 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

We consider the real–time forecast performance of models with three different measures of output ($y$), two measures of inflation ($\pi$), 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; using the federal funds rate 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

3

(BLS). 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, 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.[2] For the T-bill rate, we use a series obtained from FAME.

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 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 $t-1$.[3] 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 1965:Q4 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 ($h = 1Q$), four quarters ahead ($h = 1Y$), and eight quarters ahead ($h = 2Y$). 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 are referred to as 1–year ahead and 2–year ahead forecasts are really 5– and 9–step ahead forecasts. In keeping with common central bank practice, the 1– and 2–year ahead forecasts for GDP/GNP growth and inflation are four–quarter rates of change (the 1–year ahead forecast is the percent

---

[2]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.

[3]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.

4

change from period $t + 1$ through $t + 4$; the 2–year ahead forecast is the percent change from period $t + 5$ through $t + 8$). The 1– and 2–year ahead forecasts for output gaps and interest rates are quarterly levels in periods $t + 4$ and $t + 8$, respectively. For computational simplicity in our extensive real–time analysis, all of the multi–step forecasts are obtained by iterating the 1–step ahead models.

As discussed in such sources as Romer and Romer (2000), Sims (2002), and Croushore (2006), evaluating the accuracy of real time forecasts requires a difficult decision on what to take as the actual data in calculating forecast errors. 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 $h = 0Q$, 1Q, 1Y, and 2Y) forecasts made for period $t + h$ with vintage $t$ data ending in period $t - 1$, the second 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, for these series the real time data correspond to the final vintage data.

## 3    Forecast methods

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: sequentially updating lag orders, using various observation windows for estimation, working in differences rather than levels, making intercept corrections (as in Clements and Hendry (1996)), allowing stochastic time variation in model parameters, allowing discrete breaks in parameters identified with break tests, 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 ($\pi$), 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 the univariate model is an AR(2). In the case of inflation, we follow Stock and Watson (2006) and use an MA(1) process for the change in inflation ($\Delta\pi$), estimated with a rolling window of 40 observations. 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 $(\Delta i)$.[4]

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 50 VARs listed in Table 1. With an eye towards making this model average robust to individual forecasts that might be considered outliers, we also consider the median forecast and both 10 and 20 percent trimmed means.

We include a fifth average forecast approach motivated by the results of Clark and Mc-Cracken (2005b), 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 (2005a) 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 with detrending).

While these pairwise average forecasts may seem *ad hoc* from a Bayesian model averaging perspective, our aforementioned results, based on frequentist methods, suggest they

---

[4]After completing the results and analysis presented below, we went back and compared the IMA(1) for the interest rate to various AR alternatives. The IMA(1) generally dominated these alternatives.

may be effective, especially in the face of considerable parameter estimation noise associated with VARs. As an example, suppose that, in truth, output, inflation, and the interest rate can be modeled as a VAR(4). The frequentist results in our prior work (theory, Monte Carlo experiments, and empirics in Clark and McCracken (2005a)) imply that, unless the VAR(4) is estimated with great precision, combining forecasts from the VAR(4) with forecasts from univariate models will likely improve forecast accuracy. Similar arguments suggest averaging a DVAR(4) (or a VAR(4) in detrended data) with univariate forecasts and averaging a rolling estimate of the VAR with forecasts based on recursive estimates. In each case, combination improves forecast accuracy by shrinking the larger model forecast with relatively high sampling error and arguably less bias to a smaller model forecast with less sampling error but greater bias.

## 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) 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 50 VAR forecasts, shrinking the coefficients toward equal weights. Our implementation follows that of Stock and Watson (1999): letting $Z_{t+h|t}$ denote the vector of 50 forecasts of variable $z_{t+h}$ made in period $t$ and $\beta^{equal}$ denote a $50 \times 1$ vector filled with 1/50, the combination coefficient vector estimate is

$$\hat{\beta} = (cI_{50} + \sum_t Z_{t+h|t}Z'_{t+h|t})^{-1}(c\beta^{equal} + \sum_t Z_{t+h|t}z_{t+h}), \tag{1}$$

where $c = k \times \text{trace}(50^{-1} \sum_t Z_{t+h|t} Z'_{t+h|t})$. 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 (1999), 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 (1999, 2004) 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 50 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

We also consider several average forecasts based on inverse MSE weights. At each forecast origin $t$, historical MSEs of the 50 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 (2004), 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 $\delta^{t-s}$). We use 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

8

future.

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 (2006) 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. 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 Bayesian model averages

Following Wright (2003) and Koop and Potter (2004), among others, we also consider forecasts obtained by Bayesian model averaging (BMA). At each forecast origin $t$, for each equation of the 50 models listed in Table 1, we calculate a posterior probability using the conventional formula

$$
\begin{aligned}
\text{Prob}(M_i|\text{data}) &= \frac{\text{Prob}(\text{data}|M_i) \times \text{Prob}(M_i)}{\sum_i \text{Prob}(\text{data}|M_i) \times \text{Prob}(M_i)} \\
\text{Prob}(M_i) &\equiv \text{prior probability on model } i = 1/50 \\
\text{Prob}(\text{data}|M_i) &\equiv \text{marginal likelihood for model } i.
\end{aligned}
\tag{2}
$$

We consider several different measures of the marginal likelihood, each of which yields a different BMA forecast. The three measures are the AIC, BIC, and Phillips' (1996) PIC.[5] The BIC is well known to be proportional to the marginal likelihood of models estimated by OLS or, equivalently, diffuse priors. BMA applications such as Koop, Potter, and Strachan (2005) and Garratt, Koop, and Vahey (2006) have also used BIC to estimate the marginal likelihood and in turn average models. The AIC can be viewed as another measure of the marginal likelihood for models estimated by OLS. Phillips (1996) develops another criterion, PIC, as a measure of marginal likelihood appropriate for comparing VARs in levels, VARs in differences, and VARs estimated with informative priors (BVARs). Specifically, at each forecast origin $t$, for each of the model estimates listed in Table 1, we compute the AIC, BIC, and PIC for each equation of the model.[6] For each criterion, we then form a BMA forecast using $-.5T$ times the information criterion value as the marginal likelihood of each equation.

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 50 underlying model forecasts. In the case of models estimated with a rolling sample of data, we calculate the AIC, BIC, and PIC based on a model that allows a discrete break in all the model coefficients at the point of the beginning of the rolling sample. 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 the AIC and BIC applied to BVAR models, for simplicity we abstract from the prior and calculate the criteria based on the residual sums of squares and simple parameter count (PIC is calculated for VARs and BVARs, to take account of priors, as described in Phillips (1996)).[7] As Phillips (1996) notes, the prior is asymptotically irrelevant in the sense that, as the sample grows, sample information dominates the prior. For marginal likelihood measures other than PIC, taking (proper Bayesian) account of the

---

[5]Note that our BMA forecasts are numerically equivalent (with equal prior weights on each model) to those that would be obtained under the information criteria–weighting approach developed in Kapetanios, Labhard, and Price (2005). These authors, however, suggest a frequentist, rather than Bayesian, interpretation of the information criterion–weighted forecast.

[6]In calculating PIC for the univariate IMA models for inflation and interest rates, we simply approximate the MA fits with AR(1) models estimated for $\Delta\pi$ and $\Delta i$ (estimating separate models for the rolling sample and the earlier sample), and calculate PIC values using these AR(1) approximations.

[7]For BVARs with TVP, at each point in time $t$ we calculate the model residuals as a function of the period $t$ coefficients and use these residuals to compute the residual sums of squares.

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.[8]

## 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 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 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). In light of the evidence in Stock and Watson

---

[8]As Koop and Potter (2004) note, 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 with AIC and BIC 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.

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

(2003) and others of instabilities in forecast performance over time, we examine accuracy over forecast samples of 1970-84 and 1985-2005, to ensure our general results are robust across sample periods.[11]

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 GDP growth and inflation than forecasts of the output gap measures or interest rates. We also focus on a few forecast horizons — those for $h = 0Q$, $h = 1Q$, and $h = 1Y$ — but do present results for the $h = 2Y$ horizon.

Tables 3 and 4 report forecast accuracy (RMSE) results for GDP growth and either GDP price index-based or CPI-based inflation using 38 forecast methods. In each case we use the 3-month T–bill as the interest rate, and present results for horizons $h = 0Q$, $h = 1Q$, and $h = 1Y$. Table 5 reports the same but for the $h = 2Y$ horizon. In Table 6 we report forecast accuracy results for the T-bill rate at all horizons, from models using GDP growth 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.

In Table 7 we take another approach to broadly determining which methods tend to perform better than the benchmark. Across each variable, model and forecast horizon, we compute the average rank of the methods included in Tables 3-6. We present average rankings for every method we consider across each variable, forecast horizon, and the 1970-84 and 1985-05 samples (spanning all columns of Tables 3-6 plus unreported results for forecasts from models using an output gap as well as forecasts of the T-bill rate from models using our various measures of output and inflation).

To determine the statistical significance of differences in forecast accuracy, we use a non–parametric bootstrap patterned after White's (2000) to calculate $p$–values for each RMSE ratio in Tables 3-6. 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. To determine whether

---

[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-$h$.

a best forecast in each column of the tables is significantly better than the benchmark once the data snooping or search involved in selecting a best forecast is taken into account, we apply Hansen's (2005) (bootstrap) SPA test to differences in MSEs (for each model relative to the benchmark). Hansen shows that, if the variance of the forecast loss differential of interest differs widely across models, his SPA test will typically have much greater power than White's (2000) reality check test. For each column, if the SPA test yields a $p$–value of 10 percent or less, we report the associated RMSE ratio in bold font. Because the SPA test is based on $t$–statistics for equal MSE instead of just differences in MSE (that is, takes MSE variability into account), the forecast identified as being significantly best by SPA may not be the forecast with the lowest RMSE ratio.[12]

We implement the bootstrap procedures by sampling from the time series of forecast errors underlying the entries in Tables 3-6. 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 (2000) and Hansen (2005); 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 $h = 1Q$, 5 for $h = 1Y$, and 9 for $h = 2Y$.[13] In addition, in light of the potential for changes over time 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 declined by roughly two-thirds across the 1970-84 and 1985-05 samples (Tables 3-5). The reduced volatility continues to be evident for the inflation measures (Tables 3-5). At the shorter horizons, $h = 0Q$ and $h = 1Q$, the benchmark RMSEs fell by roughly half, but at the longer $h = 1Y$ and $h = 2Y$ horizons the variability declines nearly two-thirds. The reverse is true for the interest rate forecasts (Table 6). At the shorter horizons the benchmark RMSEs fell by roughly two-thirds but at the longer horizons the variability declines by less

---

[12]For multi–step forecasts, we compute the variance entering the $t$–test using the Newey and West (1987) estimator with a lag length of $1.5 * h$, where $h$ denotes the number of forecast periods.

[13]For a forecast horizon of $\tau$ periods, forecast errors from a properly specified model will follow an MA($\tau-1$) process. Accordingly, we use a moving block size of $\tau$ for a forecast horizon of $\tau$.

than half.

## 4.2 Declining predictability

Consistent with the results in Campbell (2005), D'Agostino, et al. (2005), Stock and Watson (2006), 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, at forecast horizons of $h = 1Y$ or less, most methods or models beat the accuracy of the univariate forecast of GDP growth during the 1970-84 period (Tables 3 and 4). In fact, many do so at a level that is statistically significant; at each horizon Hansen's (2005) SPA test identifies a statistically significant best performer. But over the 1985-2005 period, for $h = 0Q$ and $h = 1Q$ forecasts 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. Interestingly, at the longest $h = 2Y$ horizon (Table 5), it appears that it has become modestly easier to predict GDP growth, though again, few are statistically significant.

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 Table 4, 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.[14]

Predictability of the T-bill rate has not so much declined as it has shifted to a longer horizon. In Table 6 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 about 5 percent. But at the $h = 1Y$ and $h = 2Y$ horizons, not only do a

---

[14]Some the change in CPI predictability at the $h = 0Q$ horizon could be due to the 1983 change in the CPI's treatment of housing. Prior to 1983, changes in mortgage interest costs were included in the CPI.

larger number of methods improve upon the benchmark, they do so with maximal gains that are substantial and statistically significant, at about 12 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. 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 (2006b) 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 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.

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 GDP growth and CPI inflation (Table 4), these pairwise averages' RMSE ratios are less than 1 in 8 of 12 columns, and only slightly to modestly above 1 in the exceptions. The VAR(4)–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 7 confirm that, from a broad perspective, the best forecasts are simple averages. 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 6.4; the next–best forecast, the average of the univariate and VAR(4) forecasts, has an average ranking of 12.0. 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) and inflation (column 4 shows it ranks first). As to sample stability, the univariate–inflation detrended VAR(4) average is best in each of the 1970-84 and 1985-05 samples (columns 6-7).

## 4.4 Averaging methods that 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 in most cases fails to beat the benchmark in the 1985-05 sample. Similarly, in the case of T-bill forecasts from the same model (Table 6, 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

16

of averaging the univariate and inflation detrended VAR(4) forecasts beats the univariate 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 3-6 there seems to be no advantage of a median or trimmed mean forecast 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 .972 (growth) and 1.023 (inflation), compared to the simple average's ratios of, respectively, .962 and 1.036 (Table 3).

Similarly, MSE–weighted forecasts are 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 .957 (growth) and 1.028 (inflation), compared to the simple average's ratios of, respectively, .962 and 1.036 (Table 3). In 1-year ahead forecasts of CPI inflation (Table 4), the RMSE ratio of the recursively MSE–weighted forecast is .951 for 1970-84 and 1.055 for 1985-05, compared to the simple average forecast's RMSE ratios of .950 and 1.066, respectively.

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 4'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 .780 for 1970-84 and 1.017 for 1985-05, compared to the simple average forecast's RMSE ratios of, respectively, .839 and .997. Similarly, for Table 4'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 for 1970-84.

The rankings in Table 7 confirm the broad similarity of the above methods — the simple average, MSE–weighted averages, and best quartile forecasts. For example, the simple average forecast has an overall average ranking of 14.5, compared to rankings of 12.0 for the recursive MSE–weighted forecast and 12.6 for the recursive best quartile forecast. By comparison, the best forecast, the univariate–inflation detrended VAR(4) average, has an

---

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

overall ranking of 6.4. 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 17.3 (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 13.4, compared to the univariate ranking of 21.8 (column 6). But for 1985-05 forecasts, the all–model average has an average accuracy rank of 16.6, compared to the univariate ranking of 13.9 (column 7).

## 4.5   Averaging methods that rarely outperform the benchmark

Many 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 4), the RMSE ratio of the recursively estimated ridge regression with shrinkage parameter of .001 is 1.458. 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 samples — are generally 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 1.125 (10 year rolling) and 1.110 (5 year rolling), respectively, compared to the all–average forecast's RMSE ratio of 1.066.

While using more shrinkage improves the accuracy of forecast combinations estimated with generalized ridge regression, even the combinations based on ridge regression with non–trivial shrinkage are generally less accurate than the univariate benchmarks and simple average forecasts. For example, in 1985-05 forecasts of GDP growth from models using the GDP inflation measure (Table 3), the RMSE ratios of the $k = 1$ recursive ridge regression forecast are all above those of the simple average forecast. While the ridge forecasts are more commonly beaten by the simple average, there are, to be sure, a number of instances (as in the same example, but with a forecast sample of 1970-84) in which ridge forecasts are more accurate. On balance, though, the ridge combinations seem to be inferior to alternatives such as the simple average forecast.

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 1.021 (growth) and 1.536 (inflation), compared to the simple average's ratios of, respectively, .962 and 1.036 (Table 3). The same is true for the PLS forecasts: although PLS forecasts are sometimes more accurate than the simple average, they are often worse. In the same example, the recursive PLS forecast's RMSE ratios are 1.108 and 1.011, respectively.

The BMA forecasts are also generally, although not universally, dominated by the simple average. For example, in Table 6's forecasts of the T-bill rate, the RMSE ratios of the BMA: BIC forecast are consistently above the ratios of the simple average forecast. However, in Table 3's results for GDP growth and GDP inflation, the accuracy of the BMA: BIC forecast is generally comparable to that of the simple average forecast. Among the alternative BMA forecasts, there are times when those using AIC or PIC to measure the marginal likelihood are more accurate than those using BIC. But more typically, the BMA: BIC forecast is more accurate than the BMA: AIC and BMA: PIC forecasts — the pattern is especially clear in 1985-05 forecasts.

The rankings in Table 7 provide a clear and convenient listing of the forecast methods that are generally dominated by 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 consistently in the low- to mid-30s. OLS combinations of quartile forecasts also fare quite poorly when based on rolling samples, with ranks generally in the mid 20s to low 30s. The factor–based combination forecasts are also consistently ranked in the bottom tier, with average rankings generally in the mid-20s. While not necessarily in the bottom tier, the BMA forecasts are generally dominated by the simple average forecast. The overall rankings of the BMA: BIC, BMA: PIC, and BMA: AIC forecasts are 22.0, 25.4, and 29.0, respectively, compared with the simple average forecast's ranking of 14.5 (first column). The average ranks of the PLS forecasts are consistently around 20 (or much worse in the 5 year rolling case). The ridge–based combination forecasts with the highest degree of shrinkage ($k = 1$) fare much better than the OLS–approximating ridge

19

combinations, but consistently rank below the simple average forecast. For example, as shown in the first column, the 10–year rolling ridge regression with $k = 1$ has an average ranking of 16.9.

## 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 3-4: 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 7. Across all variables, horizons, and samples, the best–ranked single model is the BVAR(4) with inflation detrending, which is out–ranked by 4 different average forecasts. 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 32 average forecasts and all 50 single model forecasts across the 1970-84 and 1985-05 periods, based on the inflation and output results covered in columns 6-7 of Table 7 (using rankings including T-bill rate forecasts yields essentially the same correlations). The correlation of single model forecast rankings is 53 percent; the correlation of the average forecast rankings is 92 percent. The implication is

20

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 (1999, 2004) and 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 when the 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 use Bayesian shrinkage in estimating the VAR with inflation detrending. 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 and Potter (2004) note that such model averaging can be viewed as a form of shrinkage for addressing both parameter estimation noise and

---

[16]As discussed in Stock and Watson (2006), 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.

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]

# 5  Conclusion

In this paper we provide empirical evidence on the ability of several forms of forecast 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 obtained with a variety of primitive methods for managing model instability. These primitive methods include sequentially updating lag orders, using various observation windows for estimation, working in differences rather than levels, making intercept corrections (as in Clements and Hendry (1996)), allowing stochastic time variation in model parameters, allowing discrete breaks in parameters identified with break tests, discounted least squares estimation, Bayesian shrinkage, and detrending of inflation and interest rates. The forecast averages 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 forecasts), 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. 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 and factor model–based combination rank among the worst.

---

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

## References

Aiolfi, Marco and Allan Timmermann (2006), "Persistence in Forecasting Performance and Conditional Combination Strategies," *Journal of Econometrics* 135, 31-54.

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* 38, 1149-1173.

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

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

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

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

Clark, Todd E. and Michael W. McCracken (2006b), "The Predictive Content of the Output Gap for Inflation: Resolving In–Sample and Out–of–Sample Evidence," *Journal of Money, Credit, and Banking* 38, 1127-1148.

Clements, Michael P. and David F. Hendry (1996), "Intercept corrections and structural change," *Journal of Applied Econometrics* 11, 475-494.

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, 331-373.

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 (2006), "Forecasting with Real–Time Macroeconomic Data," *Handbook of Forecasting*, G. Elliott, C.W.J. Granger, and A. Timmermann, eds., North Holland.

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, 643-673.

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.

Garratt, Anthony, Gary Koop and Shaun P. Vahey (2006), "Forecasting Substantial Data Revisions in the Presence of Model Uncertainty," Reserve Bank of New Zealand Discussion Paper 2006/02.

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.

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, 841-858.

Hansen, Peter R. (2005), "A Test for Superior Predictive Ability," *Journal of Business and Economic Statistics* 23, 365-380.

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.

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 (2004), "Forecasting in Dynamic Factor Models Using Bayesian Model Averaging," *Econometrics Journal* 7, 550-565.

Koop, Gary, Simon M. Potter and Rodney W. Strachan (2005), "Reexamining the Consumption–Wealth Relationship: the Role of Model Uncertainty," Federal Reserve Bank of New York Staff Report no. 202.

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, 613-652.

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, 1217-1241.

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.

Nelson, Charles R. (1972), "The Predictive Performance of the FRB-MIT-PENN Model of the U.S. Economy," *American Economic Review* 62, 902-917.

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, 478-486.

Newey, Whitney K. and Kenneth D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

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, Davide Pettenuzzo and Allan Timmermann (2006), "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies* 73, 1057-1084.

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.

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, 324-330.

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, 429-457.

Rudebusch, Glenn D. (2005), "Assessing the Lucas Critique in Monetary Policy Models," *Journal of Money, Credit, and Banking* 37, 245-272.

Rudebusch, Glenn D. and Lars E.O. Svensson (1999), "Policy Rules for Inflation Targeting." in J. Taylor, ed., *Monetary Policy Rules*, University of Chicago Press: Chicago, 203-246.

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 (1999), "A Dynamic Factor Model Framework for Forecast Combination," *Spanish Economic Review* 1, 91-121.

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, 405-430.

Stock, James H. and Mark W. Watson (2006), "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit, and Banking*, forthcoming.

Timmermann, Allan (2006), "Forecast Combinations," *Handbook of Forecasting*, G. Elliott, C.W.J. Granger, and A. Timmermann, eds., North Holland.

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

Webb, Roy H. (1995), "Forecasts of Inflation from VAR Models," *Journal of Forecasting* 14, 267-285.

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

| method | details |
|---|---|
| VAR(4) | VAR in $y$, $\pi$, $i$ 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 $y$, $\pi$, $i$ 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 $y$, $\Delta\pi$, $\Delta i$ with fixed lag of 4 |
| DVAR(2) | same as above with fixed lag of 2 |
| DVAR(AIC) | VAR in $y$, $\Delta\pi$, $\Delta i$ with system lag determined at each $t$ by AIC |
| DVAR(BIC) | VAR in $y$, $\Delta\pi$, $\Delta i$ with system lag determined at each $t$ by BIC |
| DVAR(AIC, by eq.&var.) | VAR in $y$, $\Delta\pi$, $\Delta i$ 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 $y$, $\pi$, $i$, est. with Minnesota priors, using $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$ |
| BDVAR(4) | VAR(4) in $y$, $\Delta\pi$, $\Delta i$, est. with Minnesota priors, using $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$ |
| VAR(4), rolling | VAR in $y$, $\pi$, $i$ 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 $y$, $\Delta\pi$, $\Delta i$ 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 $y$, $\pi$, $i$ with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$, est. with a rolling sample |
| BDVAR(4), rolling | BVAR(4) in $y$, $\Delta\pi$, $\Delta i$ with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$, est. with a rolling sample |
| DLS, VAR(4) | VAR(4) in $y$, $\pi$, $i$, est. by DLS, using dis. rates of .01 for $y$ eq. and .05 for $\pi$ 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 $y$, $\Delta\pi$, $\Delta i$, est. by DLS using dis. rates of .01 for $y$ eq. and .05 for $\Delta\pi$ and $\Delta i$ 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 $y$, $\pi$, $i$, 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 $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$ |
| BVAR(4) with TVP | TVP BVAR(4) in $y$, $\pi$, $i$ 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 $y$, $\pi$, $i$ 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 $y$, $\pi$, $i$ 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 $y$, $\pi$, $i$ with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$, $\lambda = .0001$ |
| BVAR(4) with intercept TVP | BVAR(4) in $y$, $\pi$, $i$, 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 $y$, $\pi$, $i$, 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 $\Delta i$ |

*Notes*:

1. The variables $y$, $\pi$, 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 $\Delta i$ 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 $\lambda_1$ (general tightness), $\lambda_2$ (tightness of lags of other variables compared to lags of the dependent variable), $\lambda_3$ (tightness of longer lags compared to shorter lags), and $\lambda_4$ (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 $\sigma_j$ and $\sigma_m$ 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 $\pi$ 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 time variation in the coefficients of the TVP BVARs takes a random walk form. The variance matrix of the coefficient innovations is set to $\lambda$ 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 $\lambda$ 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 $y$, $\pi$, and $i$ |
| avg. of infl. detr. VAR(4), univariate | 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 $\Delta y$, $\Delta \pi$, and $i$ |
| avg. of VAR(4), rolling VAR(4) | average of forecasts from recursive and rolling estimates of VAR(4) in $y$, $\pi$, 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 (1), $k = .001$ |
| 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 20 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 quartile 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 |
| BMA: AIC | BMA of model forecasts, using AIC as measure of marginal likelihood |
| BMA: BIC | BMA of model forecasts, using BIC as measure of marginal likelihood |
| BMA: PIC | BMA of model forecasts, using Phillips' (1996) PIC as measure of marginal likelihood |

*Notes*:
1. All averages are based on the 50 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)*

| forecast method | GDP growth forecasts | | | | | | GDP inflation forecasts | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | 1970-84 | | | 1985-2005 | | | 1970-84 | | | 1985-2005 | | |
| | $h=0Q$ | $h=1Q$ | $h=1Y$ | $h=0Q$ | $h=1Q$ | $h=1Y$ | $h=0Q$ | $h=1Q$ | $h=1Y$ | $h=0Q$ | $h=1Q$ | $h=1Y$ |
| univariate | 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), univariate | .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 | .944 | .877 | .836 | 1.049 | 1.060 | .962 | .933 | .992 | .979 | 1.023 | .998 | 1.036 |
| median | .937 | .879 | .867 | 1.042 | 1.076 | .991 | .954 | 1.012 | 1.005 | 1.017 | .993 | 1.016 |
| trimmed mean, 10% | .946 | .881 | .844 | 1.050 | 1.061 | .967 | .940 | .994 | .985 | 1.022 | .993 | 1.027 |
| trimmed mean, 20% | .948 | .884 | .848 | 1.050 | 1.062 | .972 | .943 | .996 | .989 | 1.021 | .992 | 1.023 |
| ridge: recursive, .001 | 2.169 | 1.590 | 1.589 | 1.295 | 1.353 | 1.588 | 1.254 | 1.532 | 1.867 | 1.147 | 1.101 | 1.755 |
| ridge: recursive, .25 | .957 | .876 | .773 | 1.077 | 1.132 | 1.015 | .968 | 1.044 | 1.091 | 1.020 | .988 | 1.039 |
| ridge: recursive, 1. | .939 | .859 | .773 | 1.073 | 1.125 | .994 | .963 | 1.039 | 1.087 | 1.018 | .984 | .981 |
| ridge: 10y rolling, .001 | 2.332 | 1.871 | 1.693 | 1.563 | 1.762 | 1.722 | 1.219 | 1.517 | 1.889 | 1.177 | 1.174 | 1.713 |
| ridge: 10y rolling, .25 | .968 | .881 | .791 | 1.080 | 1.126 | 1.081 | .967 | 1.036 | 1.077 | 1.027 | .979 | 1.050 |
| ridge: 10y rolling, 1. | .944 | .861 | .784 | 1.077 | 1.117 | 1.021 | .962 | 1.032 | 1.075 | 1.023 | .983 | .997 |
| factor, recursive | .987 | .930 | .929 | 1.095 | 1.111 | 1.021 | 1.006 | 1.087 | .992 | 1.031 | 1.073 | 1.536 |
| factor, 10y rolling | .993 | .936 | .935 | 1.112 | 1.108 | 1.093 | 1.021 | 1.124 | 1.039 | 1.018 | 1.011 | 1.419 |
| MSE weighting, recursive | .943 | .876 | .824 | 1.046 | 1.058 | .957 | .935 | .992 | .984 | 1.022 | .996 | 1.028 |
| MSE weighting, 10y rolling | .943 | .877 | .825 | 1.046 | 1.057 | .952 | .935 | .992 | .984 | 1.022 | .994 | 1.029 |
| MSE weighting, 5y rolling | .941 | .875 | .831 | 1.044 | 1.056 | .957 | .934 | .989 | .974 | 1.023 | .998 | 1.026 |
| MSE weighting, discounted | .943 | .877 | .829 | 1.045 | 1.057 | .958 | .934 | .990 | .981 | 1.022 | .995 | 1.023 |
| PLS, recursive | .973 | .915 | .806 | 1.106 | 1.094 | 1.108 | .993 | 1.071 | 1.203 | .980 | .978 | 1.011 |
| PLS, 10y rolling | .973 | .915 | .810 | 1.092 | 1.151 | 1.040 | 1.043 | 1.048 | 1.253 | 1.036 | .963 | 1.092 |
| PLS, 5y rolling | 1.023 | .933 | .828 | 1.103 | 1.121 | 1.074 | 1.052 | 1.086 | 1.301 | 1.113 | 1.009 | 1.132 |
| best quartile, recursive | .943 | .879 | .778 | 1.028 | 1.063 | .940 | .946 | .994 | 1.002 | 1.015 | .982 | 1.059 |
| best quartile, 10y rolling | .951 | .884 | .784 | 1.046 | 1.061 | .962 | .945 | .988 | 1.011 | 1.011 | .977 | 1.019 |
| best quartile, 5y rolling | .935 | .878 | .823 | 1.042 | 1.088 | 1.020 | .938 | .977 | .970 | 1.030 | .984 | 1.018 |
| OLS comb. of quartiles, recursive | .978 | .917 | .910 | 1.169 | 1.150 | .940 | 1.004 | 1.039 | 1.095 | 1.034 | .992 | 1.396 |
| OLS comb. of quartiles, 10y rolling | .976 | .928 | .927 | 1.090 | 1.114 | 1.026 | 1.022 | 1.081 | 1.163 | 1.020 | .939 | 1.328 |
| OLS comb. of quartiles, 5y rolling | .968 | .914 | .951 | 1.268 | 1.413 | 1.557 | 1.014 | 1.135 | 1.370 | 1.083 | 1.116 | 1.025 |
| BMA: AIC | 1.008 | .966 | .901 | 1.120 | 1.106 | .978 | .991 | 1.023 | 1.053 | 1.113 | 1.126 | 1.502 |
| BMA: BIC | .946 | .909 | .964 | 1.047 | 1.038 | .898 | .972 | 1.004 | .991 | 1.014 | 1.004 | 1.005 |
| BMA: PIC | .902 | .837 | .849 | 1.100 | 1.087 | .926 | .962 | 1.046 | 1.075 | 1.080 | 1.082 | 1.128 |

*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 a forecast is significantly better (in MSE) than the benchmark according to data snooping-robust $p$-values (bootstrapped as in Hansen (2005)), the associated 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 GDP growth and CPI inflation**
*(RMSEs in first row, RMSE ratios in all others)*

| forecast method | GDP growth forecasts | | | | | | CPI inflation forecasts | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | 1970-84 | | | 1985-2005 | | | 1970-84 | | | 1985-2005 | | |
| | $h=0Q$ | $h=1Q$ | $h=1Y$ | $h=0Q$ | $h=1Q$ | $h=1Y$ | $h=0Q$ | $h=1Q$ | $h=1Y$ | $h=0Q$ | $h=1Q$ | $h=1Y$ |
| univariate | 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), univariate | .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 | .947 | .903 | .839 | 1.070 | 1.076 | .997 | .826 | .927 | .950 | .984 | 1.012 | 1.066 |
| median | .945 | .895 | .871 | 1.059 | 1.074 | 1.018 | .863 | .914 | .940 | .983 | 1.000 | 1.039 |
| trimmed mean, 10% | .947 | .904 | .850 | 1.070 | 1.075 | .999 | .834 | .925 | .945 | .981 | 1.011 | 1.053 |
| trimmed mean, 20% | .947 | .905 | .855 | 1.070 | 1.075 | 1.002 | .840 | .923 | .944 | .979 | 1.007 | 1.048 |
| ridge: recursive, .001 | 1.888 | 1.337 | 1.335 | 1.229 | 1.477 | 1.301 | 1.103 | 1.374 | 2.013 | 1.113 | 1.353 | 1.458 |
| ridge: recursive, .25 | .951 | .852 | .780 | 1.055 | 1.096 | .977 | .845 | .971 | 1.081 | .986 | .997 | 1.071 |
| ridge: recursive, 1. | .946 | .871 | .799 | 1.059 | 1.085 | .969 | .845 | .966 | 1.074 | .978 | .992 | 1.012 |
| ridge: 10y rolling, .001 | 1.923 | 1.331 | 1.422 | 1.311 | 1.665 | 1.483 | 1.056 | 1.253 | 1.929 | 1.195 | 1.593 | 1.659 |
| ridge: 10y rolling, .25 | .965 | .862 | .796 | 1.028 | 1.090 | 1.047 | .840 | .966 | 1.067 | .990 | .999 | 1.014 |
| ridge: 10y rolling, 1. | .949 | .874 | .803 | 1.053 | 1.093 | 1.013 | .840 | .959 | 1.063 | .982 | .993 | .964 |
| factor, recursive | 1.003 | .981 | .913 | 1.089 | 1.093 | 1.023 | .833 | .945 | .951 | .991 | 1.061 | 1.249 |
| factor, 10y rolling | 1.011 | .991 | .925 | 1.101 | 1.093 | 1.087 | .851 | .977 | .989 | .957 | 1.005 | 1.134 |
| MSE weighting, recursive | .947 | .901 | .830 | 1.065 | 1.073 | .995 | .830 | .929 | .951 | .982 | 1.007 | 1.055 |
| MSE weighting, 10y rolling | .948 | .902 | .831 | 1.064 | 1.074 | .992 | .830 | .929 | .950 | .982 | 1.007 | 1.055 |
| MSE weighting, 5y rolling | .946 | .901 | .845 | 1.060 | 1.073 | .991 | .828 | .928 | .949 | .980 | 1.005 | 1.050 |
| MSE weighting, discounted | .948 | .903 | .836 | 1.064 | 1.073 | .998 | .828 | .930 | .953 | .982 | 1.006 | 1.051 |
| PLS, recursive | .902 | .850 | .766 | 1.191 | 1.210 | 1.092 | .894 | 1.081 | .908 | 1.005 | 1.016 | 1.183 |
| PLS, 10y rolling | .902 | .842 | .766 | 1.141 | 1.202 | 1.092 | .859 | 1.077 | .910 | 1.005 | 1.037 | 1.078 |
| PLS, 5y rolling | .884 | .849 | .924 | 1.131 | 1.202 | 1.113 | .923 | 1.106 | .901 | 1.020 | 1.049 | 1.066 |
| best quartile, recursive | .938 | .898 | .773 | 1.038 | 1.069 | .995 | .868 | .933 | .922 | .974 | 1.010 | 1.079 |
| best quartile, 10y rolling | .943 | .896 | .780 | 1.043 | 1.064 | 1.017 | .867 | .941 | .928 | .976 | 1.004 | 1.078 |
| best quartile, 5y rolling | .966 | .907 | .816 | 1.058 | 1.085 | 1.013 | .838 | .938 | .961 | .975 | 1.014 | 1.036 |
| OLS comb. of quartiles, recursive | .949 | .942 | .778 | 1.186 | 1.275 | .904 | .895 | .960 | .945 | .968 | 1.019 | 1.071 |
| OLS comb. of quartiles, 10y rolling | .963 | .955 | .814 | 1.041 | 1.125 | 1.074 | .876 | 1.032 | .970 | .964 | 1.026 | 1.125 |
| OLS comb. of quartiles, 5y rolling | 1.094 | .966 | .893 | 1.309 | 1.350 | 1.709 | .779 | 1.059 | 2.331 | .995 | 1.058 | 1.110 |
| BMA: AIC | .921 | .879 | .914 | 1.154 | 1.190 | 1.147 | .898 | 1.014 | 1.056 | .991 | 1.057 | 1.154 |
| BMA: BIC | .959 | .867 | .887 | 1.055 | 1.077 | 1.024 | .878 | .962 | 1.003 | .978 | 1.027 | 1.103 |
| BMA: PIC | .873 | .796 | .812 | 1.120 | 1.148 | 1.132 | .910 | 1.076 | 1.157 | 1.007 | 1.010 | 1.030 |

*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 5: Real-time RMSE results for 2-year ahead forecasts of GDP growth and inflation**

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

| forecast method | Models with GDP growth and GDP inflation | | | | Models with GDP growth and CPI inflation | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | GDP growth forecasts | | GDP inflation forecasts | | GDP growth forecasts | | CPI inflation forecasts | |
| | 1970-84 | 1985-2005 | 1970-84 | 1985-2005 | 1970-84 | 1985-2005 | 1970-84 | 1985-2005 |
| univariate | 3.679 | 1.390 | 3.677 | .981 | 3.679 | 1.390 | 4.635 | 1.452 |
| VAR(4) | 1.268 | 1.010 | 1.145 | 1.054 | 1.081 | .997 | 1.132 | 1.246 |
| DVAR(4) | 1.042 | .975 | .900 | 1.099 | 1.045 | .991 | .882 | 1.095 |
| BVAR(4) | 1.238 | .968 | 1.175 | 1.351 | 1.188 | .924 | 1.178 | 1.038 |
| BVAR(4) with TVP | 1.170 | .954 | 1.237 | 1.065 | 1.134 | .952 | 1.215 | .894 |
| BVAR(4), inflation detrending | 1.077 | .918 | 1.015 | 1.580 | 1.030 | .984 | .972 | 1.212 |
| avg. of VAR(4), univariate | 1.090 | .971 | 1.044 | .985 | .998 | .967 | .983 | 1.048 |
| avg. of DVAR(4), univariate | 1.004 | .984 | **.941** | 1.033 | 1.005 | .992 | .926 | 1.006 |
| avg. of infl. detr. VAR(4), univariate | 1.050 | .947 | .966 | 1.075 | .965 | 1.001 | .863 | 1.020 |
| avg. of VAR(4), rolling VAR(4) | 1.299 | 1.048 | 1.138 | 1.331 | 1.162 | 1.032 | 1.179 | 1.299 |
| average of all forecasts | 1.095 | .964 | 1.040 | 1.134 | 1.055 | .998 | 1.053 | 1.057 |
| median | 1.120 | .966 | 1.040 | 1.114 | 1.066 | .995 | 1.012 | 1.012 |
| trimmed mean, 10% | 1.091 | .962 | 1.040 | 1.126 | 1.046 | .993 | 1.030 | 1.035 |
| trimmed mean, 20% | 1.091 | .962 | 1.039 | 1.123 | 1.044 | .991 | 1.021 | 1.031 |
| ridge: recursive, .001 | 2.029 | 2.517 | 1.772 | 3.027 | 2.150 | 1.546 | 3.355 | 4.169 |
| ridge: recursive, .25 | 1.189 | 1.235 | 1.175 | 1.247 | 1.378 | 1.079 | 1.539 | 1.135 |
| ridge: recursive, 1. | 1.064 | 1.146 | 1.114 | 1.060 | 1.143 | 1.040 | 1.159 | .985 |
| ridge: 10y rolling, .001 | 1.969 | 2.334 | 1.744 | 3.149 | 1.945 | 1.973 | 2.785 | 3.426 |
| ridge: 10y rolling, .25 | 1.218 | 1.149 | 1.161 | 1.243 | 1.405 | 1.025 | 1.518 | 1.121 |
| ridge: 10y rolling, 1. | 1.075 | 1.032 | 1.096 | 1.067 | 1.158 | .993 | 1.141 | .948 |
| factor, recursive | 1.335 | 1.029 | .843 | 2.266 | 1.341 | 1.011 | .856 | 1.723 |
| factor, 10y rolling | 1.338 | 1.047 | .864 | 2.161 | 1.337 | 1.019 | .878 | 1.645 |
| MSE weighting, recursive | 1.067 | .955 | 1.031 | 1.115 | 1.037 | .984 | 1.022 | 1.038 |
| MSE weighting, 10y rolling | 1.065 | .950 | 1.032 | 1.109 | 1.036 | .984 | 1.021 | 1.042 |
| MSE weighting, 5y rolling | 1.057 | .953 | 1.029 | 1.116 | 1.034 | .986 | 1.018 | 1.054 |
| MSE weighting, discounted | 1.072 | .954 | 1.030 | 1.107 | 1.041 | .985 | 1.020 | 1.041 |
| PLS, recursive | 1.119 | .976 | 1.069 | 1.295 | 1.058 | 1.058 | .889 | 1.353 |
| PLS, 10y rolling | 1.118 | 1.006 | 1.086 | 1.266 | 1.053 | 1.031 | **.846** | 1.251 |
| PLS, 5y rolling | 1.132 | 1.030 | 1.069 | 1.362 | 1.089 | 1.068 | .819 | 1.314 |
| best quartile, recursive | 1.042 | .991 | .971 | 1.202 | 1.040 | .986 | .942 | 1.085 |
| best quartile, 10y rolling | 1.040 | .955 | .981 | 1.155 | 1.034 | .984 | .940 | 1.128 |
| best quartile, 5y rolling | 1.044 | .953 | .982 | 1.143 | 1.047 | .983 | .928 | 1.115 |
| OLS comb. of quartiles, recursive | 1.654 | 1.142 | .934 | 1.943 | 1.851 | 1.055 | .842 | 1.689 |
| OLS comb. of quartiles, 10 year rolling | 1.667 | 1.053 | .943 | 2.234 | 1.858 | 1.052 | .916 | 1.706 |
| OLS comb. of quartiles, 5 year rolling | 1.839 | 1.562 | 1.369 | 1.863 | 1.883 | 1.662 | 1.205 | 2.329 |
| BMA: AIC | 1.029 | .943 | 1.100 | 1.775 | 1.112 | 1.113 | 1.154 | 1.155 |
| BMA: BIC | 1.194 | .895 | 1.059 | 1.031 | 1.176 | 1.009 | 1.151 | 1.165 |
| BMA: PIC | 1.139 | .881 | 1.105 | 1.164 | 1.186 | 1.030 | 1.282 | 1.082 |

*Notes:*
1. In the results in the left half of the table, the model variables are GDP growth, GDP inflation, and the T-bill rate. In the results in the right half of the table, the model variables are the GDP growth, CPI inflation, and the T-bill rate.
2. 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 $t$ through $t+8$ ($h = 2Y$). The forecasts of GDP growth and inflation for the $h = 2Y$ horizon correspond to annual percent changes: average growth and average inflation from $t+5$ through $t+8$.
3. See the notes to Table 3.

31

**Table 6: Real-time RMSE results for the T-bill rate, from models with GDP growth and GDP inflation**

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

| forecast method | 1970-84 | | | | 1985-2005 | | | |
|---|---|---|---|---|---|---|---|---|
| | $h=0Q$ | $h=1Q$ | $h=1Y$ | $h=2Y$ | $h=0Q$ | $h=1Q$ | $h=1Y$ | $h=2Y$ |
| univariate | 1.305 | 2.098 | 2.821 | 3.678 | .378 | .778 | 1.625 | 2.327 |
| VAR(4) | .940 | .954 | 1.108 | 1.120 | 1.084 | 1.027 | .892 | .852 |
| DVAR(4) | .933 | .917 | .981 | .939 | 1.137 | 1.131 | 1.083 | 1.108 |
| BVAR(4) | .949 | .926 | 1.027 | 1.085 | 1.067 | .986 | .907 | .852 |
| BVAR(4) with TVP | .949 | .933 | 1.054 | 1.187 | 1.078 | 1.006 | .944 | .912 |
| BVAR(4), inflation detrending | .930 | .860 | .908 | .856 | 1.150 | 1.022 | .892 | .777 |
| avg. of VAR(4), univariate | .934 | .930 | 1.030 | 1.036 | .988 | .966 | .925 | .901 |
| avg. of DVAR(4), univariate | .931 | .909 | .966 | .951 | 1.027 | 1.036 | 1.032 | 1.045 |
| avg. of infl. detr. VAR(4), univariate | .924 | .910 | .964 | **.908** | .982 | .957 | **.908** | **.862** |
| avg. of VAR(4), rolling VAR(4) | .967 | .995 | 1.198 | 1.232 | 1.182 | 1.098 | .934 | .897 |
| average of all forecasts | .911 | .927 | 1.018 | 1.048 | 1.034 | .989 | .944 | .928 |
| median | .928 | .942 | .994 | .990 | 1.033 | .989 | .946 | .922 |
| trimmed mean, 10% | .911 | .924 | 1.007 | 1.020 | 1.031 | .986 | .947 | .929 |
| trimmed mean, 20% | .916 | .924 | 1.006 | 1.013 | 1.030 | .986 | .948 | .930 |
| ridge: recursive, .001 | .981 | .821 | 1.821 | 1.227 | 1.114 | 1.188 | 1.151 | 1.493 |
| ridge: recursive, .25 | .955 | .966 | 1.128 | 1.078 | 1.031 | .985 | .936 | .939 |
| ridge: recursive, 1. | .942 | .964 | 1.136 | 1.129 | 1.037 | .991 | .932 | .907 |
| ridge: 10y rolling, .001 | 1.124 | .954 | 1.713 | 1.344 | 1.037 | 1.118 | 1.126 | 1.211 |
| ridge: 10y rolling, .25 | .970 | .988 | 1.168 | 1.078 | 1.028 | .977 | .903 | .925 |
| ridge: 10y rolling, 1. | .952 | .979 | 1.165 | 1.119 | 1.033 | .978 | .889 | .883 |
| factor, recursive | .960 | .957 | 1.182 | 1.210 | 1.139 | 1.129 | 1.108 | 1.085 |
| factor, 10y rolling | .964 | .957 | 1.250 | 1.320 | 1.213 | 1.207 | 1.308 | 1.183 |
| MSE weighting, recursive | .917 | .926 | 1.008 | 1.020 | 1.036 | .990 | .942 | .916 |
| MSE weighting, 10y rolling | .919 | .928 | 1.009 | 1.018 | 1.030 | .988 | .944 | .909 |
| MSE weighting, 5y rolling | .922 | .931 | 1.011 | 1.027 | 1.026 | .986 | .949 | .906 |
| MSE weighting, discounted | .921 | .929 | 1.013 | 1.026 | 1.032 | .989 | .943 | .907 |
| PLS, recursive | .962 | .985 | .980 | .964 | 1.104 | .984 | .911 | .780 |
| PLS, 10y rolling | 1.010 | .955 | 1.028 | .957 | 1.080 | .942 | .925 | .720 |
| PLS, 5y rolling | .998 | 1.057 | 1.310 | .955 | 1.249 | 1.086 | 1.060 | .891 |
| best quartile, recursive | .957 | .940 | .981 | .923 | 1.052 | 1.002 | .950 | .923 |
| best quartile, 10y rolling | .955 | .948 | .993 | .931 | 1.024 | .979 | .944 | .908 |
| best quartile, 5y rolling | .959 | .977 | 1.011 | .939 | 1.062 | 1.021 | .984 | .918 |
| OLS comb. of quartiles, recursive | 1.032 | .993 | 1.292 | 1.263 | 1.137 | 1.168 | 1.110 | 1.095 |
| OLS comb. of quartiles, 10y rolling | 1.089 | 1.080 | 1.404 | 1.255 | 1.232 | 1.205 | 1.444 | 1.248 |
| OLS comb. of quartiles, 5y rolling | 1.110 | 1.134 | 1.255 | 1.310 | 1.362 | 1.537 | 1.883 | 1.671 |
| BMA: AIC | 1.007 | 1.054 | 1.259 | 1.386 | 1.434 | 1.332 | 1.138 | 1.060 |
| BMA: BIC | .958 | 1.003 | 1.038 | 1.032 | 1.316 | 1.194 | .995 | .958 |
| BMA: PIC | 1.042 | 1.053 | 1.139 | 1.227 | 1.332 | 1.250 | 1.067 | 1.001 |

*Notes:*
1. The model variables are GDP growth, GDP inflation, and the T-bill rate.
2. See the notes to Tables 3 and 5.

**Table 7: Average RMSE rankings in real-time forecasts**

| | all | all $y,\pi$ | all $y$ | all $\pi$ | all $i$ | all $y,\pi$ 70-84 | all $y,\pi$ 85-05 | all $y,\pi$ $h=0Q$ | all $y,\pi$ $h=1Q$ | all $y,\pi$ $h=1Y$ | all $y,\pi$ $h=2Y$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| avg. of infl. detr. VAR(4), univar. | 6.4 | 8.0 | 8.9 | 7.1 | 3.3 | 10.4 | 5.5 | 11.3 | 6.2 | 7.0 | 7.5 |
| avg. of VAR(4), univariate | 12.0 | 13.6 | 15.8 | 11.4 | 8.9 | 17.6 | 9.5 | 14.8 | 11.4 | 13.8 | 14.3 |
| MSE weighting, recursive | 12.0 | 12.4 | 11.7 | 13.1 | 11.3 | 11.8 | 13.0 | 13.2 | 12.6 | 12.2 | 11.5 |
| MSE weighting, 10y rolling | 12.3 | 12.8 | 12.6 | 12.9 | 11.3 | 11.8 | 13.7 | 14.0 | 13.2 | 12.4 | 11.5 |
| BVAR(4), inflation detrending | 12.5 | 13.8 | 8.2 | 19.4 | 9.8 | 12.8 | 14.8 | 16.2 | 14.1 | 13.8 | 10.9 |
| best quartile, recursive | 12.6 | 11.8 | 10.4 | 13.1 | 14.3 | 11.0 | 12.6 | 11.6 | 11.7 | 11.8 | 12.0 |
| avg. of DVAR(4), univariate | 12.7 | 12.1 | 14.5 | 9.6 | 13.9 | 11.6 | 12.5 | 15.5 | 10.0 | 11.0 | 11.7 |
| MSE weighting, 5y rolling | 12.7 | 12.5 | 12.8 | 12.3 | 13.1 | 10.7 | 14.4 | 12.8 | 12.9 | 12.8 | 11.6 |
| MSE weighting, discounted | 12.8 | 13.0 | 13.3 | 12.7 | 12.4 | 12.5 | 13.5 | 13.4 | 13.6 | 13.1 | 11.9 |
| best quartile, 10y rolling | 13.2 | 12.9 | 13.2 | 12.6 | 13.9 | 11.8 | 14.0 | 14.0 | 13.0 | 12.2 | 12.4 |
| trimmed mean, 20% | 13.8 | 14.8 | 15.5 | 14.0 | 11.9 | 15.0 | 14.6 | 15.5 | 14.6 | 14.7 | 14.3 |
| trimmed mean, 10% | 13.9 | 14.8 | 15.0 | 14.6 | 12.2 | 14.2 | 15.4 | 15.9 | 14.6 | 14.4 | 14.3 |
| average of all forecasts | 14.5 | 15.0 | 14.8 | 15.2 | 13.4 | 13.4 | 16.6 | 15.4 | 14.8 | 14.1 | 15.7 |
| median | 15.2 | 15.7 | 17.0 | 14.4 | 14.1 | 16.3 | 15.1 | 16.2 | 14.9 | 16.0 | 15.7 |
| best quartile, 5y rolling | 15.9 | 14.5 | 15.8 | 13.1 | 18.9 | 12.3 | 16.6 | 14.4 | 15.1 | 14.8 | 13.5 |
| ridge: 10y rolling, 1. | 16.9 | 16.8 | 17.9 | 15.8 | 17.0 | 18.4 | 15.2 | 15.6 | 14.9 | 15.2 | 21.5 |
| univariate | 17.3 | 17.8 | 16.7 | 18.9 | 16.1 | 21.8 | 13.9 | 18.0 | 18.1 | 20.1 | 15.2 |
| BVAR(4) with TVP | 18.2 | 18.7 | 15.2 | 22.1 | 17.2 | 26.0 | 11.3 | 16.9 | 17.4 | 20.6 | 19.8 |
| ridge: recursive, 1. | 18.3 | 18.3 | 17.8 | 18.7 | 18.3 | 18.6 | 17.9 | 16.0 | 17.5 | 16.3 | 23.2 |
| PLS, recursive | 18.5 | 20.1 | 20.1 | 20.1 | 15.4 | 18.7 | 21.6 | 20.0 | 21.6 | 19.7 | 19.1 |
| BVAR(4) | 18.6 | 22.0 | 20.8 | 23.2 | 11.8 | 27.0 | 17.0 | 20.3 | 19.6 | 23.9 | 24.1 |
| DVAR(4) | 19.4 | 18.2 | 21.4 | 15.0 | 21.8 | 14.4 | 22.1 | 24.9 | 21.1 | 14.5 | 12.4 |
| PLS, 10y rolling | 19.6 | 20.9 | 19.9 | 21.9 | 17.1 | 19.1 | 22.6 | 25.1 | 20.6 | 19.4 | 18.5 |
| ridge: 10y rolling, .25 | 20.0 | 20.7 | 21.3 | 20.1 | 18.7 | 22.1 | 19.3 | 17.6 | 15.5 | 21.1 | 28.5 |
| ridge: recursive, .25 | 20.9 | 22.0 | 20.3 | 23.8 | 18.7 | 21.7 | 22.4 | 18.3 | 18.4 | 21.8 | 29.6 |
| BMA: BIC | 22.0 | 19.0 | 17.5 | 20.5 | 27.9 | 20.3 | 17.8 | 17.0 | 18.1 | 20.6 | 20.4 |
| VAR(4) | 23.0 | 24.6 | 29.2 | 20.0 | 19.8 | 28.1 | 21.1 | 25.8 | 24.0 | 24.4 | 24.3 |
| PLS, 5y rolling | 25.0 | 24.4 | 21.7 | 27.2 | 26.2 | 22.1 | 26.7 | 26.9 | 26.1 | 23.7 | 21.1 |
| OLS comb. of quartiles, recursive | 25.2 | 22.9 | 22.3 | 23.5 | 29.8 | 23.1 | 22.7 | 20.2 | 24.9 | 22.1 | 24.5 |
| BMA: PIC | 25.4 | 21.9 | 20.4 | 23.5 | 32.3 | 18.8 | 25.0 | 24.4 | 23.7 | 20.9 | 18.7 |
| factor, recursive | 25.9 | 25.2 | 25.1 | 25.3 | 27.2 | 22.0 | 28.4 | 23.3 | 28.3 | 26.9 | 22.1 |
| factor, 10y rolling | 26.7 | 25.2 | 27.1 | 23.2 | 29.8 | 24.6 | 25.8 | 20.3 | 28.1 | 28.8 | 23.6 |
| OLS comb. of quartiles, 10 year rolling | 27.8 | 25.2 | 26.1 | 24.3 | 33.2 | 25.3 | 25.1 | 20.0 | 26.2 | 26.2 | 28.4 |
| BMA: AIC | 29.0 | 26.3 | 26.3 | 26.3 | 34.3 | 23.6 | 29.0 | 27.9 | 28.6 | 26.6 | 22.0 |
| avg. of VAR(4), rolling VAR(4) | 29.2 | 29.8 | 31.8 | 27.9 | 28.0 | 30.0 | 29.7 | 30.3 | 29.9 | 30.4 | 28.8 |
| OLS comb. of quartiles, 5 year rolling | 33.0 | 31.3 | 32.2 | 30.5 | 36.3 | 30.0 | 32.7 | 29.4 | 32.0 | 31.1 | 32.8 |
| ridge: recursive, .001 | 33.7 | 35.6 | 34.3 | 36.9 | 29.9 | 35.9 | 35.3 | 32.6 | 36.3 | 36.4 | 37.1 |
| ridge: 10y rolling, .001 | 34.8 | 36.4 | 36.1 | 36.7 | 31.7 | 36.1 | 36.8 | 35.9 | 37.4 | 36.0 | 36.4 |
| **# of ranking observations** | 144 | 96 | 48 | 48 | 48 | 48 | 48 | 24 | 24 | 24 | 24 |

*Notes:*
1. The table reports average RMSE rankings of the full set of forecast methods or models included in Tables 3-6. The average rankings in the first column of figures are calculated, for each forecast method, across a total of 144 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 $h=2Y$ 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 96 forecasts of just output and inflation, with forecasts of interest rates omitted from the average ranking calculation.
2. See the notes to Table 3.