Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Forecasting with Small Macroeconomic VARs in the Presence of Instabilities

Todd E. Clark and Michael W. McCracken

2007-41

Forecasting with Small Macroeconomic VARs in the Presence of Instabilities *

Todd E. Clark
Federal Reserve Bank of Kansas City

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

November 2006

Abstract

Small–scale VARs are widely used in macroeconomics for forecasting U.S. output, prices, and interest rates. However, recent work suggests these models may exhibit instabilities. As such, a variety of estimation or forecasting methods might be used to improve their forecast accuracy. These include using different observation windows for estimation, intercept correction, time–varying parameters, break dating, Bayesian shrinkage, model averaging, etc. This paper compares the effectiveness of such methods in real time forecasting. We use forecasts from univariate time series models, the Survey of Professional Forecasters and the Federal Reserve Board's Greenbook as benchmarks.

JEL Nos.: C53, E17, E37
Keywords: Real-time data, prediction, structural change

Clark (corresponding author): Economic Research Dept.; Federal Reserve Bank of Kansas City; 925 Grand; Kansas City, MO 64198; todd.e.clark@kc.frb.org. *McCracken*: Board of Governors of the Federal Reserve System; 20th and Constitution N.W.; Mail Stop #61; Washington, D.C. 20551; michael.w.mccracken@frb.gov.

1 Introduction

In this paper we provide empirical evidence on the ability of several different methods to improve the real–time forecast accuracy of small-scale macroeconomic VARs in the presence of potential model instabilities. The 18 distinct trivariate VARs that we consider are each comprised of one of three measures of output, one of three measures of inflation, and one of two measures of short-term interest rates. For each of these models we construct real time forecasts of each variable (with particular emphasis on the output and inflation measures) using real–time data. For each of the 18 variable combinations, we consider 86 different forecasting methods or models, incorporating different choices of lag selection, observation windows used for estimation, levels or differences, intercept corrections, stochastically time–varying parameters, break dating, discounted least squares, Bayesian shrinkage, detrending of inflation and interest rates, and model averaging. We compare our results to those from simple baseline univariate models as well as forecasts from the Survey of Professional Forecasters and the Federal Reserve Board's Greenbook.

We consider this problem to be important for two reasons. The first is simply that small-scale VARs are widely used in macroeconomics. Examples of VARs used to forecast output, prices, and interest rates are numerous, including Sims (1980), Doan, et al. (1984), Litterman (1986), Brayton et al. (1997), Jacobson et al. (2001), Robertson and Tallman (2001), Del Negro and Schorfheide (2004), and Favero and Marcellino (2005). More recently these VARs have been used to model expectations formation in theoretical models. Examples are increasingly common and include Evans and Honkapohja (2005) and Orphanides and Williams (2005).

The second reason is that there is an increasing body of evidence suggesting that these VARs may be prone to instabilities.[1] Examples include Webb (1995), Boivin (1999, 2006), Kozicki and Tinsley (2001b, 2002), and Cogley and Sargent (2001, 2005). Still more studies have examined instabilities in smaller models, such as AR models of inflation or Phillips curve models of inflation. Examples include Stock and Watson (1996, 1999, 2003, 2006),

[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.

Levin and Piger (2003), Roberts (2006), and Clark and McCracken (2006b). 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.

Given the widespread use of small-scale macro VARs and the evidence of instability, it seems important to consider whether any statistical methods for managing structural change might be gainfully used to improve the forecast accuracy of the models. Of course, while structural changes might occur during the forecast horizon, in this paper we focus on the potential for breaks occurring in the estimation sample. Our results indicate that some of the methods do consistently improve forecast accuracy in terms of root mean square errors (RMSE). Not surprisingly, the best method 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 model averaging and Bayesian shrinkage methods consistently perform among the best methods. At the other extreme, the approaches of using a fixed rolling window of observations to estimate model parameters and discounted least squares estimation consistently rank among the worst.

The remainder of the paper proceeds as follows. Section 2 provides a synopsis of the methods used to forecast in the presence of potential structural changes. Section 3 describes the real-time data used as well as specifics on model estimation and evaluation. Section 4 presents our results on forecast accuracy, including rankings of the methods used. Given the large number of models and methods used we provide only a subset of the results in tables and use the text to provide further information. Section 5 concludes. Additional tables can be found in a longer working paper version, Clark and McCracken (2006a).

2 Methods Used

This section describes the various methods we use to construct forecasts from trivariate VARs in the face of potential structural change. Table 1 provides a comprehensive list, with some detail, and the method acronyms we use in presenting results in section 4. For each model — defined as being a baseline VAR in one measure of output (y), one measure of inflation (π), and one short–term interest rate (i) — we apply each of the methods described below. Output is defined as either a growth rate of GDP (or GNP) or an output gap (we defer explanation of the measurement of output and prices to section 3). Unless otherwise noted, once the specifics of the model have been chosen, the parameters of the VAR are estimated using OLS.

We begin with the perhaps naïve method of ignoring structural change. That is, we construct iterated multi-step forecasts from recursively estimated — that is, estimated with all of the data available up to the time of the forecast construction — VARs with fixed lag lengths of 2 and 4. While this approach may seem naïve, it may have benefits. As shown in Clark and McCracken (2005b), depending on the type and magnitude of the structural change, ignoring evidence of structural change can lead to more accurate forecasts. This possibility arises from a simple bias-variance trade-off. While a fixed parameter model is obviously misspecified if breaks have occurred, by using all of the data to estimate the model one might be able to reduce the variance of the parameter estimates enough to more than offset the errors associated with ignoring the coefficient shifts.

A second approach constructs forecasts in much the same way but permits updating of the lag structure as forecasting moves forward. This method, also used in such studies as Stock and Watson (2003), Giacomini and White (2005), and Orphanides and van Norden (2005), permits time variation in the number of lags in the model. We do this four separate ways. The first two consist of using either the AIC or BIC to select the number of model lags in the entire system. In two additional specifications, we allow the lag orders of each variable in each equation to differ (as is done in some of the above studies, as well as Keating (2000)), and use the AIC and BIC to determine the optimal lag combinations.

For each of the above methods, we repeat the process but with at least some of the variables in differences rather than in levels. One reason for taking this approach is based upon the observation that inflation and interest rates are sometimes characterized as being I(1), while each of the output-type variables is generally considered I(0) and hence in the absence of cointegration the predictive equations are likely to be unbalanced. A second is that, as noted in Clements and Hendry (1996), forecasting in differences rather than in levels can provide some protection against mean shifts in the dependent variable. As such, for each model considered above, we construct forecasts based upon two separate collections of the variables: one that keeps the output variable in levels but takes the first difference of the inflation and interest variables (we refer to these models as DVARs) and a second that takes the first difference of all variables (denoted as DVARs with output differenced). See Allen and Fildes (2006) for a recent discussion of forecasting in levels vs. differences.

We also consider select Bayesian forecasting methods. Specifically, we construct forecasts using Bayesian estimates of fixed lag VARs, based on Minnesota–style priors as described in Litterman (1986).[2] We consider both BVARs in "levels" (in y, π, i) and BVARs in partial–differences (in y, $\Delta\pi$, Δi), referring to the latter as BDVARs.

For our particular applications, we generally use prior means of zero for all coefficients, with prior variances that are tighter for longer lags than shorter lags and looser for lags of the dependent variable than for lags of other variables in each equation. However, in setting prior means, in select cases we use values other than zero: in BVARs, the prior means for own first lags of π and i are set at 1; in BVARs with an output gap, the prior mean for the own first lag of y is set at 0.8; and in BVARs with output growth that incorporate an informative prior variance on the intercept, the prior mean for the intercept of the output equation is set to the historical average growth rate.[3] Using the notation of Robertson and Tallman (1999), the prior variances are determined by hyperparameters λ_1 (general tightness), λ_2 (tightness of lags of other variables compared to lags of the dependent variable), λ_3 (tightness of longer lags compared to shorter lags), and λ_4 (tightness of intercept). The prior standard deviation of the coefficient on lag k of variable j in equation j is set to $\frac{\lambda_1}{k^{\lambda_3}}$. The prior standard deviation of the coefficient on lag k of variable m in equation j is $\frac{\lambda_1 \lambda_2}{k^{\lambda_3}} \frac{\sigma_j}{\sigma_m}$, where σ_j and σ_m denote the residual standard deviations of univariate autoregressions estimated for variables j and m. The prior standard deviation of the intercept in equation j is set to $\lambda_4 \sigma_j$. In our BVARs and BDVARs, we use generally conventional hyperparameter settings of $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, and $\lambda_4 = 1000$ (making the intercept prior flat).

Another common approach to estimating predictive models in the presence of structural change consists of using a rolling window of the most recent N $(N < t)$ observations to estimate the model parameters. The logic behind this approach is that for models exhibiting structural change, older observations are less likely to be relevant for the present incarnation of the DGP. In particular, using older observations implies a type of model misspecification (and perhaps bias in the forecasts) that can be alleviated by simply dropping those observations. We implement this methodology, recently advocated in Giacomini and

[2] We estimate the models with the common mixed approach applied on an equation–by–equation basis. As indicated in Geweke and Whiteman (2006), estimating the system of equations with the same Minnesota priors would require Monte Carlo simulation.

[3] In model estimates for vintage t, used for forecasting in period t and beyond, the average is calculated using data from the beginning of the available sample through period $t-1$ — data that would have been available to the forecaster at that time.

White (2005), for each of the above methods using a constant window of the past $N = 60$ quarters of observations to estimate the model parameters. Of course, it is possible that using a sample window based on break test estimates could yield better model estimates and forecasts. In practice, however, difficulties in identifying breaks and their timing may rule out such improvements (see, for example, the results in Clark and McCracken (2005b)).

While the logic behind the rolling windows approach has its appeal, it might be considered a bit extreme in its dropping of older observations. That is, while older observations might be less relevant for the present incarnation of the DGP, they may not be completely irrelevant. A less extreme approach would be to use discounted least squares (DLS) to estimate the model parameters. This method uses all of the data to estimate the model parameters but weights the observations by a factor λ^{t-j}, $0 < \lambda < 1$, that places full weight on the most recent observation ($j = t$) but gradually shrinks the weights to zero for older observations ($j < t$). While this methodology is less common in economic forecasting than is the rolling scheme, recent work by Stock and Watson (2004) and Branch and Evans (2006) suggests it might work well for macroeconomic forecasting. With this in mind we consider four separate models estimated by DLS. The first two are the baseline VARs in y, π, i and DVARs in y, $\Delta\pi$, Δi with a fixed number of lags. The second two are VARs and DVARs with the number of model lags estimated using the AIC for the system. Our setting of the discount factor roughly matches the suggestions of Branch and Evans (2006): .99 for the output equation and .95 for the inflation and interest rate equations.

Despite the appeal of both the rolling and DLS methods, one drawback they share is that they reduce the (effective) number of observations used to estimate each of the model parameters regardless of whether they have exhibited any significant structural change. There are any number of ways to avoid this problem. One would be to attempt to identify structural change in every variable in each equation. To do so one could use any number of approaches, including those proposed in Andrews (1993), Bai and Perron (1998, 2003), and many others. However, in the context of VARs (for which there are numerous parameters), these tests can be poorly sized and exhibit low power, particularly in samples of the size often observed when working with quarterly macroeconomic data. This is precisely the conclusion reached by Boivin (1999). Instead, in light of the importance of mean shifts highlighted in such studies as Clements and Hendry (1996), Kozicki and Tinsley (2001a,b), and Levin and Piger (2003), we focus attention on identifying structural change in the intercepts of the model.

To capture potential structural change in the intercepts, we consider several different methods of what might loosely be called 'intercept corrections'. The most straightforward is to use pretesting procedures to identify shifts in the intercepts, introduce dummy variables to capture those shifts, estimate the augmented model and proceed to forecasting. In particular, we follow Yao (1988) and Bai and Perron (1998, 2003) in using information criteria to identify break dates associated with the model intercepts. Specifically, at each forecast origin we first choose the number of lags in the system using the AIC and then use an information criterion to select up to two structural breaks in the set of model intercepts. For computational tractability, we use a simple sequential approach — a partial version of Bai's (1997) sequential method — to identifying multiple breaks. We first use the information criterion to determine if one break has occurred. If the criterion identifies one break, we then search for a second break that occurred between the time of the first break and the end of the sample.[4] The model with up to two intercept breaks is then estimated by OLS and used to forecast. We use two such models, one with breaks identified by the AIC and a second with breaks identified using the BIC.

While this approach might prove useful for identifying structural change in the interior of the sample, it is likely to be less well behaved when the structural change occurs at the very end of the sample.[5] Motivated by this observation, Clements and Hendry (1996) discuss several approaches to 'correcting' intercepts for structural change occurring at the very end of the sample. The approach we implement is directly related to one of theirs. Specifically, the intercept correction consists of adding the average of the past 4 residuals to the model (for each equation) at each step across the forecast horizon. Equivalently, the forecast is constructed by adding a weighted average of the past 4 residuals (with weights that depend upon the parameters of the VAR and the forecast horizon) to the baseline forecast that ignores any structural change.[6] We apply intercept correction to four different VAR systems. Two of the systems use a fixed lag order, and the other two use a lag order determined by applying AIC to the system. For each of these two baseline lag orders, we then construct intercept corrections once for the entire system of three equations and once making adjustments to only the inflation and interest rate equations.

Our final variant of intercept correction draws on the approach developed by Kozicki

[4] In the break identification, we impose a minimum segment length of 16 quarters.

[5] We leave as a topic for future research the possibility that methods designed to identify breaks at the end of a sample, such as those of Hendry, et al. (2004) and Andrews (2006), could yield better results.

[6] See equation (40) of Clements and Hendry (1996) for details.

and Tinsley (2001a,b). In their 'moving endpoints' structure, the baseline VAR is modeled as having time varying intercepts that allow continuous variation in the long run expectations of the corresponding variables. Our precise method, though, is perhaps more closely related to Kozicki and Tinsley (2002).[7] In the context of a small-scale macro VAR, the variables in their model are modeled as deviations from latent time varying steady states (trends). However, whereas they use the Kalman filter to extract estimates of this unknown trend, for tractability we use simple exponential smoothing methods to get estimates. Cogley (2002) develops a model in which exponential smoothing provides an estimate of a time–varying inflation target of the central bank, a target that the public doesn't observe but does learn about over time. With exponential smoothing, the trend estimate can be easily constructed in real time and updated over the multi–step forecast horizon to reflect forecasts of inflation. As indicated in Figure 1, exponential smoothing yields a trend estimate quite similar to an estimate of long–run inflation expectations based on 1981-2005 data from the Hoey survey of financial market participants and the Survey of Professional Forecasters (for a 10–year ahead forecast of CPI inflation) and 1960-1981 estimates of long–run inflation expectations developed by Kozicki and Tinsley (2001a). We construct two different sets of forecasts using the exponential smoothing approach.[8] Following Kozicki and Tinsley (2001b, 2002), in the first we use our exponentially smoothed inflation series to detrend both inflation and the interest rate measure. In the second we detrend the inflation and interest rate series separately. In either case we do not detrend the output variable.

Another approach to managing structural change in model parameters is to integrate the structural change directly into the VAR.[9] Following Doan, et al. (1984) and more recent

[7] In some supplemental analysis, we have considered models of the error correction form used in, among others, Brayton, et al. (1997) and Kozicki and Tinsley (2001b). These models relate y_t, $\Delta\pi_t$, and Δi_t to lags and error correction terms $\pi_{t-1} - \pi^*_{t-1}$ and $i_{t-1} - \pi^*_{t-1}$, where π^* denotes trend inflation (long–run expected inflation). We estimated the models with fixed lags of 2 and 4 and with Bayesian methods using a fixed lag of 4 (and flat priors on the error correction coefficients). We also considered Bayesian estimates of our VAR with inflation detrending. None of these methods proved to consistently beat the forecast accuracy of the best performing methods we describe below. For the applications covered in Tables 2-5, all of these supplemental methods delivered average RMSE ratios (corresponding to the averages in Table 7) above 1.000.

[8] We use a smoothing parameter of .07 for the interest rate and core PCE inflation series and a smoothing parameter of .05 for the GDP and CPI inflation series. Each trend was initialized using the sample mean of the first 20 observations available (since 1947) from the present vintage.

[9] As noted in Doan, et al. (1984), proper multi-step forecasting with VARs with TVP would involve taking into account the joint distribution of the residuals in the VAR equations and the coefficient equations. In light of the difficulty of doing so, we follow conventional practice and treat the coefficients as fixed at their period $t-1$ values for forecasting in periods t and beyond.

work by Brainard and Perry (2000) and Cogley and Sargent (2001, 2005), we model the structural change in the parameters of a VAR in y, π, i with random walks.[10] However, in light of the potentially adverse effects of parameter estimation noise on forecast accuracy and the potentially unique importance of time variation in intercepts (see above), we consider two different scopes of parameter change. In the first we allow time variation in all coefficients — both the model intercepts and slope coefficients. In the second, we allow for stochastic variation in only the intercepts.[11]

We estimate each of these TVP specifications using Bayesian methods with a range of prior variances on the standard deviation of the intercepts and a range of allowed time variation in the parameters. In some cases we use informative priors on the intercepts (λ_4 = .5 or .1); in others we use flat priors (λ_4 = 1000). The variance–covariance matrix of the innovations in the random walk processes followed by the coefficients is set to λ times the prior variance of the matrix of coefficients, which is governed by the hyperparameters described above. Drawing on the settings used in such studies as Stock and Watson (1996) and Cogley and Sargent (2001), we consider λ values ranging from .0001 to .005. Note, however, that in those instances in which the intercept prior is flat, we follow Doan, et al. (1984) in setting the variance of the innovation in the intercept at λ times the prior variance of the coefficient on the own first lag instead of the prior variance of the constant. In the baseline TVP model, we use λ_4 = .1 and λ = .0005.

The final group of methods we consider all consist of some form of model averaging. While model averaging as a means of managing structural change has its historical precedents — notably Min and Zellner (1993) — the approach has become even more prevalent in the past several years. Recent examples of studies incorporating model averaging include Koop and Potter (2003), Stock and Watson (2003), Clements and Hendry (2004), Maheu and Gordon (2004), and Pesaran, et al. (2006). We consider six distinct, simple forms of model averaging, in each case using equal weights.[12] The first takes an average of all the VAR forecasts described above and the univariate forecast described below, for a given triplet of variables. More specifically, for a given combination of measures of output,

[10] Some other studies, such as Canova (2002), impose stationarity on the coefficient time variation.

[11] Allowing both the inflation and interest rate equations to have intercepts with TVP implies a non–stationary real interest rate. While some readers might prefer specifications that impose stationarity in the real interest rate, our specifications are consistent with evidence in such studies as Laubach and Williams (2003) and Clark and Kozicki (2005) on non–stationarities in real interest rates.

[12] In doing so, we leave as a topic for future research whether more sophisticated approaches to averaging, such as approaches based on historical accuracy, would yield improvements.

inflation, and an interest rate (for example, for the combination GDP growth, GDP inflation, and the T-bill rate), we construct a total of 75 different forecasts from the alternative VAR models described above. We then average these forecasts with a univariate forecast.

We include a second average forecast approach motivated by the results of Clark and McCracken (2005b), who show that the bias-variance trade-off can be managed to produce a lower MSE by combining forecasts from a recursively estimated VAR and a VAR estimated with a rolling sample. In the results we present here, for a given baseline fixed lag VAR we take an equally weighted average of the model forecast constructed using parameters estimated recursively (with all of the available data) with those estimated using a rolling window of the past 60 observations. Two other averages are motivated by the Clark and McCracken (2005a) finding that combining forecasts from nested models can improve forecast accuracy. In this paper, we consider an average of the univariate forecast described below with the fixed lag VAR forecast, and an average of the univariate forecast with the fixed lag DVAR forecast. Finally, motivated in part by general evidence of the benefits of averaging, we consider two other averages of the univariate forecasts with some of the other forecasts that prove to be relatively good. One is a simple average of the univariate forecast with the forecast of the VAR with inflation detrending. The other is a simple average of the univariate and fixed lag VAR, DVAR, and baseline BVAR with time varying parameters.

To evaluate the practical value of all these methods, we compare the accuracy of the above VAR–based forecasts against various benchmarks. In light of common practice in forecasting research, we use forecasts from univariate time series models as one set of benchmarks.[13] For output, widely modeled as following low-order AR processes, the univariate model is an AR(2). In the case of inflation, we use a benchmark suggested by Stock and Watson (2006): an MA(1) process for the change in inflation ($\Delta\pi$), estimated with a rolling window of 40 observations. Stock and Watson find that the IMA(1) generally outperforms a random walk or AR model forecasts of inflation. For simplicity, in light of some general similarities in the time series properties of inflation and short–term interest rates and the IMA(1) rationale for inflation described by Stock and Watson, the univariate benchmark for the short-term interest rate is also specified as an MA(1) in the first differ-

[13]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.

ence of the series (Δi). As described in section 4, we use the bootstrap methods of White (2000) and Hansen (2005) to determine the statistical significance of any improvements in VAR forecast accuracy relative to the univariate benchmark models. In light of our real time forecasting focus, we also include as benchmarks forecasts of growth, inflation, and interest rates from the Survey of Professional Forecasters (SPF) and forecasts of growth and inflation from the Federal Reserve Board's Greenbook.

3 Data and Model details

As noted above, we consider the real–time forecast performance of VARs with three different measures of output, three measures of inflation, and two short–term interest rates. The output measures are GDP or GNP (depending on data vintage) growth, an output gap computed with the method described in Hallman, et al. (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 very 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), CPI, and PCE price index excluding food and energy (hereafter, core PCE price index).[14] The short–term interest rate is measured as either a 3–month Treasury bill rate or the effective federal funds rate. 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 a range of sources: the Federal Reserve Bank of Philadelphia's Real–Time Data Set for Macroeconomists (RTDSM), the Board of Governor's FAME database, the website of the Bureau of Labor Statistics (BLS), the Federal Reserve Bank of St. Louis' ALFRED database, and

[14] As the univariate forecast results suggest, these competing price indices have somewhat different characteristics. Differences appear to persist over long periods of time: there is little evidence of cointegration among these and related price indexes (see, for example, Lebow, Roberts, and Stockton (1992)).

various issues of the *Survey of Current Business*. 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. For the core PCE price index, we compile a real time data set starting with the 1996:Q1 vintage by combining information from the Federal Reserve Bank of St. Louis' ALFRED database (which provides vintages from 1999:Q3 through the present) with prior vintage data obtained from issues of the *Survey of Current Business*, following the RTDSM dating conventions.[15] Because the BEA only begin publishing the core PCE series with the 1996:Q1 vintage, it is not possible to extend the real time data set further back in history with just information from the *Survey of Current Business*.

In the case of the CPI and the interest rates, for which real time revisions are small to essentially non–existent (see, for example, Kozicki (2004)), we simply abstract from real time aspects of the data. For the CPI, we follow the advice of Kozicki and Hoffman (2004) for avoiding choppiness in inflation rates for the 1960s and 1970s due to changes in index bases, and use a 1967 base year series taken from the BLS website in late 2005.[16] For the T-bill rate, we use a series obtained from FAME.

The full forecast evaluation period runs from 1970:Q1 through 2005; we use real time data vintages from 1970:Q1 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 $t-1$.[17] 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 samples of the historical data made available by the BEA. For each forecast origin t in 1970:Q1 through 2005:Q3, we use the real time data vintage t to estimate output gaps, estimate the forecast models, and then construct forecasts for periods t and beyond. The starting point of the model estimation sample is the maximum of 1955:Q1 and the earliest quarter in which all of the data included in a given model are available, plus the number of lags included in the model (plus one

[15]In putting together vintages for 1996:Q1 through 1999:Q2, we also relied on a couple of full time series we had on file from prior research, series that correspond to the vintages for 1996:Q4 and 1999:Q2, obtained from FAME at the time of the research projects.

[16]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.

[17]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.

quarter for DVARs or VARs with inflation detrending).

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

As the forecast horizon increases beyond a year, forecasts are increasingly determined by the unconditional means implied by a model. As highlighted by Kozicki and Tinsley (1998, 2001a,b), these unconditional means — or, in the Kozicki and Tinsley terminology, endpoints — may vary over time. The accuracy of long horizon forecasts (two or three years ahead, for example) depend importantly on the accuracy of the model's endpoints. As a result, we examine simple measures of the endpoints implied by real time, 1970-2005 estimates of a select subset of the forecasting models described above. For simplicity, we use 10–year ahead forecasts (forecasts for period $t+39$ made with vintage t data ending in period $t-1$) as proxies for the endpoints.

We obtained benchmark SPF forecasts of growth, inflation, and interest rates from the website of the Federal Reserve Bank of Philadelphia.[18] The available forecasts of GDP/GNP growth and inflation span our full 1970 to 2005 sample. The SPF forecasts of CPI inflation and the 3-month Treasury bill rate begin in 1981:Q3. Our benchmark Greenbook forecasts of GDP/GNP growth and inflation and CPI inflation are taken from data on the Federal Reserve Bank of Philadelphia's website and data compiled by Peter Tulip (some of the data are used in Tulip (2005)). We take 1970-99 vintage Greenbook forecasts of GDP/GNP growth and GDP/GNP inflation from the Philadelphia Fed's data set.[19] Forecasts of GDP growth and inflation for 2000 are calculated from Tulip's data set. Finally, we take 1979:Q4–2000:Q4 vintage Greenbook forecasts of CPI inflation from Tulip's data

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

[19]We derived 1–year ahead forecasts of growth and inflation by compounding the reported quarterly percent changes.

set.[20]

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. The GDP data available today for, say, 1970, represent the best available estimates of output in 1970. However, output as defined today is quite different from the definition of output in 1970. For example, today we have available chain weighted GDP; in the 1970s, output was measured with fixed weight GNP. Forecasters in 1970 could not have foreseen such changes and the potential impact on measured output. Accordingly, in our baseline results, we use the first available estimates of GDP/GNP and the GDP/GNP deflator in evaluating forecast accuracy. In particular, we define the actual value to be the first estimate available in subsequent vintages. In the case of h–step ahead (for h = 0, 1, and 4) forecasts made for period $t+h$ with vintage t data ending in period $t-1$, the first available estimate is normally taken from the vintage $t+h+1$ data set. In light of our abstraction from real time revisions in CPI inflation and interest rates, the real time data correspond to the final vintage data. In Clark and McCracken (2006a) we provide supplementary results using final vintage (2005:Q4 vintage) data as actuals. Our qualitative results remain broadly unchanged with the use of final vintage data as actuals.

4 Results

In evaluating the performance of the forecasting methods described above, we follow Stock and Watson (1996, 2003, 2006), among others, in using squared error to evaluate accuracy and considering forecast performance over multiple samples. Specifically, we measure accuracy with root mean square error (RMSE). The forecast samples are generally specified as 1970-84 and 1985-2005 (the latter sample is shortened to 1985-2000 in comparisons to Greenbook forecasts, for which publicly available data end in 2000).[21] We split the full sample in this way to ensure our general findings are robust across sample periods, in light of the evidence in Stock and Watson (2003) and others of instabilities in forecast performance over time. However, because real time data on core PCE inflation only begin

[20] Year–ahead CPI forecasts were obtained by compounding the Greenbook's quarterly percent changes.

[21] 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.

in 1996, we also present select results for a forecast sample of 1996-2005.[22]

To be able to provide broad, robust results, in total we consider a very large number of models and methods — far too many to be able to present all details of the results. Instead we use the full set of models and methods in providing only a high–level summary of the results, primarily in the form of rankings described below. In addition, we limit the presentation of detailed results to those models and variable combinations of perhaps broadest interest and note in the discussion those instances in which results differ for other specifications. Specifically, in presenting detailed results, we draw the following limitations. (1) For the most part, accuracy results are presented for just output and inflation. (2) Output is measured with either GDP/GNP growth or the HPS gap. (3) The interest rate is measured with the 3-month Treasury bill rate. We provide results for models using the federal funds rate — results qualitatively similar to those we report in the paper — in supplemental tables in Clark and McCracken (2006a). (4) The set of forecast models or methods is limited to a subset we consider to be of the broadest interest or representative of the others. For example, while we consider models estimated with a fixed number of either 2 or 4 lags, we report RMSEs associated only with those that have 4 lags.

We proceed below by first presenting forecast accuracy results based on univariate and VAR models. We then compare results for some of the better–performing methods to the accuracy of SPF and Greenbook forecasts. We conclude by examining the real–time, long–run forecasts generated by a subset of the forecast methods that yield relatively accurate short–run forecasts.

4.1 Forecast accuracy

Tables 2 through 5 report forecast accuracy (RMSE) results for four combinations of output (GDP growth and HPS gap) and inflation (GDP price index and CPI) and 27 models. In each case we use the 3-month Treasury bill as the interest rate. 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.

To determine the statistical significance of differences in forecast accuracy, we use a

[22]Specifically, the forecast sample is 1996:Q1+h to to 2005:Q3 (for forecasts dated by the end of the forecast horizon).

non–parametric bootstrap patterned after White's (2000) to calculate p–values for each RMSE ratio in Tables 2-5. 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 level are indicated with a *slanted* font. To determine whether 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_c 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_c test will typically have much greater power than White's (2000) reality check test. For each column, if the SPA_c test yields a p–value of 10 percent or less, we report the associated RMSE ratio in bold font. Because the SPA_c 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_c may not be the forecast with the lowest RMSE ratio.[23]

We implement the bootstrap procedures by sampling from the time series of forecast errors underlying the entries in Tables 2-5. 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); White notes that the moving block is also asymptotically valid. The bootstrap is applied separately for each forecast horizon, using a block size of 1 for the $h = 0Q$ forecasts, 2 for $h = 1Q$, and 5 for $h = 1Y$.[24] 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.

While there are many nuances in the detailed results, some clear patterns emerge. The RMSEs clearly show the reduced volatility of the economy since the early 1980s, particularly for output. For each horizon, the benchmark univariate RMSE of GDP growth forecasts declined by roughly 2/3 across the 1970-84 and 1985-05 samples; the benchmark RMSE for HPS gap forecasts declined by about 1/2. The reduced volatility is less extreme for the inflation measures but still evident. For each horizon, the benchmark RMSEs fell by roughly 1/2 across the two periods, with the exception that at the $h = 1Y$ horizon the

[23]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.

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

variability in GDP inflation declined nearly 2/3.

Consistent with the results in Campbell (2006), D'Agostino, et al. (2005), Stock and Watson (2006), and Tulip (2005), there is also a clear decline in the predictability of both output and inflation: it has become harder to beat the accuracy of a univariate forecast. For example, for each forecast horizon, a number of methods or models beat the accuracy of the univariate forecast of GDP growth during the 1970-84 period (Tables 2 and 4). In fact, many of these do so at a level that is statistically significant. But over the 1985-2005 period, only the BVAR(4)-TVP models are more accurate, at only the 1–year ahead horizon. The reduction in predictability is almost, but not quite, as extreme for the HPS output gap (Tables 3 and 5). While several models perform significantly better than the benchmark in the 1970-84 period, only two classes of methods, the BDVARs and the BVAR-TVPs, significantly outperform the benchmark in the 1985-05 period.

The predictability of inflation has also declined, although less dramatically than for output. For example, in models with GDP growth and GDP inflation (Table 2), 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 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 both Tables 4 and 5, most of the models convincingly outperform the univariate benchmark during the 1970-84 period, with statistically significant maximal gains of 18%. But in the following period, many fewer methods outperform the benchmark, with gains typically about 4%.

Reflecting the decline in predictability, many of the methods that perform well over 1970-84 fare much more poorly over 1985-05. The instabilities in performance are clearly evident in both output and inflation forecasts, but more dramatic for output forecasts. For example, a VAR with AIC determined lags and intercept breaks (denoted VAR(AIC), intercept breaks) forecasts both GDP growth and the HPS gap well in the 1970-84 period, with gains as large as 25% for 1–year ahead forecasts of the HPS gap. However, in the 1985-05 period, the VAR with intercept breaks ranks among the worst performers, yielding 1–year ahead output forecasts with RMSEs 60 percent higher than the univariate forecast RMSEs. In the case of inflation forecasts, a DVAR(4) estimated with Bayesian methods and a rolling sample of data (denoted BDVAR(4)) beats the benchmark, by as much as 13 percent, at every horizon during the 1970-84 period. But in the 1985-05 period, the BDVAR(4) is always beaten by the univariate benchmark model, by as much as 21%.

The change in predictability makes it difficult to identify methods that consistently improve upon the forecast accuracy of univariate benchmarks. As noted above, none of the methods consistently improve upon the GDP growth benchmark across the subperiods. For forecasts of the HPS gap, the BVAR(4)-TVP models generally outperform the benchmark over both periods. However, the 1970-84 gains are not statistically significant. In the case of inflation forecasts, though, a number of the forecasts significantly outperform the univariate benchmark in both samples. Of particular note are the forecasts that average the benchmark univariate projection with a VAR projection — either the VAR(4), DVAR(4), or VAR(4) with inflation detrending — and the average of the univariate forecast with (together) the VAR(4), DVAR(4), and TVP BVAR(4) projections. In the 1970-84 period, these averages nearly always outperform the benchmark, although without necessarily being the best performer. In the 1985-05 period, the averages continue to outperform the benchmark and are frequently among the best performers.

In Tables 6 and 7 we take another approach to determining which methods tend to perform better than the benchmark. Across each variable, model and horizon, we compute the average rank and RMSE ratio of the methods included in Tables 2-5, as well as the corresponding sample standard deviations. For example, the figures in Table 6 are obtained by: (1) ranking, for each of the 48 columns of Tables 2-5, the 27 forecast methods or models considered; and (2) calculating the average and standard deviation of each method's (48) ranks. Table 7 does the same, but using RMSEs instead of RMSE ranks. The averages in Tables 6 and 7 show that, from a broad perspective, the best forecasts are those obtained as averages across models. The best forecast, an average of the univariate benchmark with the VAR(4) with inflation detrending, has an average RMSE ratio of .943 in Tables 2-5, and an average rank of 5.1. Not surprisingly, orderings based on average RMSE ratios are closely correlated with orderings based on the average rankings. For instance, the top eight forecasts based on average rankings are the same as the top eight based on average RMSE ratios, with slight differences in orderings.

Tables 6 and 7 also show that some VAR methods consistently perform worse — much worse, in some cases — than the univariate benchmark. The univariate forecasts have the 9th best average RMSE ratio and 11th best average ranking. Thus, on average, roughly 2/3 of the VAR methods fail to beat the univariate benchmark. Moreover, some of the methods designed to overcome the difficulty of forecasting in the presence of structural change consistently rank among the worst forecasts. Most notably, VAR forecasts based on intercept corrections and DLS estimates are generally among the worst forecasts considered, yielding RMSE ratios that, on average, exceed the univariate benchmark by roughly 15 percent (we acknowledge, however, that under different implementations, the performance of these methods could improve — we leave such analysis for future research).[25] VARs estimated with rolling samples of data also perform relatively poorly: in every case, a VAR estimated with a rolling sample is, on average, less accurate than when estimated (recursively) with the full sample. In contrast, on average, standard Bayesian estimation of VARs generally dominates OLS estimation of the corresponding model. For example, the average RMSE ratio of the BVAR(4) forecast is 1.012, compared to the average VAR(4) RMSE ratio of 1.030.

Tables 8-11 report RMSE results for models including core PCE inflation. As noted above, reflecting the real time core PCE data availability, the forecast sample is limited to 1996-05. As in Tables 2-5, we report results for models with two different measures of output, GDP growth and the HPS output gap, but a single interest rate measure, the Treasury bill rate. For comparison, we also report 1996-05 results for models using GDP inflation instead of core PCE inflation. As in the case of the results for 1970-84 and 1985-05, we use White (2000) and Hansen (2005) bootstraps to determine whether any of the RMSE ratios are significantly less than one, on both a pairwise (given model against univariate) and best–in–column basis. Individual RMSE ratios that are significantly less than 1 (10% confidence level) are indicated with a *slanted* font. Note, though, that once the search involved in selecting a best forecast is taken into account, the univariate model is never beaten in the 1996-05 results (that is, none of the data snooping–robust p–values are less than .10).

Consistent with the 1985-05 results in Tables 2-5, the forecast results for 1996-05 in Tables 8-11 show that univariate benchmarks are difficult to beat. Of the inflation measures, the benchmark is harder to beat with core PCE inflation than with GDP inflation. For 1996-05, only a few forecasts (e.g., rolling VAR(4) or DVAR(4) forecasts for $h = 0Q$) beat the univariate benchmark, and none statistically significantly. A few more forecasts are able to improve (some statistically significantly) on the accuracy of the univariate benchmark for GDP inflation. Importantly, for models with GDP inflation, those methods that per-

[25] In our results, intercept corrections don't seem to work with either GDP growth or output gaps. In the case of gaps, however, the persistence and measurement error inherent in them may warrant other approaches to intercept correction.

formed relatively well over the samples covered in Tables 2-5 — such as the averages of the benchmarks with the VAR(4) or DVAR(4) models — also perform relatively well over the 1996-05 sample.

Tables 12 and 13 provide aggregate or summary information on the forecast performance of all the methods and nearly all of the data combinations considered. The summary information covers all of the variable combinations and models included in Tables 2-5, as well as variable combinations that include the HP measure of the output gap or the federal funds rate as the interest rate, models based on a fixed lag of two instead of four, and the full set of forecasting methods described in section 2 and listed in Table 1. Our summary approach follows the ranking methodology of Tables 6 and 7. That is, in Tables 12 and 13 we present average rankings for every method we consider across every forecast horizon, various subclasses of models, and the 1970-84 and 1985-05 samples. Note, however, that we exclude the 1996-05 sample (and, as a result, results from models including core PCE inflation), in part because of its overlap with the longer 1985-05 period.

While expanding coverage to all possible models and methods generates some additional nuances in results, the broad findings described above continue to hold. As shown in Table 12's first column of ranks, across all combinations of variables the most robust forecasting methods are those that average the univariate model with one or a few VAR forecasts. For example, the average of the univariate forecast with a forecast from a VAR(2) with inflation detrending has the best average ranking, of 12.9 (and the best average RMSE ratio, not reported, of 0.937). Coming in behind these averaging methods, in the broad ranking perspective, are the fixed lag BVAR, BDVAR and BVAR-TVP models. Note that the first column includes interest rate forecast results — which were omitted from previous tables for brevity. The same classes of models that on average performed best for the output and inflation series continue to perform among the best for interest rate forecasts (and is another reason why we felt comfortable omitting those results). Somewhat more formally, the Spearman rank correlation across the results in the first and second columns of Table 12 — the second of which contains the ranks of just the output and inflation forecasts — is a robust 0.97.

Columns 3 and 4 of Table 12 delineate the average impact of the choice of interest rate on forecast accuracy for the output and inflation measures. The rankings are extremely similar. The five best methods for forecasting output and inflation in models with the T-bill rate are also the five best methods in models with the federal funds rate. Moreover, the Spearman rank correlation of the results conditioned on the T-bill rate and the results conditioned on the federal funds rate is 0.98. We should emphasize that this does not imply that there weren't differences in the nominal outcomes across these two interest rate measures. Rather, in light of our goal to identify those methods that are most robust in forecasting, the choice between the T-bill and federal funds rates makes little difference.

Columns 1-3 in Table 13 delineate the average impact of the choice of output measure in forecasts of output and inflation. These rankings are quite similar across output measures, although not quite as similar as those comparing the impact of the interest rate measures. In each case the best methods generally continue to be averages of univariate benchmarks with VAR forecasts and BVARs with TVP. For example, in models with GDP growth, on average the best forecasts of output and inflation are obtained with an average of the univariate, VAR(4), DVAR(4), and TVP BVAR(4) forecasts. Perhaps the largest distinction among the three sets of rankings is that moving from GDP growth to HPS gap to HP gap, the concentration of best methods shifts from the averaging group to the BVAR-TVP with tight intercept priors group to the BVAR-TVP with loose intercept priors group. Even so, the rank correlations among the three columns are very high, between 0.85 and 0.93.

Similarly, columns 4 and 5 of Table 13 provide average rankings of forecasts for output and inflation that condition on the inflation measure, GDP inflation or CPI inflation. Again, the top performing methods remain the averages of univariate forecasts with select VAR forecasts and BVAR TVP forecasts. And, the results are very similar across inflation measures. In the average rankings, the top seven methods for models including GDP inflation are the same as the top seven for models including CPI inflation, with slight differences in orderings. The rank correlation across all methods is 0.94.

The last two columns of Table 12 compare the performance of methods across the 1970-84 and 1985-05 periods. As in the above detailed comparisons of a subset of results, across the two subperiods there are some sharp differences in the performance of many of even the better performing methods.[26] Only four methods have an average ranking of less than 20 over the 1970-84 period (in order from smallest to largest): the average of all forecasts, the average of the univariate and VAR(4) with inflation detrending forecasts, the VAR(2) with full exponential smoothing detrending, and the average of the univariate, VAR(4),

[26] In addition, the average RMSE ratios (not reported) associated with each of the top–performing methods reflect the sharp reduction in predictability in 1985-05 compared to 1970-84. The best average RMSE ratio for 1970-84 is 0.873, from a VAR(2) with full exponential smoothing. The best average RMSE ratio for 1985-05 is 0.998, for the baseline TVP BVAR(4).

DVAR(4), and TVP BVAR(4) forecasts. For the 1985-05 sample, a total of 11 methods have average rankings below 20, but only two of them — the average of the univariate and VAR(4) with inflation detrending forecasts and the average of the univariate, VAR(4), DVAR(4), and TVP BVAR(4) forecasts — correspond to the four methods that produce average rankings of less than 20 in the 1970-84 sample. Some of the models that perform relatively well in 1970-84 fare much more poorly in the second sample. For example, the average ranking of the VAR(2) with full exponential smoothing detrending plummets from 17.7 in 1970-84 to 63.9 in 1985-05. Not surprisingly, the rank correlation between these two columns is relatively low, at just 0.58.

In Clark and McCracken (2006a) we provide still more detailed information on which methods work the best individually for forecasting each output measure and the GDP and CPI measures of inflation. Perhaps not surprisingly, this further disaggregation of the results leads to modestly more heterogeneity in rankings of the best methods. This is particularly true for output forecast rankings compared to inflation rankings. For example, a DVAR with AIC–determined lags has an average ranking of 15.4 in forecasts of GDP inflation and an average ranking of 48.5 in forecasts of GDP growth. The Spearman correlations of output rankings with inflation rankings range from 0.46 (for GDP growth and CPI inflation) to 0.57 (for the HPS gap and CPI inflation). By comparison, the correlations of output forecast rankings across measures of output average 0.7, while the correlation for GDP and CPI inflation rankings is 0.86. Despite the greater heterogeneity of these more disaggregate rankings, there are similarities among the best performers. Among the output variables, on average, the best forecasts are typically the averages of univariate forecasts with VAR forecasts and the BVAR-TVP forecasts. For the two inflation measures, the averaging methods continue to perform the best, followed by BVAR-TVP and DVAR forecasts.

Just as Tables 12 and 13 provide aggregate evidence on the best methods, they also show what methods consistently perform the worse in the full set of models, methods, and horizons. Perhaps most simply, not a single method on the second pages of the tables has an average rank less than 20! Consistent with the subset of results summarized in Tables 6 and 7, the lowest–ranked methods include: DLS estimation of VARs or DVARs, DVARs with output, in addition to inflation and the interest rate, differenced; and VARs with intercept correction. The consistency of the rankings for these worst–performing methods may be considered impressive. In addition, the average rankings of forecasts based on rolling estimation of VARs (and DVARs, BVARs, etc.) are generally considerably lower than the average rankings of the corresponding VARs estimated with the full sample of data. For example, the average ranking of rolling DVAR(2) forecasts is 41.2, compared to the recursively estimated DVAR(2)'s average ranking of 32.8. While those methods generally falling in the middle ranks (between an average rank of, say, 20 and 50) may not be considered robust approaches to forecasting with the VARs of interest, in particular instances some of these methods may perform relatively well. For example, the DVAR with AIC lags determined for each equation has an average ranking of 39.4, but yields relatively accurate forecasts of GDP inflation in 1985-05 (Tables 2 and 4).

Table 14 compares the accuracy of some of the better time series forecasting methods with the accuracy of SPF projections. The variables we report are those for which SPF forecasts exist: GDP growth, GDP inflation, and CPI inflation (in the case of CPI inflation, the SPF forecasts don't begin until 1981, so we only report CPI results for the 1985-05 period). We also report results for forecasts of the T-bill rate from the SPF and the selected time series models. In particular, Table 14 provides, for the 1970-84 and 1985-05 samples, RMSEs for forecasts from the SPF and a select set of the better–performing time series forecasts: the best forecast RMSE for each variable in each period from those methods included in Table 2 (Table 4 for CPI inflation forecasts), the univariate benchmark forecast, several of the average forecasts, and the baseline TVP BVAR(4). To be sure, comparing forecasts from a source such as SPF against the best forecast from Table 2 or 4 gives the time series models an unrealistic advantage, in that, in real time, a forecaster wouldn't know which method is most accurate. However, as the results presented below make clear, our general findings apply to all of the individual forecasts included in the comparison.

Perhaps not surprisingly, the SPF forecasts generally dominate the time series model forecasts. For example, in $h = 0Q$ forecasts of GDP growth for 1970-84, the RMSE for the SPF is 2.571, compared to the best time series RMSE of 3.735 (in which case the best forecast is the all forecast average reported in Table 2). In $h = 0Q$ forecasts of GDP inflation for 1970-84, the RMSE for the SPF is 1.364, compared to the best time series RMSE of 1.727 (again, from the all–forecast average in Table 2). At such short horizons, of course, the SPF has a considerable information advantage over simple time series methods. As described in Croushore (1993), the SPF forecast is based on a survey taken in the second month of each quarter. Survey respondents then have considerably more information, on variables such as interest rates and stock prices, than is reflected in time series forecasts that don't include the same information (as reflected in the bottom panel of Table 14, that information gives the SPF its biggest advantage in near-term interest rates). However, the SPF's advantage over time series methods generally declines as the forecast horizon rises. For instance, in $h = 1Y$ forecasts of GDP growth for 1970-84, the SPF and best time series RMSEs are, respectively, 2.891 and 2.775; for forecasts of GDP inflation, the corresponding RMSEs are 2.192 and 2.141.

Moreover, the SPF's advantage is much greater in the 1970-84 sample than the 1985-05 sample. Campbell (2006) notes the same for SPF growth forecasts compared to AR(1) forecasts of GDP growth, attributing the pattern to declining predictability (other recent studies documenting reduced predictability include D'Agostino, et al. (2005), Stock and Watson (2006), and Tulip (2005)). In this later period, the RMSEs of $h = 0Q$ forecasts of GDP growth from the SPF and best time series approach are 1.384 and 1.609, respectively. The RMSEs of $h = 0Q$ forecasts of GDP inflation from the SPF and best time series approach are 0.831 and 0.926, respectively. Reflecting the declining predictability of output and inflation and the reduced advantage of the SPF at longer horizons, for 1–year ahead forecasts in the 1985-05 period, the advantage of the SPF over time series methods is quite small. For instance, in 1–year ahead forecasts of GDP growth, the TVP BVAR(4) using GDP growth, GDP inflation, and the T-bill rate beats the SPF (RMSE of 1.218 vs. 1.274); in forecasts of GDP inflation, the TVP BVAR again beats the SPF (RMSE of 0.764 vs. 0.804).

In light of the more limited availability of Greenbook (GB) forecasts (the public sample ends in 2000), in lieu of comparing VAR forecasts directly to GB forecasts, we simply compare the GB forecasts to SPF forecasts. As long as the GB and SPF forecasts are broadly comparable in RMSE accuracy, our findings for VARs compared to SPF will also apply to VARs compared to GB forecasts. Table 15 reports RMSEs of forecasts of GDP growth, GDP inflation, and CPI inflation, for samples of 1970-84 and 1985-2000 (we omit an interest rate comparison because, for much of the sample, GB did not include an unconditional interest rate forecast). Consistent with evidence in such studies as Romer and Romer (2000) and Sims (2002), GB forecasts tend to be more accurate, especially for inflation. For instance, the 1970-84 RMSEs of 1–year ahead forecasts of GDP inflation are 2.192 for SPF and 1.653 for GB. However, perhaps reflecting declining predictability, any advantage of GB over SPF is generally smaller in the second sample than the first. Regardless, the accuracy differences between SPF and GB forecasts are modest enough that comparing VAR forecasts against GB instead of SPF wouldn't alter the findings described above.

4.2 Long–run forecasts

As noted in section 3, as the forecast horizon increases beyond the one year period considered above, the so-called endpoints come to play an increasingly important role in determining the forecast. Kozicki and Tinsley (1998, 2001a,b), among others, have shown that these endpoints can vary significantly over time. In this section we examine which, if any of the forecast methods considered above, imply reasonable endpoints. For simplicity, we use a 10–year ahead forecast (the forecast in period t+39, from a forecast origin of t using data through $t-1$) as a proxy for the endpoint estimate. Kozicki and Tinsley (2001b) use a similar metric (Kozicki and Tinsley compare 10 year–ahead forecasts to survey measures of long-term inflation expectations).

Of course, an immediate question is, what is reasonable? Trend GDP growth is generally thought to have evolved slowly over time, (at least) declining in the 1970s and rising in the 1990s. The available real–time estimates of potential GDP from the CBO, taken from Kozicki (2004), show some variation in trend growth. CBO estimates of potential output growth rose from about 2.1 percent in 1991 vintage data to 3.2 percent in 2001 and 2.75 percent in 2004 vintage data.[27] At the same time, the implicit inflation goal of monetary policymakers is thought to have trended up from the 1970s through the mid-1980s, and then trended down (see Figure 1 and the associated discussion in section 2). The trend in inflation implies a comparable trend in short-term interest rates. Accuracy in longer-term forecasting is likely to require forecast endpoints that roughly match up to variation in such trends in growth and inflation.

For simplicity, in assessing the ability of VAR forecast methods to generate reasonable endpoints, we compare the estimated endpoint proxies to trends in growth, inflation, and interest rates estimated in real time with exponential smoothing. As noted above, exponential smoothing applied to inflation yields a trend quite similar to the shifting endpoint (or implicit target) estimate of Kozicki and Tinsley (2001a,b). Exponential smoothing applied to GDP growth (with a smoothing parameter of 0.015) yields a trend measure that, in line with many economists' beliefs, shows trend growth gradually slowing over the 1970s

[27] For each each vintage t, we calculate trend growth as the projected percent change in potential GDP in year $t+5$. We use a five–year horizon because, for some years, the CBO data on potential output extend only five, rather than 10, years into the future.

and 1980s before rising in the 1990s. Reflecting real time data availability, trends in each vintage t are estimated using data through period $t-1$.

In light of space limitations, we present endpoint proxy results for just GDP growth and GDP inflation, for a limited set of forecasting methods likely to be of the most interest. The reported forecasts are obtained from models in GDP growth, GDP inflation, and the T-bill rate. Qualitatively, results are similar across other measures of output, inflation, and the interest rate. We omit endpoint results for the T-bill rate because they are qualitatively very similar to those for inflation. The forecast methods or models include the univariate benchmarks, VAR(4), DVAR(4), VAR(4) with inflation detrending, BVAR(4), BDVAR(4), rolling BDVAR(4), BVAR(4) with TVP, BVAR(4) with intercept TVP, the average of univariate and VAR(4) forecasts, and the average of the univariate and VAR(4) with inflation detrending. In light of the general value of shrinkage in forecasting and the potential success of inflation detrending in pinning down reasonable endpoints, we also include an approach not considered above: a VAR(4) with inflation detrending estimated with BVAR methods (BVAR(4) with inflation detrending).[28] This set of methods is intended to include those that work relatively well in shorter-term forecasting and particular approaches, such as differencing and rolling estimation, that are sometimes used in practice to try to capture non–stationarities such as moving endpoints.

The results provided in Figures 2 (GDP growth) and 3 (GDP inflation) show that some forecast approaches fare very poorly, yielding endpoint proxies that are far too volatile to be considered reasonable (note that, in these charts, the scales differ between those methods that work reasonably well and those that don't). These exceedingly volatile methods include the VAR, BVAR, BVAR with TVP, BVAR with intercept TVP, and the average of the univariate and VAR(4). For example, in the case of the VAR(4), the 10–year ahead forecast of GDP growth plummets to -15.2 percent in (vintage) 1975:Q1 and -12.8 percent in 1981:Q3; the forecast of inflation soars to 34.2 percent in 1981:Q3. In (vintage) 1980:Q2, the BVAR(4) forecasts of GDP growth and inflation reach the extremes of -9.4 and 25.8 percent, respectively. In the case of the BVAR(4) with TVP, the long–term projections of growth and inflation are -20.9 percent and 64.5 percent in 1980:Q2. Such extremes in forecasts of course suggest explosive roots in the autoregressive systems, which are indeed evident in the system estimates. For example, the VAR(4) system has a largest root of

[28] We obtain these estimates using the BVAR prior variances described in section 2 and prior means of 0 for all coefficients.

1.005 in the 1975:Q1 estimates, 1.002 in the 1980:Q2 estimates, and 1.031 in the 1981:Q3 estimates. The BVAR(4) system has a largest root of 1.011 in the 1981:Q3 estimates. As a result, for a practitioner interested in using these methods for forecasting in real time, some care in adjusting estimates to avoid explosive roots would be required to improve the endpoint and long–term forecast accuracy of the methods.

The other forecast methods — univariate, DVAR, VAR with inflation detrending, BVAR with inflation detrending, BDVAR, rolling BDVAR, and the average of the univariate and VAR with inflation detrending — produce much less volatile and therefore more reasonable endpoint estimates. For example, the univariate and BDVAR(4) 10–year ahead forecasts of GDP growth correspond pretty closely (at least in relative terms) to the exponentially smoothed trend. Of course, the exponentially smoothed measure may not be the best estimate of trend. However, any better estimate of trend growth is not likely to be significantly more volatile over time. As a result, even among this relatively better set of forecast methods, a smooth long–term forecast like that from the univariate model may be preferred to a modestly more volatile one, like the forecast from the VAR(4) with inflation detrending. Among inflation forecasts, the endpoint proxies from the univariate and BVAR with inflation detrending models provide the closest match to trend inflation. The endpoint proxy from the BVAR with inflation detrending includes less high frequency variation than does the estimate from the univariate model, but is farther from trend inflation in the 1970s.

Two other results are worth noting. First, for both growth and inflation, rolling estimation of the BDVAR implies endpoints that are more volatile than the endpoints implied by the recursively estimated BDVAR. Second, compared to OLS estimation, Bayesian estimation of the VAR with inflation detrending helps to dampen volatility in the endpoint proxies (although not included in the RMSE results above, Bayesian estimation also helped to modestly improve the forecast accuracy of VARs with inflation detrending).

5 Conclusion

In this paper we provide empirical evidence on the ability of several different methods to improve the real–time forecast accuracy of small-scale macroeconomic VARs in the presence of model instability. The 18 distinct trivariate VARs that we consider are each comprised of one of three measures of output, one of three measures of inflation, and one of two measures of short-term interest rates. For each of these models we construct real time forecasts of each variable (with particular emphasis on the output and inflation measures). For each of the 18 variable combinations, we consider 86 different forecast models or methods, incorporating different choices of lag selection, observation windows used for estimation, levels or differences, intercept corrections, stochastically time–varying parameters, break dating, discounted least squares, Bayesian shrinkage, detrending of inflation and interest rates, and model averaging. We compare our results to those from simple baseline univariate models as well as forecasts from the Survey of Professional Forecasters and the Federal Reserve Board's Greenbook.

Our results indicate that some of the methods do consistently improve forecast accuracy in terms of root mean square errors (RMSE). Not surprisingly, the best method often varies with the variable being forecasted, but several patterns do emerge. After aggregating across all models, horizons and variables being forecasted, it is clear that model averaging and Bayesian shrinkage methods consistently perform among the best methods. At the other extreme, the approaches of using a fixed rolling window of observations to estimate model parameters and discounted least squares estimation consistently rank among the worst. Of course, estimation methods that are unsuccessful in forecasting may nonetheless prove useful for other purposes. Perhaps not surprisingly, out–of–sample forecast accuracy does not seem to be strongly related to in–sample fit. For models in GDP growth, GDP inflation, and the T-bill rate, Figure 4 compares real time forecast RMSEs to in–sample fit estimates (for each forecasting model, in–sample fit is measured as the standard error of estimate, averaged over the forecasting sample). Except for some outlier observations, in–sample fit has little relationship (and sometimes a negative relationship) with forecast accuracy, at least in the VAR models and methods we consider.

Acknowledgments

We gratefully acknowledge helpful comments from Taisuke Nakata, participants in the conference associated with the book, seminar participants at the Federal Reserve Bank of Kansas City and Board of Governors, and an anonymous reviewer. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City, Board of Governors, Federal Reserve System, or any of its staff.

References

Allen, P.G. and R. Fildes (2006). Levels, differences and ECMs - Principles for Improved Econometric Forecasting. *Oxford Bulletin of Economics and Statistics*, forthcoming.

Andrews, D.W.K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica 61*. 821-56.

Andrews, D.W.K. (2006). End-of-sample Instability Tests. *Econometrica 71*. 1661-94.

Bai, J. (1997). Estimation of a Change Point in Multiple Regression Models. *Review of Economics and Statistics 79*. 551-63.

Bai, J. and P. Perron (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica 66*. 47-78.

Bai, J. and P. Perron (2003). Computation and Analysis of Multiple Structural-Change Models. *Journal of Applied Econometrics 18*. 1-22.

Boivin, J. (1999). Revisiting the Evidence on the Stability of Monetary VAR's. Manuscript, Columbia University.

Boivin, J. (2006). Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data. *Journal of Money, Credit and Banking 38*. 1149-73.

Brainard, W.C. and G.L. Perry (2000). Making Policy in a Changing World. In G. Perry and J. Tobin (Eds.), *Economic Events, Ideas, and Policies: The 1960s and After.* (pp. 43-69). Harrisonburg VA: R.R. Donnelley and Sons.

Branch, W. and G. Evans (2006). A Simple Recursive Forecasting Model. *Economics Letters 91*. 158-66.

Brayton, F., E. Mauskopf, D. Reifschneider, P. Tinsley and J. Williams (1997). The Role of Expectations in the FRB/US Macroeconomic Model. *Federal Reserve Bulletin*, April. 227-45.

Campbell, S.D. (2006). Macroeconomic Volatility, Predictability, and Uncertainty in the Great Moderation: Evidence from the Survey of Professional Forecasters. *Journal of Business and Economic Statistics*, forthcoming.

Canova, Fabio (2002). "G-7 Inflation Forecasts." CEPR Discussion Paper No. 3283.

Clark, T.E. and S. Kozicki (2005). Estimating Equilibrium Real Interest Rates in Real Time. *North American Journal of Economics and Finance 16*. 395-413.

Clark, T.E. and M.W. McCracken (2005a). Combining Forecasts from Nested Models. Manuscript, Federal Reserve Bank of Kansas City.

Clark, T.E. and M.W. McCracken (2005b). Improving Forecast Accuracy by Combining

Recursive and Rolling Forecasts. Manuscript, Federal Reserve Bank of Kansas City.

Clark, T.E. and M.W. McCracken (2006a). Forecasting with Small Macroeconomic VARs in the Presence of Instabilities. Research Working Paper No. 06-09, 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*. 1127-48.

Clements, M.P. and D.F. Hendry (1996). Intercept Corrections and Structural Change. *Journal of Applied Econometrics 11*. 475-94.

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

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

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

Cogley, T. and T.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, D. (1993). Introducing: the Survey of Professional Forecasters. *Federal Reserve Bank of Philadelphia Business Review*, Nov./Dec. 3-13.

Croushore, D. (2006). Forecasting with Real–Time Macroeconomic Data. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting* (pp. 961-82). Amsterdam The Netherlands: North-Holland.

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

D'Agostino, A., D. Giannone and P. Surico (2005). (Un)Predictability and Macroeconomic Stability. Manuscript, ECARES.

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

Doan, T., R. Litterman and C. Sims (1984). Forecasting and Conditional Prediction Using Realistic Prior Distributions. *Econometric Reviews 3*. 1-100.

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

Evans, G. and S. Honkapohja (2005). Policy Interaction, Expectations and the Liquidity

Trap. *Review of Economic Dynamics 8*. 303-23.

Favero, C. and M. Marcellino (2005). Modelling and Forecasting Fiscal Variables for the Euro Area. *Oxford Bulletin of Economics and Statistics 67*. 755-83.

Geweke, J. and C. Whiteman (2006). Bayesian Forecasting. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting* (pp. 3-80). Amsterdam The Netherlands: North-Holland.

Giacomini, R. and H. White (2005). Tests of Conditional Predictive Ability. *Econometrica*, forthcoming.

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

Hansen, P.R. (2005). A Test for Superior Predictive Ability. *Journal of Business and Economics Statistics 23*. 365-80.

Hendry, D.F., S. Johansen and C. Santos (2004). Selecting a Regression Saturated by Indicators. Manuscript, University of Oxford.

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

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

Keating, J.W. (2000). Macroeconomic Modeling with Asymmetric Vector Autoregressions. *Journal of Macroeconomics 22*. 1-28.

Koop, G. and S. Potter (2003). Forecasting in Dynamic Factor Models using Bayesian Model Averaging. *The Econometrics Journal 7*, 550-65.

Kozicki, S. (2004). How Do Data Revisions Affect the Evaluation and Conduct of Monetary Policy? *Federal Reserve Bank of Kansas City Economic Review*, First Quarter. 5-38.

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

Kozicki, S. and P.A. Tinsley (1998). Moving Endpoints and the Internal Consistency of Agents' ex ante Forecasts. *Computational Economics 11*. 21-40.

Kozicki, S. and P.A. Tinsley (2001a). Shifting Endpoints in the Term Structure of Interest Rates. *Journal of Monetary Economics 47*. 613-52.

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

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

Laubach, T. and J.C. Williams (2003). Measuring the Natural Rate of Interest. *Review of Economics and Statistics*, *85*. 1063-70.

Lebow, D.E., J.M. Roberts and D.J. Stockton (1992). Economic Performance Under Price Stability. Finance and Economics Discussion Series Paper No. 125, Board of Governors of the Federal Reserve System.

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

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

Liu, R.Y. and K. Singh (1992). Moving Blocks Jackknife and Bootstrap Capture Weak Dependence. In R. Lepage and L. Billiard (Eds.), *Exploring the Limits of Bootstrap* (pp. 22-148). New York NY: Wiley.

Maheu, J.M. and S. Gordon (2004). Learning, Forecasting and Structural Breaks. Manuscript, University of Toronto.

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

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

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

Orphanides, A. and S. 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.

Orphanides, A. and J. Williams (2005). Inflation Scares and Forecast-based Monetary

Policy. *Review of Economic Dynamics 8*. 498-527.

Pesaran, M.H., D. Pettenuzzo and A. Timmermann (2006). Bayesian Regime Averaging for Time Series Subject to Structural Breaks. *Review of Economic Studies 73*. 1057-84.

Roberts, J.M. (2006). Monetary Policy and Inflation Dynamics. *International Journal of Central Banking 2*. 193-230.

Robertson, J. and E. Tallman (1999). Vector Autoregressions: Forecasting and Reality. *Federal Reserve Bank of Atlanta Economic Review*, First Quarter. 4-18.

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

Rogoff, K. (2003). Globalization and Global Disinflation. *Monetary Policy and Uncertainty: Adapting to a Changing Economy*. Federal Reserve Bank of Kansas City Jackson Hole Symposium. 77-112.

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

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

Rudebusch, G.D. and L.E.O. Svensson (1999). Policy Rules for Inflation Targeting. In J. Taylor (Ed.), *Monetary Policy Rules* (pp. 203-46). Chicago IL: University of Chicago Press.

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

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

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

Stock, J.H. and M.W. Watson (1999). Forecasting Inflation. *Journal of Monetary Economics 44*. 293-335.

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

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

Stock, J.H. and M.W. Watson (2006). Why Has Inflation Become Harder to Forecast? *Journal of Money, Credit, and Banking*, forthcoming.

Tulip, P. (2005). Has Output Become More Predictable? Changes in Greenbook Forecast Accuracy. Finance and Economics Discussion Series Paper No. 2005-31, Board of Governors of the Federal Reserve System.

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

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

Yao, Y. (1988). Estimating the Number of Change-Points Via Schwarz' Criterion. *Statistics and Probability Letters 6*. 181-89.

Figure 1: Alternative estimates of CPI inflation trends

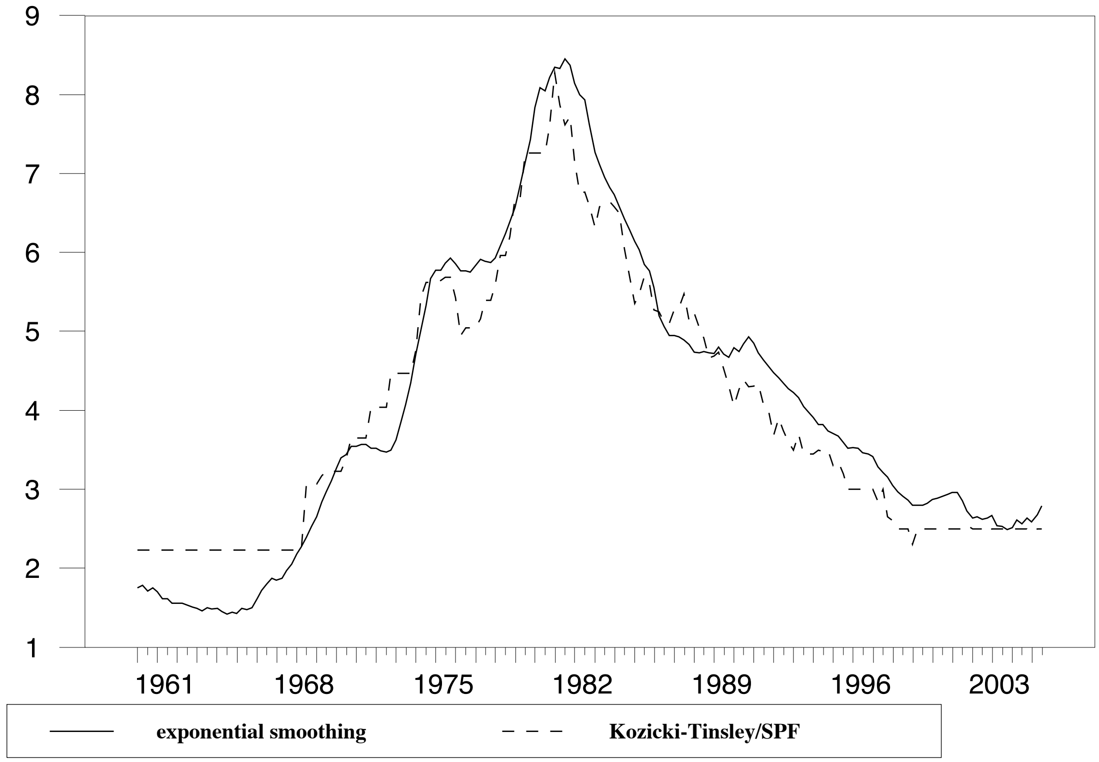
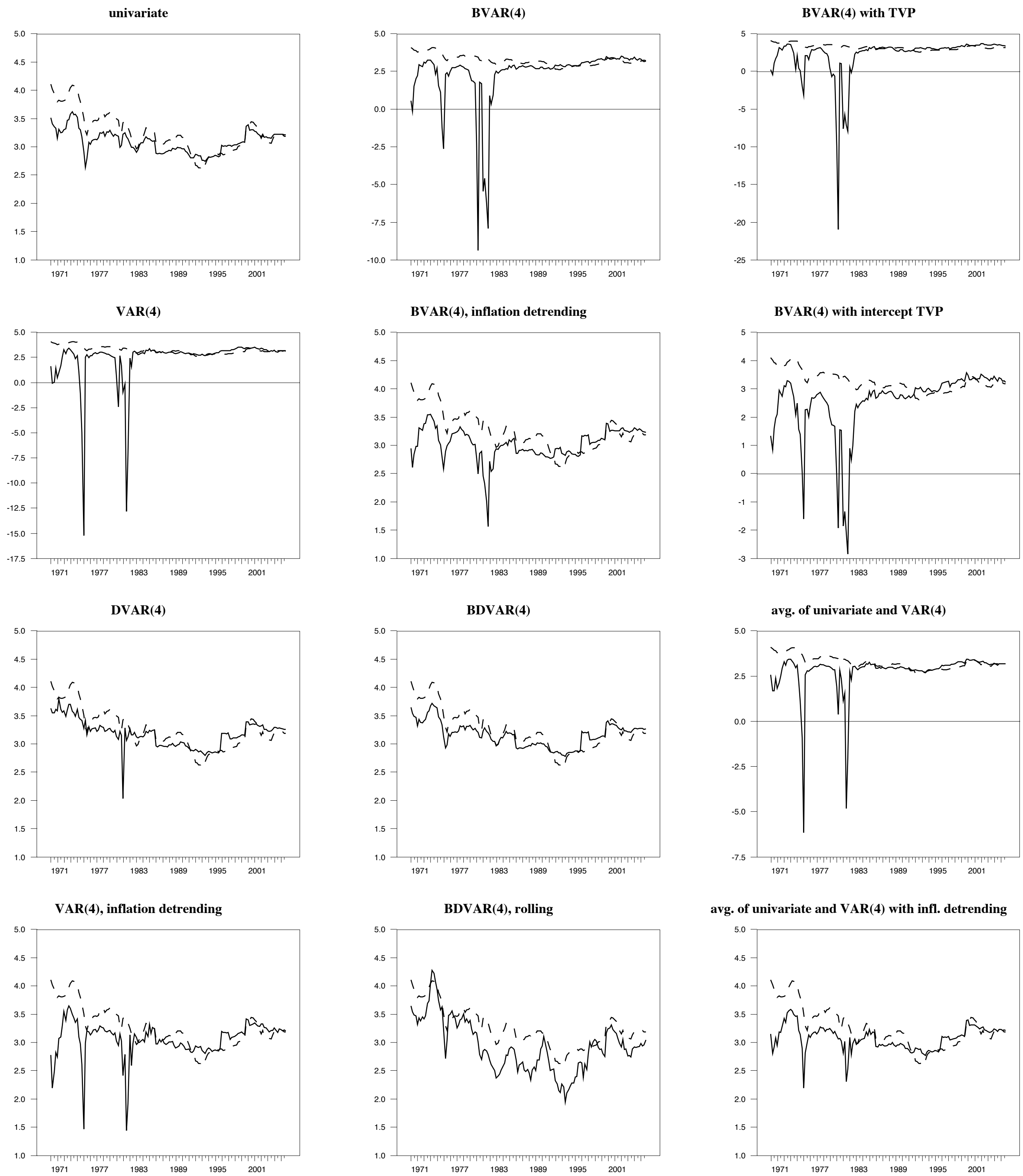


Figure 2: 10-year ahead forecasts of GDP growth

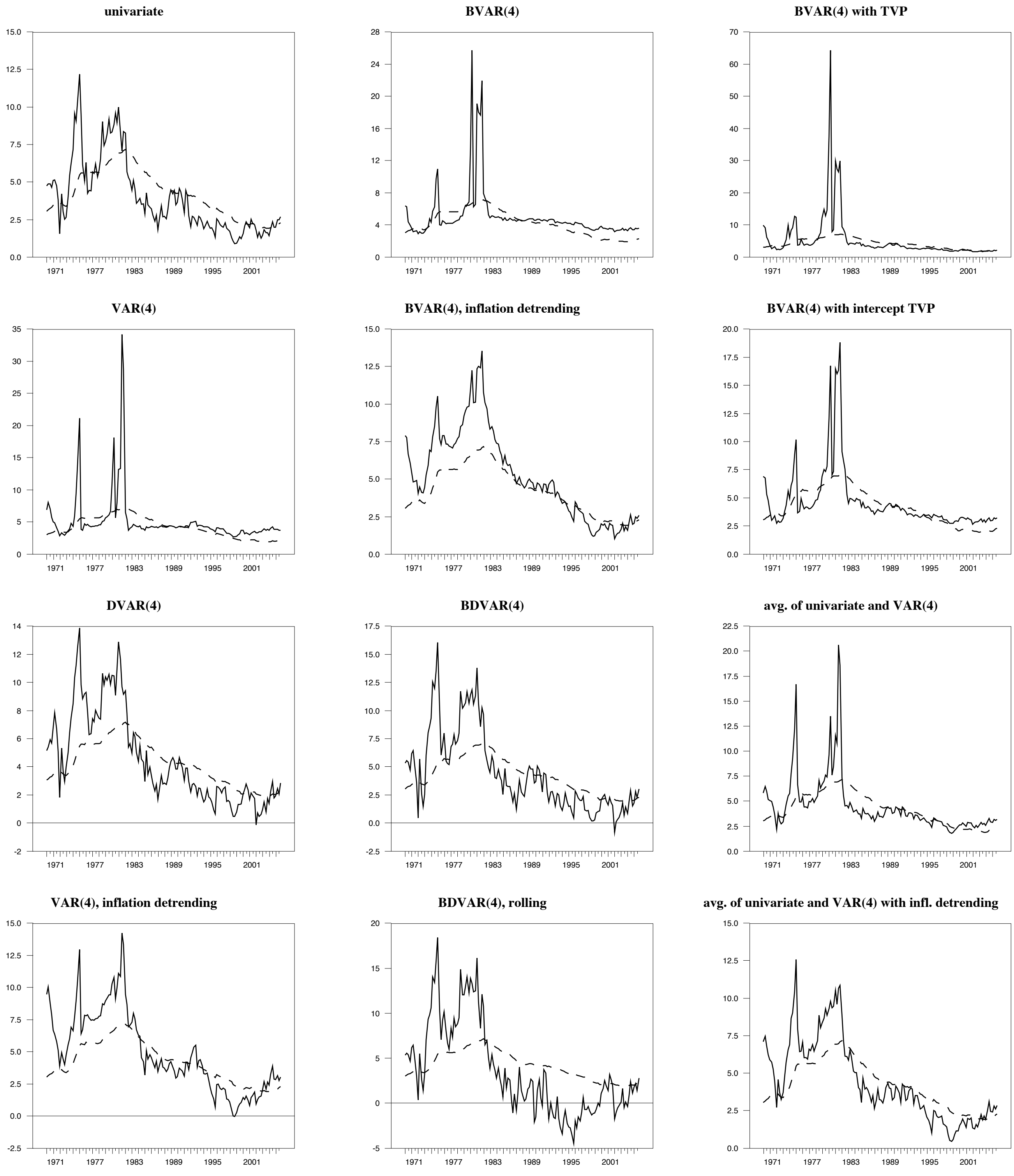
(VAR in GDP growth, GDP inflation, and T-bill rate)



solid lines: forecasts   dotted lines: exponentially smoothed trends

Figure 3: 10-year ahead forecasts of GDP inflation

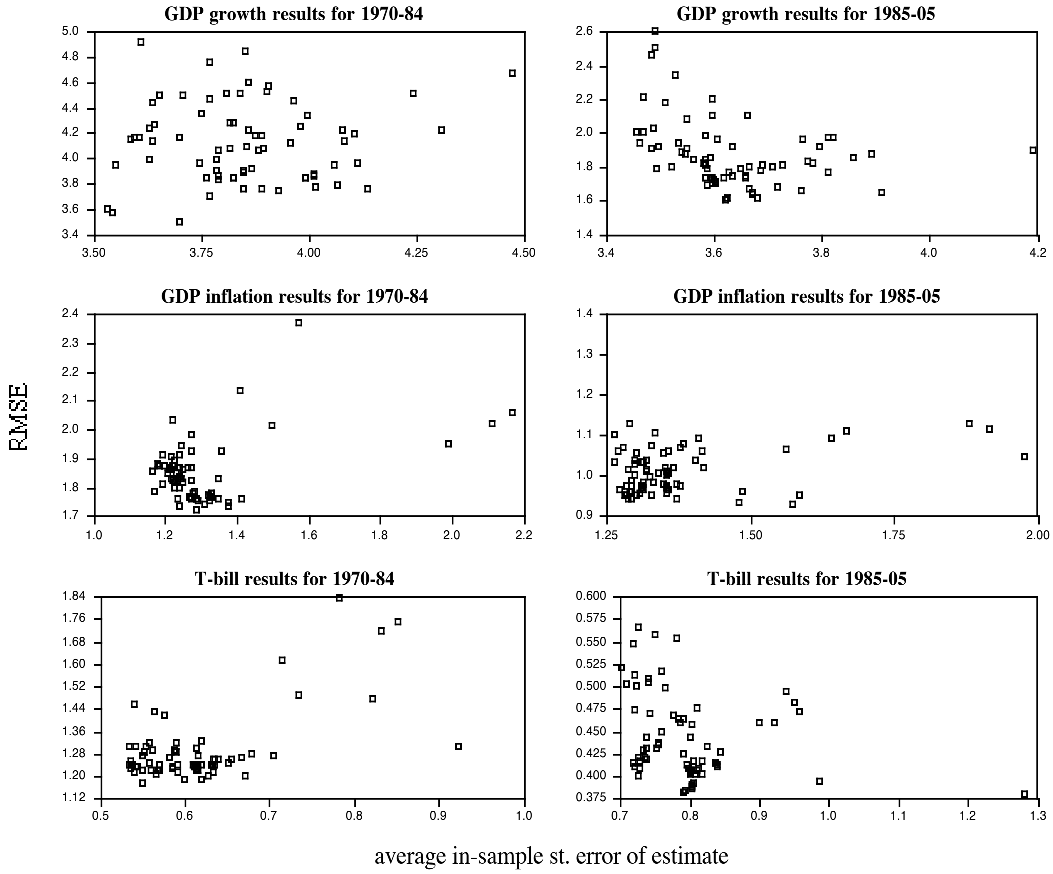
(VAR in GDP growth, GDP inflation, and T-bill rate)



solid lines: forecasts **dotted lines: exponentially smoothed trends**

Figure 4. In-sample fit vs. forecast RMSE

(VAR in GDP growth, GDP inflation, and T-bill rate)



Notes:

1. The figures compare forecast RMSEs for the indicated variable and sample to corresponding measures of in-sample fit.
2. All results are based on models in GDP growth, GDP inflation, and the T-bill rate. The forecast methods are listed in Table 1. The figures exclude results for the intercept correction methods of Clements and Hendry (1996), because it is not clear how best to measure in-sample fit for the associated forecasts.
3. The forecast RMSEs are based on the $h = 0Q$ horizon. Starting with t = 1970:Q1, the in-sample fit of each model used to forecast is estimated as the conventional standard error of estimate (with the conventional degrees of freedom adjustment). For each model, the time series of in-sample fit estimates is averaged over the 1970-84 and 1985-05 forecast samples. The charts use these average estimates of in-sample fit.
4. In the case of forecasts based on rolling sample model estimates, we fit the same model to the sample of data preceding the rolling sample (assuming, in effect, a break in the model's coefficients at the time of the rolling sample start). We then estimate in-sample fit as the (square root of the) sum of squared residuals over the whole period divided by the total sample size less the total number of parameters.

Table 1: Forecasting methods

method	details
VAR(4)	VAR in y, π, i with fixed lag order of 4
VAR(2)	same as above with fixed lag order of 2
VAR(AIC)	VAR with system lag determined by AIC
VAR(BIC)	VAR with system lag determined by BIC
VAR(AIC, by eq.&var.)	VAR in y, π, i allowing different, AIC-det. lags for each var. in each eq.
VAR(BIC, by eq.&var.)	same as above, with BIC-determined lags
DVAR(4)	VAR in y, $\Delta\pi$, Δi with fixed lag order of 4
DVAR(2)	same as above with fixed lag order of 2
DVAR(AIC)	VAR in y, $\Delta\pi$, Δi with system lag set by AIC
DVAR(BIC)	VAR in y, $\Delta\pi$, Δi with system lag set by BIC
DVAR(AIC, by eq.&var.)	VAR in y, $\Delta\pi$, Δi allowing different, AIC-det. lags for each var. in each eq.
DVAR(BIC, by eq.&var.)	same as above, with BIC-determined lags
DVAR(4), output diff.	VAR in Δy, $\Delta\pi$, Δi with fixed lag order of 4
DVAR(2), output diff.	same as above with fixed lag order of 2
DVAR(AIC), output diff.	VAR in Δy, $\Delta\pi$, Δi with system lag set by AIC
DVAR(BIC), output diff.	VAR in Δy, $\Delta\pi$, Δi with system lag set by BIC
BVAR(4)	VAR(4) in y, π, i est. with Minnesota priors, using $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$
BVAR(2)	same as above with fixed lag order of 2
BDVAR(4)	VAR(4) in y, $\Delta\pi$, Δi est. with Minnesota priors, using $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$
BDVAR(2)	same as above with fixed lag order of 2
VAR(4), rolling	VAR in y, π, i with fixed lag order of 4, est. with a rolling window of 60 observations
VAR(2), rolling	same as above with fixed lag order 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	VAR in y, π, i allowing different, AIC-det. lags for each var. in each eq., est. with a rolling sample of 60 obs.
VAR(BIC, by eq.&var.), rolling	same as above with BIC-determined lags
DVAR(4), rolling	VAR in y, $\Delta\pi$, Δi with fixed lag order of 4, est. with a rolling sample of 60 observations
DVAR(2), rolling	same as above with fixed lag order of 2
DVAR(AIC), rolling	same as above with AIC–determined lag
DVAR(BIC), rolling	same as above with BIC–determined lag

Table 1, continued: Forecasting methods

method	details
DVAR(AIC, by eq.&var.), rolling	VAR in y, $\Delta\pi$, Δi allowing different, AIC-det. lags for each var. in each eq., est. with a rolling sample of 60 obs.
DVAR(BIC, by eq.&var.), rolling	same as above with BIC-determined lags
DVAR(4), output diff., rolling	VAR in Δy, $\Delta\pi$, Δi with fixed lag order of 4, est. with a rolling sample of 60 observations
DVAR(2), output diff., rolling	same as above with fixed lag order of 2
DVAR(AIC), output diff., rolling	same as above with AIC–determined lag
DVAR(BIC), output diff., rolling	same as above with BIC–determined lag
BVAR(4), rolling	BVAR(4) in y, π, i with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$, est. with a rolling sample of 60 obs.
BVAR(2), rolling	same as above with fixed lag order of 2
BDVAR(4), rolling	BVAR(4) in y, $\Delta\pi$, Δi with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$, est. with a rolling sample of 60 obs.
BDVAR(2), rolling	same as above with fixed lag order of 2
DLS, VAR(4)	VAR(4) in y, π, i, est. with discounted least squares (DLS), using dis. rates of .99 for y eq. .95 for π and i eq.
DLS, VAR(2)	same as above with fixed lag of 2
DLS, VAR(AIC)	same as above with lag order det. from AIC applied to OLS estimates of system
DLS, DVAR(4)	VAR(4) in y, $\Delta\pi$, Δi, est. with DLS, using dis. rates of .99 for y eq., .95 for $\Delta\pi$ and Δi eq.
DLS, DVAR(2)	same as above with fixed lag of 2
DLS, DVAR(AIC)	same as above with lag order set by AIC applied to OLS estimates of system
VAR(AIC), AIC intercept breaks	VAR in y, π, i with AIC-det. lags, allowing up to two breaks in the set of intercepts, with the number and dates that minimize the AIC
VAR(AIC), BIC intercept breaks	same as above, using the BIC to determine the breaks
VAR(4), intercept correction	VAR(4) forecasts adjusted by the average of the last 4 residuals (Clements and Hendry (1996), eq. 40)
VAR(2), intercept correction	same as above with fixed lag order of 2
VAR(AIC), intercept correction	VAR(AIC lag) forecasts adjusted by the average of the last 4 residuals (Clements and Hendry (1996), eq. 40)
VAR(4), partial int. corr.	VAR(4) forecasts of π and i adjusted by the average of the last 4 residuals (y residuals treated as 0)
VAR(2), partial int. corr.	same as above with fixed lag order of 2
VAR(AIC), partial int. corr.	VAR(AIC lag) forecasts of π and i adjusted by the average of the last 4 residuals (y residuals treated as 0)

Table 1, continued: Forecasting methods

method	details
VAR(4), inflation detrending	VAR(4) in y, $\pi - \pi^*_{-1}$, and $i - \pi^*_{-1}$, where $\pi^* = \pi^*_{-1} + \alpha(\pi - \pi^*_{-1})$, α = .05 for GDP and CPI inflation, .07 for core PCE inflation
VAR(2), inflation detrending	same as above with fixed lag of 2
VAR(AIC), inflation detrending	same as above with AIC–det. lag for the y, $\pi - \pi^*_{-1}$, and $i - \pi^*_{-1}$ system
VAR(BIC), inflation detrending	same as above with BIC–det. lag for the y, $\pi - \pi^*_{-1}$, and $i - \pi^*_{-1}$ system
VAR(4), full ES detrending	VAR(4) in y, $\pi - \pi^*_{-1}$, and $i - i^*_{-1}$, where $\pi^* = \pi^*_{-1} + \alpha(\pi - \pi^*_{-1})$ (α = .05 or .07, depending on π measure), $i^* = i^*_{-1} + .07(i - i^*_{-1})$
VAR(2), full ES detrending	same as above with fixed lag of 2
VAR(AIC), full ES detrending	same as above with AIC–det. lag for the y, $\pi - \pi^*_{-1}$, and $i - i^*_{-1}$ system
VAR(BIC), full ES detrending	same as above with BIC–det. lag for the y, $\pi - \pi^*_{-1}$, and $i - i^*_{-1}$ system
TVP BVAR(4)	TVP BVAR(4) in y, π, i with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = .1$, $\lambda = .0005$
TVP BVAR(2)	same as above with fixed lag of 2
TVP BVAR(4), $\lambda_4 = .5, \lambda = .0025$	TVP BVAR(4) in y, π, i with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = .5$, $\lambda = .0025$
TVP BVAR(2), $\lambda_4 = .5, \lambda = .0025$	same as above with fixed lag of 2
TVP BVAR(4), $\lambda_4 = 1000, \lambda = .005$	TVP BVAR(4) in y, π, i with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$, $\lambda = .005$
TVP BVAR(2), $\lambda_4 = 1000, \lambda = .005$	same as above with fixed lag of 2
TVP BVAR(4), $\lambda_4 = 1000, \lambda = .0001$	TVP BVAR(4) in y, π, i with $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = 1000$, $\lambda = .0001$
TVP BVAR(2), $\lambda_4 = 1000, \lambda = .0001$	same as above with fixed lag of 2
Intercept TVP BVAR(4)	BVAR(4) in y, π, i, TVP in only intercepts, $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = .1$, $\lambda = .0005$
Intercept TVP BVAR(2)	same as above with fixed lag of 2
Intercept TVP BVAR(4), $\lambda_4 = .5, \lambda = .0025$	BVAR(4) in y, π, i, TVP in only intercepts, $\lambda_1 = .2$, $\lambda_2 = .5$, $\lambda_3 = 1$, $\lambda_4 = .5$, $\lambda = .0025$
Intercept TVP BVAR(2), $\lambda_4 = .5, \lambda = .0025$	same as above with fixed lag of 2

Table 1, continued: Forecasting methods

method	details
average of all forecasts	simple average of all of the above forecasts
avg. of VAR(4), rolling VAR(4)	average of forecasts from recursive and rolling estimates of VAR(4) in y, π, and i
avg. of VAR(2), rolling VAR(2)	same as above using VARs with fixed lag of 2
avg. of univariate, VAR(4)	average of forecasts from univariate model and VAR(4) in y, π, and i
avg. of univariate, VAR(2)	same as above using VAR with fixed lag of 2
avg. of univariate, DVAR(4)	average of forecasts from univariate model and VAR(4) in Δy, $\Delta\pi$, and i
avg. of univariate, DVAR(2)	same as above using VAR with fixed lag of 2
avg. of univ., IDTR VAR(4)	average of forecasts from univariate model and VAR(4) with inflation detrending
avg. of univ., IDTR VAR(2)	same as above using VAR with fixed lag of 2
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	simple average of univariate, VAR(4), DVAR(4), and TVP BVAR(4) ($\lambda_4 = .1, \lambda = .0005$) forecasts
avg. of univ., VAR(2), DVAR(2), TVP BVAR(2)	same as above using VARs with fixed lag of 2
univariate	AR(2) for y, rolling MA(1) for $\Delta\pi$, rolling MA(1) for Δi

Notes:

1. The variables y, π, and i refer to, respectively, output (GDP growth, the HPS gap, or the HP gap), inflation (GDP inflation, CPI inflation, or core PCE inflation), and the interest rate (T-bill or federal funds).

2. Unless otherwise noted, all models are estimated recursively, using all data (starting in 1955 or later) available up to the forecasting date.

3. The rolling estimates of the univariate models for $\Delta\pi$ and Δi use 40 observations.

4. The AIC and BIC lag orders range from 0 (the minimum allowed) to 4 (the maximum allowed).

5. Section 2 details the hyperparameterization (and λ notation above) used in BVAR estimation. In BVAR estimation, prior means for all coefficients are generally set at 0, with the following exceptions: (a) prior means for own first lags of π and i are set at 1 in models with levels of inflation and interest rates; (b) prior means for own first lags of y are set at 0.8 in models with an output gap; and (c) prior means for the intercept of GDP growth equations are set to the historical average of growth in BVAR estimates that impose informative priors (λ_4 = .1 or .5) on the constant term.

6. The time variation in the coefficients of the TVP BVARs takes a random walk form. In time–varying BVARs with flat priors on the intercepts (λ_4 = 1000), the variation of the innovation in the intercept is set at λ times the prior variance of the coefficient on the own first lag instead of the prior variance of the constant.

7. The exponential smoothing used in the models with detrending is initialized with the average value of inflation over the first five years of each sample.

Table 2: Real-time RMSE results for GDP growth and GDP inflation

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

	GDP growth forecasts					
	1970-84			**1985-2005**		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univariate	4.183	4.761	3.652	1.609	1.668	1.293
VAR(4)	1.022	.912	.936	1.184	1.200	1.110
VAR(4), intercept correction	1.038	.944	1.047	1.177	1.209	1.325
VAR(AIC)	1.024	.921	.969	1.169	1.188	1.105
DVAR(4)	1.039	.932	*.760*	1.260	1.298	1.152
DVAR(AIC)	.974	*.847*	*.798*	1.208	1.240	1.108
VAR(AIC, by eq.&var.)	.948	.902	.989	1.113	1.122	.998
DVAR(AIC, by eq.&var.)	1.019	.943	*.783*	1.204	1.260	1.155
BVAR(4)	*.919*	*.875*	.949	1.077	1.090	1.005
BDVAR(4)	.988	*.956*	*.956*	1.045	1.045	1.013
VAR(4), inflation detrending	.956	*.837*	*.797*	1.247	1.283	1.162
VAR(AIC), intercept breaks	.994	.894	.891	1.378	1.478	1.562
VAR(4), rolling	1.175	1.062	1.091	1.222	1.306	1.385
DVAR(4), rolling	1.077	1.003	*.773*	1.115	1.221	1.143
VAR(AIC, by eq.&var.), rolling	1.014	.943	1.019	1.296	1.301	1.321
BVAR(4), rolling	.945	*.880*	1.004	1.196	1.220	1.193
BDVAR(4), rolling	1.008	.993	1.003	1.024	1.040	1.066
TVP BVAR(4)	*.927*	*.896*	.955	1.025	1.024	.941
Intercept TVP BVAR(4)	*.922*	*.891*	.940	1.019	1.013	*.914*
DLS, VAR(4)	1.081	1.005	1.068	1.154	1.183	1.143
DLS, DVAR(4)	1.078	1.028	.949	1.167	1.208	1.159
average of all forecasts	*.893*	*.815*	*.816*	1.078	1.093	1.015
avg. of VAR(4), rolling VAR(4)	1.070	.957	.953	1.158	1.212	1.210
avg. of univariate, VAR(4)	.958	*.901*	.900	1.057	1.056	.988
avg. of univariate, DVAR(4)	.945	**.882**	**.796**	1.086	1.096	1.027
avg. of univ., IDTR VAR(4)	*.931*	*.871*	*.849*	1.060	1.061	.952
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	*.922*	*.850*	*.804*	1.078	1.084	.995

Table 2, continued: RMSE results for GDP growth and GDP inflation

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

	GDP inflation forecasts					
	1970-84			1985-2005		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univariate	1.825	2.153	2.389	.951	1.016	.760
VAR(4)	1.022	1.033	1.061	1.001	*.948*	.959
VAR(4), intercept correction	1.020	1.054	1.142	1.133	1.134	1.439
VAR(AIC)	1.037	1.066	1.057	1.024	*.977*	.982
DVAR(4)	1.007	.946	*.896*	.989	*.946*	1.006
DVAR(AIC)	.964	.955	*.912*	.994	*.950*	.985
VAR(AIC, by eq.&var.)	1.028	1.085	1.120	1.014	*.965*	.992
DVAR(AIC, by eq.&var.)	1.027	1.033	.998	1.003	.965	1.031
BVAR(4)	.971	1.047	1.093	1.023	1.039	1.161
BDVAR(4)	.969	.985	*.936*	1.030	1.034	1.069
VAR(4), inflation detrending	1.024	1.013	1.006	1.011	.979	1.081
VAR(AIC), intercept breaks	1.032	1.013	.996	1.085	1.098	1.438
VAR(4), rolling	1.016	1.083	1.080	1.156	1.128	1.407
DVAR(4), rolling	1.026	1.000	*.900*	1.066	.990	1.151
VAR(AIC, by eq.&var.), rolling	1.016	1.165	1.212	1.159	1.152	1.504
BVAR(4), rolling	.950	1.022	1.050	1.090	1.174	1.482
BDVAR(4), rolling	.965	.991	*.939*	1.075	1.101	1.191
TVP BVAR(4)	.975	1.053	1.108	.992	.977	1.006
Intercept TVP BVAR(4)	.975	1.047	1.081	1.007	1.004	1.079
DLS, VAR(4)	1.129	1.334	1.290	1.173	1.132	1.243
DLS, DVAR(4)	1.300	1.251	1.070	1.170	1.109	1.161
average of all forecasts	.946	.989	.970	1.025	1.015	1.057
avg. of VAR(4), rolling VAR(4)	1.009	1.052	1.063	1.055	1.014	1.131
avg. of univariate, VAR(4)	.967	.985	.996	.980	*.958*	*.942*
avg. of univariate, DVAR(4)	.967	*.952*	*.931*	*.974*	*.954*	.967
avg. of univ., IDTR VAR(4)	.971	.979	.974	.985	*.969*	.980
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	.959	.978	.980	.977	*.951*	.953

Notes:

1. The variables in each multivariate model are GDP growth, GDP inflation, and the T-bill rate.
2. 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.
3. 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.
4. The forecast errors are calculated using the first–available (real–time) estimates of output and inflation as the actual data on output and inflation.
5. In each quarter t from 1970:Q1 through 2005:Q4, vintage t data 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 $t+4$.
6. See Table 1 for detail on each forecast method.

Table 3: Real-time RMSE results for the HPS output gap and GDP inflation

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

	HPS output gap forecasts					
	1970-84			1985-2005		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univariate	1.039	1.988	3.891	.702	1.028	2.044
VAR(4)	1.051	.960	.944	1.110	1.159	1.204
VAR(4), intercept correction	1.079	1.010	1.110	1.066	1.048	1.060
VAR(AIC)	1.016	.966	.991	1.108	1.155	1.207
DVAR(4)	1.068	.942	*.743*	1.102	1.127	1.084
DVAR(AIC)	1.039	.947	*.866*	1.099	1.105	1.059
VAR(AIC, by eq.&var.)	.985	.946	.995	1.077	1.133	1.176
DVAR(AIC, by eq.&var.)	1.088	1.005	*.880*	1.071	1.110	1.085
BVAR(4)	1.012	.931	.922	1.077	1.151	1.176
BDVAR(4)	1.064	1.002	.994	1.002	.997	.991
VAR(4), inflation detrending	1.030	.920	.892	1.060	1.077	1.012
VAR(AIC), intercept breaks	1.008	.929	*.754*	1.189	1.320	1.267
VAR(4), rolling	1.190	1.110	1.032	1.116	1.237	1.305
DVAR(4), rolling	1.103	.993	*.802*	1.029	1.074	1.008
VAR(AIC, by eq.&var.), rolling	1.170	1.129	1.064	1.099	1.181	1.211
BVAR(4), rolling	1.060	.968	.986	1.087	1.172	1.186
BDVAR(4), rolling	1.093	1.047	1.059	.993	1.005	.995
TVP BVAR(4)	1.020	.957	.947	.982	.970	*.921*
Intercept TVP BVAR(4)	1.015	.944	.923	*.977*	*.957*	*.908*
DLS, VAR(4)	1.100	1.041	.935	1.053	1.067	1.108
DLS, DVAR(4)	1.106	1.020	.919	1.061	1.066	1.056
average of all forecasts	.948	*.872*	*.824*	1.025	1.036	1.000
avg. of VAR(4), rolling VAR(4)	1.091	1.005	.931	1.089	1.162	1.218
avg. of univariate, VAR(4)	.974	*.912*	*.876*	1.034	1.041	1.028
avg. of univariate, DVAR(4)	.973	*.904*	**.804**	1.038	1.045	1.024
avg. of univ., IDTR VAR(4)	.954	*.878*	*.841*	1.011	1.003	.950
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	.966	*.888*	*.809*	1.028	1.027	.992

Table 3, continued: RMSE results for the HPS output gap and GDP inflation

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

	GDP inflation forecasts					
	1970-84			**1985-2005**		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univariate	1.825	2.153	2.389	.951	1.016	.760
VAR(4)	1.020	1.037	1.075	.973	*.923*	.933
VAR(4), intercept correction	1.017	1.043	1.116	1.109	1.108	1.436
VAR(AIC)	1.020	1.046	1.050	.992	.968	.975
DVAR(4)	1.003	.942	*.904*	.990	.960	1.132
DVAR(AIC)	.941	*.931*	*.879*	.992	.967	1.130
VAR(AIC, by eq.&var.)	1.054	1.112	1.130	.989	*.934*	1.007
DVAR(AIC, by eq.&var.)	1.008	.993	*.906*	.992	.972	1.202
BVAR(4)	.967	1.026	1.048	.993	.986	1.042
BDVAR(4)	*.960*	*.954*	*.879*	1.031	1.047	1.209
VAR(4), inflation detrending	.982	.978	.942	.970	*.910*	.897
VAR(AIC), intercept breaks	.975	.973	.930	1.022	1.014	1.101
VAR(4), rolling	1.024	1.108	1.139	1.136	1.134	1.437
DVAR(4), rolling	1.013	1.017	.942	1.059	.971	1.123
VAR(AIC, by eq.&var.), rolling	1.017	1.166	1.167	1.145	1.152	1.579
BVAR(4), rolling	.958	1.010	1.022	1.088	1.190	1.525
BDVAR(4), rolling	.966	.978	*.917*	1.076	1.107	1.261
TVP BVAR(4)	.959	1.010	1.043	.996	1.001	1.169
Intercept TVP BVAR(4)	.958	1.004	1.018	.998	1.000	1.153
DLS, VAR(4)	1.139	1.311	1.322	1.208	1.176	1.368
DLS, DVAR(4)	1.350	1.257	1.236	1.166	1.100	1.251
average of all forecasts	*.935*	.957	*.907*	1.005	.991	1.035
avg. of VAR(4), rolling VAR(4)	1.014	1.065	1.098	1.023	.986	1.081
avg. of univariate, VAR(4)	.968	.982	.990	*.967*	*.944*	*.930*
avg. of univariate, DVAR(4)	.963	*.947*	*.926*	*.966*	*.946*	.982
avg. of univ., IDTR VAR(4)	*.954*	*.957*	*.924*	*.963*	**.934**	*.894*
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	*.951*	.960	.954	*.966*	*.942*	.983

Notes:

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

Table 4: Real-time RMSE results for GDP growth and CPI inflation

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

	GDP growth forecasts					
	1970-84			**1985-2005**		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univariate	4.183	4.761	3.652	1.609	1.668	1.293
VAR(4)	1.039	.945	.926	1.155	1.172	1.103
VAR(4), intercept correction	1.054	.952	1.004	1.156	1.197	1.367
VAR(AIC)	.981	.948	1.031	1.142	1.157	1.084
DVAR(4)	1.093	.959	*.767*	1.236	1.264	1.159
DVAR(AIC)	1.058	.983	.947	1.236	1.264	1.159
VAR(AIC, by eq.&var.)	.937	*.873*	.926	1.113	1.121	.974
DVAR(AIC, by eq.&var.)	1.043	.944	*.773*	1.200	1.254	1.151
BVAR(4)	*.919*	*.871*	.917	1.061	1.071	.982
BDVAR(4)	.987	*.958*	*.958*	1.035	1.041	1.014
VAR(4), inflation detrending	.977	.863	*.793*	1.324	1.380	1.341
VAR(AIC), intercept breaks	.935	.925	.963	1.413	1.504	1.498
VAR(4), rolling	1.135	1.061	1.049	1.363	1.348	1.333
DVAR(4), rolling	1.114	1.019	*.813*	1.179	1.190	1.178
VAR(AIC, by eq.&var.), rolling	1.011	.976	1.078	1.343	1.297	1.311
BVAR(4), rolling	.935	*.872*	.971	1.224	1.236	1.211
BDVAR(4), rolling	1.009	.991	1.004	1.036	1.045	1.066
TVP BVAR(4)	*.925*	*.893*	.929	1.009	1.015	.952
Intercept TVP BVAR(4)	*.921*	*.888*	.916	1.007	1.007	*.919*
DLS, VAR(4)	1.071	1.077	1.017	1.170	1.170	1.129
DLS, DVAR(4)	1.104	1.041	.909	1.191	1.182	1.186
average of all forecasts	*.904*	**.843**	*.826*	1.090	1.100	1.037
avg. of VAR(4), rolling VAR(4)	1.067	.982	.952	1.210	1.225	1.192
avg. of univariate, VAR(4)	.969	*.914*	*.879*	1.044	1.042	.985
avg. of univariate, DVAR(4)	.976	*.909*	**.807**	1.075	1.080	1.031
avg. of univ., IDTR VAR(4)	*.937*	*.873*	*.807*	1.083	1.091	1.019
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	*.944*	*.872*	*.800*	1.063	1.070	1.003

Table 4, continued: RMSE results for GDP growth and CPI inflation

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

	CPI inflation forecasts					
	1970-84			1985-2005		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univariate	2.117	2.733	2.970	1.347	1.475	1.247
VAR(4)	*.866*	.957	1.016	.975	1.028	1.078
VAR(4), intercept correction	.885	1.046	1.152	1.188	1.393	1.963
VAR(AIC)	.895	1.001	1.045	.975	1.022	1.064
DVAR(4)	*.847*	.888	*.854*	.952	1.006	1.095
DVAR(AIC)	*.868*	.917	*.889*	.952	1.006	1.095
VAR(AIC, by eq.&var.)	.907	.993	1.045	.970	1.022	1.095
DVAR(AIC, by eq.&var.)	*.851*	*.894*	*.869*	.952	.982	1.066
BVAR(4)	*.926*	1.037	1.120	.986	.985	.999
BDVAR(4)	*.848*	*.912*	*.933*	.977	1.009	1.065
VAR(4), inflation detrending	*.824*	*.889*	*.822*	.985	1.054	1.191
VAR(AIC), intercept breaks	.895	1.024	1.063	1.025	1.081	1.208
VAR(4), rolling	.880	1.020	1.094	1.127	1.242	1.430
DVAR(4), rolling	*.847*	.939	.916	1.025	1.093	1.255
VAR(AIC, by eq.&var.), rolling	.950	1.099	1.181	1.113	1.173	1.383
BVAR(4), rolling	*.928*	1.026	1.066	1.028	1.056	1.170
BDVAR(4), rolling	*.869*	*.933*	*.955*	1.005	1.042	1.114
TVP BVAR(4)	*.914*	1.014	1.090	.979	.970	*.936*
Intercept TVP BVAR(4)	*.914*	1.001	1.043	.986	.981	.979
DLS, VAR(4)	1.007	1.357	1.603	1.262	1.264	1.407
DLS, DVAR(4)	1.031	1.153	1.082	1.194	1.216	1.451
average of all forecasts	*.831*	.931	.962	.989	1.025	1.099
avg. of VAR(4), rolling VAR(4)	*.863*	.983	1.047	1.011	1.075	1.138
avg. of univariate, VAR(4)	*.868*	*.920*	*.935*	*.959*	.989	.997
avg. of univariate, DVAR(4)	*.862*	*.898*	*.894*	*.944*	.980	1.013
avg. of univ., IDTR VAR(4)	**.857**	**.895**	**.863**	*.962*	.993	1.021
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	*.851*	*.915*	*.933*	*.950*	.978	.990

Notes:

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

Table 5: Real-time RMSE results for the HPS output gap and CPI inflation

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

	HPS output gap forecasts					
	1970-84			1985-2005		
forecast method	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
univariate	1.039	1.988	3.891	.702	1.028	2.044
VAR(4)	1.066	.980	.943	1.097	1.142	1.162
VAR(4), intercept correction	1.096	1.030	1.092	1.054	1.038	1.063
VAR(AIC)	.991	.979	1.017	1.086	1.135	1.155
DVAR(4)	1.146	1.014	*.790*	1.088	1.123	1.091
DVAR(AIC)	1.036	.992	.998	1.077	1.097	1.043
VAR(AIC, by eq.&var.)	1.011	1.007	1.003	1.074	1.124	1.136
DVAR(AIC, by eq.&var.)	1.064	.945	*.760*	1.069	1.109	1.085
BVAR(4)	1.022	.941	.906	1.060	1.123	1.127
BDVAR(4)	1.065	1.005	.997	.995	.996	.992
VAR(4), inflation detrending	1.037	.916	.839	1.070	1.111	1.089
VAR(AIC), intercept breaks	.990	.961	*.778*	1.132	1.233	1.186
VAR(4), rolling	1.206	1.170	1.163	1.143	1.228	1.274
DVAR(4), rolling	1.163	1.062	*.909*	1.076	1.098	1.056
VAR(AIC, by eq.&var.), rolling	1.170	1.097	1.133	1.119	1.190	1.190
BVAR(4), rolling	1.068	.983	.998	1.093	1.173	1.189
BDVAR(4), rolling	1.093	1.049	1.063	.999	1.008	.995
TVP BVAR(4)	1.031	.971	.953	*.972*	*.961*	*.916*
Intercept TVP BVAR(4)	1.025	.957	.926	*.972*	*.959*	*.913*
DLS, VAR(4)	1.089	1.085	.973	1.055	1.059	1.084
DLS, DVAR(4)	1.156	1.094	.961	1.076	1.084	1.055
average of all forecasts	.967	.909	*.864*	1.026	1.036	1.003
avg. of VAR(4), rolling VAR(4)	1.108	1.051	1.022	1.100	1.157	1.190
avg. of univariate, VAR(4)	.992	.937	*.892*	1.029	1.035	1.015
avg. of univariate, DVAR(4)	1.012	.944	***.828***	1.032	1.043	1.028
avg. of univ., IDTR VAR(4)	.964	*.885*	*.817*	1.012	1.013	.985
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	.998	.927	*.838*	1.020	1.021	.988

Table 5, continued: RMSE results for the HPS output gap and CPI inflation

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

	CPI inflation forecasts					
	1970-84			**1985-2005**		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univariate	2.117	2.733	2.970	1.347	1.475	1.247
VAR(4)	.906	1.012	1.108	.967	1.012	1.016
VAR(4), intercept correction	.910	1.077	1.187	1.180	1.372	1.856
VAR(AIC)	*.874*	.937	1.027	.960	.987	.960
DVAR(4)	.902	.959	.946	.964	1.005	1.055
DVAR(AIC)	*.821*	*.896*	.934	.974	1.007	1.063
VAR(AIC, by eq.&var.)	.943	1.002	1.087	.962	1.021	1.052
DVAR(AIC, by eq.&var.)	.880	.921	.938	.955	1.006	1.075
BVAR(4)	*.925*	1.021	1.089	.981	.976	.949
BDVAR(4)	*.847*	*.901*	.928	.983	1.030	1.123
VAR(4), inflation detrending	*.860*	.932	.900	.944	.964	*.865*
VAR(AIC), intercept breaks	.889	.943	.963	1.021	1.101	1.176
VAR(4), rolling	.912	1.105	1.250	1.141	1.237	1.331
DVAR(4), rolling	.912	1.046	1.072	1.018	1.052	1.116
VAR(AIC, by eq.&var.), rolling	.946	1.064	1.135	1.067	1.113	1.200
BVAR(4), rolling	*.940*	1.029	1.062	1.019	1.041	1.127
BDVAR(4), rolling	*.883*	.946	.992	1.007	1.050	1.143
TVP BVAR(4)	*.916*	1.010	1.102	.996	1.019	1.069
Intercept TVP BVAR(4)	*.915*	.998	1.060	.997	1.016	1.058
DLS, VAR(4)	1.062	1.375	1.623	1.287	1.331	1.523
DLS, DVAR(4)	1.132	1.216	1.258	1.178	1.202	1.399
average of all forecasts	*.834*	.920	.939	.984	1.011	1.038
avg. of VAR(4), rolling VAR(4)	.897	1.052	1.167	1.017	1.065	1.050
avg. of univariate, VAR(4)	*.882*	.945	.968	*.957*	.985	.981
avg. of univariate, DVAR(4)	*.886*	*.932*	*.931*	*.944*	.972	.976
avg. of univ., IDTR VAR(4)	**.861**	*.894*	*.828*	*.941*	*.950*	**.877**
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	*.873*	.948	.981	*.951*	.977	.972

Notes:

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

Table 6: Average forecast accuracy rankings, across applications and methods in Tables 2-5

(sorted low to high)

method	*average*	*st. dev.*
avg. of univ., IDTR VAR(4)	5.1	2.8
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	5.7	2.6
avg. of univariate and DVAR(4)	6.8	3.1
avg. of univariate and VAR(4)	7.7	2.9
average of all forecasts	8.0	4.9
Intercept TVP BVAR(4)	9.8	6.4
BDVAR(4)	10.7	6.4
TVP BVAR(4)	10.8	6.9
VAR(4), inflation detrending	10.8	7.5
DVAR(AIC)	11.2	6.6
univariate	12.1	6.7
DVAR(4)	12.2	7.9
DVAR(AIC, by eq.&var.)	12.5	6.2
BVAR(4)	12.6	6.3
BDVAR(4), rolling	14.2	7.1
VAR(AIC, by eq.&var.)	14.4	6.2
VAR(4)	14.8	5.6
VAR(AIC)	15.0	5.8
DVAR(4), rolling	15.9	6.3
VAR(AIC), AIC intercept breaks	17.3	7.9
BVAR(4), rolling	18.5	5.9
avg. of VAR(4) and rolling VAR(4)	19.1	3.7
VAR(4), intercept correction	21.0	4.6
DLS, DVAR(4)	21.4	5.2
DLS, VAR(4)	22.3	5.4
VAR(AIC, by eq.&var.), rolling	23.9	2.6
VAR(4), rolling	24.4	2.8

Notes:

1. The figures in the table are obtained by: (1) ranking, for each of the 48 columns of Tables 2-5, the 27 forecast methods or models considered; and (2) calculating the average and standard deviation of each method's (48) ranks.

Table 7: Average RMSEs, across applications and methods in Tables 2-5

(sorted low to high)

method	*average*	*st. dev.*
avg. of univ., IDTR VAR(4)	.943	.070
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	.955	.068
avg. of univariate and DVAR(4)	.960	.072
average of all forecasts	.967	.082
avg. of univariate and VAR(4)	.968	.050
Intercept TVP BVAR(4)	.981	.056
TVP BVAR(4)	.987	.058
BDVAR(4)	.995	.064
univariate	1.000	.000
VAR(4), inflation detrending	1.001	.143
DVAR(AIC)	1.004	.109
DVAR(4)	1.009	.130
DVAR(AIC, by eq.&var.)	1.011	.117
BVAR(4)	1.012	.076
BDVAR(4), rolling	1.025	.072
VAR(AIC, by eq.&var.)	1.025	.074
VAR(4)	1.030	.087
VAR(AIC)	1.031	.078
DVAR(4), rolling	1.036	.107
avg. of VAR(4) and rolling VAR(4)	1.068	.088
BVAR(4), rolling	1.081	.132
VAR(AIC), AIC intercept breaks	1.088	.196
DLS, DVAR(4)	1.141	.113
VAR(4), intercept correction	1.149	.204
VAR(AIC, by eq.&var.), rolling	1.157	.132
VAR(4), rolling	1.173	.128
DLS, VAR(4)	1.184	.156

Notes:

1. The figures in the table are simple averages and standard deviations, across the 48 columns of Tables 2-5, of each forecast method's RMSE ratios. Note that the RMSE ratio of the univariate forecast is always 1.

Table 8: Real-time 1996-2005 RMSE results for GDP growth and GDP inflation

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

	GDP growth forecasts			**GDP inflation forecasts**		
forecast method	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
univariate	1.624	1.691	1.283	.762	.841	.717
VAR(4)	1.228	1.223	1.104	.964	.965	.973
VAR(4), intercept correction	1.249	1.260	1.295	1.105	1.156	1.476
VAR(AIC)	1.228	1.223	1.104	.964	.965	.973
DVAR(4)	1.254	1.231	1.065	1.013	1.032	1.084
DVAR(AIC)	1.245	1.230	1.065	1.011	1.028	1.085
VAR(AIC, by eq.&var.)	1.176	1.193	1.052	.970	.976	.970
DVAR(AIC, by eq.&var.)	1.184	1.193	1.049	1.015	1.023	1.077
BVAR(4)	1.102	1.132	1.065	1.015	1.038	1.110
BDVAR(4)	1.053	1.032	.981	1.017	1.040	1.097
VAR(4), inflation detrending	1.222	1.228	1.057	.974	.968	.953
VAR(AIC), intercept breaks	1.263	1.314	1.204	.977	.980	.995
VAR(4), rolling	1.000	1.044	1.051	1.117	1.115	1.184
DVAR(4), rolling	1.058	1.099	1.176	1.069	1.022	1.047
VAR(AIC, by eq.&var.), rolling	1.125	1.131	1.039	1.105	1.081	1.201
BVAR(4), rolling	1.033	1.052	.986	1.042	1.094	1.255
BDVAR(4), rolling	1.036	1.037	1.080	1.030	1.064	1.156
TVP BVAR(4)	1.065	1.083	1.012	.998	1.001	1.016
Intercept TVP BVAR(4)	1.058	1.072	.987	1.008	1.019	1.051
DLS, VAR(4)	1.178	1.183	1.091	1.200	1.129	1.173
DLS, DVAR(4)	1.180	1.179	1.089	1.215	1.176	1.160
average of all forecasts	1.082	1.084	.991	1.000	1.004	1.046
avg. of VAR(4), rolling VAR(4)	1.073	1.101	1.049	1.018	1.021	1.058
avg. of univariate, VAR(4)	1.084	1.069	.975	*.949*	*.949*	*.933*
avg. of univariate, DVAR(4)	1.100	1.080	.989	.967	.975	.979
avg. of univ., IDTR VAR(4)	1.070	1.057	.925	*.952*	*.951*	*.923*
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	1.108	1.098	.991	.963	.968	.970

Notes:

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

Table 9: Real-time 1996-2005 RMSE results for GDP growth and core PCE inflation

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

forecast method	GDP growth forecasts $h=0Q$	GDP growth forecasts $h=1Q$	GDP growth forecasts $h=1Y$	core PCE forecasts $h=0Q$	core PCE forecasts $h=1Q$	core PCE forecasts $h=1Y$
univariate	1.624	1.691	1.283	.646	.602	.460
VAR(4)	1.223	1.174	1.077	1.233	1.339	1.630
VAR(4), intercept correction	1.238	1.237	1.180	1.316	1.599	2.301
VAR(AIC)	1.223	1.174	1.077	1.233	1.339	1.630
DVAR(4)	1.171	1.134	.976	1.200	1.297	1.322
DVAR(AIC)	1.171	1.134	.976	1.200	1.297	1.322
VAR(AIC, by eq.&var.)	1.251	1.239	1.151	1.253	1.455	1.949
DVAR(AIC, by eq.&var.)	1.204	1.173	1.019	1.186	1.252	1.264
BVAR(4)	1.175	1.165	1.130	1.224	1.376	1.819
BDVAR(4)	1.049	1.007	.958	1.167	1.234	1.243
VAR(4), inflation detrending	1.231	1.195	1.061	1.212	1.284	1.394
VAR(AIC), intercept breaks	1.425	1.536	1.604	1.222	1.384	1.578
VAR(4), rolling	1.014	1.034	1.076	.981	1.166	1.580
DVAR(4), rolling	.982	1.002	1.137	.938	1.077	1.060
VAR(AIC, by eq.&var.), rolling	1.157	1.115	1.174	1.024	1.261	1.670
BVAR(4), rolling	1.067	1.071	1.053	1.176	1.314	1.764
BDVAR(4), rolling	1.024	1.034	1.079	1.105	1.159	1.162
TVP BVAR(4)	1.090	1.081	1.028	1.161	1.257	1.459
Intercept TVP BVAR(4)	1.089	1.073	1.001	1.198	1.319	1.624
DLS, VAR(4)	1.168	1.146	1.051	1.122	1.458	1.551
DLS, DVAR(4)	1.150	1.108	1.072	1.123	1.505	1.387
average of all forecasts	1.093	1.068	.988	1.117	1.199	1.326
avg. of VAR(4), rolling VAR(4)	1.081	1.072	1.052	1.052	1.172	1.489
avg. of univariate, VAR(4)	1.074	1.042	.947	1.089	1.137	1.260
avg. of univariate, DVAR(4)	1.064	1.038	.955	1.076	1.120	1.108
avg. of univ., IDTR VAR(4)	1.069	1.038	.921	1.081	1.117	1.156
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	1.091	1.061	.960	1.123	1.187	1.275

Notes:

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

Table 10: Real-time 1996-2005 RMSE results for the HPS output gap and GDP inflation

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

forecast method	HPS gap forecasts $h=0Q$	HPS gap forecasts $h=1Q$	HPS gap forecasts $h=1Y$	GDP inflation forecasts $h=0Q$	GDP inflation forecasts $h=1Q$	GDP inflation forecasts $h=1Y$
univariate	.714	1.036	2.075	.762	.841	.717
VAR(4)	1.121	1.131	1.155	.981	.990	1.073
VAR(4), intercept correction	1.091	1.052	1.114	1.098	1.163	1.527
VAR(AIC)	1.118	1.130	1.158	.976	.985	1.067
DVAR(4)	1.119	1.086	1.084	1.029	1.056	1.226
DVAR(AIC)	1.130	1.088	1.089	1.025	1.062	1.246
VAR(AIC, by eq.&var.)	1.075	1.117	1.137	.980	.995	1.085
DVAR(AIC, by eq.&var.)	1.077	1.075	1.098	1.003	1.053	1.287
BVAR(4)	1.057	1.116	1.147	1.013	1.040	1.166
BDVAR(4)	1.007	.985	.983	1.031	1.075	1.274
VAR(4), inflation detrending	1.088	1.097	1.073	.976	.959	1.005
VAR(AIC), intercept breaks	1.066	1.041	.959	1.047	1.073	1.217
VAR(4), rolling	1.040	1.147	1.234	1.116	1.179	1.335
DVAR(4), rolling	1.010	1.024	1.060	1.081	1.046	1.161
VAR(AIC, by eq.&var.), rolling	1.037	1.071	1.160	1.116	1.171	1.322
BVAR(4), rolling	1.043	1.128	1.246	1.052	1.124	1.324
BDVAR(4), rolling	.991	1.000	1.063	1.043	1.091	1.259
TVP BVAR(4)	.994	1.004	.997	1.032	1.078	1.276
Intercept TVP BVAR(4)	.989	.991	.971	1.030	1.071	1.266
DLS, VAR(4)	1.062	1.047	1.060	1.270	1.309	1.483
DLS, DVAR(4)	1.067	1.022	1.050	1.282	1.245	1.402
average of all forecasts	1.028	1.029	1.049	1.004	1.022	1.123
avg. of VAR(4), rolling VAR(4)	1.052	1.091	1.134	1.025	1.059	1.178
avg. of univariate, VAR(4)	1.047	1.038	1.026	.956	.961	.982
avg. of univariate, DVAR(4)	1.051	1.032	1.031	.971	.979	1.028
avg. of univ., IDTR VAR(4)	1.033	1.028	1.003	*.948*	*.937*	*.926*
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	1.044	1.030	1.019	.978	.995	1.082

Notes:

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

Table 11: Real-time 1996-2005 RMSE results for the HPS output gap and core PCE inflation

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

	HPS gap forecasts			core PCE forecasts		
forecast method	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
univariate	.714	1.036	2.075	.646	.602	.460
VAR(4)	1.053	1.078	1.165	1.162	1.216	1.384
VAR(4), intercept correction	1.020	1.006	1.088	1.267	1.472	2.071
VAR(AIC)	1.071	1.110	1.190	1.129	1.200	1.429
DVAR(4)	1.033	1.021	1.041	1.161	1.289	1.409
DVAR(AIC)	1.050	1.025	1.027	1.153	1.242	1.362
VAR(AIC, by eq.&var.)	1.083	1.128	1.206	1.198	1.315	1.703
DVAR(AIC, by eq.&var.)	1.071	1.055	1.078	1.147	1.230	1.232
BVAR(4)	1.071	1.145	1.219	1.172	1.275	1.553
BDVAR(4)	.985	*.974*	.987	1.153	1.248	1.358
VAR(4), inflation detrending	1.061	1.084	1.102	1.117	1.161	1.093
VAR(AIC), intercept breaks	1.055	1.112	1.161	1.252	1.423	1.905
VAR(4), rolling	.999	1.104	1.312	1.006	1.155	1.622
DVAR(4), rolling	*.938*	.954	1.023	.925	1.081	1.087
VAR(AIC, by eq.&var.), rolling	1.075	1.138	1.312	1.018	1.165	1.529
BVAR(4), rolling	1.054	1.138	1.307	1.196	1.355	1.894
BDVAR(4), rolling	.975	.996	1.061	1.110	1.175	1.210
TVP BVAR(4)	.985	.997	1.009	1.151	1.260	1.515
Intercept TVP BVAR(4)	.981	.986	.989	1.160	1.265	1.499
DLS, VAR(4)	.997	1.005	1.046	1.165	1.504	1.689
DLS, DVAR(4)	.994	.987	1.040	1.115	1.572	1.505
average of all forecasts	1.007	1.019	1.070	1.093	1.174	1.290
avg. of VAR(4), rolling VAR(4)	.999	1.048	1.187	1.044	1.141	1.452
avg. of univariate, VAR(4)	1.010	1.011	1.028	1.057	1.074	1.114
avg. of univariate, DVAR(4)	1.008	.999	1.012	1.048	1.086	1.032
avg. of univ., IDTR VAR(4)	1.014	1.015	1.015	1.032	1.042	.937
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	1.003	.998	1.011	1.089	1.139	1.182

Notes:

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

Table 12: Average rankings of all methods in 1970-84 and 1985-2005 forecasts, across all models and data

	all	all y, p	all y, p using Tbill	all y, p using FFR	all y, p 70-84	all y, p 85-05
avg. of univ., IDTR VAR(2)	12.9	16.7	15.5	18.0	21.1	12.4
avg. of univ., IDTR VAR(4)	13.2	13.4	12.7	14.1	15.1	11.6
avg. of univ., VAR(2), DVAR(2), TVP BVAR(2)	15.7	19.0	18.8	19.1	22.2	15.7
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	17.6	16.2	16.6	15.7	17.9	14.4
avg. of univariate, VAR(2)	18.8	23.7	22.3	25.1	31.3	16.0
average of all forecasts	19.7	18.8	19.1	18.5	11.5	26.1
avg. of univariate, VAR(4)	20.3	20.6	20.9	20.3	26.2	15.0
avg. of univariate, DVAR(4)	21.3	19.9	19.8	20.1	21.3	18.6
avg. of univariate, DVAR(2)	22.9	24.1	23.9	24.2	27.1	21.0
Intercept TVP BVAR(4)	25.1	28.1	27.4	28.9	38.4	17.9
VAR(2), inflation detrending	25.2	29.0	27.0	31.0	21.8	36.1
Intercept TVP BVAR(4), $\lambda_4 = .5, \lambda = .0025$	26.4	27.4	27.4	27.3	27.1	27.7
BDVAR(4)	27.0	28.7	27.2	30.1	30.8	26.5
TVP BVAR(4), $\lambda_4 = .5, \lambda = .0025$	28.2	23.8	23.5	24.0	30.4	17.1
TVP BVAR(4), $\lambda_4 = 1000, \lambda = .005$	29.1	23.4	22.9	23.9	30.0	16.8
TVP BVAR(4), $\lambda_4 = 1000, \lambda = .0001$	29.4	31.2	31.0	31.3	36.2	26.1
TVP BVAR(4)	29.7	29.3	28.7	29.9	42.8	15.8
BVAR(4)	30.1	32.9	33.0	32.8	37.3	28.4
Intercept TVP BVAR(2), $\lambda_4 = .5, \lambda = .0025$	30.6	36.4	35.3	37.5	35.1	37.7
Intercept TVP BVAR(2)	31.1	38.5	36.9	40.1	50.3	26.7
TVP BVAR(2), $\lambda_4 = .5, \lambda = .0025$	31.8	31.6	31.0	32.2	36.4	26.9
BDVAR(2)	32.0	34.4	33.2	35.6	36.9	31.9
TVP BVAR(2), $\lambda_4 = 1000, \lambda = .005$	32.2	30.2	29.2	31.1	36.3	24.0
VAR(4), inflation detrending	32.6	31.6	31.2	32.0	25.8	37.4
DVAR(2)	32.8	31.3	31.1	31.5	25.0	37.6
avg. of VAR(2), rolling VAR(2)	33.3	40.0	38.9	41.0	38.3	41.6
TVP BVAR(2), $\lambda_4 = 1000, \lambda = .0001$	33.3	40.0	38.9	41.0	45.7	34.2
univariate	33.6	36.5	34.0	38.9	52.2	20.8
BVAR(2)	34.2	41.6	40.5	42.8	46.9	36.3
TVP BVAR(2)	34.7	38.6	37.1	40.2	53.5	23.7
DVAR(AIC)	34.8	33.1	32.5	33.7	29.1	37.1
VAR(AIC), inflation detrending	34.9	32.8	32.9	32.7	25.7	39.9
VAR(BIC), inflation detrending	35.0	40.0	38.7	41.2	36.3	43.6
BDVAR(4), rolling	35.2	38.4	36.9	39.9	40.3	36.5
VAR(2)	35.5	41.0	37.7	44.3	48.7	33.3
DVAR(BIC, by eq.&var.)	37.6	33.5	33.6	33.4	36.6	30.4
VAR(AIC, by eq.&var.)	38.8	37.2	38.6	35.8	44.8	29.6
DVAR(BIC)	38.9	37.2	37.6	36.8	34.7	39.7
DVAR(AIC, by eq.&var.)	39.4	34.4	35.3	33.5	33.1	35.7
DVAR(4)	39.4	35.2	35.9	34.6	31.0	39.5
BDVAR(2), rolling	39.6	44.3	42.9	45.7	45.5	43.1
VAR(4)	40.8	41.0	41.9	40.1	45.6	36.4
DVAR(2), output diff.	41.0	42.8	44.2	41.3	39.9	45.6
DVAR(2), rolling	41.2	39.2	39.8	38.7	32.9	45.6

Table 12, continued: Average rankings across all results

	all	all y,p	all y,p using Tbill	all y,p using FFR	all y,p 70-84	all y,p 85-05
VAR(AIC)	41.3	40.0	40.8	39.3	45.1	35.0
VAR(2), full ES detrending	41.3	40.8	41.4	40.2	17.7	63.9
VAR(BIC, by eq.&var.)	43.8	44.8	43.7	45.9	55.7	33.8
DVAR(AIC), output diff.	44.2	45.1	45.6	44.6	43.4	46.9
DVAR(4), output diff.	45.9	45.1	45.5	44.6	43.5	46.6
VAR(BIC), full ES detrending	46.1	46.4	46.9	45.9	29.0	63.8
VAR(BIC)	46.2	52.2	50.1	54.3	61.5	42.9
DVAR(AIC), rolling	46.5	40.2	39.1	41.3	37.2	43.2
VAR(AIC), full ES detrending	47.6	42.3	45.2	39.4	24.6	60.1
VAR(4), full ES detrending	48.8	45.6	49.6	41.7	27.9	63.4
BVAR(4), rolling	49.3	51.9	52.8	51.0	41.5	62.4
BVAR(2), rolling	49.6	54.8	54.9	54.7	43.5	66.1
DVAR(BIC), rolling	49.6	49.1	50.0	48.2	45.0	53.2
DVAR(BIC), output diff.	49.9	52.2	54.6	49.9	55.2	49.3
DVAR(AIC, by eq.&var.), rolling	50.3	41.1	44.2	38.1	39.6	42.7
DVAR(2), output diff., rolling	51.3	54.4	56.9	51.9	51.5	57.3
DVAR(BIC, by eq.&var.), rolling	51.8	46.9	48.7	45.0	48.3	45.5
VAR(AIC), BIC intercept breaks	52.7	47.5	47.7	47.3	30.7	64.4
avg. of VAR(4), rolling VAR(4)	53.6	52.0	52.7	51.3	53.8	50.1
VAR(AIC), AIC intercept breaks	55.4	49.0	47.7	50.3	32.9	65.1
DVAR(4), rolling	55.9	47.9	47.3	48.6	44.7	51.2
DLS, VAR(2)	56.1	56.8	54.8	58.7	65.2	48.3
VAR(2), intercept correction	56.9	60.0	59.5	60.6	61.2	58.8
DVAR(BIC), output diff., rolling	57.8	64.9	66.9	63.0	63.8	66.1
DVAR(AIC), output diff., rolling	59.0	57.3	57.7	56.8	55.2	59.3
DLS, DVAR(2)	59.4	55.7	56.7	54.8	57.2	54.3
VAR(2), rolling	62.5	65.3	65.3	65.3	56.4	74.2
DVAR(4), output diff., rolling	63.5	60.5	59.4	61.6	56.3	64.7
DLS, DVAR(AIC)	63.9	59.3	59.3	59.3	63.7	54.9
VAR(AIC), intercept correction	64.0	63.6	62.7	64.5	61.1	66.2
VAR(4), intercept correction	64.9	64.4	64.3	64.5	63.8	65.1
DLS, VAR(AIC)	65.5	63.4	64.0	62.7	73.2	53.6
VAR(BIC), rolling	65.7	69.9	71.8	68.0	66.0	73.8
DLS, DVAR(4)	68.7	63.6	64.5	62.6	67.9	59.3
VAR(BIC, by eq.&var.), rolling	68.8	69.3	69.7	68.9	65.5	73.1
VAR(AIC, by eq.&var.), rolling	69.2	69.2	71.4	67.1	64.7	73.8
VAR(AIC), rolling	69.7	69.7	70.2	69.2	66.5	72.8
DLS, VAR(4)	69.8	66.7	68.5	64.9	75.8	57.5
VAR(2), partial int. corr.	72.1	67.4	67.1	67.7	72.2	62.6
VAR(4), rolling	72.4	71.8	71.8	71.9	68.1	75.5
VAR(AIC), partial int. corr.	76.4	72.9	72.8	73.0	74.8	71.0
VAR(4), partial int. corr.	76.5	73.1	73.9	72.2	74.9	71.3
# of ranking observations	216	144	72	72	72	72

Notes:
1. The table reports average rankings of the full set of forecast methods or models listed in Table 1. The average rankings in the first column of figures are calculated, for each forecast method, across a total of 216 (= $3 \times 2 \times 2 \times 3 \times 2 \times 3$) forecasts of output (3: GDP growth, HPS gap, HP gap), inflation (2: GDP inflation, CPI inflation), and interest rates (2: T-bill rate, federal funds rate) at horizons (3) of $h = 0Q$, $h = 1Q$, and $h = 1Y$ and sample periods (2) of 1970-84 and 1985-05. The average rankings in remaining columns are based on forecasts with models that include particular variables or forecasts of a particular variable, etc. For example, the average rankings in the second column are based on 144 forecasts of just output and inflation, with forecasts of interest rates omitted from the average ranking calculation.
2. See the notes to Table 2.

Table 13: Average rankings in 1970-84 and 1985-2005 forecasts, conditioned on output and inflation measures

	using $\Delta\ln$ *GDP*	using HPS gap	using HP gap	using GDP π	using CPI π
avg. of univ., IDTR VAR(2)	21.1	12.7	16.3	16.7	16.7
avg. of univ., IDTR VAR(4)	16.8	8.9	14.4	13.0	13.8
avg. of univ., VAR(2), DVAR(2), TVP BVAR(2)	16.9	19.0	20.9	18.8	19.2
avg. of univ., VAR(4), DVAR(4), TVP BVAR(4)	13.3	15.6	19.6	14.1	18.2
avg. of univariate, VAR(2)	23.6	24.6	22.9	24.1	23.2
average of all forecasts	19.9	17.7	18.8	16.0	21.6
avg. of univariate, VAR(4)	18.5	20.6	22.7	19.7	21.5
avg. of univariate, DVAR(4)	17.9	18.4	23.6	18.0	21.9
avg. of univariate, DVAR(2)	23.4	22.6	26.2	23.7	24.5
Intercept TVP BVAR(4)	24.7	30.3	29.4	27.1	29.1
VAR(2), inflation detrending	37.4	20.2	29.3	33.5	24.5
Intercept TVP BVAR(4), $\lambda_4 = .5, \lambda = .0025$	30.1	29.6	22.5	27.1	27.6
BDVAR(4)	28.2	32.3	25.6	29.8	27.6
TVP BVAR(4), $\lambda_4 = .5, \lambda = .0025$	28.2	22.8	20.3	22.0	25.5
TVP BVAR(4), $\lambda_4 = 1000, \lambda = .005$	28.3	20.7	21.3	21.2	25.7
TVP BVAR(4), $\lambda_4 = 1000, \lambda = .0001$	30.5	39.1	23.9	31.5	30.8
TVP BVAR(4)	25.4	34.1	28.5	27.8	30.8
BVAR(4)	31.5	41.7	25.4	33.3	32.4
Intercept TVP BVAR(2), $\lambda_4 = .5, \lambda = .0025$	35.2	40.6	33.4	36.7	36.1
Intercept TVP BVAR(2)	32.9	41.7	40.9	38.3	38.7
TVP BVAR(2), $\lambda_4 = .5, \lambda = .0025$	31.5	33.7	29.7	29.8	33.4
BDVAR(2)	29.2	40.8	33.2	34.7	34.1
TVP BVAR(2), $\lambda_4 = 1000, \lambda = .005$	30.6	30.1	29.8	27.0	33.4
VAR(4), inflation detrending	36.0	25.8	33.1	31.9	31.3
DVAR(2)	28.8	31.9	33.1	33.0	29.6
avg. of VAR(2), rolling VAR(2)	36.2	46.6	37.1	43.8	36.1
TVP BVAR(2), $\lambda_4 = 1000, \lambda = .0001$	36.0	49.5	34.4	41.5	38.4
univariate	35.8	37.0	36.6	34.3	38.6
BVAR(2)	36.9	52.0	36.0	43.2	40.0
TVP BVAR(2)	32.2	43.9	39.8	37.9	39.4
DVAR(AIC)	31.9	32.4	34.9	30.6	35.5
VAR(AIC), inflation detrending	40.5	25.7	32.3	36.6	29.1
VAR(BIC), inflation detrending	47.7	33.5	38.7	43.9	36.0
BDVAR(4), rolling	38.2	42.9	34.1	38.6	38.2
VAR(2)	37.2	47.2	38.5	47.3	34.7
DVAR(BIC, by eq.&var.)	40.8	37.0	22.8	34.8	32.2
VAR(AIC, by eq.&var.)	31.6	44.2	35.9	39.0	35.5
DVAR(BIC)	41.0	34.9	35.7	38.6	35.8
DVAR(AIC, by eq.&var.)	35.9	35.1	32.2	37.4	31.4
DVAR(4)	34.1	35.3	36.3	33.4	37.1
BDVAR(2), rolling	40.3	51.7	41.0	43.7	45.0
VAR(4)	37.5	45.4	40.0	40.0	41.9
DVAR(2), output diff.	43.6	37.2	47.4	44.5	41.0
DVAR(2), rolling	39.2	40.4	38.1	40.0	38.5

Table 13, continued: average rankings, conditioned on output and inflation measures

	using $\Delta \ln$ *GDP*	using HPS gap	using HP gap	using GDP π	using CPI π
VAR(AIC)	40.1	44.1	36.0	45.4	34.7
VAR(2), full ES detrending	43.2	32.6	46.6	41.6	40.0
VAR(BIC, by eq.&var.)	45.5	53.2	35.7	47.1	42.5
DVAR(AIC), output diff.	47.9	38.4	49.1	38.5	51.7
DVAR(4), output diff.	56.4	35.9	42.9	42.5	47.7
VAR(BIC), full ES detrending	47.7	42.3	49.2	44.5	48.2
VAR(BIC)	47.1	63.8	45.7	57.3	47.0
DVAR(AIC), rolling	41.2	38.4	41.2	33.9	46.6
VAR(AIC), full ES detrending	42.0	35.8	49.2	43.3	41.4
VAR(4), full ES detrending	41.4	42.0	53.5	44.5	46.8
BVAR(4), rolling	48.8	57.0	49.9	50.8	53.0
BVAR(2), rolling	47.5	60.1	56.7	53.5	56.1
DVAR(BIC), rolling	50.4	49.8	47.0	49.7	48.4
DVAR(BIC), output diff.	50.3	47.7	58.7	55.1	49.4
DVAR(AIC, by eq.&var.), rolling	42.9	39.8	40.7	42.0	40.2
DVAR(2), output diff., rolling	55.1	47.2	60.9	54.3	54.4
DVAR(BIC, by eq.&var.), rolling	50.0	49.2	41.5	49.2	44.5
VAR(AIC), BIC intercept breaks	51.8	44.4	46.4	53.1	41.9
avg. of VAR(4), rolling VAR(4)	50.4	57.2	48.3	49.6	54.3
VAR(AIC), AIC intercept breaks	56.2	45.7	45.1	51.5	46.5
DVAR(4), rolling	45.3	46.9	51.7	45.9	50.0
DLS, VAR(2)	56.5	57.5	56.3	56.5	57.0
VAR(2), intercept correction	59.1	51.0	70.0	59.0	61.0
DVAR(BIC), output diff., rolling	64.9	62.2	67.8	65.5	64.4
DVAR(AIC), output diff., rolling	61.5	47.8	62.5	48.8	65.7
DLS, DVAR(2)	55.4	57.7	54.1	54.4	57.0
VAR(2), rolling	61.6	67.7	66.6	65.3	65.3
DVAR(4), output diff., rolling	67.6	49.7	64.2	58.5	62.5
DLS, DVAR(AIC)	62.3	60.3	55.3	56.2	62.4
VAR(AIC), intercept correction	63.6	59.0	68.3	64.7	62.5
VAR(4), intercept correction	64.3	60.0	68.9	62.9	65.9
DLS, VAR(AIC)	66.8	62.5	60.9	62.6	64.2
VAR(BIC), rolling	64.4	72.8	72.6	72.0	67.8
DLS, DVAR(4)	64.8	62.9	63.0	61.7	65.4
VAR(BIC, by eq.&var.), rolling	64.3	74.8	68.8	69.8	68.8
VAR(AIC, by eq.&var.), rolling	66.6	72.6	68.5	71.0	67.5
VAR(AIC), rolling	70.5	72.9	65.5	69.0	70.3
DLS, VAR(4)	68.8	64.5	66.8	65.5	67.8
VAR(2), partial int. corr.	66.8	57.8	77.5	68.0	66.8
VAR(4), rolling	70.7	75.0	69.7	70.5	73.1
VAR(AIC), partial int. corr.	70.7	67.4	80.5	74.2	71.6
VAR(4), partial int. corr.	72.1	66.1	81.2	73.7	72.5
# of ranking observations	48	48	48	72	72

Notes:
1. The results in this table are based on just forecasts of output and inflation (excluding forecast results for interest rates).
2. See the notes to Tables 2 and 12.

Table 14: Accuracy of select VAR forecasts compared to SPF forecasts

(RMSEs in all cases)

	GDP growth forecasts					
	1970-84			1985-2005		
	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
SPF	2.571	3.699	2.891	1.384	1.635	1.274
best forecast from Table 2	3.735	3.878	2.775	1.609	1.668	1.182
univariate forecast	4.183	4.761	3.652	1.609	1.668	1.293
TVP BVAR(4)	3.876	4.267	3.487	1.650	1.708	1.218
avg. of all Table 2 forecasts	3.735	3.878	2.978	1.734	1.824	1.312
avg. of univ., DVAR(4)	3.953	4.199	2.906	1.747	1.828	1.328
avg. of univ., IDTR VAR(4)	3.893	4.145	3.101	1.705	1.770	1.232

	GDP inflation forecasts					
	1970-84			1985-2005		
	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
SPF	1.364	1.917	2.192	.831	.922	.804
best forecast from Table 2	1.727	2.036	2.141	.926	.961	.716
univariate forecast	1.825	2.153	2.389	.951	1.016	.760
TVP BVAR(4)	1.779	2.267	2.646	.944	.993	.764
avg. of all Table 2 forecasts	1.727	2.129	2.318	.974	1.032	.803
avg. of univ., DVAR(4)	1.764	2.051	2.224	.926	.970	.735
avg. of univ., IDTR VAR(4)	1.772	2.108	2.328	.937	.985	.744

	CPI inflation forecasts					
	1970-84			1985-2005		
	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
SPF				.823	1.278	.969
best forecast from Table 4	1.744	2.427	2.441	1.272	1.431	1.167
univariate forecast	2.117	2.733	2.970	1.347	1.475	1.247
TVP BVAR(4)	1.935	2.772	3.238	1.319	1.431	1.167
avg. of all Table 4 forecasts	1.758	2.544	2.856	1.333	1.511	1.370
avg. of univ., DVAR(4)	1.825	2.456	2.656	1.272	1.446	1.262
avg. of univ., IDTR VAR(4)	1.815	2.447	2.564	1.296	1.465	1.273

	T-bill rate forecasts					
	1970-84			1985-2005		
	$h = 0Q$	$h = 1Q$	$h = 1Y$	$h = 0Q$	$h = 1Q$	$h = 1Y$
SPF	.310	1.436	2.589	.104	.460	1.543
best forecast from Table 2	1.173	1.879	2.669	.371	.742	1.418
univariate forecast	1.305	2.098	2.821	.379	.777	1.633
TVP BVAR(4)	1.239	1.959	2.981	.407	.781	1.529
avg. of all Table 2 forecasts	1.182	1.920	2.834	.386	.764	1.555
avg. of univ., DVAR(4)	1.215	1.908	2.725	.389	.805	1.680
avg. of univ., IDTR VAR(4)	1.206	1.910	2.719	.371	.742	1.473

Notes:
1. The forecast errors are calculated using the first–available (real–time) estimates of output and inflation as the actual data on output and inflation.
2. RMSEs for SPF forecasts of CPI inflation are not reported for the 1970-84 sample because the SPF data don't begin until 1981.
3. See the notes to Table 2.

Table 15: Accuracy of SPF forecasts compared to Greenbook forecasts, in real time data

(RMSEs in all cases)

	GDP growth forecasts					
	1970-84			**1985-2000**		
	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
SPF	2.571	3.699	2.891	1.334	1.543	1.352
Greenbook	2.434	3.783	2.832	1.309	1.650	1.485
	GDP inflation forecasts					
	1970-84			**1985-2000**		
	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
SPF	1.364	1.917	2.192	.849	.932	.834
Greenbook	1.330	1.626	1.653	.691	.852	.670
	CPI inflation forecasts					
	1970-84			**1985-2000**		
	$h=0Q$	$h=1Q$	$h=1Y$	$h=0Q$	$h=1Q$	$h=1Y$
SPF				.700	1.206	.984
Greenbook				.603	1.160	.949

Notes:
1. The forecast errors are calculated using the first–available (real–time) estimates of output and inflation as the actual data on output and inflation.
2. RMSEs for forecasts of CPI inflation are not reported for the 1970-84 sample because the SPF and Greenbook data don't begin until circa 1980.
3. See the notes to Table 2.