

# Do DSGE Models Forecast More Accurately Out-of-Sample than VAR Models?\*

Refet S. Gürkaynak,<sup>†</sup>Burçin Kısacıkoglu,<sup>‡</sup>and Barbara Rossi<sup>§</sup>

April 20, 2013

## Abstract

Recently, it has been suggested that macroeconomic forecasts from estimated DSGE models tend to be more accurate out-of-sample than random walk forecasts or Bayesian VAR forecasts. Del Negro and Schorfheide (2013) in particular suggest that the DSGE model forecast should become the benchmark for forecasting horse-races. We compare the real-time forecasting accuracy of the Smets and Wouters DSGE model with that of several reduced-form time series models. We first demonstrate that none of the forecasting models is efficient. Our second finding is that there is no single best forecasting method. For example, typically simple AR models are most accurate at short horizons and DSGE models are most accurate at long horizons when forecasting output growth, while for inflation forecasts the results are reversed. Moreover, the relative accuracy of all models tends to evolve over time. Third, we show that there is no support the common practice of using large-scale Bayesian VAR models as the forecast benchmark when evaluating DSGE models. Indeed, low-dimensional unrestricted AR and VAR forecasts may forecast more accurately.

**Keywords:** Forecasting, forecast comparison, DSGE, Bayesian VAR, forecast optimality, regression-based tests of forecasting ability

**J.E.L. Codes:** C22, C52, C53

---

\***Acknowledgements:** We are grateful to Gergely Ganics and Lutz Kilian for comments on an earlier draft. The views expressed in this paper are those of the authors.

<sup>†</sup>Dept. of Economics, Bilkent University, 06800 Ankara, Turkey. E-mail: refet@bilkent.edu.tr.

<sup>‡</sup>Dept. of Economics, Johns Hopkins University, Baltimore, MD 21218, USA. E-mail: bkisaci1@jhu.edu

<sup>§</sup>Corresponding author at: ICREA-UPF BGSE, and CREI, Carrer Ramon Trias Fargas 25-27, Mercè Rodoreda Building, 08009 Barcelona, Spain. E-mail: barbara.rossi@upf.edu.

# 1 Introduction

The goal of this paper is to study what the benchmark should be in assessing macroeconomic forecasts from Dynamic Stochastic General Equilibrium (DSGE) models. This question has gained importance as DSGE models are increasingly used in forecasting and indeed are judged by their forecasting performance – see the reviews by Del Negro and Schorfheide (2012) and Christoffel, Coenen and Warne (2011).<sup>1</sup> In particular, Del Negro and Schorfheide (2012) argued that DSGE models should be the new benchmark in forecasting. We focus on a benchmark DSGE model and evaluate both its forecast efficiency (or optimality) as well as its forecasting ability relative to reduced-form models. Our main objective is to determine whether one forecasting method systematically outperforms the others.

We find that, on average over the sample, the DSGE model forecasts output better than autoregressive methods, especially at longer horizons. When forecasting inflation, in contrast, the DSGE model is less accurate on average and, in several instances, the DSGE model’s performance worsens towards the end of the sample relative to its competitors. In general, we find that no single model is most accurate at all times. We conclude that the forecasting method with the lowest root mean squared forecast error (RMSFE) depends on the sample and the forecast horizon, implying there is no obvious forecast benchmark. Our analysis does not support the common practice of using the random walk forecasts or Bayesian VAR model forecasts as benchmarks for judging the accuracy of DSGE model forecasts. For example, we find that standard large Bayesian VARs appear over-parameterized, making them poor benchmarks against which to judge other forecasting models. In fact, unrestricted small-scale reduced-form models typically outperform the large-scale Bayesian VAR model. This is the case for interest rate forecasts as well as forecasts of inflation and output growth. The random walk model forecasts well both inflation (at horizons greater or equal to two) and the interest rate (up to four quarter-ahead) but not output growth.

Our analysis is based on the DSGE model of Smets and Wouters (2007). We focus on this particular model due to its prominence in the literature and its wide use at central banks for forecasting. We evaluate the out-of-sample accuracy of forecasts generated from this model for the period 1992-2007. As competitors, we consider a series of reduced-form models: univariate Autoregressions (ARs), Vector Autoregressions (VARs), Bayesian VARs (BVARs)

---

<sup>1</sup>See also Wieland and Wolters (2012) for an analysis of forecasting in policy making settings, Bache et al. (2011) for an analysis of DSGE models’ forecast densities, and Lees, Matheson and Smith (2011) for an analysis of the forecasting performance of an open economy DSGE model for New Zealand.

and forecast combinations. VAR, AR and random walk models are considered because they are a natural benchmark and are often used in practice when forecasting, although not in conjunction with DSGE model-based forecasts. We report results for AR and VAR forecasts produced via a direct method, although the main results are qualitatively robust to using an iterated forecast method.<sup>2</sup> We also explore forecast combinations, which often, but not always, perform well (e.g. Stock and Watson, 2004 and Timmermann, 2006, show that forecast combinations work well in forecasting several macroeconomic and financial data; Rossi, 2013, shows that forecast combinations perform well in forecasting inflation and output growth in the U.S.; on the other hand, Inoue and Kilian, 2008, show that bagging predictors provides more accurate inflation forecasts than equally weighted forecasts). Bayesian VARs are the benchmark used in Smets and Wouters (2007).

The most important macroeconomic variables to be forecast are inflation, output growth and the interest rate, and we focus on these variables in this paper. Medium scale DSGE models have many more model-defined variables (some of which are unobservable) and also use a relatively large number of observables, making it difficult to include the same information in a VAR, due to overfitting concerns. For example, the Smets and Wouters (SW) model uses seven observables in estimation, which would mean 105 parameters to estimate in a second order, seven variable reduced-form VAR. Hence, the literature has employed Bayesian VARs as the forecast benchmark to improve the out-of-sample forecasting accuracy of the VAR model. We show in this paper that moving to a smaller, three variable VAR or to an AR model actually reduces the MSFE of the macroeconomic aggregates we consider compared with larger BVAR models.

The accuracy of the forecasts is evaluated using standard tests of forecast efficiency as well predictive accuracy. In particular, we use Mincer-Zarnowitz's (1969) regressions to evaluate the forecast efficiency of the models. We also compare the models' relative forecasting ability by comparing models' relative mean squared forecast errors. Importantly, we focus on studying how the forecasting ability of the models has evolved over time, both in terms of their efficiency and their relative predictive ability.

Our main question is whether there is a forecasting method based on reduced-form time series models or DSGE models that has better properties than others in all samples and for all forecast horizons. Such a model would be the "model to beat" and a natural benchmark. We find, however, that no such model exists among the standard candidates. The models'

---

<sup>2</sup>For a discussion of iterated versus direct forecasts, see Marcellino, Stock and Watson (2005), Ravazzolo and Rothman (2010) and Vigfusson and Kilian (2013).

forecast accuracy in our study depends on the sample and the forecast horizon.

Regarding the models' relative forecasting performance, we find that, with some exceptions, a simple autoregression performs best at very short horizons (i.e. horizons of one quarter) whereas DSGE models perform the best at long horizons (i.e. horizons of two years) when forecasting output growth, and vice versa for inflation. When looking at how the forecasting ability evolved over time, however, several interesting aspects are worth noting. First, when forecasting output growth, the better forecasting ability of the AR relative to the DSGE model is evident in the 1990s but disappears towards the end of the sample (in mid-2000), when the difference between the two models first becomes insignificant and then the DSGE model begins to dominate. Results are similar for forecast combinations. Regarding inflation, results are different: there is a clear tendency for the DSGE model's forecasts to worsen over time relative to that of reduced-form models at most horizons. Similar results hold for interest rates at short and medium horizons.

Regarding the models' forecast efficiency, we find that, with few exceptions, the models' forecasts of output growth and inflation are not efficient – i.e., either the forecasts are biased (i.e. on average the forecast is different from the realized value), or the forecasts are not correlated one-to-one with the actual realizations, or both. Only in the case of output growth, the DSGE model forecasts are efficient at long horizons at some point over the sample. In the case of inflation, the DSGE model forecasts' lack of efficiency shows up consistently over time, whereas that of the ARs and forecast combinations shows up mainly in the early 2000s at some forecast horizons, while they are more efficient in the most recent period. Similarly, interest rate forecasts suffer from lack of efficiency, although VAR-based interest rate forecasts suffer less from lack of efficiency than the DSGE model at short horizons and more at long horizons.

An important finding of this paper is the comparative advantage that a simple three variable VAR has in forecasting inflation and output compared with the seven variable BVAR. The BVAR-based forecast, both because it employs the same observables as the DSGE model and because it is estimated using Bayesian methods similar to those used in the estimation of the DSGE model, is the standard reduced-form benchmark of choice against the DSGE model forecast.<sup>3</sup> We show that while this benchmark makes intuitive

---

<sup>3</sup>Note however that, even though both the Bayesian VAR and the DSGE models are estimated with Bayesian methods, their priors might be quite different. Note also that, if DSGE model satisfies the invertibility condition, it may be interpreted as a VAR model subject to cross-equation restrictions; however, it is not clear whether a VAR with a finite lag structure is a good approximation for the true underlying VAR.

sense for those reasons, the large-scale BVAR appears over-parameterized and constitutes a weak benchmark for the evaluation of other forecasts. If the variables of interest are inflation and GDP growth, a three variable VAR may be a more suitable benchmark for forecasting than the BVAR. Similarly, the AR model also forecasts better than the BVAR.

Another interesting finding we present is the sample dependency of the most accurate forecast (in terms of root mean squared forecast error, RMSFE). In particular, in several cases the performance of the DSGE model deteriorates towards the end of the sample relative to its competitors, possibly because it corresponds to a truly out-of-sample period, after the publication of Smets and Wouters (2007). Note that our goal is to evaluate the true out-of-sample forecasting ability of the DSGE model. Even when the data are real time, as the priors are chosen at the time the model was constructed, some ex-post information affects the model forecasts. More subtly, even if one had a way of using real-time priors, as the model is built to fit the data in a certain period (the data about up to 2004 in the Smets-Wouters model's case) modeling choices are made to maximize this fit. Hence, the real out-of-sample period begins at the time the model is built, regardless of the real time nature of the data used to estimate the model. In the present case, this leaves a short true out of sample period as the financial crisis began shortly after the inception of this model and limited the data available for the forecasting exercise. But the deterioration in the model forecasting ability in the true out-of-sample period is nonetheless striking.

Our paper is related to the recent contributions by Adolfson, Linde and Villani (2007), Edge and Gürkaynak (2010), Edge, Kiley and Laforde (2010), Edge, Gürkaynak and Kısacıkoglu (2013), Giacomini and Rossi (2012), and Del Negro and Schorfheide (2012). Adolfson, Linde and Villani (2007) compare the forecasting ability of an open economy DSGE model for the Euro area with that of several reduced-form models; however they focus on the Euro area and only study the MSFE of the competing forecasts, so it is unclear whether the differences in forecast performance of competing models are statistically significant. In contrast, we perform statistical tests of relative forecast comparisons as well as forecast efficiency. Moreover, unlike Adolfson et al. (2007), we evaluate the forecasting ability of the DSGE model using a real-time database, rather than ex-post revised data.

Edge and Gürkaynak (2010), have been the first to perform a thorough analysis of the actual forecasting ability of DSGE models using real-time data based on vintages. They show that the performance of DSGE models is comparable or slightly superior to that of a constant mean model, but both DSGE models' and Blue Chip forecasts are biased: the reason why they perform similarly is because, during the Great Moderation period, volatility

was low and most variables were therefore unpredictable. Edge, Gürkaynak and Kısacıköğlü (2013), whose data we utilize, extend the sample of Edge and Gürkaynak and study the model's forecasting ability against the model's own implication of how well it should forecast by generating data from the model and examining the model's ability to forecast model generated data and its ability to forecast real data. Edge, Kiley and Laforte (2010) find that the Federal Reserve Board DSGE model produces competitive forecasts relative to those of the Federal Reserve Board staff; however, they do not consider the Smets and Wouters (2005) model nor the same sample period we consider.

Giacomini and Rossi (2012) compare the performance of DSGE models and BVARs over time, from an in-sample fit perspective. They show that the DSGE model's performance seems to deteriorate over time, once the trend in the model is estimated in real time.<sup>4</sup> Unlike Giacomini and Rossi (2012), we focus on the models' out-of-sample forecasting performance and focus on U.S. rather than Euro area data. Del Negro and Schorfheide (2012) consider DSGE models and Blue Chip forecasts in their forecasting tests, whereas we compare the forecasts of the DSGE model to those of several reduced-form models. In addition, Del Negro and Schorfheide (2012) only forecast over the full out-of-sample period; we instead analyze the forecast performance of the models as it evolves over time. This allows us to identify periods in time where each of the models were the best in forecasting specific macroeconomic variables.

The remainder of the chapter is organized as follows. Section 2 describes the data, the methods and the models we use for forecasting, as well as the test statistics we rely upon to evaluate the models' forecasting ability. Section 3 reports empirical results for three of the most important macroeconomic variables in the Smets and Wouters DSGE model: output growth, inflation and the interest rate. We discuss both relative forecasting performance of the models as well as their forecast efficiency. Section 4 concludes.

---

<sup>4</sup>That is, when the trend estimate does not use information from periods to be forecasted. This is an issue especially for models that use detrended data as the filter for detrending may be two sided and impart ex-post information into real time data.

## 2 Data, Forecasting Methods and Models, and Test Statistics

### 2.1 The Data

The data used in this paper come from Edge, Gürkaynak and Kısacıkoglu (2013), who extend and improve the quarterly dataset of Edge and Gürkaynak (2010), which itself closely follows the data choices of Smets and Wouters (2007). We refer the readers to the detailed data appendix of Edge, Gürkaynak and Kısacıkoglu (2013) for details of the data, and provide only a brief outline here.

Edge and Gürkaynak (2010) collected real-time vintages of U.S. data on the same series used in Smets and Wouters (2007). The series are the following: the per capita, real GDP growth rate (quarter on quarter, non-annualized); the GDP deflator inflation rate (quarter on quarter, non-annualized); the interest rate level; real consumption; real investment; the real wage; hours worked (in log); the growth rate of the GDP deflator; and the Federal Funds rate. They use real GDP, the GDP price deflator, the nominal personal consumption expenditure and the nominal fixed private investment from the national income and product accounts (NIPA); compensation per hour in the non-farm business sector from the Bureau of Labor Statistics' quarterly Labor Productivity and Costs (LPC) release; average weekly hours of production, civil employment and population from the Bureau of Labor Statistics Employment Situation Summary (ESS). The realized data, or the "true" outcome that is to be forecast is the first final release corresponding to the last observation in the third release of NIPA data and the second release of LPC data (which are the final release of the data before they are revised in either an annual or a comprehensive revision). For ESS releases, they use the last available observation in the second revision of the data. Edge, Gürkaynak and Kısacıkoglu (2013) extend this data set in sample coverage and also provide a better treatment of the population, which makes an important difference in the output growth forecasts.

The data we use is in the form of 20-year rolling windows. Rolling window estimates help capture changes in parameters (regime shifts) and forecasting tests we will use below have desirable properties when the underlying forecasts are made using rolling window estimates. In the extended dataset we use, the out-of-sample period used for forecast evaluation starts in 1992:I and ends in 2005:IV for one-step ahead forecasts, 2006:I for two-step-ahead forecasts, and 2007:III for eight-step-ahead forecasts. The end of the sample is dictated by the financial

crisis, which a model without a housing sector or a financial sector cannot be expected to forecast. The sample includes the Great Moderation period as well as the recession of 2001. The Great Moderation is characterized by lower permanent (predictable) and higher temporary (unpredictable) fluctuations in macroeconomic aggregates, which makes inference about model validity based on forecasting ability a tricky affair, as discussed by Edge and Gürkaynak (2010). In this paper, our aim is to evaluate the forecast accuracy of DSGE models in a comparative setting without drawing conclusions about the models themselves.

In short, our data are real time versions of the Smets and Wouters (2007) database, employing not only their data series choices but also their priors for the Bayesian estimation of the DSGE model and the BVAR. The only material difference is in the population series, the proper treatment of which helps the DSGE model output forecasts as the model produces per capita output forecasts which are then multiplied with population growth to obtain aggregate output growth forecasts.

## 2.2 Forecasting Models

Smets and Wouters' (2007) DSGE model is a real business cycle model with nominal rigidities (sticky prices and wages) as well as real rigidities (such as habit formation in consumption and cost of adjustment in investment). A detailed description of the model is provided in Section 1 in Smets and Wouters (2007), to which we refer the interested reader. The model includes seven observable variables and seven shocks.<sup>5</sup> In the Bayesian estimation of the model parameters we use the same priors as Smets and Wouters.

In addition to the Smets and Wouters (2007) model, we consider several reduced-form time series models. The first reduced-form model we consider is a three-variable VAR with output growth ( $\Delta y_t$ ), inflation ( $\pi_t$ ) and the interest rate ( $i_t$ ). Let  $Y_t = (\Delta y_t, \pi_t, i_t)'$ ; then the VAR is:

$$A(L)Y_t = A_0 + u_{VAR,t} \tag{1}$$

This represents the basic VAR specification, which we use to forecast output growth, inflation and the interest rate. This basic specification follows Stock and Watson (2001).<sup>6</sup>

---

<sup>5</sup>Recall that the out-of-sample output growth forecasts from the DSGE model are per capita. We transform them in aggregate values by adjusting for realized population growth. Note that the reduced-form models' forecast of output growth are aggregate, instead.

<sup>6</sup>The number of lags used in the VAR and the autoregressive specifications is determined by the Bayesian Information criterion (BIC), recursively applied each time the model is estimated. See the next sub-section for details. Appendix A evaluates the robustness of the results when the AIC is used instead of the BIC.



Let the  $k$ -th variable in  $Y_t$  be denoted by  $Y_t^{(k)}$ , for  $k = 1, 2, 3$ . We also consider an AR model for the variables of interest; for example, for  $Y_t^{(k)}$ , we estimate:

$$Y_t^{(k)} = a_0 + a_1 Y_{t-1}^{(k)} + \dots + a_p Y_{t-p}^{(k)} + u_{AR,t}. \quad (2)$$

Rather than relying on iterated AR and VAR forecasts from eqs. (1, 2), we construct direct forecasts instead, as described in section 2.4. We will label these direct forecasts as AR and VAR forecasts (even though they are technically not obtained from an AR or VAR model) and we demonstrate the robustness of the qualitative results to iterated AR and VAR forecasts in Appendix B.<sup>7</sup>

The Random Walk (labeled "RW") model for  $Y_t^{(k)}$  is the following:

$$Y_t^{(k)} = Y_{t-1}^{(k)} + u_{RW,t}. \quad (3)$$

We also consider equal-weighted forecast averaging, which has been shown to provide competitive forecasts for output growth and inflation (see Rossi, 2013). The equal-weighted forecast combinations ("FC") are obtained by averaging (using equal weights) the forecasts of the autoregressive model with those of additional univariate models that contain an additional predictor,  $x_{i,t}$  (the procedure follows Stock and Watson, 2003). For example, when forecasting  $Y_t^{(k)}$ , we estimate:

$$Y_t^{(k)} = a_0 + a_1 Y_{t-1}^{(k)} + \dots + a_p Y_{t-p}^{(k)} + b_1 x_{i,t-1} + \dots + b_k x_{i,t-k} + u_{ARX_i,t} \quad (4)$$

for  $i = 1, \dots, 6$ , and then combine the forecasts of models (4) for the set of variables in Smets and Wouters (2007),  $x_{1,t}, \dots, x_{6,t}$ , and that of the AR model, eq. (2), giving each of them a weight equal to  $1/7$ .

## 2.3 Forecasting Methods

The forecasts are based on a model that is characterized by the  $(k \times 1)$  parameter vector  $\gamma$ . The forecasts are obtained by dividing the sample of size  $T+h$  observations into an in-sample portion of size  $R$  and an out-of-sample portion of size  $P$ . The size of the estimation window,  $R$ , equals 20 years, corresponding to 80 quarterly observations that provide information

---

<sup>7</sup>Direct forecasts for the AR(p) model are obtained as follows. We estimate  $\Delta y_t = a_0 + a_1 \Delta y_{t-h} + \dots + a_p \Delta y_{t-p-h+1} + u_{AR,t,t+h}$  using information up to time  $t$ , and then use the parameter estimate and the realized values of  $\Delta y_t, \dots, \Delta y_{t-p}$  to forecast directly  $\Delta y_{t+h}$ .

for the model estimation at each forecast date.<sup>8</sup> The sequence of  $P$  out-of-sample forecast errors depends on the realizations of the forecasted variable and on the in-sample parameter estimates,  $\widehat{\gamma}_{t,R}$ . These parameters are re-estimated at each  $t = R, \dots, T$  over a window of  $R$  observations including data indexed  $t - R + 1, \dots, t$  (also known as the "rolling scheme").<sup>9</sup>

The forecasts of the VAR and AR models are obtained using a direct forecast method. The number of lags in the VAR and AR specifications is recursively chosen by the Bayesian information criterion (BIC).<sup>10</sup> The number of lags in eq. (4) is estimated conditional on selecting  $p$  by BIC in the univariate AR specification, eq. (2), and then applying the BIC to (4) to select the number of lags of the extra predictor,  $k$ . To guarantee parsimony, the maximum number of lags imposed in the BIC is two.

It is important to note that it has been shown that the BIC does not necessarily select the best forecasting model (see Inoue and Kilian, 2006). An alternative criterion is the Akaike Information Criterion (AIC). In Appendix A, we discuss results based on the AIC and show that our results are robust to the choice of the criterion.<sup>11</sup>

## 2.4 Forecast Evaluation Test Statistics

Let the estimated forecast error associated with the  $h$ -step-ahead direct forecast made at time  $t$  be denoted by  $\varepsilon_{t+h}$ . For example, in the case of a simple linear AR(1) regression model with  $h$ -period lagged regressors, we estimate the regression:  $y_t = \gamma_0 + \gamma_1 y_{t-h}$  using the most recent  $R$  data, and obtain the parameter estimates at time  $t$ :  $\widehat{\gamma}_{0,t,R}, \widehat{\gamma}_{1,t,R}$ . The forecast made at time  $t$  is equal to:  $y_{t+h}^f = \widehat{\gamma}_{0,t,R} + \widehat{\gamma}_{1,t,R} y_t$  and the forecast error associated with the direct forecast is:  $\widehat{\varepsilon}_{t+h} = y_{t+h} - \widehat{\gamma}_{0,t,R} - \widehat{\gamma}_{1,t,R} y_t$ , and  $\widehat{\gamma}_{t,R} \equiv (\widehat{\gamma}_{0,t,R}, \widehat{\gamma}_{1,t,R})'$ . We perform a series of tests on the sequence of the forecast errors,  $\{\widehat{\varepsilon}_{t+h}\}_{t=R}^T$ .

First, we focus on testing forecast efficiency (under quadratic loss), based on the following regression:

$$\widehat{\varepsilon}_{t+h} = \alpha + \beta y_{t+h}^f + \eta_{t+h}, \quad t = R, \dots, T, \quad (5)$$

---

<sup>8</sup>It is possible that the empirical results might be different if another window size is chosen. See for example Inoue and Rossi (2012) and Hansen and Timmermann (2012) for a discussion and methods for preventing data mining over the estimation window size.

<sup>9</sup>As we show in section 3.4 below, results based on the recursive estimation scheme (i.e. using observations  $1, \dots, t$ ) are qualitatively similar.

<sup>10</sup>That is, each time a VAR or an AR specification is estimated, the number of lags is estimated as well, according to the BIC.

<sup>11</sup>Note that, while the BIC does not necessarily select the best model, neither does the AIC. Under Inoue and Kilian's (2006) assumptions, the AIC is dominated by the BIC for one-step-ahead direct forecasts.

where  $\widehat{\varepsilon}_{t+h}$  is the estimated forecast error. We test whether  $\alpha = 0$  (forecast unbiasedness),  $\beta = 0$ , and  $\alpha = \beta = 0$ . We refer to the latter test (the joint test for  $\alpha, \beta$  equal to zero) as the Mincer and Zarnowitz (1969) forecast efficiency test.

Second, we note that, as discussed in Rossi and Sekhposyan (2012), traditional forecast efficiency test results suffer from the drawback that they focus on whether forecasts are efficient on average over the whole out-of-sample period. Thus, we follow Rossi and Sekhposyan (2012) and also report a sequence of forecast efficiency tests obtained by estimating regression (5) using data from  $t - m + 1$  up to  $t$ , and construct a *sequence* of Mincer and Zarnowitz (1969) tests at time  $t$ , for  $t = m, \dots, P$ .<sup>12</sup>

Third, we consider tests of relative forecasting performance. We report results for comparing models' forecasts based on the difference of the mean square forecast errors (MSFEs) of competing models using the Diebold and Mariano (1995) test statistic and using the critical values proposed by Giacomini and White (2006).

To conclude, we also report results for comparing models' forecasting ability over time based on the fluctuation test introduced by Giacomini and Rossi (2010). The latter propose to calculate the Diebold and Mariano (1995) test statistic over rolling windows in an effort to keep track of the models' relative forecasting ability over time. In other words, we calculate a sequence of Diebold and Mariano's (1995) test statistics over time by estimating regression (5) using data from  $t - m + 1$  up to  $t$ , and use the results to perform the test at time  $t$ , for  $t = m, \dots, P$ . To evaluate whether the models are statistically significantly different, we rely on the critical values in Giacomini and Rossi (2010).

### 3 Empirical Results on Forecasting Output Growth, Inflation and the Interest Rate

In this section we focus on forecasting output growth, inflation and the short-term interest rate, showing the results for output growth and inflation (which policy makers are most interested in) first, and showing the results for interest rates separately later. The forecasts are reported in Figure 1. The figure reports the realized value of the target variable (output growth in panel (a) and inflation in panel (b)), labeled "Actual Data", together with forecasts of the following models: the VAR (labeled "VAR"), the AR (labeled "AR"), Smets and Wouters (2007, labeled "DSGE"), equal-weighted forecast combinations (labeled "FC"), the

---

<sup>12</sup>We select  $m = P/2$ , i.e. approximately seven years for most forecast horizons.

Bayesian VAR (labeled "BVAR") and the random walk (labeled "RW"). Each graph reports results for different horizons, reported in each graph's title. From Figure 1(a), it appears that AR forecasts of output growth track the realized values more closely than the DSGE model at the one-quarter-ahead forecast horizon, but they worsen as the forecast horizon becomes larger. It also appears that the longer the horizon, the more the VAR under-estimates output growth. In this section, we will use formal statistical tools to investigate whether this visual impression is correct.

### 3.1 Empirical Results Based on Tests of Forecast Efficiency

In this sub-section we investigate whether the forecasts of the models that we study are efficient. It is worthwhile to reflect briefly on what it means for the forecast of an atheoretical reduced-form model or a DSGE model to be efficient. Could a practitioner who is trying to forecast, say, output improve over the forecast of any one of these models by using information available in real time? If so, the practitioner would not use the forecast of the model as is, but modify the model. This is the sense in which we use the term forecast efficiency for the models under study.

Table 1 reports results of forecast efficiency tests for several models and target variables (output growth on the left and inflation on the right). Each panel reports, for several horizons (reported in the first column), the p-values of the Mincer and Zarnowitz (1969) forecast efficiency test (i.e. the joint test that  $\alpha = \beta = 0$  in eq. (5), labeled "MZ p-value"),  $\hat{\alpha}$  in regression (5) together with the p-value of the test that the constant equals zero (in parenthesis),  $\hat{\beta}$  in regression (5) together with the p-value of the test that the slope equals zero (in parenthesis). Table 1 shows, in general, that the forecasts are not efficient, no matter which model we consider. There are some exceptions, however. The tests do not reject joint forecast efficiency for output growth at some intermediate horizons (four or five quarter-ahead) in some models. Note that most reduced-form forecasts under-predict the target variables.

To investigate whether the efficiency of the forecasts has changed over time, Figure 2 reports the p-value of the sequence of Mincer and Zarnowitz (1969) tests over time, using a rolling window equal to half of the number of out-of-sample forecasts (i.e. approximately 7 years) in for each test. This window size is used in all of the sequence of tests of forecast comparisons and efficiency *over time* that we use below.<sup>13</sup> The figure displays results for

---

<sup>13</sup>Note that we always use a rolling window of 20 years in estimating the model and generating forecasts.

forecasts at various horizons (reported in the legend) and several models (reported in the title). The target variable is output growth in panel (a) and inflation in panel (b). Figure 2(a) shows that the large rejections of forecast efficiency for output growth are mainly driven by the end of the sample for the statistical models, whereas those of the DSGE model are mainly driven by both the beginning and the end of the sample, although the bias is more pronounced at the end. The intermediate horizon forecasts of the Smets and Wouters (2007) model are consistently efficient over time, although the p-values of the forecast efficiency test are very small at the end of the sample, raising concerns of lack of efficiency in the late 2000s. Efficiency is uniformly rejected over time for the random walk at all horizons and for the BVAR forecasts at most horizons.

Results are similar for inflation, except that, for inflation, the DSGE model forecasts are not efficient, no matter which sample period or forecast horizon we consider (except for one-step-ahead forecasts in the late 1990s). Forecasts of the VAR, AR and forecast combination models are efficient towards the end of the sample, instead.

Our findings are related to Edge and Gürkaynak (2010), who find that the DSGE model forecasts poorly over the sample period they consider. They attribute this finding to their focus on the Great Moderation sample, a period of stability and low macroeconomic volatility, where most macroeconomic variables have become largely unforecastable (see Stock and Watson, 2007). Our results show that not only the DSGE model, but also the majority of reduced-form models had difficulties forecasting output and inflation over the sample that we consider.

### **3.2 Empirical Results Based on Tests of Relative Forecasting Performance**

Table 2 reports results for comparing the models' forecasting ability. The columns labeled "RMSFE" in the table report the RMSFE of Smets and Wouters (2007) DSGE model's forecasts (labeled "DSGE"). The remaining columns labeled "RMSFE" in the table report the RMSFE of the following models *relative to* that of the DSGE model: autoregression ("AR"), forecast combinations ("FC"), vector autoregression ("VAR"), Bayesian VAR ("BVAR"),

---

But these forecasts themselves may be evaluated as a single sample or in different sub-samples using a rolling window. The window size of 7 years refers to the size of the rolling window used to smooth the forecast errors when we study the forecasts in sub-samples to see how the relative and absolute forecasting abilities of models changed over time.

and the random walk ("RW"). For example, the column labeled "VAR" reports the RMSFE of the VAR model divided by the RMSFE of the DSGE model. Values less than unity denote forecasts that have lower RMSFE than the DSGE benchmark. The columns labeled "DM" in the table report the p-value of the Diebold and Mariano's (1995) test for comparing the specified forecasts (e.g. the column labeled "DM, FC vs. DSGE" reports the p-value of the Diebold and Mariano (1995) test for comparing the forecast of the forecast combination with those of the Smets and Wouters (2007) model).

Table 2 shows that, when considering output growth, the DSGE model forecasts are less accurate than the AR model at the shortest horizon, and more accurate at longer horizons (although differences are not significant in either case). The VAR forecasts are always worse than the DSGE ones and also worse than the forecasts of the AR model. Forecast combinations produce forecasts that are better than the DSGE model at the one-quarter-ahead horizon, but worse at longer horizons, again insignificantly so. The BVAR forecasts worse than the DSGE model at all horizons. When forecasting inflation, the AR model has lower RMSFE than the DSGE model at most horizons, except for short ones and the differences are statistically significant. Note that, overall, the performance of the BVAR is worse than that of most models; in particular it is worse than the VAR and DSGE models: the latter is statistically more accurate than the BVAR in forecasting both inflation and output growth at horizons up to five quarters. Note also that the random walk forecasts inflation better than the DSGE model at medium to long horizons, very much like the AR.

The models' relative forecasting ability has, however, changed significantly over time. Figure 3 reports results based on the fluctuation test, which is a rolling Diebold-Mariano test with Giacomini and Rossi's (2007) critical values. Regarding output growth, Figure 3 shows that the VAR performs similarly to or better than the DSGE model up to the mid-1990s, after which the DSGE model performs significantly better. The BVAR, instead, is typically worse than the DSGE model over most time periods. Interestingly, the performance of forecast combinations and the AR models changes drastically across both time and forecast horizons: at long horizons, they perform significantly worse than the DSGE model, especially at the end of the sample; at shorter horizons, however, the DSGE model seems to outperform AR and forecast combinations only later in the sample.

Regarding inflation, we observe the opposite pattern. The forecasting performance of the VAR, AR and forecast combination models mostly improve over time relative to the DSGE model at most forecast horizons; the DSGE model is significantly better only in the early 1990s, for short horizons. Regarding the performance of the BVAR, it performs significantly

worse than the DSGE model over most of the sample and for most forecast horizons.

These empirical findings are related to those in Edge and Gürkaynak (2010) and Giacomini and Rossi (2012). Edge and Gürkaynak (2010) show that the DSGE model forecasts are better than those of the BVAR (in the sense that the former have lower RMSFE). What we find is that the large-scale BVAR may be not be a good benchmark to compare the DSGE model forecasts, as a reduced-form AR model (estimated by unrestricted least squares) as well as other reduced-form time series models perform significantly better. Giacomini and Rossi (2012) compare a DSGE model estimated for the Euro area with a BVAR, and find that the in-sample fit of the DSGE model worsens towards the end of the sample. In the most recent period they have available, the DSGE model fits the data significantly worse than the reduced-form model. However, there are two important differences between this paper and Giacomini and Rossi (2012). The first, and the conceptually important difference is that the latter consider the in-sample performance of the models, whereas this paper considers out-of-sample forecasting performance. The second is that the latter estimates the DSGE model for the Euro area, whereas we consider U.S. data.

### 3.3 Empirical Results on Forecasting Interest Rates

The results reported in the previous sub-sections show that none of the models produce efficient forecasts for output growth and inflation, with rare exceptions. Do these results carry over to interest rate forecasts?

We begin by showing interest rate forecasts of various models together with the realized interest rates in Figure 4, which is the interest rate analogue of Figure 1. To judge the quality of these forecasts, we then do the same tests we did for output growth and inflation. Table 3 shows that interest rate forecasts from the VAR and RW models are efficient at short horizons; the DSGE model interest rate forecasts are instead efficient at long but not at short horizons, according to the joint Mincer-Zarnowitz (1969) test. Figure 5, however, shows that VAR, forecast combinations and DSGE forecasts were efficient in the early 2000 but became less efficient towards the end of the sample. In contrast, BVAR forecasts were, instead, essentially never efficient.

Which model forecasts interest rates best? Table 4 shows that RW and combinations forecast interest rates more accurately than the DSGE model at the shortest horizons, whereas the DSGE model forecasts better at longer horizons (although results are not statistically significant).

Figure 6(a) shows that the interest rates' MSFEs of the DSGE model tends to be higher than those of the VAR model towards the beginning of the sample, signalling a worsening of forecast accuracy of the DSGE model forecasts. The differences in the MSFEs are statistically significant. Unreported results show that similar findings hold for the autoregressive model.<sup>14</sup> The BVAR instead is always significantly worse than the DSGE model (see Figure 6b).

### 3.4 Summary of Empirical Findings

In general, most models' forecasts are not efficient, no matter which model and time period we consider. When considering the models' relative forecasting performance, however, several interesting patterns emerge from the data. When forecasting output growth, DSGE models forecast better than the AR model and forecast combinations at longer horizons but slightly worse at shorter horizons, although the differences are not significant. When forecasting inflation the DSGE model forecasts better only at the shortest horizon, with AR, random walk and the combined forecast all outperforming the DSGE model forecasts at longer horizons, significantly so at intermediate horizons. VAR and BVAR are always outperformed by the DSGE model for both output and inflation forecasting.

Interestingly, the relative forecasting performance does change over time. ARs forecast output growth better than the DSGE model in the late 1990s at shorter horizons, and significantly so; the DSGE model instead performs significantly better at longer horizons most of the time, although its performance worsens towards the end of the sample. Regarding inflation, we observe the opposite pattern. The forecasting performance of the VAR, AR and forecast combinations models improves over time relative to the DSGE model at most forecast horizons, and the DSGE model is significantly better only in the late 1990s, but not in the most recent period.

The finding that relative forecasting performance changes over time is related to Stock and Watson (2003) and Rossi and Sekhposyan (2010). The former demonstrate the existence of widespread instabilities in the parameters of models describing output growth and inflation in the U.S.; the latter show that most predictors for output growth lost their predictive ability in the mid-1970s, whereas, when forecasting inflation, fewer predictors are significant, and their predictive ability significantly worsened around the time of the Great Moderation. This paper shows that the instabilities significantly affect DSGE and VAR models' performance as well.

---

<sup>14</sup>See Appendix C for detailed results.



Table 5 investigates how the empirical results would change using a recursive (rather than a rolling) window estimation scheme. The table shows that the results are qualitatively similar, using rolling or recursive data sets to estimate the models and make forecasts do not change the relative abilities of the models much.

## 4 Conclusions

In this paper we studied the forecasting performance of several standard reduced-form models and the current baseline DSGE model for macroeconomic aggregates. While forecasting has gained importance in the applied literature, no comprehensive study of the merits of DSGE model forecasts and reduced-form model forecasts exists. Nor have existing studies examined how the forecast accuracy of these models has evolved over time. Preliminary results in the literature suggested that, during the 1990s and the early 2000s, the DSGE model outperformed large-scale BVAR models. This has led to the DSGE model forecast to be taken more seriously and prompted Del Negro and Schorfheide (2012) to call for the DSGE model forecast to be used as the benchmark for other forecasts' evaluation.

This paper uses a variety of standard reduced-form models as well as the DSGE model to generate forecasts for the 1992-2007 sample period and studies how their forecast efficiency as well as their relative forecasting performance changed with sample periods and forecast horizons. We showed that all forecasting methods fail the “efficiency” test – their forecast errors are forecastable with current information at some forecast horizon or at some point in time. This means judgmental forecasters can always adjust these forecasts and have better outcomes, which is important to know for practitioners and policymakers who use these models' forecasts. For example, in principle, by adjusting the mean of the forecast with current quarter's information, judgemental forecasters might be able to improve the forecast bias; they might also have access to other information that is available in real time, although it is not included in the models that we consider (e.g. flash estimates of current economic conditions). However, it is not clear whether this approach would succeed in practice, as judgemental forecasts might be biased themselves and may add variability to the forecast.

The relative forecast performance, as measured by the relative RMSFE of various forecasts, is important to establish a benchmark non-judgmental forecast. We look at whether one of the forecasting methods routinely outperforms others as we change sample periods and forecasting horizons. We find that there is no single best forecasting method. The large-scale Bayesian VAR is almost never the best forecast, calling into question its use as the

standard competitor for the estimated DSGE model forecast.<sup>15</sup> On the other hand, forecasts generated by the AR model outperform the DSGE model forecasts at different horizons for different macroeconomic aggregates and to varying degrees of statistical significance over different sample periods. This makes us conclude that among the current batch of standard forecasting methods, none qualifies as the forecast benchmark alone. Typically, AR models forecast better than VAR models, although not always.

A series of caveats are in order. First, we only looked at a particular VAR with inflation, output growth and the interest rate. Second, we use traditional forecast evaluation tests that do not take into account the real-time nature of the data. Third, we investigate large dimensional BVAR models; it might be that small dimensional BVAR models forecast better than large dimensional ones. However, we focus on the latter because they are currently used in the literature; in addition, BVARs have been introduced exactly to address the parameter proliferation in large-dimensional VARs.<sup>16</sup> We leave it to future work to study why certain reduced-form models' forecasts are better than the DSGE model forecast for some sample periods. It would be valuable to learn what features of the data favor what forecasting method. Perhaps it would be possible to select the most accurate forecasting method in real time. Finally, it may be that there is another forecasting method that consistently outperforms all others but omitted in this study.

---

<sup>15</sup>An interesting avenue for research is to think about the priors used for the BVAR estimation. The Smets and Wouters priors shrink towards random walks, which may not be helping in inflation and output growth forecasting. As we show in this paper, random walk is itself a poor forecast, especially for output growth.

<sup>16</sup>Note also that Smets and Wouters (2007) shrink the priors towards a random walk. This may not be appropriate for inflation and GDP growth in the Great Moderation period. Another option would be to shrink towards a white noise prior, which might be better suited for this period.

## 5 References

- Adolfson, M., J. Linde, and M. Villani (2007), “Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model,” *Econometric Reviews* 26(2-4), 289-328.
- Bache, I.W., A.S. Jore, J. Mitchell, S.P. Vahey (2011), “Combining VAR and DSGE Forecast Densities,” *Journal of Economic Dynamics and Control* 35(10), 1659-1670.
- Christoffel, K., G. Coenen and A. Warne (2011), “Forecasting with DSGE Models,” in: M.P. Clements and D.F. Hendry (eds.), *Handbook of Economic Forecasting*, Oxford University Press, 89-127.
- Del Negro, M. and F. Schorfheide (2012), “Forecasting with DSGE Models: Theory and Practice,” in: G. Elliott and A. Timmermann (eds.), *Handbook of Forecasting*, Vol. 2, Elsevier.
- Diebold, F.X. and R.S. Mariano (1995), “Comparing Predictive Accuracy,” *Journal of Business and Economic Statistics* 13, 253-263.
- Edge, R.M. and R.S. Gürkaynak (2010), “How Useful Are Estimated DSGE Model Forecasts for Central Bankers?,” *Brookings Papers on Economic Activity* 41(2), 209-259.
- Edge, R.M., R.S. Gürkaynak, and B. Kısacıkoglu (2013), “Judging the DSGE Model by Its Forecast,” *mimeo*.
- Edge, R.M., M.T. Kiley and J.P. Laforte (2010), “A Comparison of Forecast Performance Between Federal Reserve Staff Forecasts, Simple Reduced-form Models, and a DSGE Model,” *Journal of Applied Econometrics* 25(4), 720-754.
- Giacomini, R. and B. Rossi (2010), “Forecast Comparisons in Unstable Environments,” *Journal of Applied Econometrics* 25(4), 595-620.
- Giacomini, R. and B. Rossi (2012), “Model Comparisons in Unstable Environments,” *mimeo*.
- Giacomini, R. and H. White (2006), “Tests of Conditional Predictive Ability,” *Econometrica*, 74, 1545-1578.
- Hansen, P.R. and A. Timmermann (2012), “Choice of Sample Split in Out-of-Sample Forecast Evaluation,” *mimeo*.
- Inoue, A. and L. Kilian (2006), “On the Selection of Forecasting Models,” *Journal of Econometrics* 130(2), 273-306.
- Inoue, A. and L. Kilian (2008), “How Useful Is Bagging in Forecasting Economic Time Series? A Case Study of U.S. Consumer Price Inflation,” *Journal of the American Statistical*

*Association* 103, 511-522.

Inoue, A. and B. Rossi (2012), “Out-of-Sample Forecast Tests Robust to the Choice of Window Size,” *Journal of Business and Economic Statistics* 30(3), 432-453.

Kilian, L. and R.J. Vigfusson (2013), “Do Oil Prices Help Forecast U.S. Real GDP? The Role of Nonlinearities and Asymmetries”, *Journal of Business and Economic Statistics* 31(1), 78-93.

Lees, K., T. Matheson and C. Smith (2011), “Open Economy Forecasting with a DSGE-VAR: Head to Head with the RBNZ Published Forecasts,” *International Journal of Forecasting* 27(2), 512-528.

Marcellino, M., J.H. Stock and M.W. Watson (2005), “A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series”, *Journal of Econometrics* 135(1-2), 499-526.

Mincer, J. and V. Zarnowitz (1969), “The Evaluation of Economic Forecasts,” in *Economic Forecasts and Expectations*, ed. J. Mincer, New York: National Bureau of Economic Research, 81–111.

Ravazzolo, F. and P. Rothman (2010), “Oil and U.S. GDP: A Real-time Out-of-Sample Examination”, *Norges Bank Working Paper* 2010/18.

Rossi, B. (2013), “Advances in Forecasting under Model Instability,” in: G. Elliott and A. Timmermann (eds.), *Handbook of Forecasting*, Volume 2, Elsevier Publications, forthcoming.

Rossi, B. and T. Sekhposyan (2010), “Has Models’ Forecasting Performance Changed Over Time, and When?,” *International Journal of Forecasting* 26(4).

Rossi, B. and T. Sekhposyan (2012), “Forecast Optimality Tests in the Presence of Instabilities,” *mimeo*.

Smets, F. and R. Wouters (2007), “Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach,” *American Economic Review* 97(3), 586-607.

Stock, J.H. and M.W. Watson (2001), “Vector Autoregressions,” *Journal of Economic Perspectives* 15(4), 101-115.

Stock, J.H. and M.W. Watson (2003), “Forecasting Output and Inflation: The Role of Asset Prices,” *Journal of Economic Literature* 41(3), 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, 405-430.

Stock, J.H. and M.W. Watson (2007), “Why Has Inflation Become Harder to Forecast?,” *Journal of Money, Credit and Banking* 39 (1), 3–34.

Timmermann, A. (2006), “Forecast Combinations,” in: G. Elliott, C. Granger and A.

Timmermann, *Handbook of Economic Forecasting Vol. 1*, North Holland: Elsevier.

Wieland, V. and M.H. Wolters (2012), “Forecasting and Policy Making,” in: G. Elliott and A. Timmermann (eds.), *Handbook of Forecasting*, Vol. 2, Elsevier.

# Tables

**Table 1. Forecast Optimality Tests: Forecasting Output Growth and Inflation**

**Panel A: VAR Model**

$h$	Output Growth Forecasts					Inflation Forecasts				
	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.00	0.55	(0.10)	0.24	(0.00)	0.00	0.33	(0.00)	0.30	(0.00)
2	0.00	0.94	(0.01)	-0.09	(0.00)	0.00	0.41	(0.00)	0.15	(0.00)
3	0.00	1.07	(0.00)	-0.19	(0.00)	0.00	0.37	(0.00)	0.23	(0.00)
4	0.00	0.78	(0.00)	0.04	(0.00)	0.00	0.41	(0.00)	0.14	(0.00)
5	0.00	0.32	(0.02)	0.43	(0.00)	0.00	0.48	(0.00)	0.00	(0.00)
6	0.00	0.67	(0.00)	0.14	(0.00)	0.00	0.54	(0.00)	-0.08	(0.00)
7	0.00	0.74	(0.00)	0.09	(0.00)	0.00	0.52	(0.00)	-0.05	(0.00)
8	0.00	0.87	(0.00)	-0.03	(0.00)	0.00	0.50	(0.00)	-0.02	(0.00)

**Panel B: AR Model**

$h$	Output Growth Forecasts					Inflation Forecasts				
	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.00	0.63	(0.00)	0.22	(0.00)	0.01	0.30	(0.03)	0.36	(0.01)
2	0.00	0.83	(0.00)	0.01	(0.00)	0.01	0.40	(0.01)	0.17	(0.01)
3	0.00	1.22	(0.00)	-0.46	(0.00)	0.03	0.30	(0.04)	0.37	(0.03)
4	0.00	1.02	(0.00)	-0.22	(0.00)	0.00	0.38	(0.00)	0.21	(0.00)
5	0.23	0.50	(0.29)	0.41	(0.38)	0.00	0.56	(0.00)	-0.13	(0.00)
6	0.00	0.93	(0.00)	-0.13	(0.00)	0.00	0.65	(0.00)	-0.30	(0.00)
7	0.00	0.83	(0.00)	0.00	(0.00)	0.00	0.69	(0.00)	-0.35	(0.00)
8	0.00	0.85	(0.00)	-0.01	(0.00)	0.00	0.60	(0.00)	-0.20	(0.00)

**Panel C: DSGE Model**

$h$	Output Growth Forecasts					Inflation Forecasts				
	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.00	0.46	(0.12)	0.36	(0.00)	0.01	0.24	(0.06)	0.50	(0.02)
2	0.00	0.23	(0.53)	0.62	(0.10)	0.00	0.47	(0.00)	0.02	(0.00)
3	0.00	0.30	(0.51)	0.57	(0.20)	0.00	0.68	(0.00)	-0.34	(0.00)
4	0.05	0.27	(0.60)	0.62	(0.35)	0.00	0.71	(0.00)	-0.39	(0.00)
5	0.01	0.50	(0.27)	0.36	(0.10)	0.00	0.78	(0.00)	-0.50	(0.00)
6	0.05	0.60	(0.17)	0.26	(0.10)	0.00	0.78	(0.00)	-0.47	(0.00)
7	0.00	0.96	(0.00)	-0.15	(0.00)	0.00	0.98	(0.00)	-0.76	(0.00)
8	0.00	1.03	(0.00)	-0.23	(0.00)	0.00	0.93	(0.00)	-0.68	(0.00)

**Panel D: Random Walk Model**

$h$	Output Growth Forecasts					Inflation Forecasts				
	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.00	0.46	(0.12)	0.36	(0.00)	0.01	0.24	(0.06)	0.50	(0.02)
2	0.00	0.23	(0.53)	0.62	(0.10)	0.00	0.47	(0.00)	0.02	(0.00)
3	0.00	0.30	(0.51)	0.57	(0.20)	0.00	0.68	(0.00)	-0.34	(0.00)
4	0.05	0.27	(0.60)	0.62	(0.35)	0.00	0.71	(0.00)	-0.39	(0.00)
5	0.01	0.50	(0.27)	0.36	(0.10)	0.00	0.78	(0.00)	-0.50	(0.00)
6	0.05	0.60	(0.17)	0.26	(0.10)	0.00	0.78	(0.00)	-0.47	(0.00)
7	0.00	0.96	(0.00)	-0.15	(0.00)	0.00	0.98	(0.00)	-0.76	(0.00)
8	0.00	1.03	(0.00)	-0.23	(0.00)	0.00	0.93	(0.00)	-0.68	(0.00)

**Panel E: Forecast Combinations Model**

$h$	Output Growth Forecasts					Inflation Forecasts				
	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.00	0.62	(0.01)	0.22	(0.00)	0.00	0.25	(0.05)	0.45	(0.01)
2	0.00	0.87	(0.00)	-0.04	(0.00)	0.00	0.42	(0.01)	0.13	(0.00)
3	0.00	1.26	(0.00)	-0.47	(0.00)	0.01	0.32	(0.02)	0.32	(0.02)
4	0.00	1.04	(0.00)	-0.24	(0.00)	0.00	0.33	(0.00)	0.28	(0.00)
5	0.21	0.53	(0.22)	0.34	(0.22)	0.00	0.52	(0.00)	-0.06	(0.00)
6	0.00	0.90	(0.00)	-0.08	(0.00)	0.00	0.68	(0.00)	-0.33	(0.00)
7	0.00	0.81	(0.00)	0.03	(0.00)	0.00	0.70	(0.00)	-0.35	(0.00)
8	0.00	0.89	(0.00)	-0.06	(0.00)	0.00	0.59	(0.00)	-0.17	(0.00)

**Panel F: Bayesian VAR Model**

$h$	Output Growth Forecasts					Inflation Forecasts				
	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.00	0.81	(0.00)	0.02	(0.00)	0.00	0.44	(0.00)	0.00	(0.00)
2	0.00	0.80	(0.00)	0.07	(0.00)	0.00	0.48	(0.00)	0.01	(0.00)
3	0.00	0.74	(0.00)	0.16	(0.00)	0.00	0.49	(0.00)	-0.00	(0.00)
4	0.00	0.70	(0.00)	0.20	(0.00)	0.00	0.44	(0.00)	0.06	(0.00)
5	0.00	0.91	(0.00)	-0.11	(0.00)	0.00	0.45	(0.00)	0.04	(0.00)
6	0.00	0.62	(0.01)	0.24	(0.00)	0.00	0.52	(0.00)	-0.05	(0.00)
7	0.00	0.66	(0.00)	0.20	(0.00)	0.00	0.57	(0.00)	-0.11	(0.00)
8	0.00	0.93	(0.00)	-0.10	(0.00)	0.00	0.53	(0.00)	-0.05	(0.00)

Note to the table. The table reports results for forecasting output (on the left) and inflation (on the right). Each panel reports, for several horizons ( $h$ , reported in the first column), the values of the Mincer and Zarnowitz (1969) forecast efficiency test (labeled "MZ p-value"), the estimate of  $\hat{\alpha}$  in regression (5) together with the p-value of the test that the constant equals zero (in parentheses), and the estimate of  $\hat{\beta}$  in regression (5) together with the p-value of the test that the slope equals zero (in parentheses).

**Table 2. The Diebold and Mariano Test:  
Forecasting Output Growth and Inflation**

$h$	Output					Inflation				
	RMSFE			DM: Model-DSGE		RMSFE			DM: Model-DSGE	
	DSGE	VAR	AR	VAR	AR	DSGE	VAR	AR	VAR	AR
1	0.54	1.09	0.94	0.45	0.61	0.19	1.12	1.06	0.31	0.46
2	0.46	1.31	1.11	0.04	0.20	0.23	1.05	0.93	0.64	0.22
3	0.46	1.41	1.16	0.05	0.20	0.26	0.95	0.80	0.34	0.00
4	0.46	1.46	1.15	0.06	0.16	0.26	1.09	0.83	0.42	0.00
5	0.48	1.24	0.98	0.06	0.87	0.28	1.20	0.85	0.30	0.01
6	0.48	1.39	1.09	0.22	0.29	0.29	1.38	0.90	0.20	0.11
7	0.50	1.45	1.04	0.21	0.50	0.30	1.31	0.89	0.26	0.02
8	0.51	1.59	1.06	0.16	0.35	0.31	1.37	0.88	0.25	0.01

$h$	Output						Inflation					
	RMSFE			DM: Model-DSGE			RMSFE			DM: Model-DSGE		
	FC	BVAR	RW	FC	BVAR	RW	FC	BVAR	RW	FC	BVAR	RW
1	0.95	1.50	1.06	0.67	0.01	0.51	1.03	2.66	1.06	0.69	0.00	0.44
2	1.10	1.58	1.40	0.25	0.00	0.00	0.95	2.25	1.06	0.36	0.01	0.50
3	1.14	1.41	1.37	0.29	0.03	0.01	0.81	1.95	0.92	0.00	0.02	0.31
4	1.14	1.31	1.43	0.17	0.02	0.02	0.82	1.80	0.83	0.00	0.03	0.07
5	1.00	1.33	1.48	0.98	0.06	0.00	0.87	1.70	0.85	0.01	0.04	0.15
6	1.11	1.09	1.40	0.21	0.51	0.02	0.94	1.69	0.88	0.43	0.09	0.34
7	1.07	1.05	1.37	0.34	0.63	0.01	0.92	1.63	0.92	0.17	0.12	0.53
8	1.13	1.11	1.45	0.20	0.33	0.00	0.91	1.56	0.84	0.15	0.13	0.17

Note to the table. The columns labeled RMSFE in the table report the RMSFE of the DSGE model forecasts (labeled "DSGE") as well as the RMSFE of the autoregressive ("AR"), forecast combinations ("FC"), VAR ("VAR"), Bayesian VAR ("BVAR") and the random walk ("RW") relative to that of the DSGE model (e.g. the column labeled "VAR" reports the RMSFE of the VAR model divided by the RMSFE of the DSGE model). The columns labeled "DM" in the table report the p-value of the Diebold and Mariano (1996) test (calculated using Giacomini and White's (2006) critical values) for comparing the specified forecasts (e.g. the column labeled "DM: Model-DSGE" and "VAR" reports the p-value for comparing the forecast of the VAR with those of Smets and Wouters' (2007) model).



**Table 3. Forecast Optimality Tests: Forecasting Interest Rates**

Panel A: VAR Model						Panel B: DSGE Model					
$h$	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		
1	0.66	-0.03	(0.68)	1.04	(0.69)	0.00	-0.14	(0.00)	1.11	(0.03)	
2	0.97	-0.02	(0.97)	1.02	(0.97)	0.01	-0.24	(0.03)	1.18	(0.21)	
3	0.76	0.00	(1.00)	0.95	(0.94)	0.04	-0.28	(0.26)	1.18	(0.62)	
4	0.36	0.01	(1.00)	0.89	(0.84)	0.07	-0.26	(0.60)	1.12	(0.90)	
5	0.17	0.05	(0.98)	0.81	(0.74)	0.11	-0.17	(0.89)	1.00	(1.00)	
6	0.05	0.20	(0.84)	0.65	(0.48)	0.11	-0.03	(1.00)	0.85	(0.93)	
7	0.01	0.39	(0.63)	0.47	(0.23)	0.08	0.19	(0.92)	0.66	(0.71)	
8	0.00	0.43	(0.56)	0.42	(0.12)	0.05	0.53	(0.63)	0.37	(0.38)	
Panel C: AR Model						Panel D: Random Walk Model					
$h$	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		
1	0.01	-0.03	(0.48)	0.97	(0.58)	0.68	0.02	(0.85)	0.97	(0.69)	
2	0.01	-0.05	(0.59)	0.94	(0.54)	0.38	0.08	(0.49)	0.90	(0.38)	
3	0.01	-0.06	(0.83)	0.90	(0.56)	0.20	0.17	(0.26)	0.81	(0.21)	
4	0.01	-0.05	(0.96)	0.85	(0.56)	0.12	0.27	(0.15)	0.70	(0.13)	
5	0.00	-0.00	(1.00)	0.78	(0.53)	0.06	0.40	(0.07)	0.58	(0.07)	
6	0.00	0.06	(0.98)	0.69	(0.43)	0.03	0.51	(0.04)	0.48	(0.04)	
7	0.00	0.12	(0.94)	0.61	(0.33)	0.01	0.62	(0.02)	0.38	(0.02)	
8	0.00	0.12	(0.96)	0.60	(0.32)	0.01	0.71	(0.01)	0.29	(0.01)	
Panel E: Forecast Combinations Model						Panel F: Bayesian VAR Model					
$h$	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		
1	0.01	-0.03	(0.44)	0.98	(0.74)	0.00	0.38	(0.40)	0.33	(0.00)	
2	0.01	-0.05	(0.51)	0.96	(0.75)	0.00	0.38	(0.46)	0.33	(0.00)	
3	0.02	-0.06	(0.75)	0.92	(0.70)	0.00	0.38	(0.51)	0.34	(0.00)	
4	0.01	-0.05	(0.94)	0.87	(0.64)	0.00	0.39	(0.53)	0.34	(0.01)	
5	0.01	-0.03	(0.99)	0.82	(0.66)	0.00	0.44	(0.47)	0.32	(0.01)	
6	0.00	0.02	(1.00)	0.74	(0.57)	0.00	0.50	(0.39)	0.30	(0.01)	
7	0.00	0.13	(0.94)	0.62	(0.38)	0.00	0.56	(0.28)	0.27	(0.00)	
8	0.00	0.15	(0.93)	0.59	(0.32)	0.00	0.61	(0.17)	0.24	(0.00)	

Note to the table. The table reports results for forecasting interest rates. Each panel reports, for several horizons ( $h$ , reported in the first column), the values of the Mincer and Zarnowitz (1969) forecast efficiency test (labeled "MZ p-value"), the estimate of  $\hat{\alpha}$  in regression (5) together with the p-value of the test that the constant equals zero (in parentheses), and the estimate of  $\hat{\beta}$  in regression (5) together with the p-value of the test that the slope equals zero (in parentheses).

**Table 4. Diebold and Mariano Test: Forecasting Interest Rates**

h	RMSFE			DM: Model-DSGE		RMSFE			DM: Model-DSGE		
	DSGE	VAR	AR	VAR	AR	FC	BVAR	RW	FC	BVAR	RW
1	0.11	1.15	1.02	0.35	0.87	0.95	8.99	0.93	0.64	0.02	0.40
2	0.20	1.07	1.06	0.61	0.71	0.99	4.88	0.95	0.93	0.03	0.54
3	0.28	1.02	1.09	0.85	0.54	1.03	3.41	0.97	0.84	0.04	0.77
4	0.34	1.01	1.13	0.93	0.37	1.08	2.68	0.98	0.62	0.06	0.90
5	0.39	1.02	1.15	0.87	0.31	1.09	2.25	1.01	0.57	0.08	0.93
6	0.43	1.05	1.19	0.71	0.28	1.12	1.96	1.02	0.49	0.10	0.89
7	0.47	1.10	1.23	0.56	0.24	1.17	1.76	1.02	0.38	0.13	0.91
8	0.51	1.12	1.22	0.50	0.22	1.18	1.72	1.01	0.34	0.14	0.96

Note to the table. The columns labeled RMSFE in the table report the RMSFE of the DSGE model forecasts (labeled "DSGE") as well as the RMSFE of the autoregressive ("AR"), forecast combinations ("FC"), VAR ("VAR"), Bayesian VAR ("BVAR") and the random walk ("RW") relative to that of the DSGE model (e.g. the column labeled "VAR" reports the RMSFE of the VAR model divided by the RMSFE of the DSGE model). The columns labeled "DM" in the table report the p-value of the Diebold and Mariano (1996) test (calculated using Giacomini and White's (2006) critical values) for comparing the specified forecasts (e.g. the column labeled "DM: Model-DSGE" and "VAR" reports the p-value for comparing the forecast of the VAR with those of Smets and Wouters' (2007) model).

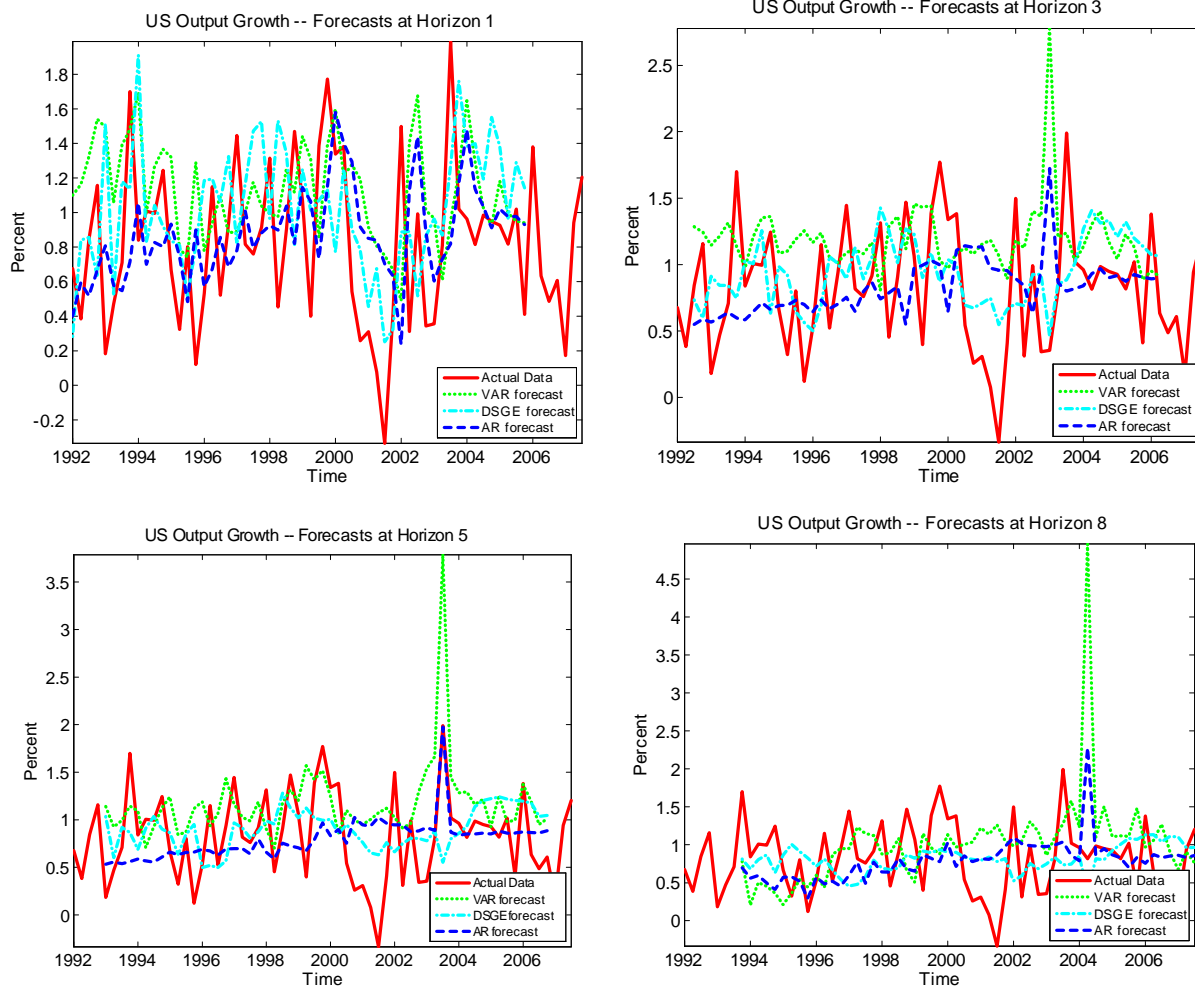
**Table 5. Comparisons of MSFEs: Rolling vs. Recursive Estimation**

<b>Forecasting Output</b>												
Panel A. Recursive							Panel B. Rolling					
$h$	DSGE	VAR	AR	FC	BVAR	RW	DSGE	VAR	AR	FC	BVAR	RW
1	0.54	1.18	0.85	0.858	1.39	1.05	0.54	1.09	0.94	0.95	1.50	1.06
2	0.51	1.43	0.97	0.97	1.29	1.25	0.46	1.31	1.11	1.10	1.58	1.40
3	0.49	1.42	1.00	0.99	1.26	1.47	0.46	1.41	1.16	1.14	1.41	1.37
4	0.48	1.43	1.03	1.02	1.26	1.45	0.46	1.46	1.15	1.14	1.31	1.43
5	0.46	1.38	1.04	1.00	1.36	1.50	0.48	1.24	0.98	1.00	1.33	1.48
6	0.44	1.51	1.12	1.12	1.32	1.67	0.48	1.39	1.09	1.11	1.09	1.40
7	0.44	1.52	1.11	1.11	1.32	1.53	0.50	1.45	1.04	1.07	1.05	1.37
8	0.44	1.58	1.12	1.13	1.36	1.60	0.51	1.59	1.06	1.13	1.11	1.45
<b>Forecasting Inflation</b>												
Panel A. Recursive							Panel B. Rolling					
$h$	DSGE	VAR	AR	FC	BVAR	RW	DSGE	VAR	AR	FC	BVAR	RW
1	0.19	1.15	1.12	1.12	2.78	1.06	0.19	1.12	1.06	1.03	2.66	1.06
2	0.21	1.27	1.11	1.15	2.56	1.14	0.23	1.05	0.93	0.95	2.25	1.06
3	0.22	1.20	1.09	1.13	2.45	1.02	0.26	0.95	0.80	0.81	1.95	0.92
4	0.23	1.27	1.18	1.18	2.36	1.24	0.26	1.09	0.83	0.82	1.80	0.83
5	0.23	1.48	1.33	1.35	2.37	1.22	0.28	1.20	0.85	0.87	1.70	0.85
6	0.23	1.80	1.53	1.56	2.47	1.14	0.29	1.38	0.90	0.94	1.69	0.88
7	0.26	1.73	1.47	1.47	2.29	1.08	0.30	1.31	0.89	0.92	1.63	0.92
8	0.26	1.85	1.55	1.54	2.24	1.11	0.31	1.37	0.88	0.91	1.56	0.84
<b>Forecasting Int. Rate</b>												
Panel A. Recursive							Panel C. Rolling					
$h$	DSGE	VAR	AR	FC	BVAR	RW	DSGE	VAR	AR	FC	BVAR	RW
1	0.14	0.72	0.90	0.81	6.94	0.72	0.11	1.15	1.02	0.95	8.99	0.93
2	0.21	0.90	1.16	1.07	4.44	1.28	0.20	1.07	1.06	0.99	4.88	0.95
3	0.28	0.94	1.20	1.11	3.15	1.39	0.28	1.02	1.09	1.03	3.41	0.97
4	0.33	1.00	1.24	1.17	2.51	1.38	0.34	1.01	1.13	1.08	2.68	0.98
5	0.38	1.01	1.29	1.21	2.14	1.39	0.39	1.02	1.15	1.09	2.25	1.01
6	0.42	1.07	1.37	1.28	1.94	1.42	0.43	1.05	1.19	1.12	1.96	1.02
7	0.45	1.13	1.42	1.33	1.78	1.41	0.47	1.10	1.23	1.17	1.76	1.02
8	0.48	1.16	1.42	1.34	1.68	1.40	0.51	1.12	1.22	1.18	1.72	1.01

Note to the table. The columns labeled RMSFE in the table report the RMSFE of the DSGE model forecasts (labeled "DSGE") as well as the RMSFE of the autoregressive ("AR"), forecast combinations ("FC"), VAR ("VAR"), Bayesian VAR ("BVAR") and the random walk ("RW") relative to that of the DSGE model (e.g. the column labeled "VAR" reports the RMSFE of the VAR model divided by the RMSFE of the DSGE model). Panel A reports results for the recursive estimation scheme and panel B for the rolling scheme.

# Figures

## Figure 1(a). Output Growth Forecasts



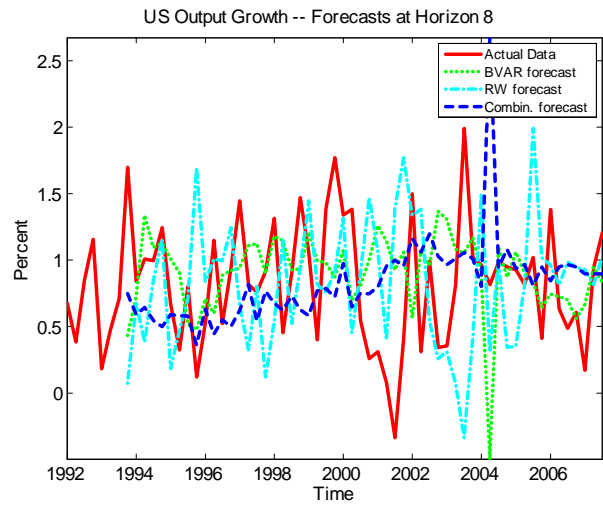
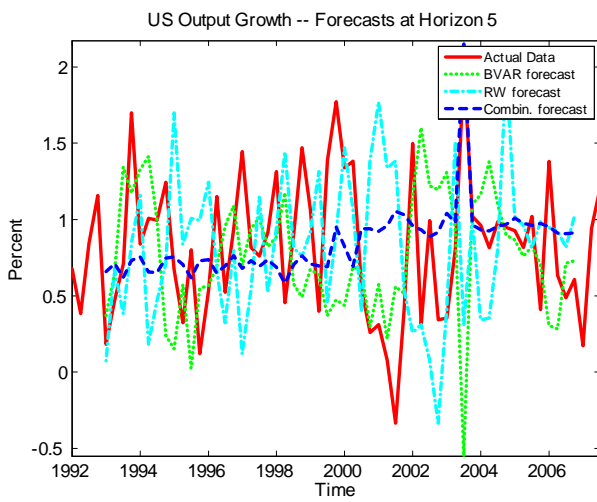
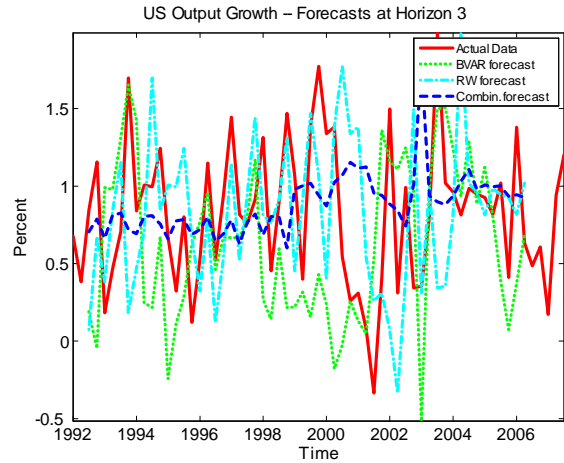
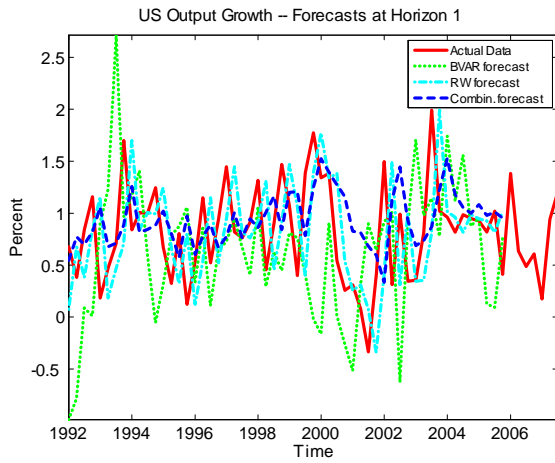
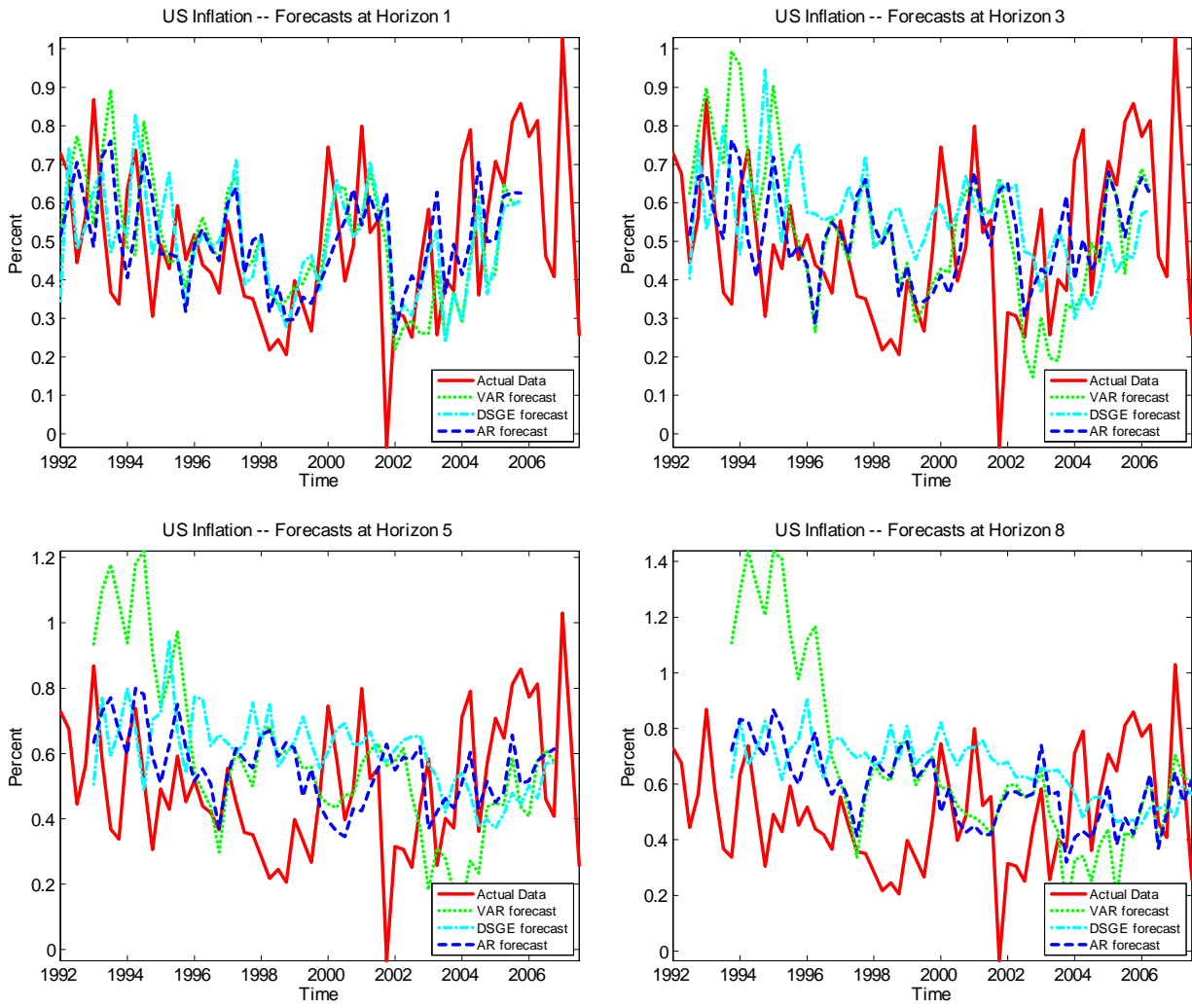
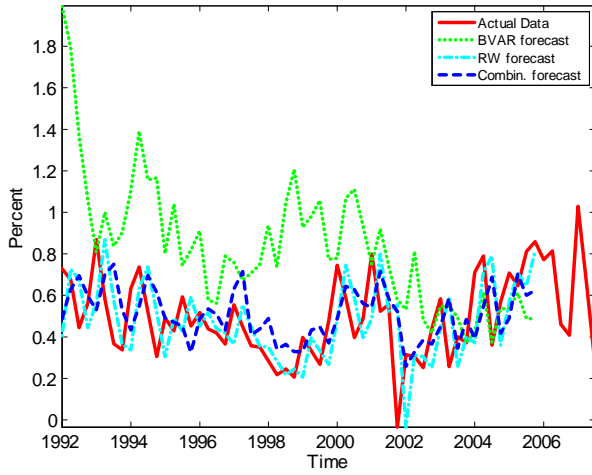


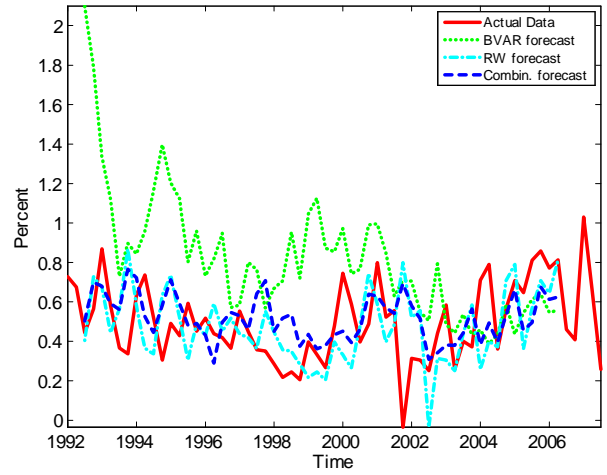
Figure 1(b). Inflation Forecasts



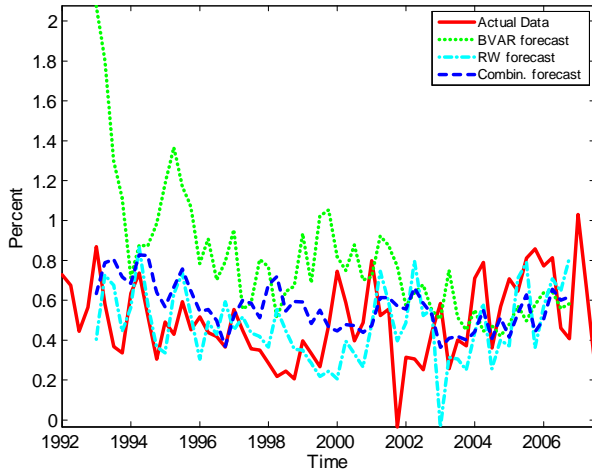
US Inflation -- Forecasts at Horizon 1



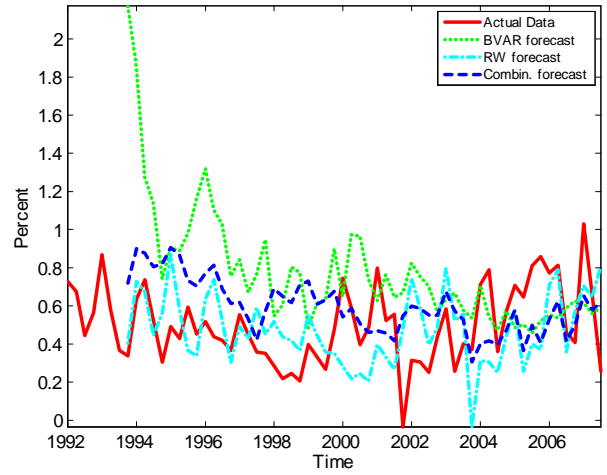
US Inflation -- Forecasts at Horizon 3



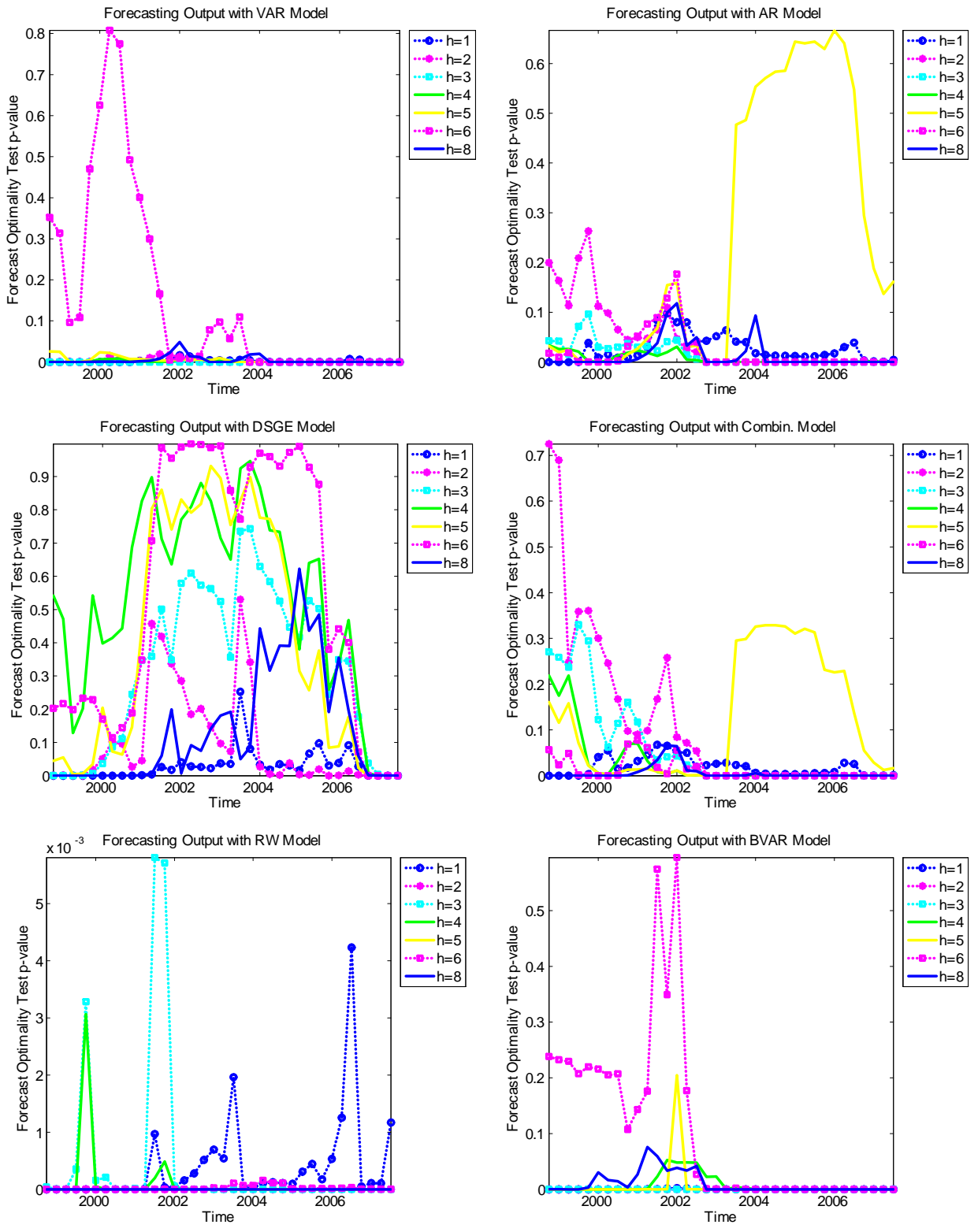
US Inflation -- Forecasts at Horizon 5



US Inflation -- Forecasts at Horizon 8



**Figure 2(a). Forecast Optimality Test Over Time: Output Growth**





**Figure 2(b). Forecast Optimality Test Over Time: Inflation**

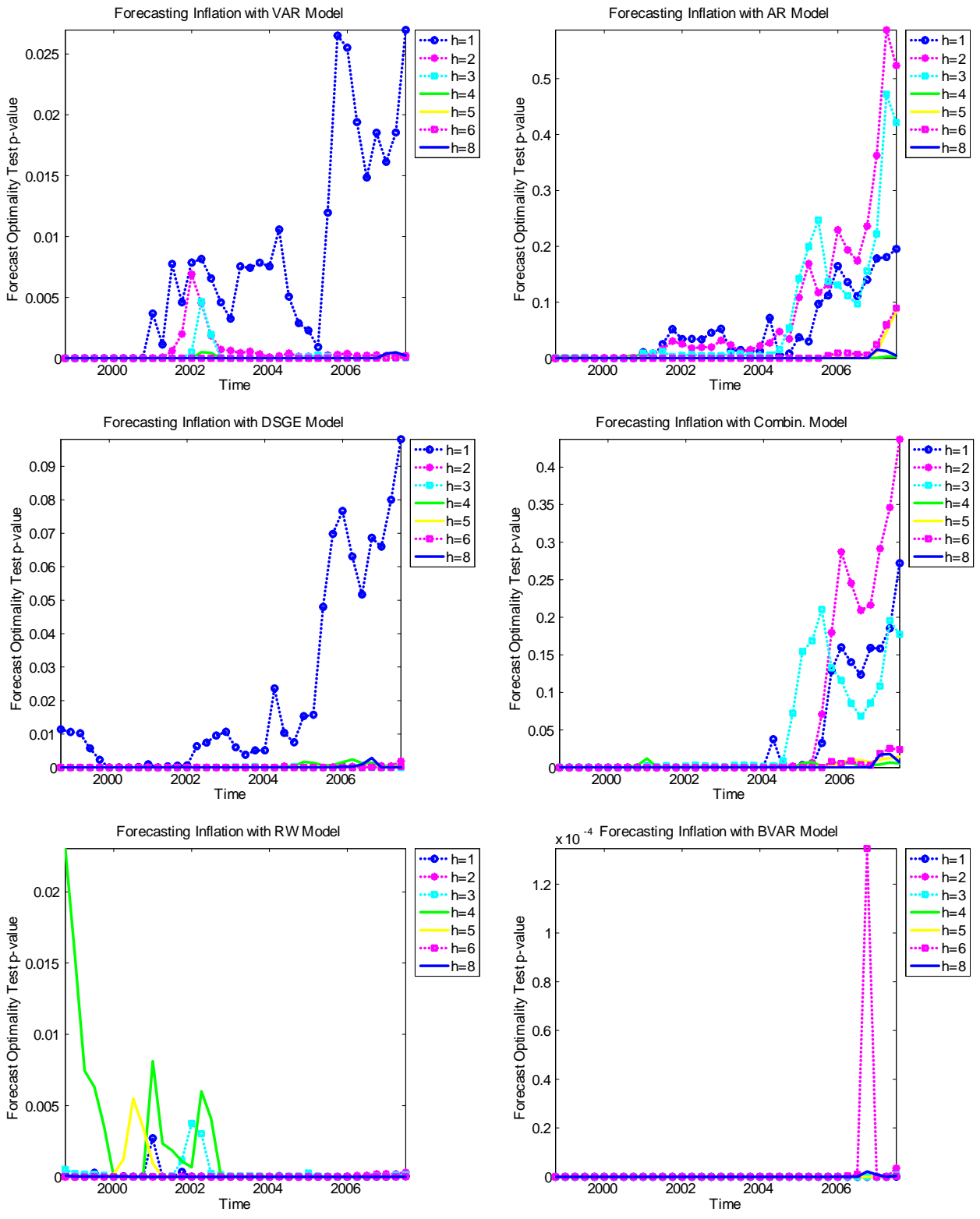


Figure 3(a). Fluctuation Test: Output

(VAR vs. DSGE Model)

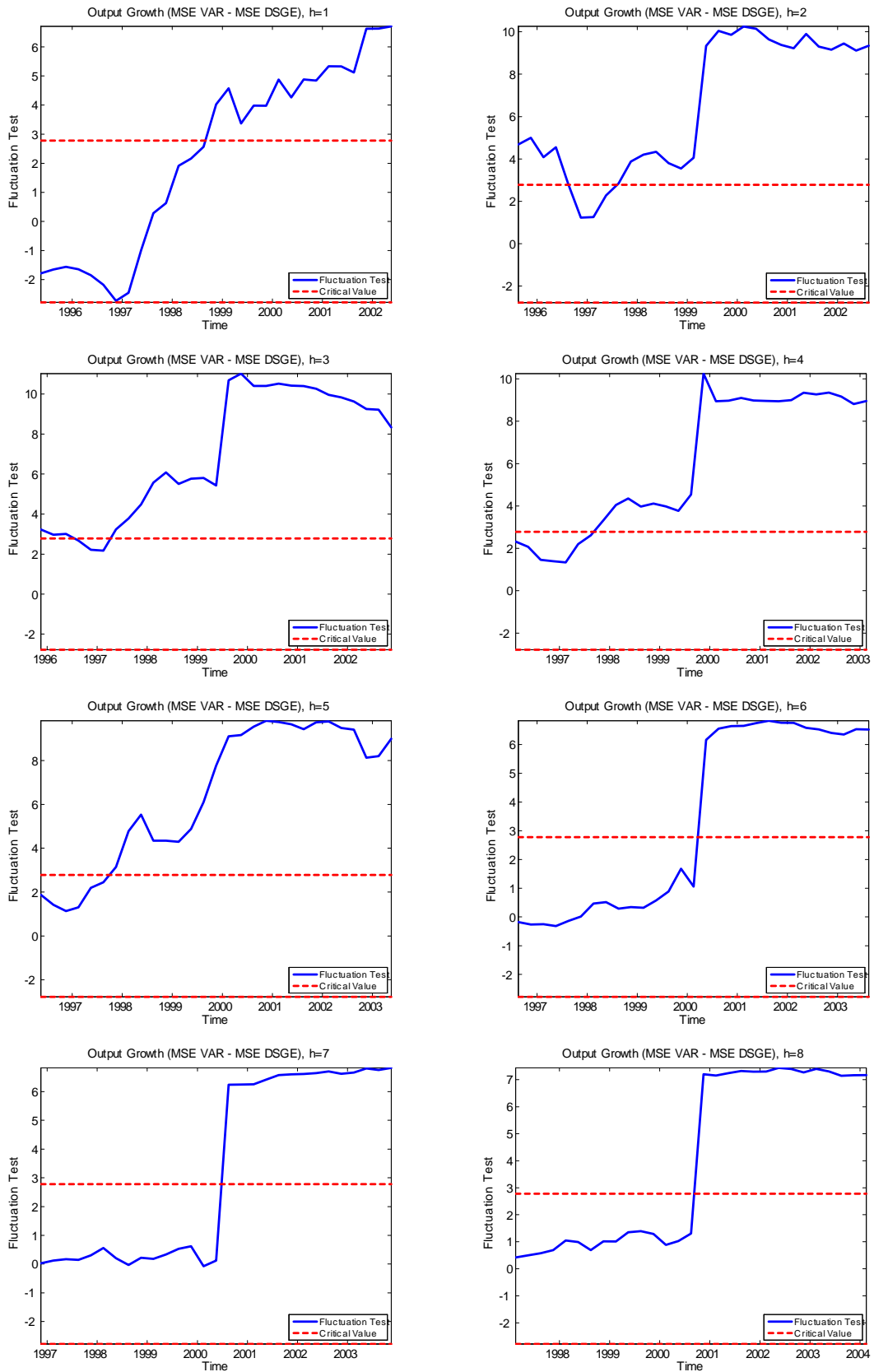
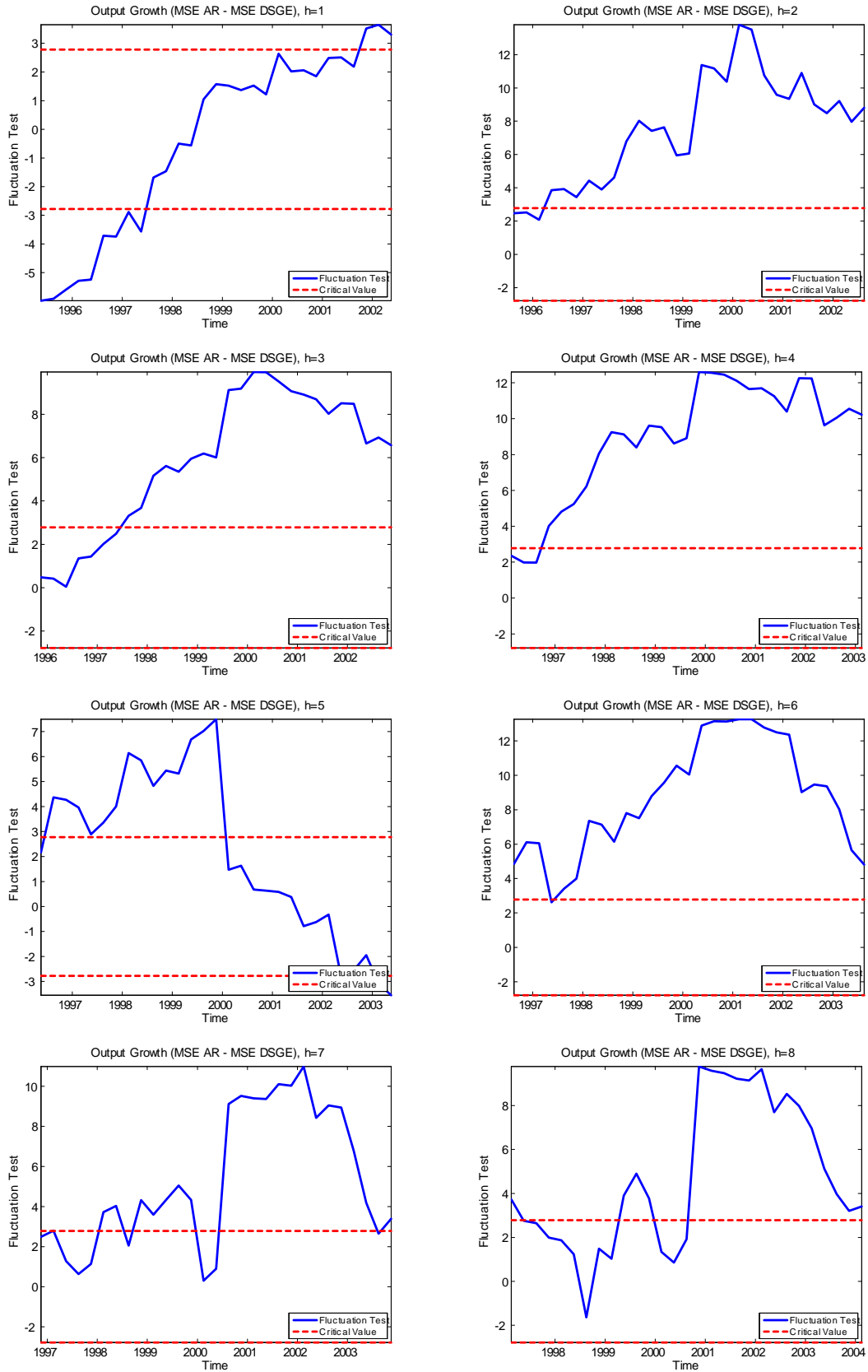


Figure 3(a). Fluctuation Test: Output

(AR vs. DSGE Model)



**Figure 3(b). Fluctuation Test: Output (Combinations vs. DSGE Model)**

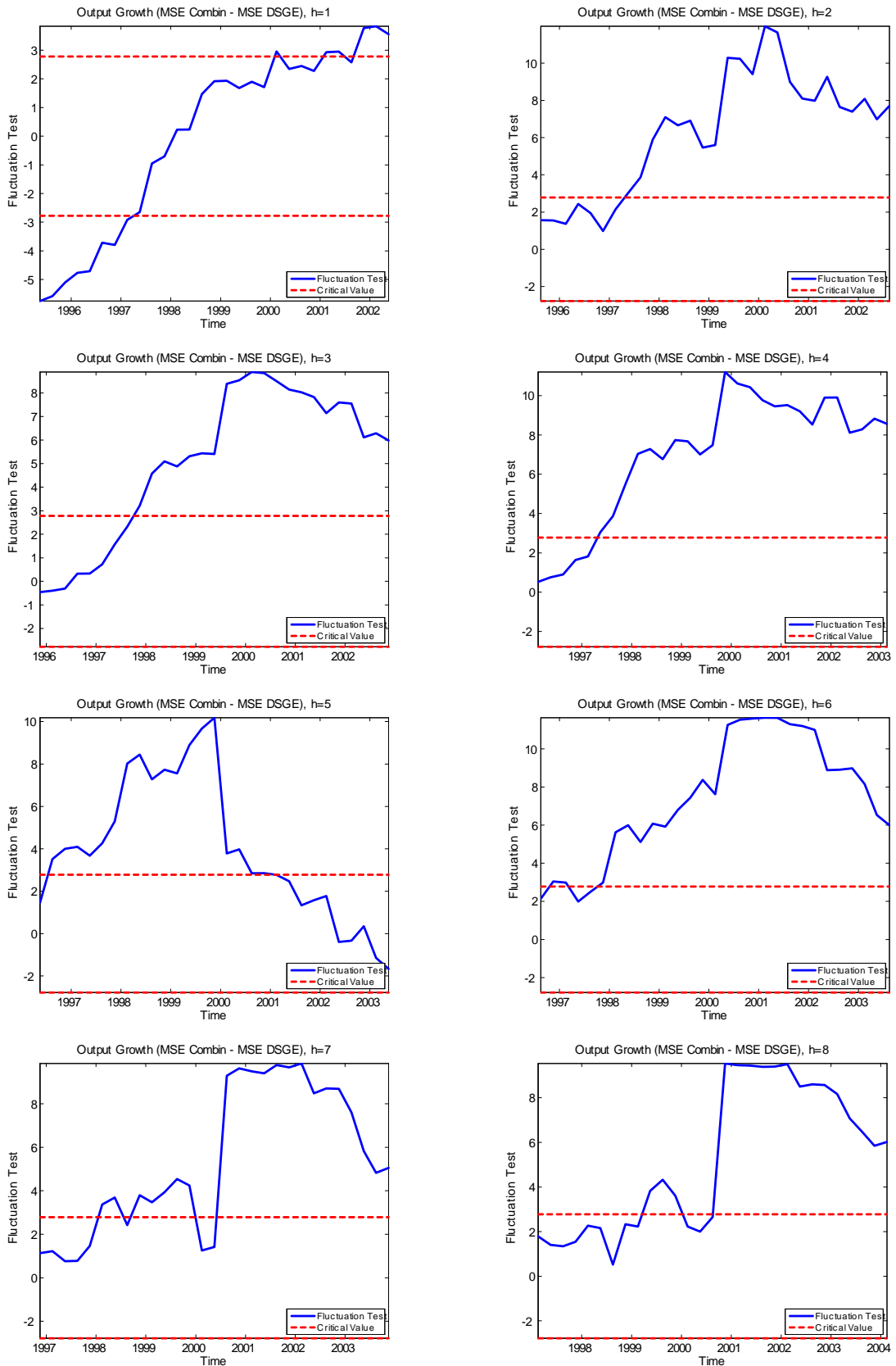


Figure 3(b). Fluctuation Test: Output

(BVAR vs. DSGE Model)

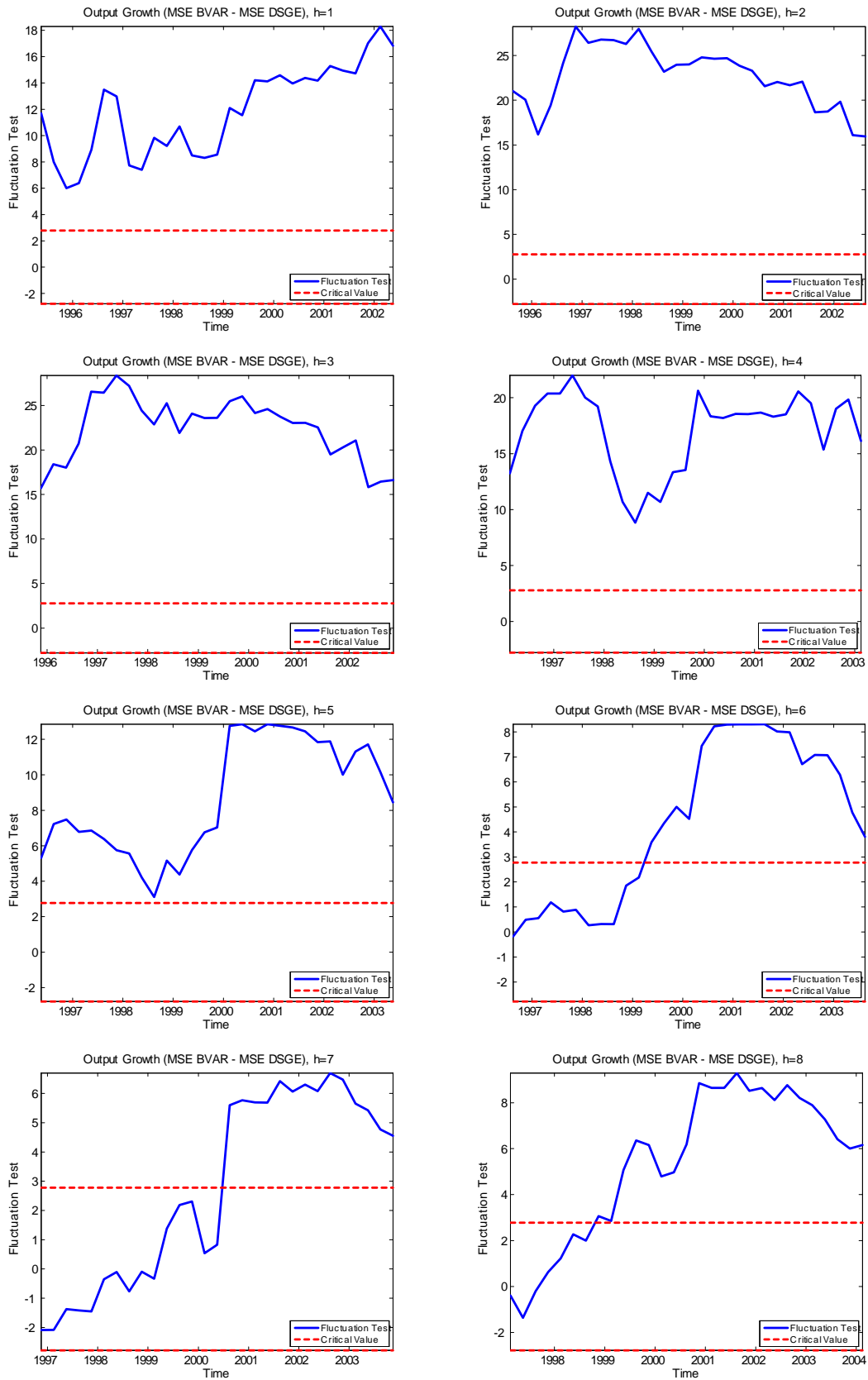


Figure 3(b). Fluctuation Test: Inflation

(VAR vs. DSGE Model)

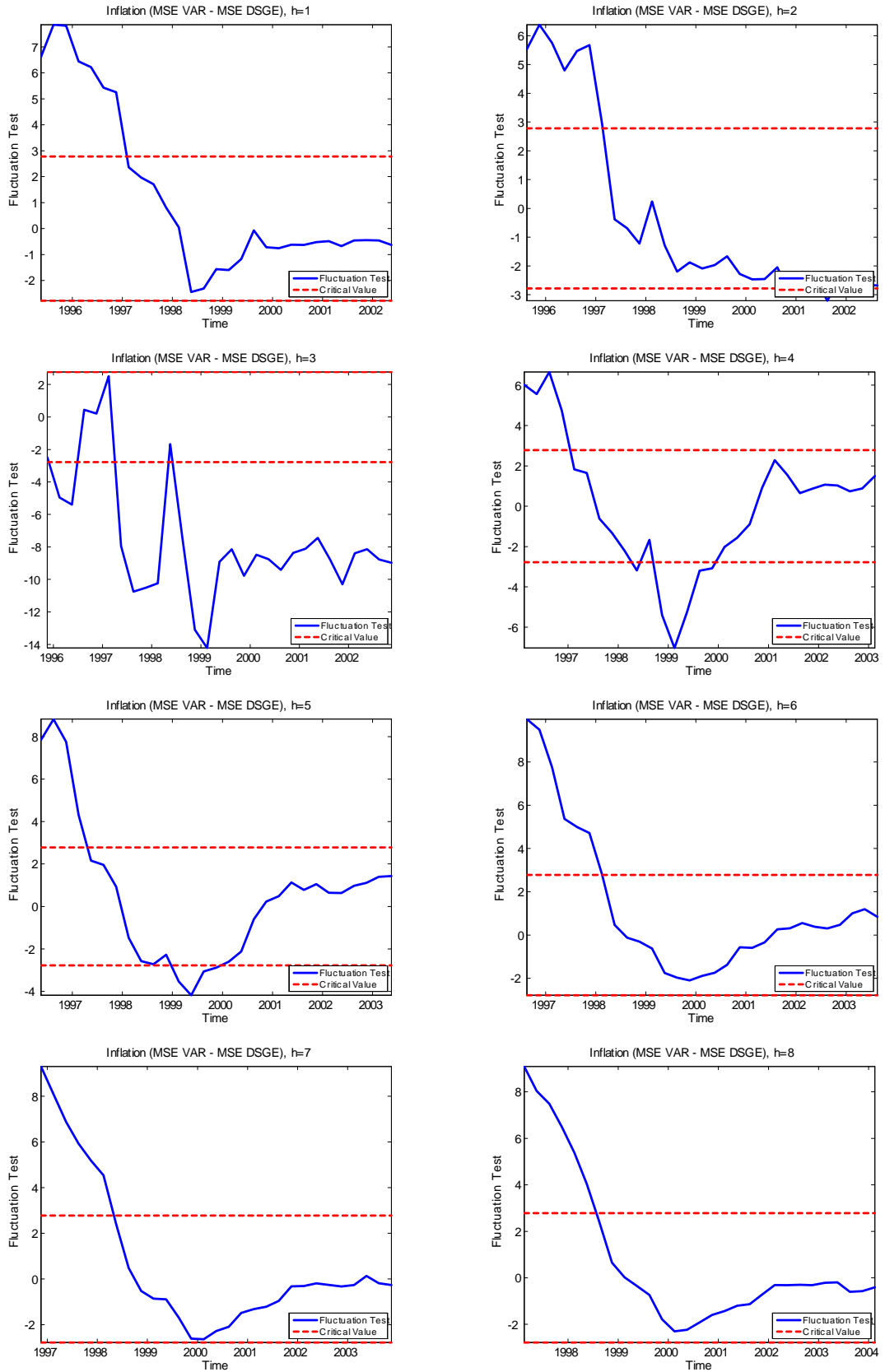


Figure 3(b). Fluctuation Test: Inflation

(AR vs. DSGE Model)

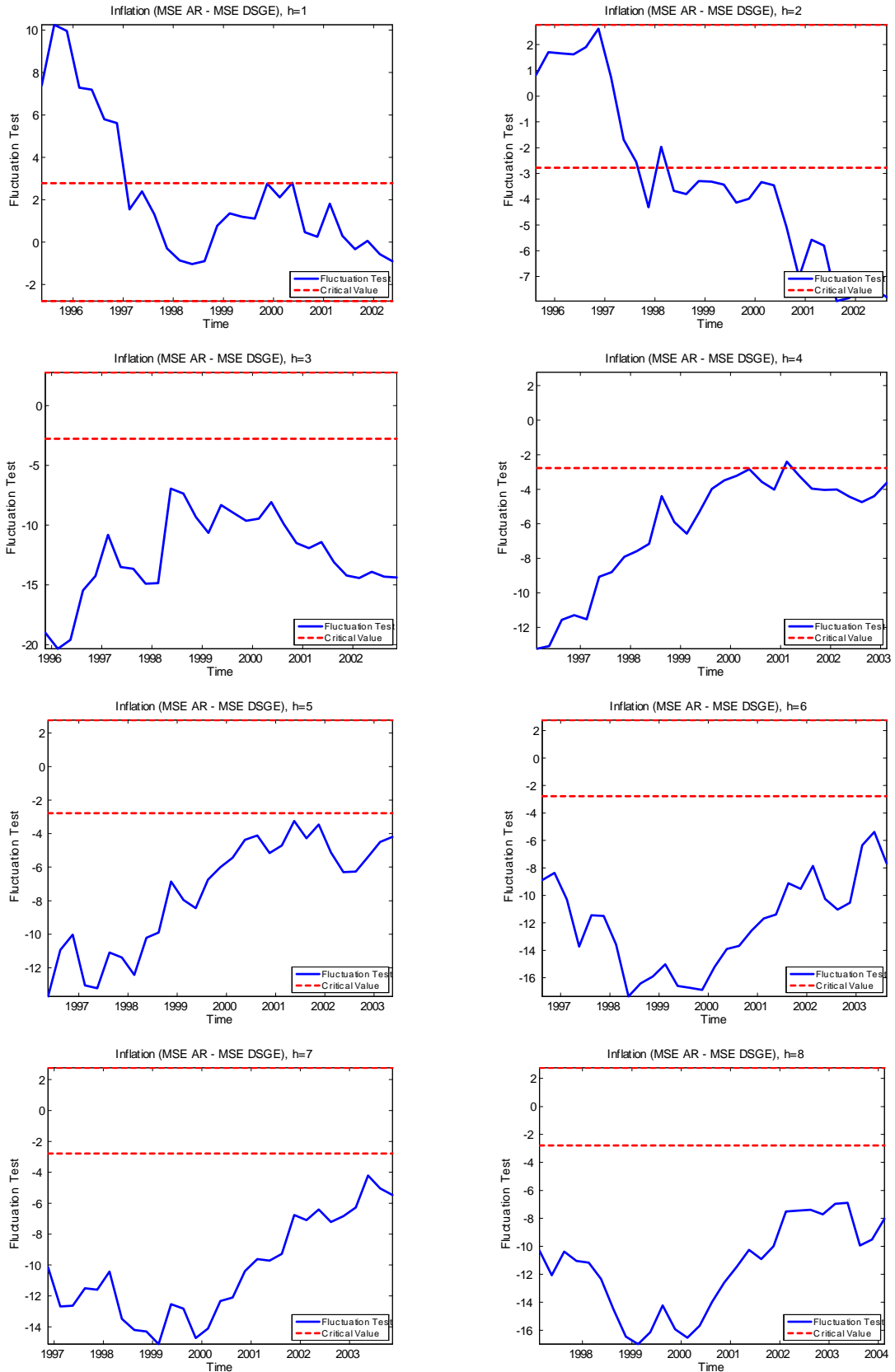


Figure 3(b). Fluctuation Test: Inflation

(Combinations vs. DSGE Model)

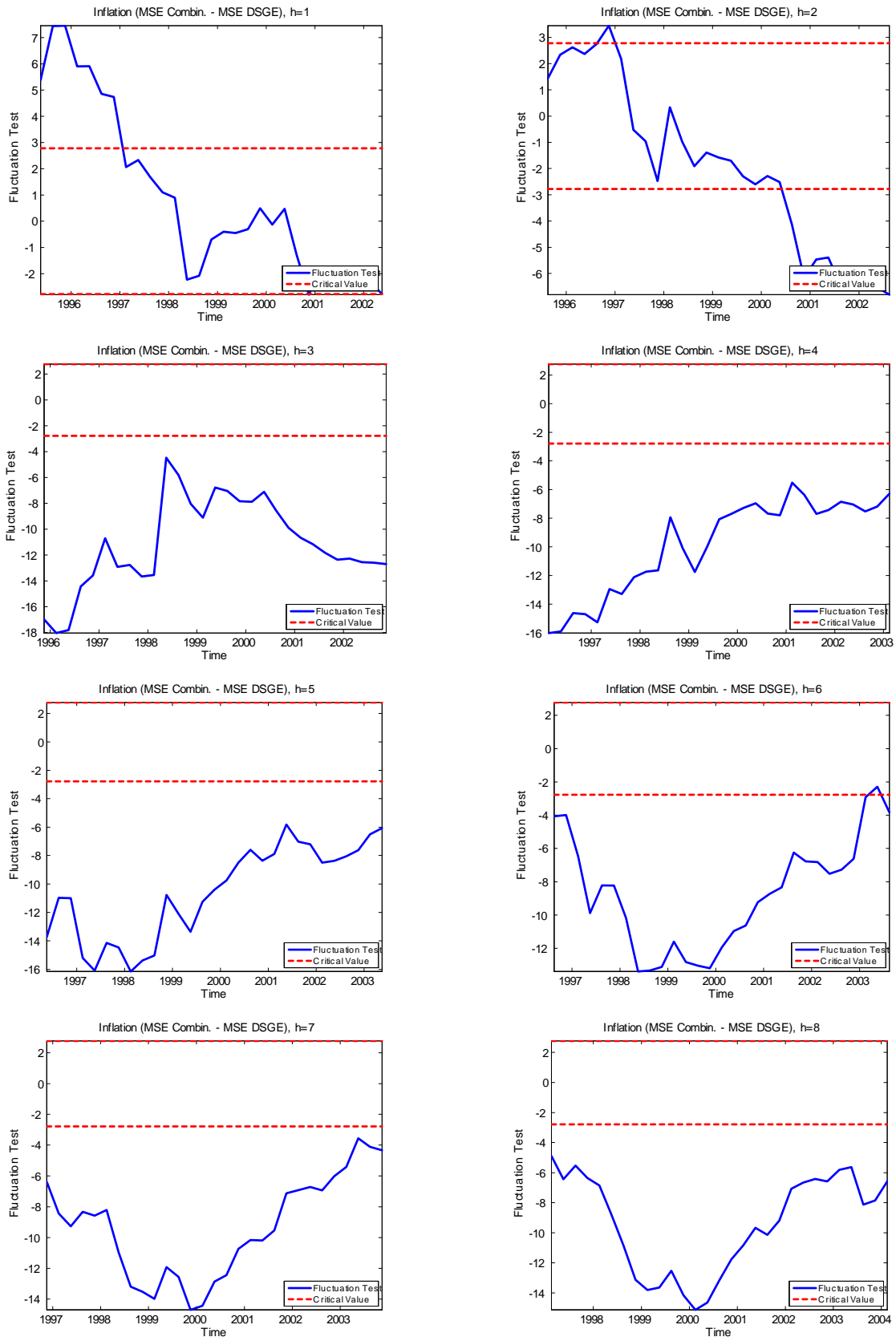




Figure 3(b). Fluctuation Test: Inflation

(BVAR vs. DSGE Model)

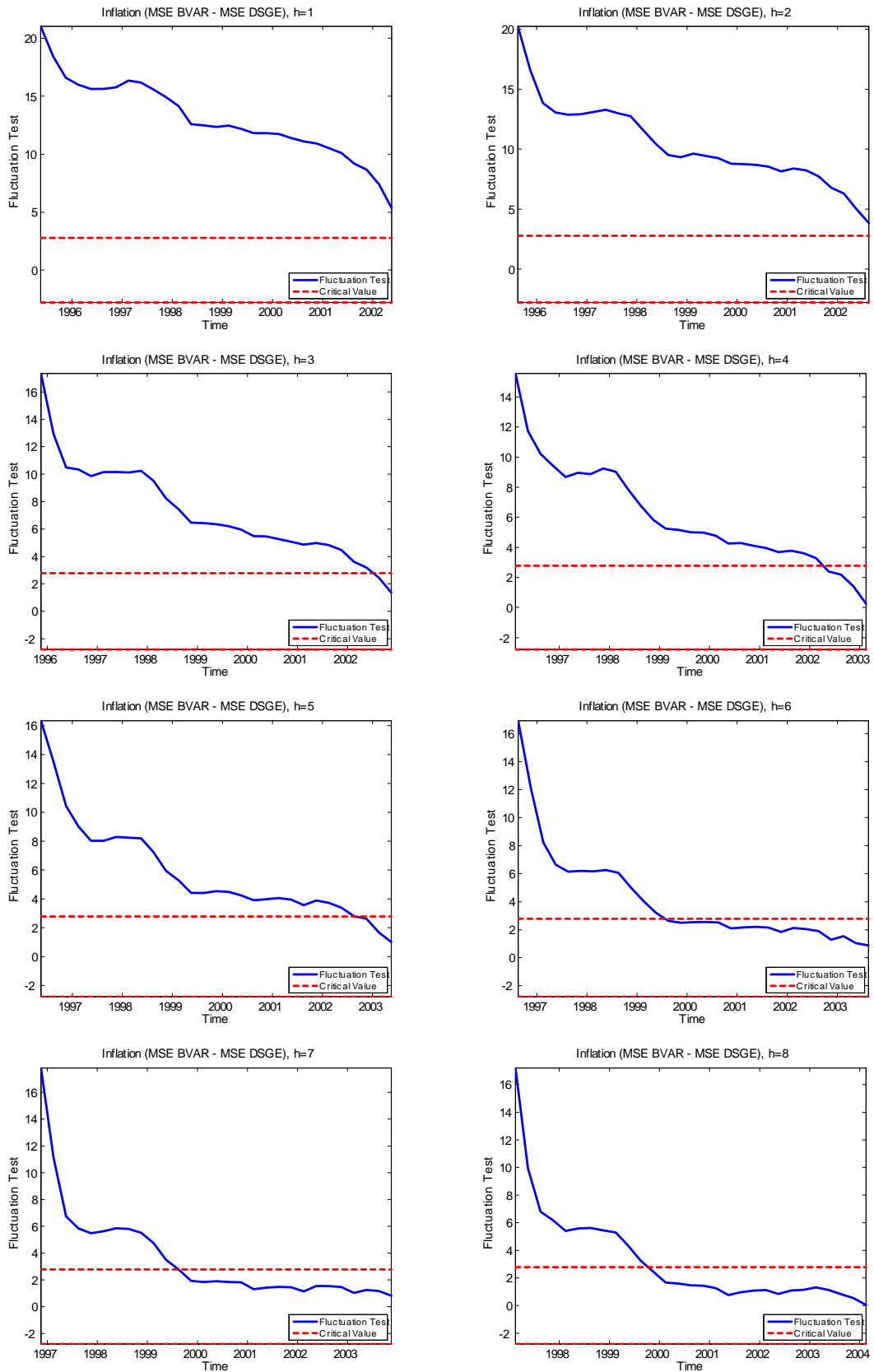
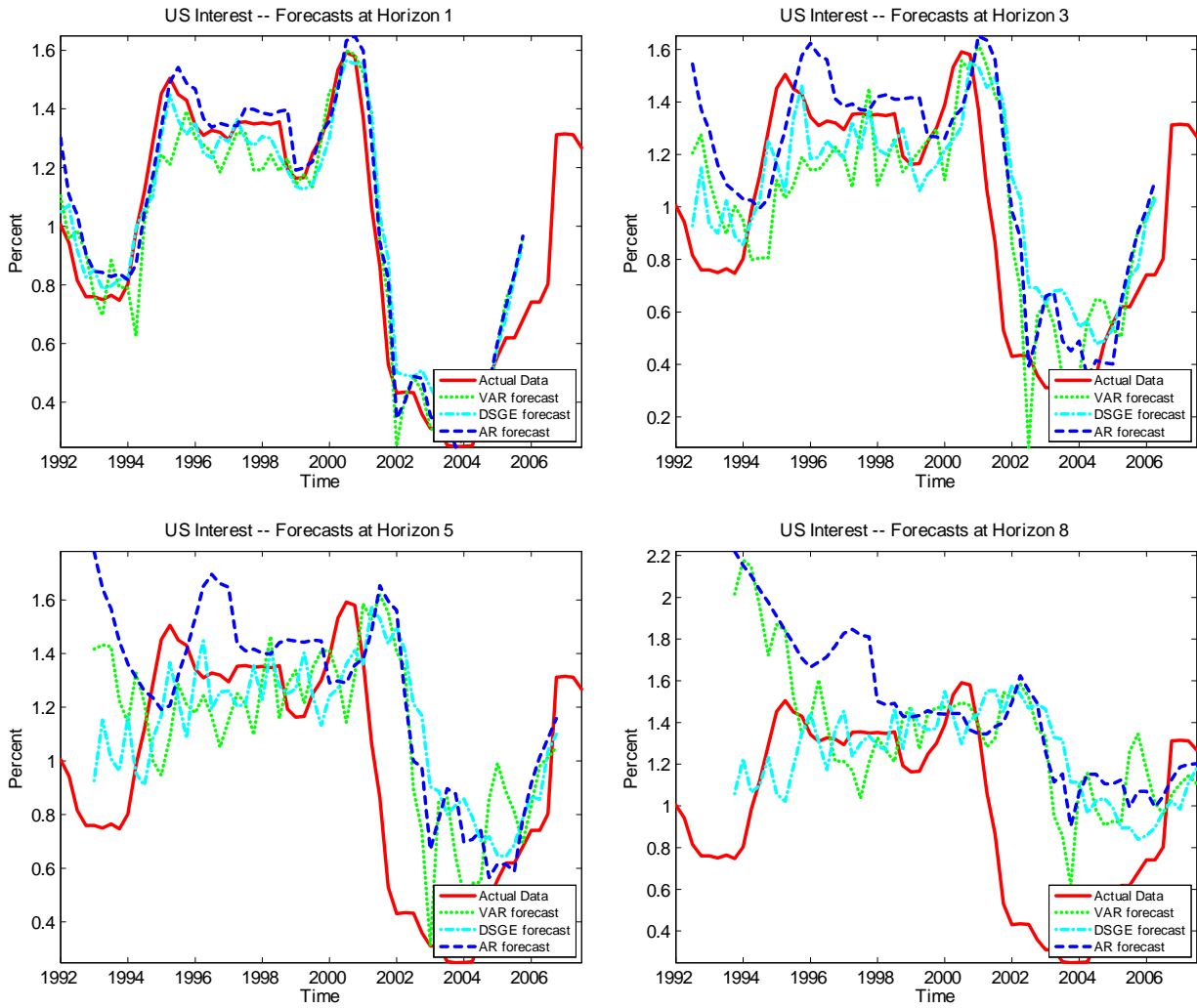


Figure 4. Interest Rate Forecasts



**Figure 5. Forecast Optimality Test Over Time: Interest Rate**

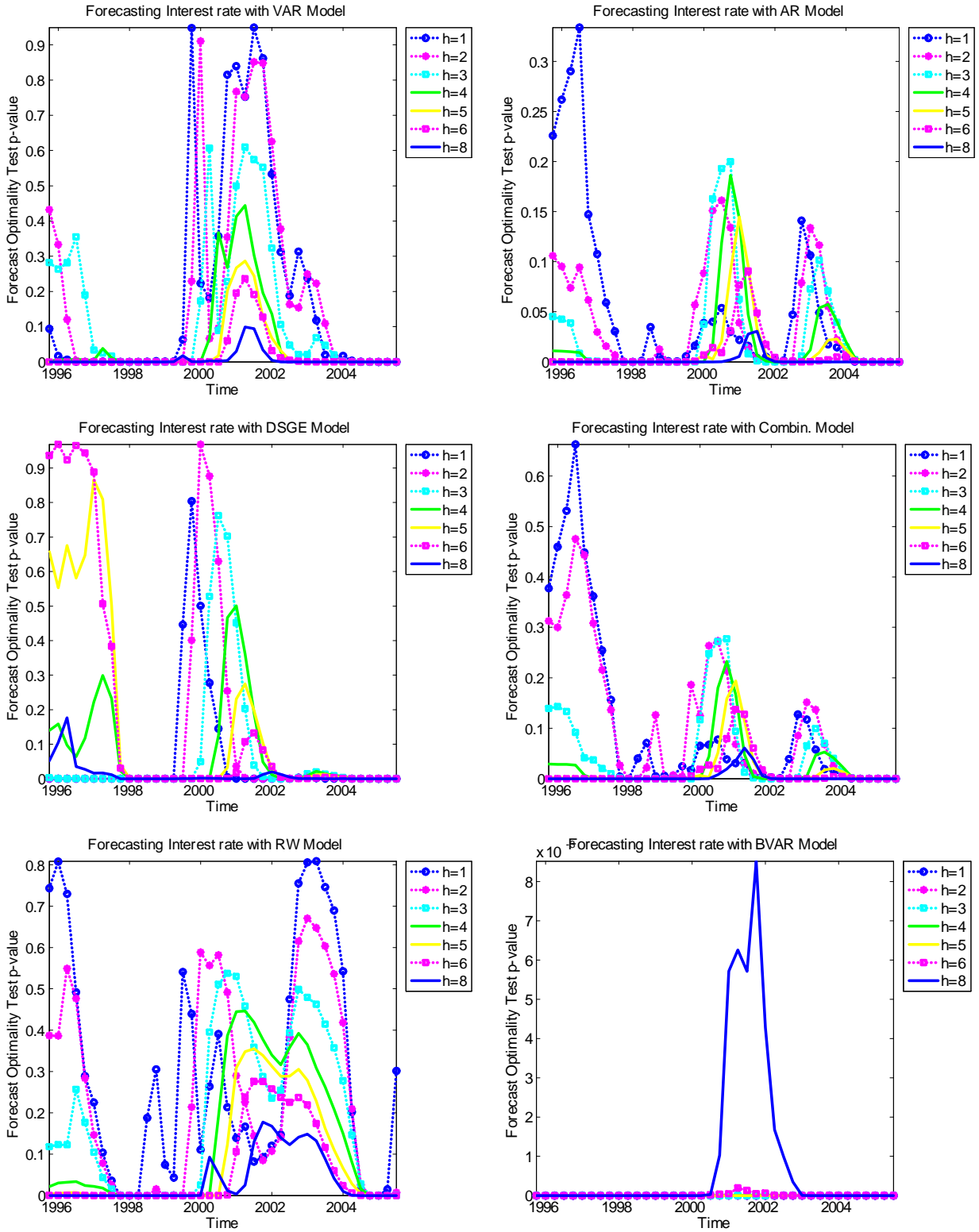


Figure 6(a). Fluctuation Test: Interest Rate

(VAR vs. DSGE)

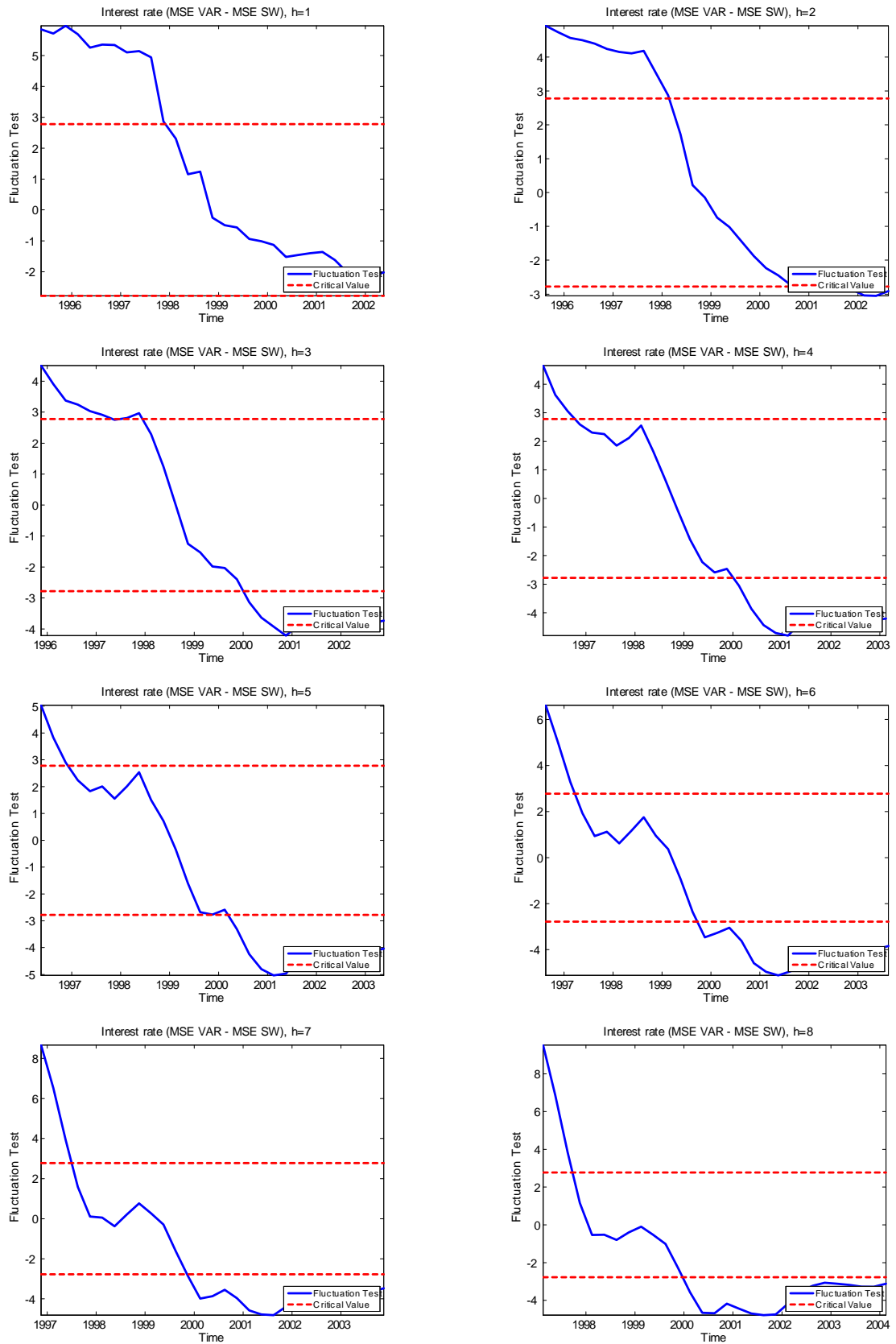
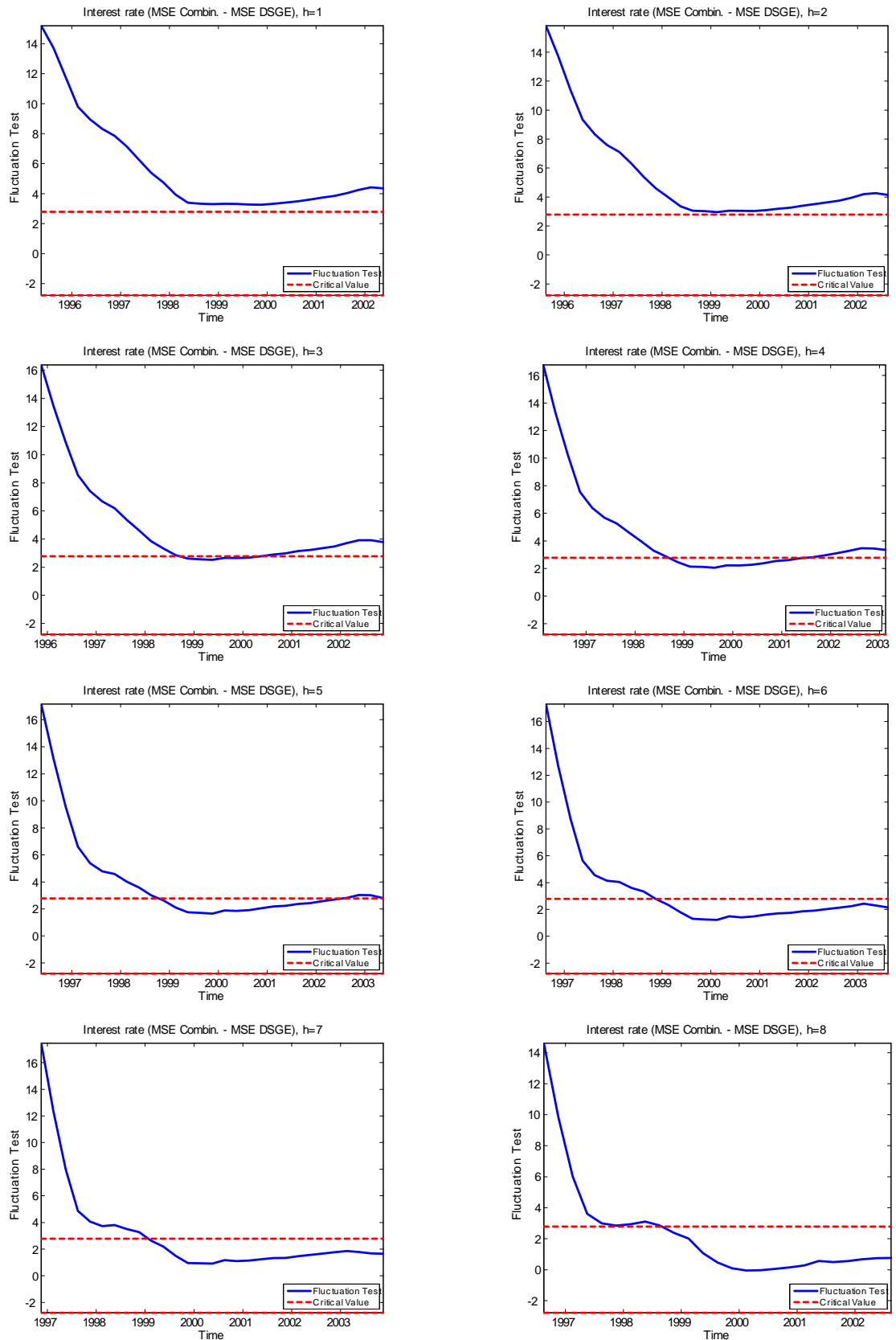


Figure 6(b). Fluctuation Test: Interest Rate

(BVAR vs. DSGE)



Notes to Figure 1. The figure reports the realized value of the target variable (labeled in the title – e.g. output growth in Figure 1(a)), labeled "Actual Data", together with forecasts of the VAR (labeled "VAR"), AR (labeled "AR"), Smets and Wouters (2007, labeled "DSGE"), Bayesian VAR (labeled "BVAR"), forecast combinations (labeled "Combin. forecast") and the random walk (labeled "RW") models. Each graph reports results for different horizons, reported in each graph's title.

Notes to Figure 2. The figure reports the p-value of the Mincer and Zarnowitz (1969) test in eq. (5) over time. The figure reports results for forecasts at various horizons (see the legend) and several models (see the title). The target variable is output growth in panel (a) and inflation in panel (b).

Notes to Figure 3. The figure reports Giacomini and Rossi's (2010) Fluctuation test. The figure reports results for forecasts at various horizons and several models (see the title). The target variable is output growth in panel (a) and inflation in panel (b).

Notes to Figure 4. The figure reports the realized value of the target variable (interest rates), labeled "Actual Data", together with forecasts of the VAR (labeled "VAR"), AR (labeled "AR"), Smets and Wouters (2007, labeled "DSGE"), Bayesian VAR (labeled "UCSV"), forecast combinations (labeled "Combin. forecast") and the random walk (labeled "RW") models. Each graph reports results for different horizons, reported in each graph's title.

Notes to Figure 5. The figure reports the p-value of the Mincer and Zarnowitz (1969) test in eq. (5) over time. The figure report results for forecasts at various horizons (see the legend) and several models (see the title). The target variable is the interest rate.

Notes to Figure 6. The figure reports Giacomini and Rossi's (2010) Fluctuation test. The figure reports results for forecasts at various horizons and several models (see the title). The target variable is the interest rate.

# Appendix A.

This appendix investigates how robust are the results in the main paper when models estimated using the AIC rather than the BIC. The maximum lag length used in the information criterion procedure is the same for AIC and BIC and is 3 lags for the AR model and forecast combinations from AR models and 2 lags for the VAR model; the minimum lag length is one in every model.<sup>17</sup> We report several tables corresponding to those in the main text. To save space, we do not report the figures corresponding to those in the text, as they are very similar to those reported in the paper.

**Table A1. Forecast Optimality Tests: Forecasting Output Growth and Inflation**

<b>Panel A: VAR Model</b>										
$h$	Output Growth Forecasts					Inflation Forecasts				
	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.00	0.69	(0.00)	0.14	(0.00)	0.00	0.34	(0.00)	0.28	(0.00)
2	0.00	0.91	(0.00)	-0.07	(0.00)	0.00	0.42	(0.00)	0.13	(0.00)
3	0.00	0.97	(0.00)	-0.12	(0.00)	0.00	0.35	(0.00)	0.26	(0.00)
4	0.00	0.85	(0.00)	-0.02	(0.00)	0.00	0.41	(0.00)	0.13	(0.00)
5	0.00	0.53	(0.00)	0.27	(0.00)	0.00	0.49	(0.00)	-0.01	(0.00)
6	0.00	0.75	(0.00)	0.08	(0.00)	0.00	0.54	(0.00)	-0.09	(0.00)
7	0.00	0.94	(0.00)	-0.11	(0.00)	0.00	0.51	(0.00)	-0.04	(0.00)
8	0.00	1.24	(0.00)	-0.39	(0.00)	0.00	0.50	(0.00)	-0.02	(0.00)

<b>Panel B: AR Model</b>										
$h$	Output Growth Forecasts					Inflation Forecasts				
	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.00	0.53	(0.05)	0.33	(0.01)	0.03	0.27	(0.06)	0.42	(0.04)
2	0.00	1.06	(0.00)	-0.25	(0.00)	0.01	0.38	(0.01)	0.20	(0.01)
3	0.00	1.36	(0.00)	-0.59	(0.00)	0.00	0.35	(0.00)	0.27	(0.00)
4	0.00	1.22	(0.00)	-0.44	(0.00)	0.00	0.38	(0.00)	0.21	(0.00)
5	0.08	0.62	(0.12)	0.23	(0.09)	0.00	0.57	(0.00)	-0.17	(0.00)
6	0.00	0.91	(0.00)	-0.10	(0.00)	0.00	0.65	(0.00)	-0.30	(0.00)
7	0.00	0.90	(0.00)	-0.08	(0.00)	0.00	0.66	(0.00)	-0.29	(0.00)
8	0.00	0.98	(0.00)	-0.16	(0.00)	0.00	0.62	(0.00)	-0.22	(0.00)

<sup>17</sup>Since the VAR model contains, by construction, more parameters than the AR model, we selected a more parsimonious maximum number of lags.

**Panel C: Forecast Combinations Model**

$h$	Output Growth Forecasts					Inflation Forecasts				
	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.00	0.53	(0.05)	0.33	(0.00)	0.00	0.26	(0.04)	0.44	(0.01)
2	0.00	0.98	(0.00)	-0.16	(0.00)	0.00	0.41	(0.01)	0.14	(0.00)
3	0.00	1.18	(0.00)	-0.37	(0.00)	0.00	0.35	(0.00)	0.26	(0.00)
4	0.00	1.06	(0.00)	-0.26	(0.00)	0.00	0.35	(0.00)	0.26	(0.00)
5	0.03	0.51	(0.12)	0.35	(0.05)	0.00	0.50	(0.00)	-0.03	(0.00)
6	0.00	0.85	(0.00)	-0.03	(0.00)	0.00	0.68	(0.00)	-0.33	(0.00)
7	0.00	0.92	(0.00)	-0.10	(0.00)	0.00	0.67	(0.00)	-0.30	(0.00)
8	0.00	1.06	(0.00)	-0.25	(0.00)	0.00	0.57	(0.00)	-0.15	(0.00)

Note to the table. The table reports results for forecasting output (on the left) and inflation (on the right). Each panel reports, for several horizons ( $h$ , reported in the first column), the values of the Mincer and Zarnowitz (1969) forecast efficiency test (labeled "MZ p-value"), the estimate of  $\hat{\alpha}$  in regression (5) together with the p-value of the test that the constant equals zero (in parentheses), and the estimate of  $\hat{\beta}$  in regression (5) together with the p-value of the test that the slope equals zero (in parentheses).

**Table A2. The Diebold and Mariano Test:  
Forecasting Output Growth and Inflation**

$h$	Output						Inflation				
	RMSFE			DM: Model-DSGE			RMSFE			DM: Model-DSGE	
	DSGE	VAR	AR	VAR	AR	DSGE	VAR	AR	VAR	AR	
1	0.54	1.04	0.90	0.78	0.34	0.19	1.13	1.03	0.25	0.72	
2	0.46	1.35	1.15	0.11	0.10	0.23	1.06	0.92	0.55	0.15	
3	0.46	1.47	1.17	0.20	0.23	0.26	0.93	0.82	0.22	0.00	
4	0.46	1.61	1.21	0.19	0.08	0.26	1.09	0.83	0.41	0.00	
5	0.48	1.33	1.04	0.16	0.69	0.28	1.20	0.87	0.28	0.01	
6	0.48	1.41	1.17	0.21	0.09	0.29	1.38	0.91	0.19	0.16	
7	0.50	1.41	1.18	0.02	0.10	0.30	1.30	0.90	0.27	0.02	
8	0.51	1.40	1.22	0.01	0.13	0.31	1.37	0.89	0.25	0.02	

$h$	Output						Inflation					
	RMSFE			DM: Model-DSGE			RMSFE			DM: Model-DSGE		
	FC	BVAR	RW	FC	BVAR	RW	FC	BVAR	RW	FC	BVAR	RW
1	0.95	1.50	1.06	0.64	0.01	0.51	1.03	2.66	1.06	0.64	0.00	0.44
2	1.11	1.58	1.40	0.25	0.00	0.00	0.95	2.25	1.06	0.39	0.01	0.50
3	1.14	1.41	1.37	0.31	0.03	0.01	0.83	1.95	0.92	0.00	0.02	0.31
4	1.15	1.31	1.43	0.17	0.02	0.02	0.83	1.80	0.83	0.00	0.03	0.07
5	1.00	1.33	1.48	1.00	0.06	0.00	0.87	1.70	0.85	0.01	0.04	0.15
6	1.12	1.09	1.40	0.21	0.51	0.02	0.95	1.69	0.88	0.51	0.09	0.34
7	1.08	1.05	1.37	0.32	0.63	0.01	0.92	1.63	0.92	0.20	0.12	0.53
8	1.13	1.11	1.45	0.20	0.33	0.00	0.91	1.56	0.84	0.14	0.13	0.17

Note to the table. The columns labeled RMSFE in the table report the RMSFE of the DSGE model forecasts (labeled "DSGE") as well as the RMSFE of the autoregressive ("AR"), forecast combinations ("FC"), VAR ("VAR"), Bayesian VAR ("BVAR") and the random walk ("RW")



relative to that of the DSGE model (e.g. the column labeled "VAR" reports the RMSFE of the VAR model divided by the RMSFE of the DSGE model). The columns labeled "DM" in the table report the p-value of the Diebold and Mariano (1996) test (calculated using Giacomini and White's (2006) critical values) for comparing the specified forecasts (e.g. the column labeled "DM: Model-DSGE" and "VAR" reports the p-value for comparing the forecast of the VAR with those of Smets and Wouters' (2007) model).

**Table A3. Forecast Optimality Tests: Forecasting Interest Rates**

<b>Panel A: VAR Model</b>					
$h$	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.57	-0.04	(0.59)	1.04	(0.61)
2	0.95	-0.02	(0.95)	1.02	(0.96)
3	0.69	0.01	(1.00)	0.94	(0.88)
4	0.32	0.02	(0.99)	0.87	(0.78)
5	0.16	0.05	(0.98)	0.81	(0.72)
6	0.04	0.20	(0.84)	0.65	(0.46)
7	0.00	0.39	(0.63)	0.47	(0.23)
8	0.00	0.44	(0.55)	0.41	(0.12)
<b>Panel B: Forecast Combinations Model</b>					
$h$	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.01	-0.03	(0.44)	0.98	(0.75)
2	0.02	-0.05	(0.51)	0.96	(0.76)
3	0.02	-0.06	(0.76)	0.92	(0.70)
4	0.01	-0.05	(0.94)	0.87	(0.65)
5	0.01	-0.03	(0.99)	0.82	(0.66)
6	0.00	0.02	(1.00)	0.74	(0.57)
7	0.00	0.14	(0.93)	0.62	(0.37)
8	0.00	0.16	(0.92)	0.58	(0.31)

Note to the table. The table reports results for forecasting interest rates. Each panel reports, for several horizons ( $h$ , reported in the first column), the values of the Mincer and Zarnowitz (1969) forecast efficiency test (labeled "MZ p-value"), the estimate of  $\hat{\alpha}$  in regression (5) together with the p-value of the test that the constant equals zero (in parentheses), and the estimate of  $\hat{\beta}$  in regression (5) together with the p-value of the test that the slope equals zero (in parentheses).

**Table A4. Diebold and Mariano Test: Forecasting Interest Rates**

$h$	RMSFE			DM: Model-DSGE		RMSFE			DM: Model-DSGE		
	DSGE	VAR	AR	VAR	AR	FC	BVAR	RW	FC	BVAR	RW
1	0.11	1.09	1.00	0.63	0.99	0.95	8.99	0.93	0.64	0.02	0.40
2	0.20	1.05	1.05	0.75	0.75	0.99	4.88	0.95	0.93	0.03	0.54
3	0.28	1.03	1.09	0.84	0.55	1.03	3.41	0.97	0.84	0.04	0.77
4	0.34	1.02	1.14	0.88	0.37	1.08	2.68	0.98	0.62	0.06	0.90
5	0.39	1.02	1.16	0.88	0.29	1.09	2.25	1.01	0.57	0.08	0.93
6	0.43	1.06	1.19	0.70	0.27	1.12	1.96	1.02	0.49	0.10	0.89
7	0.47	1.10	1.23	0.55	0.24	1.17	1.76	1.02	0.38	0.13	0.91
8	0.51	1.12	1.22	0.49	0.22	1.18	1.72	1.01	0.35	0.14	0.96

Note to the table. The columns labeled RMSFE in the table report the RMSFE of the DSGE model forecasts (labeled "DSGE") as well as the RMSFE of the autoregressive ("AR"), forecast combinations ("FC"), VAR ("VAR"), Bayesian VAR ("BVAR") and the random walk ("RW") relative to that of the DSGE model (e.g. the column labeled "VAR" reports the RMSFE of the VAR model divided by the RMSFE of the DSGE model). The columns labeled "DM" in the table report the p-value of the Diebold and Mariano (1996) test (calculated using Giacomini and White's (2006) critical values) for comparing the specified forecasts (e.g. the column labeled "DM: Model-DSGE" and "VAR" reports the p-value for comparing the forecast of the VAR with those of Smets and Wouters' (2007) model).

# Appendix B.

This appendix investigates how robust are the AR and VAR results in the main paper are to obtaining the forecasts by iteration. We report several tables corresponding to those in the main text. To save space, we do not report the figures corresponding to those in the text, as they are very similar to those reported in the paper.

Overall, qualitatively, forecasts from iterated VAR and AR models behave similarly: typically, when one has a MSFE lower than that of the DSGE model, the other does as well. In addition, iterated VAR and AR forecasts, like their direct counterparts, are not efficient.

Specifically, Tables A.5 to A.7 show that, when forecasting output growth, the iterated VAR forecasts have higher MSFE than the direct forecast (except for  $h=8$ ) and the iterated AR forecasts have a higher MSFE than the direct AR forecasts at all horizons. When forecasting inflation, the iterated VAR produces forecasts that have higher MSFEs for short to medium horizons (up to four quarters) and viceversa for medium to long horizons (longer than four quarters); similarly, iterated AR forecasts have higher MSFEs for short to medium horizons (up to two quarters) and viceversa for medium to long horizons (three quarters or longer). When forecasting interest rates, the results depend on the forecast horizon.

**Table A5. Iterated VAR Forecasts. Forecast Optimality Tests.**

Output Growth Forecasts						Inflation Forecasts				
$h$	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.00	0.55	(0.10)	0.24	(0.00)	0.00	0.33	(0.00)	0.30	(0.00)
2	0.00	1.05	(0.00)	-0.18	(0.00)	0.00	0.43	(0.00)	0.10	(0.00)
3	0.00	1.01	(0.00)	-0.14	(0.00)	0.00	0.41	(0.00)	0.15	(0.00)
4	0.00	0.80	(0.00)	0.03	(0.00)	0.00	0.43	(0.00)	0.10	(0.00)
5	0.00	0.31	(0.06)	0.42	(0.00)	0.00	0.50	(0.00)	-0.02	(0.00)
6	0.00	0.62	(0.00)	0.16	(0.00)	0.00	0.59	(0.00)	-0.17	(0.00)
7	0.00	0.74	(0.00)	0.08	(0.00)	0.00	0.60	(0.00)	-0.18	(0.00)
8	0.00	0.83	(0.00)	0.01	(0.00)	0.00	0.56	(0.00)	-0.12	(0.00)

Interest rate Forecasts					
$h$	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.66	-0.03	(0.68)	1.04	(0.69)
2	0.92	-0.03	(0.93)	1.02	(0.97)
3	0.82	0.03	(0.97)	0.93	(0.88)
4	0.53	0.12	(0.80)	0.82	(0.62)
5	0.32	0.22	(0.64)	0.70	(0.42)
6	0.17	0.34	(0.47)	0.57	(0.26)
7	0.08	0.45	(0.35)	0.47	(0.16)
8	0.04	0.54	(0.25)	0.38	(0.09)

**Table A6. Iterated AR Forecasts. Forecast Optimality Tests.**

Output Growth Forecasts						Inflation Forecasts				
$h$	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$		MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.00	0.63	(0.00)	0.22	(0.00)	0.01	0.30	(0.03)	0.36	(0.01)
2	0.00	0.90	(0.00)	-0.08	(0.00)	0.01	0.40	(0.02)	0.15	(0.01)
3	0.00	1.18	(0.00)	-0.40	(0.00)	0.07	0.26	(0.18)	0.43	(0.11)
4	0.00	1.10	(0.00)	-0.31	(0.00)	0.01	0.22	(0.19)	0.48	(0.05)
5	0.33	0.66	(0.34)	0.20	(0.37)	0.00	0.42	(0.00)	0.12	(0.00)
6	0.00	1.11	(0.00)	-0.34	(0.00)	0.00	0.64	(0.00)	-0.27	(0.00)
7	0.00	1.16	(0.00)	-0.39	(0.00)	0.00	0.60	(0.00)	-0.18	(0.00)
8	0.00	1.12	(0.00)	-0.33	(0.00)	0.00	0.42	(0.03)	0.10	(0.00)

Interest rate Forecasts					
$h$	MZ p-value	$\hat{\alpha}$		$\hat{\beta}$	
1	0.01	-0.03	(0.48)	0.97	(0.58)
2	0.01	-0.04	(0.73)	0.94	(0.51)
3	0.02	-0.03	(0.95)	0.89	(0.41)
4	0.01	0.00	(1.00)	0.83	(0.36)
5	0.01	0.05	(0.96)	0.77	(0.31)
6	0.01	0.11	(0.87)	0.70	(0.24)
7	0.00	0.17	(0.79)	0.65	(0.20)
8	0.00	0.21	(0.74)	0.60	(0.16)

**Table A7. Iterated AR and VAR Forecasts.**

**The Diebold and Mariano Test**

h	Output					Inflation				
	RMSFE			DM: Model-DSGE		RMSFE			DM: Model-DSGE	
	DSGE	VAR	AR	VAR	AR	DSGE	VAR	AR	VAR	AR
1	0.54	1.09	0.94	0.45	0.61	0.19	1.12	1.06	0.31	0.46
2	0.46	1.40	1.12	0.02	0.18	0.23	1.07	0.94	0.44	0.40
3	0.46	1.44	1.16	0.04	0.25	0.26	0.97	0.79	0.61	0.00
4	0.46	1.48	1.16	0.02	0.17	0.26	1.01	0.78	0.83	0.00
5	0.48	1.31	1.02	0.01	0.89	0.28	1.07	0.81	0.57	0.00
6	0.48	1.45	1.12	0.07	0.24	0.29	1.15	0.86	0.35	0.09
7	0.50	1.45	1.10	0.12	0.22	0.30	1.13	0.81	0.39	0.01
8	0.51	1.47	1.08	0.17	0.32	0.31	1.11	0.78	0.44	0.00

Interest Rate					
h	RMSFE			DM	
	DSGE	VAR	AR	VAR-DSGE	AR-DSGE
1	0.11	1.15	1.02	0.35	0.87
2	0.20	1.04	1.00	0.73	0.97
3	0.28	1.04	1.03	0.74	0.78
4	0.34	1.03	1.05	0.82	0.66
5	0.39	1.01	1.05	0.91	0.64
6	0.43	1.00	1.04	0.97	0.66
7	0.47	0.98	1.02	0.84	0.78
8	0.51	0.96	1.00	0.60	0.98

# Appendix C.

The following figure reports the Fluctuation test for comparing the AR model with the DSGE model when forecasting interest rates.

**Figure A.1. Fluctuation Test: Interest Rate**

**(AR versus DSGE Model)**

