CHAPTER 3

# Forecasting Output

**Marcelle Chauvet[*] and Simon Potter[†]**

[*]Department of Economics, University of California Riverside, CA 92521, USA
[†]Federal Reserve Bank of New York, 33 Liberty St., New York, NY 10045, USA

## Contents



## Abstract

This chapter surveys the recent literature on output forecasting, and examines the real-time forecasting ability of several models for U.S. output growth. In particular, it evaluates the accuracy of short-term forecasts of linear and nonlinear structural and reduced-form models, and judgmental forecasts of output growth. Our emphasis is on using solely the information that was available at the time the forecast was being made, in order to reproduce the forecasting problem facing forecasters in real-time. We find that there is a large difference in forecast performance across business cycle phases. In particular, it is much harder to forecast output growth during recessions than during expansions. Simple linear and nonlinear autoregressive models have the best accuracy in forecasting output growth during expansions, although the dynamic stochastic general equilibrium model and the vector

*Handbook of Economic Forecasting*, Volume 2A
ISSN 1574-0706, http://dx.doi.org/10.1016/B978-0-444-53683-9.00003-7

autoregressive model with financial variables do relatively well. On the other hand, we find that most models do poorly in forecasting output growth during recessions. The autoregressive model based on the nonlinear dynamic factor model that takes into account asymmetries between expansions and recessions displays the best real time forecast accuracy during recessions. Even though the Blue Chip forecasts are comparable, the dynamic factor Markov switching model has better accuracy, particularly with respect to the timing and depth of output fall during recessions in real time. The results suggest that there are large gains in considering separate forecasting models for normal times and models especially designed for periods of abrupt changes, such as during recessions and financial crises.

Keywords

Real time, Evaluating forecasts, Macroeconomic forecasting, Nonlinear, Recession, DSGE models, Markov switching, Dynamic factor model, Vector autoregressive model

## 1. INTRODUCTION

Forecasting national output is one of the main objectives of private and government forecasters. The forecasts are keys inputs to the decision making of central banks, fiscal authorities, and private sector agents. For example, in assessing fiscal sustainability it is crucial to have good forecasts of the future path of national output. A wide range of approaches are used to produce the forecasts: at one end are judgmental methods that rely on the expertise of the individual forecaster to adjust forecasts produced by a suite of models and at the other end dynamic stochastic general equilibrium (DSGE) models that use modern economic theory to produce a forecast disciplined by economic theory.

In this chapter we provide a survey of a wide range of approaches to forecast output growth with a focus on recent forecast performance and models of interest to Central Bankers and time series econometricians. We start by giving some general background on the forecasting of output, then turn to the specific focus of the chapter. We then examine the forecasts of several models for U.S. output growth in the last 50 years, and compare their accuracy in real time. In particular, we evaluate short-term forecasts of linear and non-linear structural and reduced-form models, and judgmental-based forecasts of U.S. output growth. Our emphasis is on using solely the information that was available at the time the forecast was being made, in order to reproduce the forecasting problem in real time. This exercise is most compelling for policymakers and economic agents, who wish to know the economic situation and its short-run trends as they are occurring. This is especially the case in times of uncertainty around recessions and severe crises. The question we want to answer is whether existing models proved helpful in yielding accurate short-run forecasts of the dynamics of the economy, including during recessions.

### 1.1. Background

The concept of output most forecasted is Gross Domestic Product (GDP) from the national income accounts. GDP is the monetary value of the gross production of all finished goods and services within the borders of an economy. Gross National Product (GNP) is a related concept that measures the production attributable to all members of an economy without respect to their current location. Another related measure of economic activity is Gross Domestic Income (GDI), which is the sum of all income earned while producing goods and services within an economy's borders. As accounting concepts GDI and GDP are equal. However, in practice since GDP estimates are built from expenditure data and GDI estimates are derived from income data, there is usually a discrepancy between the two. These discrepancies can be large for real-time estimates of national output, especially around business cycle turning points (see e.g., Nalewaik, 2012). We will focus on GDP but will examine carefully the issues related to real-time estimates of GDP.

GDP can be split up into expenditure categories using the standard national account identity:

$$Y_t = C_t + I_t + G_t + X_t - M_t,$$

where $Y_t$ is GDP during period $t$, $C_t$ is consumption expenditures, $I_t$ is gross investment (structures, capital equipment and change in inventories), $G_t$ is direct government expenditures, $X_t$ is exports, and $M_t$ is imports.

Most of the effort in forecasting output focuses on real GDP, that is, nominal GDP deflated so that comparisons across time involve real changes in output rather than in the number of goods and services that can be purchased with a unit of currency. There are many ways to deflate nominal GDP to obtain a real measure. The best methods use chain weighting to produce an index of real output. This index can then be quoted in the value of the unit of account for a particular year but in practice the underlying real growth data come directly from the index. Chain-weighting is also applied to the individual components to produce indices of their real levels.[1] Unlike fixed-based year deflation this means that the individual component indices are not additive to the overall real output index and thus, the concept of growth contributions is used.

For most countries estimates of GDP are now produced at a quarterly frequency but some countries still only produce annual estimates. These estimates are designed to give the average flow of GDP during the quarter or the year. Initial estimates for GDP are usually produced in the quarter following the one being estimated but are subject to continuous revision after the initial estimates. Similar revisions occur in the related concepts of GNP and GDI. The origins of the revisions are a mixture of better source data and changes in national income concepts. For example, in 1999 chain-weighting was adopted for the U.S. national income accounts, changing the whole history of the GDP series.

In addition to standard quarter-ahead forecasts, GDP forecasts are often presented as year over year. That is, the average output in year $t + j$ over average output in year

[1] The chain-weighted method of measuring GDP growth entails two calculations of growth for each year, using the year itself and the preceding year as bases. The chain-weighted GDP growth for a year is then the average of these two growth rates. Since this method continually changes the relative weights of goods over time, it corrects potential distortion in growth when there are shifts in production from goods that have similar prices.

$t+j-1$. This convention was based on the greater availability and at times greater accuracy of annual GDP estimates. The convention can produce some unusual effects if data for the current year are revised. We focus on models used to forecast quarterly U.S. GDP where the effect of data revisions is on the predictors rather than the prediction variable.

In many advanced countries the volatility of realized output growth fell in the last three decades. This has been documented by a number of authors, who find a structural break in its volatility in the mid-1980s (e.g., McConnell and Perez-Quiros, 2000; Kim and Nelson, 1999; Koop and Potter, 2000; Blanchard and Simon, 2000; Chauvet and Potter, 2001; van Dijk et al., 2002a; Chauvet and Gopli, 2003; Mills and Wang, 2003, etc.). A feature of the so-called Great Moderation in the United States was two very long business cycle expansions in the 1980s and 1990s. The 1990s expansion was followed by a very mild recession in 2001 when according to current estimates there were not two consecutive quarters of negative growth and the lowest four-quarter growth over the period including the recession was +0.4 percent. This so-called Great Moderation was associated with smaller absolute forecast errors of economic activity but in a relative sense the accuracy of many forecasts fell. As we show below this long period of relative calm made it very difficult for linear time series models to produce forecasts that captured some of the big swings in GDP growth starting in 2008. The "Great Moderation" in the advanced economies was followed by the Great Recession during which the absolute size of forecast errors increased dramatically as discussed in Section 3.3, and in more detail in Chauvet et al. (2013).

Many emerging market economies had suffered severe recessions in the last few decades. A pattern of drastic contractions and fast recoveries was repeated for many of those. However, only recently advanced economies have experienced such deep contractions, with many yet to recover to the previous level of real GDP. While linear time series models are not capable of forecasting large swings in GDP growth, some non-linear model perform well during these periods, as discussed in this chapter.

### 1.2. A Brief History and Survey of Recent Literature

The development of output forecasting was tightly related to the development of national income accounts and the availability of estimates of GDP and GNP. Prior to this the focus had been on forecasts of industrial output. Following the work of Tinbergen (1939, 1974) and Klein (1970) the approach was to estimate linear equations for different sectors of the economy that could be used to forecast aggregates measures such as GDP by use of the national income account identity.

The amount of economic theory used in the estimating equations varied and there was some debate about how to incorporate more formal time series models (Fair, 1970). The performance of these models was initially encouraging but a rigorous time series evaluation by Nelson (1972) showed that the large model for the U.S. developed at Penn, MIT, and the Federal Reserve Board was in many respects worse that the use of simple autoregressive time series models for forecasting.

In addition, the forecasting results from such large models were and continue to be judgmentally adjusted by their users. The judgment is applied to estimation of the current level of GDP (nowcasting) and to the path of GDP going forward. Some of this judgment uses expert knowledge of the construction of national income accounts to refine current quarter estimates while other forms of judgment rely on subjective assessments of the future course of the economy often informed by a suite of time series models that supplement the main model. While the judgmental approach to forecasting is very common, it is impossible to replicate so we rely on surveys of forecasters in real time to include judgmental forecasts in our assessment.

As noted above forecasts of GDP are important inputs to policy decisions. The rational expectations revolution of the 1970s highlighted a crucial conceptual issue with forecasts as inputs to policy decisions: if, as was the assumption, policy could affect outcomes then changes in policy might alter the time series properties of the data invalidating the use of the estimation sample for forecasting (The Lucas critique, Lucas, 1976). The poor forecast performance of the many large macro models in the 1970s gave support to this conceptual issue and led first to the use of vector autoregressions (VAR) and then to estimated DSGE models.

In its seminal paper, Sims (1980) criticizes large-scale macroeconometric models for failing to predict economic activity and inflation in face of the oil shocks in the 1970s. In addition, he argues that the identification of these models was based on restrictions that did not arise from economic theory or institutional facts. Sims proposes as alternative VARs to study economic data and relationships without imposing restrictions. This system of reduced form equations is used to obtain impulse response function of economic variables to shocks. Cooley and LeRoy (1985) criticize VARs arguing that identification of shocks and interpretation of impulse-response functions require structural assumptions. In response, Sims considers the possibility of weak identifying restrictions in order to achieve interpretation of impulse response functions, giving rise to structural VARs (SVARs).

Another response to the poor performance of large macroeconometric models in the 1970s was the construction of structural models based on microeconomic foundations that are not vulnerable to Lucas' Critique. Kydland and Prescott (1982) propose the Real Business Cycle (RBC) model based on principles of neoclassical growth models, in which real shocks are sources of economic fluctuations under flexible prices. The model assumes that agents' optimizing decisions follow rational expectations and are dynamically consistent. Later, Rotemberg and Woodford (1997) propose the New Keynesian DSGE model using a similar framework. Decisions are also based on microfoundations, but prices instead are set by monopolistic competitive firms and are not instantaneously adjusted.

In the last decade there has been substantial progress in the quantitative implementation and estimation of DSGE models. These models offer a coherent framework to characterize the behavior of the economy based on the interactions of microfounded decisions. The seminal work of Smets and Wouters (SW 2003, 2007) showed the feasibility of estimating large and richly specified DSGE models, and found that they can provide a good description of the U.S. macroeconomic data. This led to an increased interest by Central Banks in many countries in their application to policy analysis, particularly due to their story-telling implications.[2]

The next question was whether these models could also be used for forecasting. SW (2007) compare out-of-sample forecasts of the DSGE model with VAR, Bayesian vector autoregressive models (BVAR), and DSGE-VARs. Several authors have extended this approach to verify the forecasting ability of DSGE models for the U.S. and other countries, comparing the results also to judgmental-based forecasts, and to simple benchmarks such as univariate autoregressive processes or random walks.

Some examples are Adolfson et al. (2007a,b), Christoffel et al. (2008, 2011), and Liu et al. (2009). These studies compare out-of-sample results using revised, ex-post data. Some recent studies use real-time data such as Rubaszek and Skrzypczyński (2008), Kolasa et al. (2009), Edge et al. (2010), Wolters (2010), Edge and Gürkaynak (2010), Del Negro and Schorfheide (2013), and Wieland and Wolters (2011).

The general finding from the literature is that the DSGE forecasts are comparable or slightly superior to the ones obtained from VARs and BVAR, but not significantly different from simple benchmarks such as univariate autoregressive processes. Judgmental forecasts outperform DSGE, VAR, and BVAR or DSGE-VAR models in the short run (one or two quarters ahead). DSGE models show a better result in the medium run (three and four quarters ahead), but tests of equal forecast accuracy generally indicate that the differences in forecasts are not significantly different across models at these horizons. Interestingly, these results hold for the U.S., Euro area, and other countries.

Wolters (2010) additionally finds that structural models fail to forecast turning points (i.e., the beginning or end of a recession), large recessions, and booms, but display comparable accuracy to the judgmental and BVAR forecasts during "normal" times for medium-run horizons. This is also found by Del Negro and Schorfheide (2013) and Wieland and Wolters (2011). The former compares the real-time (pseudo out-of-sample) forecast ability of Blue Chip forecasts with the SW (2007) model and extensions that include financial frictions, or information on default risk and current interest rates in the last two decades, while the latter examines the real-time forecast accuracy of structural and reduced-form models with judgmental-based forecasts during the last five NBER-dated recessions. Both papers find that the model forecasts are outperformed by those from professional forecasts at short-run horizons but are comparable to them at medium-run horizons. Wieland and Wolters (2011), however, find that, with the exception of the 1980–1981 recession, the judgmental-based forecasts outperform the model-based ones for all other recessions, with the largest difference in forecasts being for the 2007–2009 recession and the smallest for the 2001 recession.

[2] Sims (2006), Tovar (2009), and Faust (2012) discuss some of the omissions of these models and important features that will enhance their contribution to discussion of policy analysis.

Regarding models used by Central Banks, Christoffel et al. (2011) compare the forecast accuracy of the New Area-Wide Model (NAWM), designed and used by the European Central Bank (ECB) for macroeconomic forecasts, with Bayesian DSGE-VARs, BVARs and reduced-form models.[3] They find that the DSGE-VAR model outperforms DSGE, NAWM, VAR, and BVAR models in forecasting output. Edge et al. (2010) compare the performance of Estimated Dynamic Optimization-based model (EDO) from the Federal Reserve Board (FRB) with VAR, BVAR, univariate autoregressive model, and the Greenbook and FRB forecasts. They find that the out-of-sample real-time forecasts of the EDO are comparable to the autoregressive model but generally not statistically different from it and the other models. However, as noticed by the authors, the models are evaluated for a period of relative stability, between 1996 and 2004.

Edge et al. (2010), Edge and Gürkaynak (2010), and Wang (2009) reach the same conclusions as the literature, but note the surprising evidence that the forecasting models or judgmental forecasts generally examined are very poor. This is exacerbated when long samples are considered. The models and judgmental-based results show modest nowcasting ability, but they display almost no forecasting ability from one-quarter ahead and on. As stressed by the authors, the comparison in the literature has been among poor forecasting methods.

The findings in the literature support the early evidence of Nelson (1972) that forecasts from simple autoregressive models are hard to beat by large-scale macroeconomic model. However, another strand of the literature has shown that the use of some variables in simple autoregressive processes or frontier time series models generally generate substantial gains in forecasting output growth.

Many papers find that some financial and macroeconomic series have significant predictive power for future economic activity across a number of countries.[4] Among those, the early work of Estrella and Hardouvelis (1991) and Stock and Watson (1993) find that the yield curve has the best short- and medium-run forecast power for output growth beyond the predictive power of several other variables including lagged output growth. The early literature is summarized in the comprehensive literature review of Stock and Watson (2003) and the more recent one focusing on the yield curve on the survey by Wheelock and Wohar (2009).

Some of the cutting-edge time series models and methods found to be useful to forecast economic activity are factor models, mixed frequency models, non-linear models,

[3] The literature finds that the BVAR model generally has good forecast accuracy for several variables in the system compared to other models, but not of real GDP growth.

[4] Some of the series that have been found to be good predictors of GDP growth are interest rates and spreads, stock prices, monetary aggregates, inflation, survey forecasts (e.g., NAPMC Purchasing managers' survey), the index of leading indicator and its components, such as vendor performance, contracts and orders for plants and equipment, housing permits, consumer expectations, change in manufacturers' unfilled durable goods orders, etc. Banegas (2011) finds that for emerging economies additional series that help predict output growth are portfolio investment flows, global commodity markets, and a cross-sectional firm size factor.

and forecasts combination. This list is not exhaustive as the literature is vast and dynamic, with innovations being proposed at a rapid pace. Below we discuss some of the recent developments.

A large number of papers have examined forecasts of macroeconomic variables using factor models, which are a parsimonious way of extracting large information on overall economic conditions. Stock and Watson (2002), Marcellino et al. (2003), and Forni et al. (2003) survey applications of these models in forecasting macroeconomic and financial variables. These and more recent studies find that factor models consistently beat univariate and multivariate autoregressive models and, often, judgmental forecasts both during expansions and recessions.

More recently, Wang (2009) compares the out-of-sample forecast performance for U.S. output growth of DSGE models with VARs and factor models. He finds that the factor model generally outperforms any other model in the short run, with significantly different gains. Lombardi and Maier (2011) study pseudo real-time forecasts of GDP for the euro area and its countries during the Great Moderation and Great Recession, comparing the performance of dynamic factor models and a model using survey indices (Purchasing Managers' Indices). Winter (2011) compares the short-term forecasting ability of factor models with several linear reduced-form models and private sector forecasts for recessions in general, and particularly the Great Recession in the Netherlands. Both papers conclude that the dynamic factor model displays the best forecast accuracy overall and during the recent crisis, and the difference in forecasts with the other models and judgmental forecasts is statistically significant.

Some advances in forecast methods include a recent growing literature on nowcasting and short-term forecasting GDP growth using mixed frequency models.[5] The idea is to explore information in more timely indicators that are available at a higher frequency to improve the forecast of quarterly output growth. Several papers use factor models cast in state space with mixed frequency to nowcast GDP growth in the U.S. or Euro area such as Marcellino and Schumacher (2008), Camacho et al. (2011,2012), Banbura and Runstler (2011), Giannone et al. (2004), Angelini et al. (2011), Giannone and Reichlin (2013), which are related to the methods of Trehan (1989), Mariano and Murasawa (2003), Evans (2005), Proietti and Moauro (2006), and Giannone et al. (2008). The general finding is that these models generally outperform nowcasts of GDP growth compared to models that use quarterly frequency only, and are comparable to judgmental forecasts in the United States and in the Euro area.[6]

Other authors apply the mixed frequency approach to univariate and multivariate autoregressive (VAR) processes. Clements and Galvao (2008) use the mixed data sampling

[5] Most macroeconomic data are released with at least one month lag, and many with longer delays. In addition, many series are revised substantially since their first release. This leads to a need to forecast the present and even the near past, which the literature has dubbed "nowcast."

[6] For a survey see Giannone and Reichlin (2013).

MIDAS method proposed by Ghysels et al. (2004, 2006) in autoregressive processes (AR). They find that MIDAS improves real-time forecasts of U.S. output growth compared to standard AR models at nowcasting and short horizons. Banbura et al. (2010), Kuzin et al. (2011) apply mixed frequency method to VAR and Schorfheide and Song (2012) to BVARs. They find that adding within-quarter monthly information improve VAR and BVAR forecasts. Kuzin et al. (2011) find additionally that the nowcasting and forecasting ability of the MIDAS and mixed-frequency VAR (MF-VAR) to quarterly GDP growth in the euro area are complements as the MIDAS does better for short horizons (up to five months), whereas MF-VAR are better for longer horizons (up to nine months).[7]

Aruoba et al. (2009) and Aruoba and Diebold (2010) propose a factor model with mixed frequency that include high-frequency data to measure economic activity.[8] These are the first frameworks that include frequencies higher than monthly. The approach is based on a linear small-data dynamic factor model to construct a high-frequency coincident indicator of business conditions, building on Stock and Watson (1989) and Mariano and Murasawa (2003).[9]

A popular forecasting approach is pooling of forecasts from several models. The common finding is that forecast combinations generate better results on average than forecasts from single models. Arguments in favor of pooling are that specification and selection of single forecast and nowcast models involve decisions on variable selection, model specification, estimation method, which all could lead to potential misspecification in theory. Timmermann (2006) studies the theoretical and empirical reasons behind the possible determinants of the advantages from combining forecasts, such as the correlation between forecast errors and the relative size of the forecast error variances of single models, model misspecification, and non-stationarities (see also Clements and Hendry, 2004). Recent empirical evidence favors forecast combination such as Clark and McCracken (2010), Assenmacher-Wesche and Pesaran (2008), Kuzin et al. (2013), or Aiolfi et al. (2011), amongst many others. Kuzin et al. (2013) find that forecast combination yields more robust forecasts than factor models for nowcasting and short-term forecasting output growth in several industrialized countries. Aiolfi et al. (2011) study forecast combinations of model-based forecasts from linear and non-linear univariate specifications, and multivariate factor-augmented models with judgmental survey forecasts and find that the leading forecasts are obtained from combining a simple equal-weighted average of survey and model-based forecasts.

Some recent papers show that a different type of aggregation can also be promising. In particular, aggregating forecasts of components, regions, or countries can lead to improved

[7] MIDAS are based on exponential lag polynomials coefficients, while the MF-VAR has unrestricted parameters.

[8] Real-time updates of the indicator is posted in the Philadelphia Fed website at http://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/

[9] Stock and Watson (1989) propose a widely popular low-dimensional linear dynamic factor model to construct coincident indicators of the U.S. economy.

performance compared to directly forecasting aggregated data. Frale et al. (2011) forecast output by aggregating forecasts of its components from the expenditure side, and Marcellino et al. (2003) produce nowcasts from aggregating member countries of the euro area. These papers find that forecasting aggregated components outperforms direct forecasts of aggregate output. Owyang et al. (2012) find that including state-level series to aggregate predictors at the national level improves the short-run forecast performance of the U.S. business cycle phase.

A large recent literature has shown that non-linearities in the dynamics of the economy can be quite important for forecasting. Non-linear models may reveal additional information and improve forecasts compared to frameworks that take into account only the average linear effect of one series on another. Many studies have shown that the largest forecasting errors in some series occur around the beginning or end of a recession, because it is at these times that the linear relationship may break down. This is especially the case for recessions, when most models yield large forecasts errors.

Recent methods have been advanced to provide a formal representation of non-linearities in economic series in a rigorous framework.[10] For example, the prewar emphasis on business cycles based on the idea of recurrent expansion and contraction phases has been formalized in the threshold models of Tong (1990) and Granger and Teräsvirta (1993), and in Hamilton's (1989) widely applied Markov switching model. There has also been lots of progress in modeling sophisticated versions of Probit models to forecast business cycle phases. These non-linear models are powerful tools for modeling recurrent phase changes as they capture potential asymmetries across phases, allowing expansions and contractions to display different amplitude, duration, or steepness.[11]

Diebold and Rudebusch (1996) propose a combination of the small scale dynamic factor model of Stock and Watson (1989) with Markov switching as in Hamilton (1989). Chauvet (1998) estimates such a multivariate dynamic factor model with Markov switching model to build a coincident indicator of the U.S. business cycle and probabilities that the economy is in a specific phase of the business cycle, and evaluates turning points of the 1990–1991 recession in real time.[12] Adding Markov switching to this framework allows analysis of asymmetries and recurrent breaks in a multivariate setting. The proposed model successfully signals recessions, which many models failed to predict in real time. Kim and Nelson (1998) estimate this model with Bayesian techniques and extending it to allow for time-varying probabilities.

[10] For a recent collection of papers on advances of nonlinear models see Ma and Wohar (forthcoming).

[11] In this paper we focus on forecasting output growth rather than on forecasting business cycle turning points. There is a vast literature in this area. Just to name a few see Stock and Watson (1993), Bernard and Gerlach (1998), Estrella and Mishkin (1998), Chauvet (1998), Anderson and Vahid (2001), Chauvet and Hamilton (2006), Chauvet and Piger (2008, 2013), Chauvet and Potter (2001, 2005), Chauvet et al. (2013), Chauvet and Senyuz (2012), Kauppi and Saikkonen (2008), and Nyberg (2010), etc. For a survey of threshold models see van Dijk et al. (2002b). A collection of recent papers describing techniques for building indicators of turning points can be found in Mazzi and Ozyildirim (forthcoming).

[12] The model uses the same coincident series as considered by the NBER: sales, personal income, industrial production, and employment.

Chauvet and Hamilton (2006) and Chauvet and Piger (2008) collect a large database of unrevised real-time data to reconstruct inferences that would have been generated if parameters had to be estimated based on data as they were originally released at each historical date. Chauvet and Hamilton (2006) examine the performance of the univariate Markov switching model of Hamilton (1989) and the dynamic factor model with Markov switching in Chauvet (1998) in forecasting U.S. business cycles in real time, while Chauvet and Piger (2008) compare the performance of a non-parametric algorithm and the parametric Markov-switching dynamic-factor model. These papers find that the recession probabilities from Markov-switching models perform quite well in estimation with real-time databases and are a successful tool in monitoring the U.S. business cycle.

More recently, Chauvet and Piger (2013) examine the real-time performance of the dynamic factor model with Markov switching in forecasting recessions, particularly the last one, using coincident series and different measures of employment. The recent recessions have been followed by jobless recoveries that led to great uncertainty regarding the recession end in real time. This paper evaluate the speed in identifying the beginning and end of recessions in real time when payroll employment or civilian employment are considered. They find that the model timely signaled the onset of recessions, including the Great Recession.[13] Altogether, these papers show that the dynamic factor model with regime switching is one of the most successful models in predicting business cycle phases in real time.

Recent innovations in Markov switching models include Kim et al. (2008), Chauvet (2010), Guerin and Marcellino (2011), Nalewaik (2011), Camacho et al. (2012), Chauvet and Su (2013), among others. Kim et al. (2008) extend Hamilton's model to allow for endogenous Markov regime switching and apply the framework to a volatility feedback model of stock returns. Chauvet (2010) proposes a 4-state Markov switching model to represent four phases of the US business cycle: recessions, recoveries, expansions, and slowdowns. Guerin and Marcellino (2011) propose a univariate Markov switching mixed data sampling (MS-MIDAS) extending Hamilton's (1989) model to allow the use of mixed-frequency data in Markov switching models. They find that the MS-MIDAS improve forecasts performance for U.S. output growth compared to the univariate MS model. Nalewaik (2011) uses a three state Markov switching model to capture transitions

[13] These authors find that the version of the model with payroll is quicker to call peaks, while the one with civilian employment is best for troughs. Notice that this is a model designed for identifying recessions based on coincident series, not a model for anticipating recessions based on leading series. For example, the model estimated with *coincident* series and payroll representing employment timely signaled in real time the onset of the Great Recession as December 2007 with information available in April 2008 (the earliest possible signal, given the lag in the availability of the data, would have been in March 2008). The real-time probabilities of recession were above 50% already in April 2008, and above 80% in July 2008. The probabilities stayed close to 100% during the whole financial crisis and the most of 2009, correctly signaling the intensity and duration of the recession. The real-time probabilities of recession are made publicly available on a monthly basis on Chauvet's website since 2006–2007 at: http://sites.google.com/site/marcellechauvet/probabilities-of-recession and on Piger's website at: http://pages.uoregon.edu/jpiger/us_recession_probs.htm.

in economic activity from expansion to a short stall phase, and then to recession. The model includes additional leading series as the yield curve, GDI, unemployment, and housing starts, and generates improved forecasts of output growth. Camacho et al. (2012) extend Chauvet (1998) Dynamic Factor Model with Markov Switching (DFMS) model to include ragged edges and mixed frequencies. The real-time analysis is not applied to forecast output growth, but phases of the business cycle.

Chauvet and Su (2013) propose a model with three Markov switching processes in order to simultaneously capture business cycle phases, structural breaks or outliers. Market economies undergo recurrent fluctuations and changes in the structure of aggregate activity. Models that do not take into account the evolving dynamics of the economy yield poor representation and forecast of economic activity. This has been specially the case with the Great Moderation in the mid 1980s and the recent Great Recession. Chauvet and Su's (2013) model successfully represents business cycle phases under structural and pulse breaks.

Finally, other recent models that have been used in a debate regarding the evolving dynamics of the economy related to the Great Moderation is the Time-Varying VAR (TVAR) model and the Markov switching VAR model (MS-VAR).[14] Cogley and Sargent (2001, 2005) and Primiceri (2005) use a reduced form TVAR that takes into account drifting parameters or heteroskedasticity while Sims and Zha (2006) study changes in monetary policy via MS-VAR models with discrete breaks that capture potential switching policy pre and post Volcker. The findings in these papers contrast regarding the nature of changes – whether they were abrupt as in depicted in MS-VAR models or more gradual as in TVAR models. Chauvet and Tierney (2009) use a non-parametric VAR model and find that there have been abrupt as well as gradual changes in shocks and in the dynamics of the U.S. economy in the last five decades. These models have not been widely used to predict output growth.[15]

On the other spectrum, there are judgmental methods that rely on the expertise of individual forecasters to fine-tune forecasts generated by a set of models. Most countries and some regions have survey of forecasters. For example, the ECB publishes the ECB Survey of Professional Forecasters, which is a quarterly survey of expectations of several key macroeconomic indicators, including real GDP growth. The Survey has been conducted since 1999 and it is based on forecasts from financial and non-financial institutions from the European Union.

In the U.S., the most popular ones are the Survey of Professional Forecasters (SPF) and the Blue Chip Indicators (BC). The Survey of Professional Forecasters is published

[14] Time-varying models have not been widely used to predict output growth. A possible reason suggested by Cogley in discussions of this paper is that since drifts in the parameters are gradual, TVAR models may not perform well around business cycle turning points, particularly if the recent Great Recession period is included.

[15] A possible reason in the case of TVAR models suggested by Cogley in discussions of this paper is that since drifts in the parameters are assumed to be gradual, TVAR models may not perform well around business cycle turning points, particularly if the recent Great Recession period is included.

by the Federal Reserve Bank of Philadelphia since 1990, taking over from the American Statistical Association and the National Bureau of Economic Research, which were publishing it since 1968. The SPF makes available the mean and median forecasts as well as the individual responses from each forecaster.

The Blue Chip Indicators is a poll of around top 50 forecast economists from banks, manufacturing industries, brokerage firms, and insurance companies. The poll has been conducted since 1976 and comprises several macro series, including real GDP growth. The survey contains forecasts from each member and the average (or consensus) of their forecasts. It also provides the average of the 10 highest and 10 lowest forecasts for each variable, and the median forecast. Finally, it also publishes a diffusion index that reflects changes in expectations that might take place before changes in the consensus forecast.

Another popular set of forecasts in the U.S. is the Greenbook of the Federal Reserve Board of Governors. The Greenbook is prepared as discussion material for each of the Federal Open Market Committee meetings. The Greenbook forecasts are based on assumptions about monetary policy and are put together by research economists from the Board of Governors. The forecasts are only publicly available after a lapse of 5 years.

### 1.3. Chapter Plan

In this chapter we focus on evaluating the real-time performance of several models in forecasting output growth over time as well as during expansions and recessions in a genuine out-of-sample exercise. As discussed in the previous section, there are a plethora of models that could be selected and, clearly choices have to be made in a literature this vast. In order to keep the task manageable, we focus on some popular structural, linear and non-linear multivariate models and apply them to U.S. output data.

Given that the literature has extensively compared the performance of DSGE models with VAR and BVAR models and found that these models have, on average, yielded poor forecasts, we focus also on comparing the forecast accuracy of structural models and state of the art reduced-form time series models, which could be more informative benchmarks.

We compare the forecast accuracy of the DSGE model of Smets and Wouters (2007) with linear and non-linear autoregressive models such as AR(2), Cumulative Depth of Recession by Beaudry and Koop (CDR 1993), VARs, BVARs, the univariate Markov switching model (MS), and the proposed autoregressive model associated with the Dynamic Factor Model with Markov Switching (AR-DFMS). Given the importance of the financial sector in explaining the recent crisis, we also study VARs with financial variables.[16]

The model-based forecasts are contrasted with the judgmental forecasts from the Blue Chip Indicators. The literature has found that the forecasts of U.S. output growth from

[16] For results of DSGE models with financial frictions, see the chapter by Del Negro and Schorfheide (2013) in this volume.

the Survey of Professional Forecasters and Blue Chip are similar (see e.g., Wieland and Wolters, 2011). On the other hand, the Greenbook forecasts are only available with a lag of 5 years. The latest forecasts available from the Philadelphia Fed website at the time this chapter was written were up to 2006.

We thus focus on comparisons with the Blue Chip Indicators representing judgmental forecasts, which allow analysis of forecasts using the same sample as the estimated models.[17] We also evaluate the forecast accuracy arising from equal-weight forecast combination as in Aiolfi et al. (2011).

As discussed in Edge et al. (2010), Edge and Gürkaynak (2010), and Wang (2009) the forecasts beyond two quarters from models or from professional forecasters generally examined in the literature are very poor. The models and judgmental forecasts have some nowcasting accuracy, but almost no forecasting ability from one-quarter ahead and on as they are beaten even by naïve constant growth models. In fact, these poor medium run forecast results is one reason why the nowcasting literature with mixed frequency has flourished (Giannone et al., 2008). In this chapter, we choose to focus on the comparison of informative short-run forecasts rather than discussing which models are best at scoring higher in uninformative medium or long-run results for output growth.

We find that recessions are generally harder to forecast than expansions for all models examined. The univariate AR(2) model, the MS model, and the forecast combination have the best forecast ability during expansions, significantly better than the DSGE model, and the VAR model with the term spread, but comparable to the BVAR (at two-quarter ahead horizon). The findings suggest that by using simple univariate linear autoregressive models of GDP growth, one would have gotten in real time as good as forecasts during expansions than any other model and the professional forecasters.

The Blue Chip forecasts of output growth are outperformed by the AR(2) and the MS models and the forecast combination during expansions and by the DFMS model during recessions. Although the Blue Chip forecasts are very similar to the ones from the DFMS model for the full sample, they are worse and significantly different during recession periods. Interestingly, the BC forecasts do not track well the variance of GDP growth, particularly at one-quarter horizon.

The autoregressive model associated with the non-linear dynamic factor model (AR-DFMS) displays the best real-time forecast accuracy for recessions. Even though the professional forecasters have information advantage over all models, the AR-DFMS model shows short-run improvements, particularly with respect to the timing and depth of recessions in real time. The reason for its successful performance is that this model uses not only information from GDP growth but also from monthly coincident series that signal a deterioration of the economy early on. The forecast ability of the model is also closely related to the dynamics of its real-time probabilities of recession, which increase

[17] The Greenbook forecasts are only available with a lag of 5 years. The latest forecasts available from the Philadelphia Fed website at the time this chapter was written were up to 2006.

around the beginning of recessions and remain high until around their trough. The accuracy of GDP growth forecasts from the AR-DFMS model is, thus, closely related to the ability of the model to forecast recessions in real time.

Combining all forecasts from the models and from the Blue Chip indicators using equal weight average results in slight better accuracy compared to the simple AR(2), but the differences in forecasts are not statistically significant. The forecast combination is outperformed by the AR-DFMS model and the BC forecasts for the full sample and during recessions at one and two-quarter horizons. The forecast combination is also outperformed by the CDR and MS models during recessions at the two-quarter horizon.

In summary, the results suggest that (a) it is hard to beat the univariate AR(2) model for short-term forecasts of U.S. real GDP growth, particularly for expansions. Its performance is followed by the MS model and the professional forecasts; (b) the DSGE models, VAR models, and BVAR models are commonly used but are outperformed by forecasts from simple autoregressive model during normal times; (c) it is harder to predict output growth during recessions than during expansions. Most models perform poorly in forecasting output growth during recessions. In particular, they miss the timing of output downfall and its intensity during recessions. The DFMS model yields the best forecasts as it allows for regime shifts and includes timely monthly information from several variables other than GDP; (d) pooling forecasts does not significantly improves precision compared to the autoregressive model during expansions. It is outperformed by several models during recessions, as it does not track well the volatility of GDP growth (particularly for two-quarters ahead).

Thus, we find that there are large gains in using distinct models to forecast output growth during normal times and models to forecast output growth during recessions. Although the DSGE and VAR models might be a useful story-telling tool to evaluate policy during normal times, there are substantial gains in forecasting output growth around recessions using non-linear models designed for periods of abrupt changes. By using and comparing forecasts from different models, especially those designed to handle regime changes and non-linearities, economic agents and Central Bankers can hedge against abrupt changes and obtain more informative forecasts at times of large uncertainty such as around business cycle turning points, when most linear models break down.

The chapter is organized as follows. Section 2 describes the forecasting models, and the Blue Chip Indicators. Section 3 describes the real-time data and studies the ability of the models and the professional forecasters in forecasting the economy in real time. Section 4 concludes.

## 2. FORECASTING MODELS

We examine the forecasts of GDP growth from the Blue Chip indicators, and seven linear and non-linear, structural and reduced form models. We consider a univariate linear autoregressive model (AR) as a benchmark, and two univariate non-linear models: the Cumulative Depth of Recession model (CDR) from Beaudry and Koop (1993), and the Markov Switching model (MS). We also investigate the performance of four multivariate models: the structural DSGE model, several versions of the reduced-form linear VAR, BVAR, and the non-linear Dynamic Factor Model with Regime Switching model.

## 2.1. Benchmark Univariate Linear AR Model

Let $y_t$ be the log of real GDP and $\Delta = 1 - L$, where L is the lag operator. The model is:

$$\Delta y_t = c + \phi_1 \Delta y_{t-1} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2). \tag{1}$$

Using Akaike and Schwarz information criteria, one-quarter ahead Theil inequality coefficients, and root mean squared forecast errors (RMSFE), we find that the best specification (order of $p$) for GDP growth is an AR(2) process. In addition, we estimate the model recursively with different lags using real-time data, and find that $p = 2$ is uniformly better in terms of AIC for most part of the sample. We use forecasts obtained from the AR(2) model as a benchmark to compare with the other models. This is the same benchmark used in Edge et al. (2010), Krane (2011), Wolters (2010), and Del Negro and Schorfheide (2013), and many others. The simple linear univariate AR(2) offers an interesting comparison with the more complicated models as it has been shown in several papers to have a comparable or better forecasting performance.

## 2.2. Current Depth of Recession

Beaudry and Koop (BK 1993) extend the autoregressive representation of output growth to allow for asymmetric persistence. The asymmetry is examined by allowing the depth of a current recession to have an impact in the path of future fluctuations. In BK's (1993) current depth of recession model (CDR) output growth is defined as the gap between the current level of output and its historical maximum level at horizon $j$. We extend our benchmark AR(2) model in Eq. (1) as:

$$\begin{aligned} \Delta y_t &= c + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \theta CDR_{t-1} + \varepsilon_{cdr,t} \\ CDR_t &= \max(0, \{y_{t-j} - y_t\}_{j \geq 0}). \end{aligned} \tag{2}$$

The lag $p = 2$ is also the selected specification in BK (1993), based on Akaike and Schwarz information criteria. The model implies that if current output growth is below the level value of the previous peak, the difference is positive and hence the economy is in recession. Otherwise the economy is in an expansion and the value of $CDR_t$ is zero.

## 2.3. Judgmental Forecast: Blue Chip Economic Indicators

The Blue Chip Economic Indicator (BC) is a compilation of macroeconomic forecasts of the U.S. economy from about 50 major investment and commercial banks, financial and industrial firms, universities, and economic consulting firms. The quarterly forecasts are available on a monthly basis. The BC forecasts for GDP growth is the average of the panelists' projections and is released on the tenth of each month for responses based on information for the previous month. Before the official GDP growth observation for each quarter is released by the Bureau of Economic Analysis, the BC produces three forecasts for the quarter. For example, GDP growth in the first quarter of 2011 is forecast in the February 2011 survey based on information available as of the end of January; in the March 2011 survey based on information in the end of February; and in the April 2011 survey with information up to the end of March 2011.

## 2.4. Dynamic Stochastic General Equilibrium Model

In this chapter we consider the medium-scale DSGE model of Smets and Wouters' (SW 2007) to form forecasts of GDP growth. The description of the model in this section follows closely (SW 2007), Edge et al. (2010), and Del Negro and Schorfheide (2013). SW's (2007) framework is based on Christiano et al. (2005) model, and it consists of a real business cycle model with nominal and real rigidities. In addition to sticky prices and wages, it contains real rigidities in the form of habit formation in consumption, adjustment cost in investment in capital accumulation, and variable capacity utilization.

The model comprises households, firms, and a monetary authority. Households maximize a non-separable utility function with goods and labor effort over an infinite life horizon. Consumption is related to time-varying external habit. Labor is heterogeneous across households in the sense that there is a union that allows for some monopoly power over wages. This enables introduction of Calvo rigidities in wages. Households own capital and rent its services to firms. Their investment decisions are affected by capital adjustment costs: as rental price increases, capital utilization can be more intensive but at a variable cost.

There is monopolistic competition in the markets for intermediate goods. The firms rent labor via a union and capital from households to produce differentiated goods, setting their prices according to the Calvo model. These intermediate goods are aggregated into a final good by different firms in a perfectly competitive final-good sector. In addition to Calvo setting in prices and wages, prices that are not re-optimized are assumed to be partially indexed to past inflation. Thus, prices depend on current and expected marginal costs as well as past inflation. Marginal costs depend on the price of factor inputs. Similarly, wages are a function of current and expected marginal rates of substitution between leisure and consumption and past wage inflation.

Following Del Negro and Schorfheide (2013), we assume that the series used in the model contain a stochastic trend rather than a determinist trend as in SW. Thus, all series are detrended by $Z_t = e^{\gamma t + \frac{1}{1-\alpha}\tilde{z}_t}$, where $\gamma$ is the steady state growth rate of the economy, $\alpha$ is the income share of capital net of markups and fixed costs, and

$\tilde{z}_t = \rho_z \tilde{z}_{t-1} + \sigma_z \varepsilon_{z,t}$. Hence, the growth rate of $Z_t$ in deviation from $\gamma$ is:

$$z_t = \ln\left(\frac{Z_t}{Z_{t-1}}\right) - \gamma = \frac{1}{1-\alpha}(\rho_z - 1)\tilde{z}_{t-1} + \frac{1}{1-\alpha}\sigma_z \varepsilon_{z,t}. \quad (3)$$

The detrended variables are expressed in log deviations from their non-stochastic steady state. Most of the resulting log-linearized equilibrium conditions are the same as in SW such as the Euler equation, the optimality condition for capital producers, the arbitrage condition between the return to capital and the riskless rate, and the optimality condition determining the rate of capital utilization. The only two equilibrium conditions that change under the assumption that technology has a unit root rather than a stationary trend are the equilibrium production function and the equilibrium resource constraint. These equations are reduced from the terms involving $\frac{1}{1-\alpha}\tilde{z}_t$.[18]

The model has seven observable variables. The observable variables are quarterly growth rate of real output, consumption, investment and real wage, and quarterly log hours worked, inflation, and nominal interest rates. The model is cast in state space form mapping these observable variables into the 14 endogenous variables. The stochastic behavior of the system of linear rational expectations equations is driven by seven exogenous disturbances: total factor productivity, investment-specific technology, risk premium, exogenous spending, price mark-up, wage mark-up, and monetary policy shocks.

The model is estimated using Bayesian methods, with the same priors as SW (2007). The priors are combined with the conditional density of the observables to obtain the posterior distribution. The moments and quantiles of the posterior distribution are obtained via Markov Chain Monte Carlo (MCMC) simulation, using the Random-Walk Metropolis algorithm. The sequences of draws from the posterior distribution can be used to obtain numerical approximations of the moments, and predictive density distribution. The model is estimated for a given data vintage, and the forecasts are obtained from the predictive distribution and posterior modes of each parameter.[19]

## 2.5. Vector Autoregressive Model

Let $\Delta \boldsymbol{Y}_t$ be a $nx1$ vector containing the values that $n$ variables take at date $t$. The reduced form VAR is:

$$\Delta \boldsymbol{Y}_t = \boldsymbol{a} + \boldsymbol{A}_1 \Delta \boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{A}_p \Delta \boldsymbol{Y}_{t-p} + \boldsymbol{u}_t \quad \boldsymbol{u}_t \sim (0, \Theta). \quad (4)$$

The assumption that $\Delta \boldsymbol{Y}_t$ follows a vector autoregression corresponds to the idea that $p$ lags are sufficient to summarize all of the dynamic correlations among elements of $\Delta \boldsymbol{Y}_t$. Notice that the parameters of the reduced-form VAR include contemporaneous relations among the endogenous variables. To see this, let $\boldsymbol{x}_t$ be an $[(np+1) \times 1]$ vector containing

[18] For details on the equilibrium conditions and their derivation see Del Negro and Schorfheide (2013), and for a full version of the log-linearized version of the estimated model see SW (2003, 2007).

[19] For a detailed explanation see Del Negro and Schorfheide (2013).

a constant and the $p$ lags of each of the elements of $\Delta \boldsymbol{Y}_t$ and $\mathbf{A}'$ be a $[n \times (np+1)]$ matrix of coefficients:

$$\boldsymbol{x}_t \equiv \begin{bmatrix} 1 \\ \Delta \boldsymbol{Y}_{t-1} \\ \Delta \boldsymbol{Y}_{t-2} \\ \vdots \\ \Delta \boldsymbol{Y}_{t-p} \end{bmatrix} \quad \text{and} \quad \boldsymbol{A}' \equiv [\boldsymbol{a}\ \boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_2 \cdots \boldsymbol{\alpha}_p].$$

The standard vector autoregressive system can then be written as:

$$\Delta \boldsymbol{Y}_t = \boldsymbol{A}'\boldsymbol{x}_t + \boldsymbol{u}_t, \tag{5}$$

where $\boldsymbol{u}_t$ is the vector of zero mean disturbances, which are independent of $\boldsymbol{x}_t$, or as:

$$\Delta y = (\boldsymbol{I}_M \otimes x)\alpha + u \tag{6}$$

with $\alpha = vec(A)$ and $u \sim N(0, \Theta \otimes \boldsymbol{I}_T)$. The least squares estimators (OLS) of $\mathbf{A}$ is:

$$\underset{[nx\{np+1)]}{\hat{\boldsymbol{A}}'} = \left[\sum_{t=1}^{T} \Delta \boldsymbol{Y}_t \boldsymbol{x}_t'\right] \left[\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right]^{-1},$$

From the regression of $\Delta Y_{jt}$ on $\boldsymbol{x}_t$:

$$\Delta Y_{jt} = \boldsymbol{\alpha}_j \boldsymbol{x}_t + u_{jt} \tag{7}$$

we obtain the estimated coefficient vector:

$$\underset{[1x\{np+1)]}{\hat{\boldsymbol{\alpha}}_j'} = \left[\sum_{t=1}^{T} \Delta Y_{jt} \boldsymbol{x}_t'\right] \left[\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right]^{-1},$$

which corresponds to the $j^{\text{th}}$ row of $\hat{\mathbf{A}}'$. Given that only predetermined variables are on the right side of the equations, and that the error terms are serially uncorrelated, OLS estimates of the VAR coefficients are consistent. Further, if the disturbances are normal, OLS is efficient. In fact, VAR with same right-hand side variables is a Seemingly Unrelated Model (SUR), which implies that the estimates are efficient regardless on the contemporaneous correlations among the disturbances.

VARs have been widely used as a tool to study the relationship of economic series, the dynamic impact of shocks on the system of variables, and also for forecasting. It has also been used to compare actual data with data generated by DSGE models with calibrated parameters. VAR models are one of the tools used by Central Banks to conduct policy analysis and for economic forecasting.

We estimate a baseline VAR model with three series generally used in New Keynesian VARs: growth rate of real GDP, inflation rate, unemployment rate, and interest rates.

These are the same series used in several recent papers such as Koop and Korobilis (2010), Cogley and Sargent (2005), Primiceri (2005), and Koop et al. (2009), among many others.

Given the importance of the financial sector in the recent financial crisis, we also estimate alternative VARs using additionally several measures of term and default spreads (VAR-Fin). The details of the data are described in Section 3.3. The VARs are estimated with two lags.

## 2.6. Bayesian Vector Autoregressive Model

We also consider the BVAR proposed in Koop and Korobilis (2010). The model and series used are the same as in the baseline VAR discussed in the previous section, but it is estimated with Bayesian methods. The parameters of the model are assumed to be random variables associated with prior probabilities. The likelihood function of (6) can be obtained from the sampling density, $p(y|\boldsymbol{\alpha}, \Theta)$.

Koop and Korobilis (2010) propose several alternative priors and estimation methods for BVARs. They show that all methods yield similar result.[20] We follow Koop and Korobilis (2010) and Del Negro and Schorfheide (2011) and use Minnesota priors. This implies that $\otimes$ is replaced by an estimate, and the prior for $\alpha$ assumes that:

$$\boldsymbol{\alpha} \sim N(\underline{\boldsymbol{\alpha}}_{nM}, \underline{V}_{nM}).$$

The Minnesota prior yields simple posterior distribution using the Normal distribution:

$$\boldsymbol{\alpha}|y \sim N(\overline{\alpha}_{nM}, \overline{V}_{nM}),$$

where

$$\overline{V}_{nM} = [\underline{V}_{nM}^{-1} + (\hat{\Theta}^{-1} \otimes (x'x))]^{-1}$$

and

$$\overline{\alpha}_{nM} = \overline{V}_{nM}[\underline{V}_{nM}^{-1}\underline{\alpha}_{nM} + (\hat{\Theta}^{-1} \otimes x)'y].$$

The prior coefficient $\overline{\alpha}_{nM}$ is set to zero, including the first parameter of the lag of each variable (as the data considered are stationary). The variance-covariance matrix $\otimes$ is assumed to be diagonal with elements obtained from regressing each dependent variable on an intercept and four lags of all variables.[21]

The use of Minnesota priors allows simple analytical posterior and predictive results. The model is, thus, estimated using Monte Carlo integration. At each real-time recursive estimation 1000 parameters are drawn, and for the forecasts, 50 are drawn from the predictive density for each parameter draw ($50 \times 1000$). The BVAR is estimated with four lags as in Koop and Korobilis (2010).

[20] Carriero et al. (2011) also find that choices related to priors, selection of hyper-parameters, and forecast construction do not affect substantially the point and density forecasting performance of BVAR models.

[21] The restrictions on coefficients become tighter at longer lags with prior variance depending on the inverse square of lag length.

## 2.7. Univariate Markov Switching Model

We apply the version of the univariate Markov switching model (MS) in Hamilton (1994) to predict output growth. As before, let $\Delta y_t$ be the growth rate of real GDP:

$$\begin{aligned} \Delta y_t &= \mu_{S_t} + \rho \Delta y_{t-1} + \cdots + \rho_r \Delta y_{t-r} + \varepsilon_t \\ \mu_{S_t} &= \mu_0 + \mu_1 S_t \quad \mu_0 < 0 \\ \varepsilon_t &\sim N(0, \sigma_\varepsilon^2). \end{aligned} \tag{8}$$

$S_t = \{0, 1\}$ is an unobserved state variable that enables the parameter $\mu_{S_t}$ to switch between two regimes, following a first-order Markov process with transition probabilities $p_{ij} = \Pr[S_t = j|S_{t-1} = i]$, where $\sum_{j=0}^{1} p_{ij} = 1,\ \ i, j = 0, 1$. The growth rate of economic activity switches back and forth from $\mu_0$ to $\mu_0 + \mu_1$. When $\mu_0 < 0$ and $\mu_0 + \mu_1 > 0$, the model captures business cycle phases representing economic contractions and economic expansions, respectively. The estimated model can be used to draw probabilities of the unobservable states representing business cycle phases, that is, filtered probabilities conditional on current information set $I_t$ denoted $\Pr[S_t = j|I_t]$, or smoothed probabilities obtained by backward recursion based on the full sample information set $I_T$, denoted $\Pr[S_t = j|I_T]$.

McConnell and Perez-Quiros (2000) found evidence of a structural break in the volatility of U.S. economic growth towards stabilization in the first quarter of 1984. This result has been further investigated by many authors and the period post-1984 has been dubbed the Great Moderation.[22] One implication of this break, as discussed in Chauvet and Potter (2002, 2005) and Chauvet and Su (2013), among many others, is that the smoothed probabilities from the standard Markov switching model miss the U.S. recessions post-1984.[23] We augment the model by allowing $y_t$ to follow two independent two-state Markov processes: one representing switches between economic recessions and expansions and the other that captures permanent structural breaks. The Markov process for detecting structural break has a switching drift and variance as proposed in Chib (1998):

$$\begin{aligned} \alpha_{D_t} &= \alpha_0(1 - D_t) + \alpha_1 D_t \\ \sigma^2_{D_t} &= \sigma^2_0(1 - D_t) + \sigma^2_1 D_t, \end{aligned}$$

where $D_t = 0$ if $t < t^*$ and $D_t = 1$ otherwise, and $t^*$ is the break date. The transition probabilities for the Markov process are set to capture the endogenous permanent break as:

$$\begin{aligned} \Pr[D_t = 0|D_{t-1} = 0] &= q \quad 0 < q < 1 \\ \Pr[D_t = 1|D_{t-1} = 1] &= 1. \end{aligned}$$

[22] Chauvet and Popli (2003) find evidence of multitple structural volatility breaks in many industrialized countries.

[23] Extensions of the model that include other series (e.g., the DFMS model, or Nalewaik, 2011), or that explicitly takes into account pulse or dummy breaks (e.g., Kim and Nelson, 1999; Chauvet and Su, 2013 etc.) overcome this problem. For a discussion, see Chauvet and Su (2013).

The linear autoregressive dynamics or order $r = 1$ is found to be the best specification in characterizing business cycle phases, and in minimizing loss functions such as BIC and AIC criteria.

Following Hamilton (1994), forecasts from the univariate Markov switching model are obtained as follows. At first, suppose $\{S_t\}$ is observed. Then, the $h$-period ahead forecast for $\mu_{S_t}$ is:

$$E(\mu_{St+h}|S_t\} = \mu_0 + \{\pi_1 + \lambda^m(S_t - \pi_1)\}(\mu_1 - \mu_0), \tag{9}$$

where $\lambda \equiv ( - 1 + p_{11} + p_{00}\}$ and $\pi_1 = (1 - p_{00})/(1 - p_{11} + 1 - p_{00}\}$. The optimal forecast of $z_{t+h} = \rho\Delta y_{t-1+h} + \cdots + \rho_r\Delta y_{t-r+h} + \varepsilon_{t+h}$ is:

$$E(z_{t+h}|z_t, z_{t-1}, \ldots, z_{t-r+1}) = e_1'\Phi^h[z_t \quad z_{t-1} \quad \cdots \quad z_{t-r+1}]', \tag{10}$$

where $e_1'$ corresponds to the first row of the (r x r) identity matrix and $\Phi$ is the (r x r) matrix of autoregressive coefficients. Substituting (9) and (10) in (8) we get:

$$\begin{aligned} E(y_{t+h}|S_t, I_t) = \mu_0 + \{\pi_1 + \lambda^m(S_t - \pi_1)\}(\mu_1 - \mu_0) \\ + e_1'\Phi^m[(y_t - \mu_{s_t}) \quad (y_{t-1} - \mu_{s_{t-1}}) \quad \cdots \quad (y_{t-r+1} - \mu_{s_{t-r+1}})], \end{aligned} \tag{11}$$

where $I_t$ is the set of observables variables. Applying the law of iterated expectations to (11) we obtain the $h$-ahead forecast, which is based only on observable variables:

$$E(y_{t+h}|I_t) = \mu_0 + \{\pi_1 + \lambda^m[\Pr(S_t = 1|I_t) - \pi_1]\}(\mu_1 - \mu_0) + e_1'\Phi^m\tilde{\mathbf{y}}_t, \tag{12}$$

where $\tilde{y}_{it} = y_{t-i+1} - \mu_0 \Pr(S_{t-i+1} = 0|I_t) - \mu_1 \Pr(S_{t-i+1} = 1|I_t)$ is the $i^{\text{th}}$ element of the (r x 1) vector $\tilde{\boldsymbol{y}}_t$.

## 2.8. Dynamic Factor Model with Markov Switching

We extend the dynamic factor model with regime switching approach in Chauvet (1998) to study the dynamics of output growth in a reduced-form multivariate setting, as explained below. This model takes into account the dynamic comovements of several variables and, therefore, captures pervasive cyclical fluctuations in various sectors of economic activity. Since recessions and expansions are caused by different shocks over time, the inclusion of different variables increases the ability of the model in representing and signaling phases of the business cycle. In addition, the combination of variables reduces measurement errors in the individual series and, consequently, the likelihood of false signaling turning points. Thus, this model allows representation of business cycle as the comovements of several sectors, with potential asymmetries in its phases, as suggested in Diebold and Rudebusch (1996).

The model is applied to variables that move contemporaneously with GDP. The series used are the same four coincident series used by the NBER Business Cycle Dating Committee to date recessions: employment, sales, personal income, and industrial production.

The model extracts the co-movements in these coincident series into a single unobserved common factor. This latent factor follows a two-state Markov switching process, capturing recession and expansion phases of the business cycle.

Let $y^*_{it}$ be the log level of the $i^{\text{th}}$ series that move simultaneously with GDP, and $\Delta y^*_{it}$ be the first difference of $y^*_{it}$.[24] The dynamic factor model with regime switching model (DFMS) is:

$$\begin{bmatrix} \Delta y^*_{1t} \\ \Delta y^*_{2t} \\ \Delta y^*_{3t} \\ \Delta y^*_{4t} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} F_t + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \end{bmatrix}. \tag{13}$$

That is, the first difference of each series is modeled as an unobserved component common to each series, given by the dynamic factor $F_t$, and an idiosyncratic component to each series, given by $v_{it}, \;\; i = 1, \ldots, 4$. The factor loadings $\lambda_i$ measure the sensitivity of the series to the dynamic factor.[25] The common component is assumed to follow a stationary autoregressive process:

$$\varphi(L)(F_t - \mu^*_{S_t}) = \eta_t \quad \eta_t \sim N(0, \sigma^2_\eta) \tag{14}$$

$$\mu^*_{S_t} = \mu^*_0 + \mu^*_1 S_t \qquad \mu^*_0 < 0,$$

where $\eta_t$ is the common shock and $\varphi(L)$ is a lag polynomial with all roots outside the unit circle. The model separates out common signal underlying the observed variables from individual variations in each sector of the economic activity. The dynamic factor captures widespread simultaneous downturns and upturns of several sectors of the economy, which are the most important features of business cycles as proposed by the pioneer economists Burns and Mitchell's (1946). On the other hand, if only one of the variables declines (e.g., industrial production), this would not characterize a recession in the model, and it would be captured by the industrial production idiosyncratic term. A recession (expansion) will occur when all variables decrease (increase) at about the same time. That is, $v_{it}$ and $\eta_t$ are assumed to be mutually independent at all leads and lags, for all $i = 1, \ldots, 4$ variables, and $d_i(L)$ is diagonal.

The asymmetries across different states of the business cycle is modeled by allowing the intercept of the factor to switch regimes according to the Markov variable, $S^*_t = 0, 1$. That is, the economy can be either in an expansion state($S^*_t = 1$), where the mean growth rate is positive; or in a contraction phase ($S^*_t = 0$), with a negative mean growth rate. The switches from one state to another is determined by the transition probabilities of the first-order two-state Markov process with transition probabilities $P(S^*_t = 1|S^*_{t-1} = 1) = p^*_{11}$

[24] The series used in estimating the model are the same coincident variables used by the NBER is calling recessions: sales, personal income, employment, and industrial production, as discussed in more detail in Section 3.2.

[25] The factor loading of one of the coincident series is set equal to one to provide a scale for the dynamic factor. This normalization is a necessary condition for identification of the factor. Notice that the choice of scale does not affect any of the time series properties of the dynamic factor or the correlation with its components.

and $P(S_t^* = 0|S_{t-1}^* = 0) = p_{00}^*$. Finally, the idiosyncratic components are assumed to follow a stationary autoregressive process:

$$d_i(L)v_{it} = u_{it} \quad u_{it} \sim i.i.d.N(0, \Omega). \tag{15}$$

The model yields estimated filtered and smoothed probabilities of the recessions and expansions at time $t$ conditional on current data or the full sample, denoted $P(S_t^* = j|I_t)$ and $P(S_t^* = j|I_T)$, $j = 0, 1$, respectively, and the filtered and smoothed business cycle index, denoted $E(F_t|I_t)$ and $E(F_t|I_T)$, respectively. The results from dynamic factor models with Markov regime switching, as estimated in Chauvet (1998), Chauvet and Hamilton (2006), and Chauvet and Piger (2008, 2013) are not affected by the structural break in variance.

The DFMS business cycle index can be interpreted as a nowcast of business cycle, but it is not a direct forecast of GDP growth, as it neither includes this series nor projects it forward. We augment Eq. (1) with the probabilities of recession and the business cycle index in the AR-DFMS model:

$$\begin{aligned} \Delta y_t &= c + \nu(L)\Delta y_t + \gamma(L)F_t + \delta(L)P(S_t = i|I_t) + \upsilon_t \\ \upsilon_t &\sim WN(0, \sigma_\upsilon^2), \end{aligned} \tag{16}$$

where $\nu(L)$, $\gamma(L)$, and $\delta(L)$ are lag polynomials, with the roots of $\nu(L)$ outside of the unit circle.[26]

Chauvet et al. (2013) also examine the marginal prediction of a linear version of the dynamic factor model in forecasting output growth. In this case, the AR(2) is augmented based on lags of factor that is produced as a linear combination of the coincident series $\Delta y_{it}^*$, and it does not include the term $P(S_t = i|I_t)$.

## 2.9. Forecast Combination

An interesting question is whether a combination of the model forecasts and subjective forecasts from the Blue Chip produces better results than the best single ones. This is particularly interesting, given that the set of information across some of the models are different and the judgmental and AR-DFMS model forecasts also include more timely monthly series. In addition, judgmental forecasts from the Blue Chip incorporate subjective information as well as expectations based on timely announcements of economic policy.

Aiolfi et al. (2011) find that the pooling that yields more accuracy gains is the combination model-based forecasts from linear and non-linear univariate specifications, and multivariate factor-augmented models with judgmental survey forecasts obtained achieved a simple equal-weighted average. We follow these authors and obtain the pooling of

[26] The standard errors are obtained using bootstrap.

forecasts $\hat{y}_{T+k|T}$ at horizon $k$ as:

$$\hat{y}_{T+k|T} = \sum_{k=1}^{M} \hat{y}_{k,T+k|T}, \tag{17}$$

where $M$ is the number of forecast combined.

## 3. FORECAST COMPARISON: REAL-TIME PERFORMANCE

Models that exhibit reasonable power in explaining the average linear dynamics of output over time may show poor performance during some events, such as recessions, or financial, currency, and banking crises, to name a few. Many papers have shown that the largest errors in forecasting output occur around business cycle turning points (see, e.g., Oh and Waldman, 1990; Beaudry and Koop, 1993; Chauvet and Guo, 2003). This has been particularly the case for recessions, which most models show a lesser forecast accuracy.

In this section we investigate the ability of models and professional forecasters to forecast the dynamics of output growth in real time as well as during expansions and recessions. We use unrevised real-time data that would have actually been available at any given point in time. The availability of these unrevised series allows analysis of the model performance at the time events were taking place.

We use annualized quarter-over-quarter changes in GDP growth, not annual growth rates, and focus on short horizons. Also, we obtain independent out-of-sample $k$-period ahead forecasts over the forecast period, in which the parameters of the model are recursively reestimated at each new observation $T$ and the $T + k$ periods ahead forecasts are computed every time based on information at $T$.[27] As discussed in Tashman (2000), this disentangles potential impact on the forecast errors of special events associated with a unique origin and also reduces the sensitivity of the errors to rapid changes across phases of the business cycle. Most important, this emulates the real-time forecasting procedures of economic agents and Central Banks at that the time events were occurring.

### 3.1. Forecast Evaluation

We examine GDP growth forecasts of the five models described above and the judgmental-based forecasts from the Blue Chip indicators.[28] Summing up, the models examined and their acronyms are:

Model 1 – Benchmark AR(2)
Model 2 – Current Depth of Recession (CDR)

[27] We do not evaluate pseudo-out-of-sample forecasts in this paper. For a discussion of these results see Chauvet et al. (2013).

[28] The best specifications of the models in terms of the lags of the common factor and the idiosyncratic components were selected based on the Bayesian and Akaike Information Criteria, root mean squared error and Theil coefficient.

Model 3 – Dynamic Stochastic General Equilibrium (DSGE)
Model 4 – Vector Autoregressive Model (VAR)
Model 5 – Vector Autoregressive Model with Financial Variables (VAR-Fin)
Model 6 – Bayesian Vector Autoregressive Model (BVAR)
Model 7 – Univariate Markov Switching (MS)
Model 8 – AR-Dynamic Factor Model with Markov Switching (AR-DFMS)[29]
Judgmental Forecast - Blue Chip Indicators (BC)
Forecast Combination (FC)

We consider three loss functions: the RMSFE, Theil inequality coefficient (THEIL), and Pesaran and Timmermann (1992) sign test:

$$RMSFE = \sqrt{\frac{1}{N}\sum_{t=T+k}^{T+N}(\Delta\hat{y}_t - \Delta y_t)^2}$$

$$THEIL = \frac{\sqrt{\frac{1}{N}\sum_{t=T+k}^{T+N}(\Delta\hat{y}_t - \Delta y_t)^2}}{\sqrt{\frac{1}{N}\sum_{t=T+k}^{T+N}\Delta\hat{y}_t^2} + \sqrt{\frac{1}{N}\sum_{t=T+k}^{T+N}\Delta y_t^2}},$$

where $T$ and $N$ denote the number of observations in the estimation and forecast samples, respectively, $k$ is the forecast horizon, $\Delta\hat{y}_t$ is the forecast and $\Delta y_t$ is the observation. Note that Theil coefficient ranges between zero and one. For both loss functions zero is a perfect forecast. The RMSFE is scale-dependent while Theil is scale invariant. Although the dependent variable is the same across the models studied, we report both the total RMSFE of forecasts and the relative to the benchmark AR(2) model. We compute the RMSFE for the full sample, for expansion periods, and for recession periods as determined by the NBER Business Cycle Dating Committee.

Theil inequality coefficient can be decomposed into bias, variance, and covariance proportion:

Bias Proportion

$$\frac{\left(\frac{1}{N}\sum_{t=T+k}^{T+N}\Delta\hat{y}_t - \frac{1}{N}\sum_{t=T+k}^{T+N}\Delta y_t\right)^2}{\left(\frac{1}{N}\sum_{t=T+k}^{T+N}(\Delta\hat{y}_t - \Delta y_t)^2\right)}$$

[29] A linear version of this model and its forecast are examined in Chauvet et al. (2013), and briefly discussed in Section 3.3.

Variance Proportion

$$\frac{[Stdev(\Delta\hat{y}_t) - Stdev(\Delta y_t)]^2}{\left(\frac{1}{N}\sum_{t=T+k}^{T+N}(\Delta\hat{y}_t - \Delta y_t)^2\right)}$$

Covariance Proportion

$$\frac{2[1 - corr(\Delta\hat{y}_t, \Delta y_t)]Stdev(\Delta\hat{y}_t)Stdev(\Delta y_t)}{\left(\frac{1}{N}\sum_{t=T+k}^{T+N}(\Delta\hat{y}_t - \Delta y_t)^2\right)}.$$

The bias and variance proportions measure, respectively, how far the mean and the variance of the forecast are from the mean and the variance of actual GDP growth. The covariance proportion is obtained by residual as the three components add up to one. Thus, the smaller the bias and variance proportions, the better the forecasts are. That is, ideally the largest fraction of the Theil coefficient should be from the covariance proportion.

We also consider the sign forecast test of Pesaran and Timmermann (1992). The test is based on the number of corrected predicted signs in the forecast series. Under the null hypothesis, the forecast $\Delta\hat{y}_{t+k}$ of the actual $\Delta y_{t+k}$ have independent distributions, that is, the forecast values have no power to predict the sign of $\Delta y_{t+k}$. That is, it measures whether there is a significant difference between the observed probability of a correctly signed forecast and the estimate of this probability under the null. The corresponding statistic is:

$$S_n = \frac{(\hat{p} - p_*)}{[V(\hat{p}) - V(\hat{p}_*)]^{1/2}} \qquad S_n \xrightarrow{d} N(0, 1, ),$$

where $\hat{p}$ is the sample estimate of the probability of a correctly signed forecast, $\hat{p}_*$ is the estimate of its expectation and $V(.)$ is their variance, obtained under the null.

### 3.2. Real-Time Data

In this section we provide a description of the data used in the estimation and forecasting process. The sample period used is determined by the common availability of all data.[30] All models are first estimated using data from 1964:Q2 to 1991:Q4. The models are recursively re-estimated for each quarter for the period starting in 1992:Q1 and ending in 2011:Q1 using only collected real-time realizations of the series as released at each quarter to generate $k$-quarter ahead forecasts.

The current U.S. real GDP series is obtained from the Bureau of Economic Analysis (BEA). All versions of the historical unrevised real-time GDP series released each month

[30] The real-time forecast sample is determined by limitations in the availability of some real-time variables for the DSGE model.

are collected and archived by the Federal Reserve Bank of Saint Louis and the Federal Reserve Bank of Philadelphia.[31] The quarterly real-time database used in this paper consists of realizations, or quarterly *vintages*, of the series as they would have appeared in the end of each quarter from 1992:Q1 to 2011:Q2. The sources and descriptions of other series used in the multivariate models are described below.

**DSGE Model.** We use the same series as SW (2007). Average weekly hours of production and non-supervisory employees for total private industries (HOUR), civilian employment (TCE), civilian non-institutional population (POP), and compensation per hour for the non-farm business sector (WAGE) are obtained from the Bureau of Labor Statistics (BLS). GDP deflator (GDPDEF), nominal personal consumption expenditures (NPCE), and nominal fixed private investment (NFPI) are obtained from the Bureau of Economic Analysis (BEA). The federal funds rate (FFR) is obtained from the Federal Reserve Board.

All nominal series are deflated using the GDP deflator. The series are transformed as in SW (2007). Real output, real consumption, real investment, and hours (times TCE/100) are in per capita terms obtained as 100 times the log of the ratio of these series to population. Inflation is the 100 times log first difference of the GDP deflator, and the annualized daily federal funds dates are converted to quarterly averages:

Real Output = 100xln ((GDP/GDPDEF)/POP);
Real Consumption = 100xln ((NCPE/GDPDEF)/POP)
Real Investment = 100xln ((NFPI/GDPDEF)/POP)
Real Wage = 100xln(WAGE/GDPDEF)
Hours = 100xln(WAGExTCE/100)/POP)
Inflation = 100xln(GDPDEF/GDPDEF(-1))
Interest Rates = FFR/4

The series are transformed into stationary according to the procedure described in subsection 2.4.

**VAR and BVAR Models.** In addition to the GDP series used in all other models, the VAR and BVAR models use the same GDP price index and interest rates as in the DSGE model. However, the series are further differenced: inflation is the annualized second log difference of the GDP deflator, and interest rates are the first difference of the Federal Funds rate, as in Koop and Korobilis (2010).

We also consider as a fourth series in the baseline VAR model several versions of the default premium (i.e., the difference between bond yields with different credit ratings): the Baa and Aaa, Aaa and Treasury Bond 10- year, Baa minus Treasury-10 year. We also consider the term premium Treasury 10-year minus Treasury 5-year. The data are obtained from Haver-DLX.

[31] See Croushore and Stark (2001) for a description of the data and of the collection procedure. The data and information are available at http://www.phil.frb.org/research-and-data/real-time-center/real-time-data/data-files/ROUTPUT/

**DFMS Model**. The series used to estimate the DFMS model are U.S. monthly Industrial Production (IP) obtained from the Federal Reserve Board, Real Manufacturing and Trade Sales (MTS) and Real Personal Income excluding Transfer Payments (PILTP) obtained from the BEA, Payroll Employment (ENAP), and Total Civilian Employment (TCE) obtained from the BLS. These are the same four monthly variables used by the NBER Business Cycle Dating Committee in establishing the beginning and end of recession dates.

The real-time data used to estimate the DFMS model (PILTP, MTS, ENAP and IP) were obtained from a combination of the real-time datasets collected in Chauvet (1998), Chauvet and Hamilton (2006), Chauvet and Piger (2008), the Federal Reserve Bank of Philadelphia and the Federal Reserve Bank of Saint Louis archives. Real-time data for PILTP and MTS were hand collected as part of a larger real-time data collection project at the Federal Reserve Bank of St. Louis and first used in Chauvet and Piger (2008). The ENAP and IP data series were obtained from the Federal Reserve Bank of Philadelphia real-time data archive described in Croushore and Stark (2001). The real-time data for TCE were hand-collected as part of Chauvet (1998) and Chauvet and Hamilton (2006) and Chauvet and Piger's (2008) research, and some more recent data obtained from the Federal Reserve Bank of Saint Louis ALFRED archive.

## Timing of Forecasts

The GDP series is first released based on preliminary and incomplete information, as it is the case of many macroeconomic variables. Multiple and often large revisions are implemented in subsequent releases in order to correct discrepancies caused by lags in the availability of primary data. There are three main releases of GDP for a quarter, which occur in the three subsequent months following that quarter. For example, the first release of GDP for the last quarter of a year occurs in the end of January of the following year, and is called "advance" version. The second release, named "second estimate" version, occurs in the end of February, and the "third estimate" release takes place in the end of March. After this "third estimate" release there are other revisions later onto include more complete information (annual or benchmark revisions, correction updates, etc.).

We use the "final" real-time release of GDP for each quarter.[32] Thus, the quarterly vintages are obtained from GDP data as available in the end of March, June, September and December of each year. For example, the vintage available in the first quarter of 1992 corresponds to GDP series for the fourth quarter of 1991 as available in March 1992, that is, the "final" estimate for this quarter. For each vintage the sample collected begins in the first quarter of 1964 and ends with the most recent data available for that vintage. The effective sample starts in 1964:Q2 after transforming the data in growth rates.

[32] We have also used the "advance" and "second" estimates. The results are discussed in Chauvet et al. (2013).

**Table 3.1** Blue Chip Survey Dates and Forecasts

| Blue Chip Survey Date | End of Estimation Sample T | Forecast Horizon $k=1$ | $k=2$ | $k=3$ |
|---|---|---|---|---|
| Apr 1992 | 1991:Q4 | 1992:Q1 | 1992:Q2 | 1992:Q3 |
| Jul 1992 | 1992:Q1 | 1992:Q2 | 1992:Q3 | 1992:Q4 |
| Oct 1992 | 1992:Q2 | 1992:Q3 | 1992:Q4 | 1993:Q1 |
| Jan 1992 | 1992:Q3 | 1992:Q4 | 1993:Q1 | 1993:Q2 |

As in Edge and Gürkaynak (2010), Edge et al. (2010), Krane (2011), and Del Negro and Schorfheide (2013), among others, we build vintages of real-time data available at the time of Blue Chip publication dates.[33] As explained in Section 2.3, the quarterly forecasts of the Blue Chip are available on a monthly basis and the surveys with forecasts of GDP growth are released on the tenth of each month for responses based on information for the previous month. We use the Blue Chip survey forecast for each quarter published in January, April, July, and October. For example, the forecast of GDP growth released in the April 10, 2008 survey is the $k$-quarter ahead forecast based on information as of the end of March 2008, which includes the "final" release of GDP for the fourth quarter of 2007. Hence, in this survey the "current" or "nowcast" forecast ($k=0$) corresponds to GDP forecast of the fourth quarter of 2007, the one-quarter ahead ($k=1$) is GDP growth projection for the first quarter of 2008, and the two-quarter ahead ($k=2$) is projection for the second quarter of 2008, all based on the "final" release of GDP for 2007. Note in most cases, the vintage date for which the Blue Chip professional forecasters are surveyed falls after the release of the actual GDP by the BEA. As a result, the $k=0$ forecast for the Blue Chip – which the Survey calls "one quarter ahead forecasts" in their publication – is actually the realized data for some dates. We, thus, focus instead on the comparison of the one and two quarter ahead forecasts of the BC with the $k=1, 2$ of the models. Let $E(y_{T+k}|I_{T)}$ be the $k$-quarter ahead forecast of $\Delta y_T$ made at $T$. The nowcast $k=0$ is then $E(y_T|I_T)$, the one-quarter-ahead forecast $k=1$ is $E(y_{T+1}|I_T)$, and the two quarter-ahead forecast $k=2$ is $E(y_{T+2}|I_T)$.

We align the dates of the Blue Chip forecasts with the ones from the other models. That is, at each Blue Chip forecast survey, we use the data that were available on that date to estimate the AR(2), the CDR, the MS, the DSGE, and the DFMS models (see Table 3.1). The first forecast considered in the analysis is for 1992:Q1 (end of March/April

[33] For the DSGE model, the real-time dataset is from Edge and Gürkaynak (2010), updated in Del Negro and Schorfheide (2013), which were obtained from the Saint Louis Fed. The Blue Chip data are obtained from Del Negro and Schorfheide (2013).

1992 release using information up to the end of March) and the last one is for 2011:Q2 (end of June/July 2011 release using information up to the end of June).

For the real-time series used to estimate the DFMS model (PILTP, MTS, ENAP and IP), we use the *vintages* of these time series as they would have appeared at each month from April 1992 to June 2011. The series ENAP, IP, and PILTP are released for month $t-1$ in month $t$. However, at time $t$, MTS is only available for month $t-2$. We use MTS availability to restrict the month data to be included in a vintage estimation sample. That is, even though ENAP, IP, and PILTP are available for month $t-1$ at time $t$, we only use at $t$ their data up to $t-2$ to balance it with the data for MTS. For example, ENAP, IP and PILTP are available up to February 1992 in the vintage for late March/early April 1992, but MTS is only available for January 1992. Thus, for this vintage we use all series up to January 1992.[34]

The DFMS model is estimated soon after the release of MTS data for that monthly vintage. For each vintage, the DFMS model is recursively estimated with the real-time data set, and monthly business cycle index and real-time probabilities of recessions are computed. The DFMS business cycle index can be interpreted as a nowcast of business cycle, but it is not a direct forecast of GDP growth, as it neither includes this series nor projects it forward. We use the business cycle index and the probabilities of recession as in Eq. (14) to obtain GDP growth forecasts from Eq. (17), pairing GDP with the ones used in BC and in the other models.

Note that this pairing generates information advantage for the Blue Chip forecast since the BC uses information all the way up to end of month prior to the survey date. For example, in the April 1992 survey, although the BC judgmental-based forecasts use the same GDP data as the models, they also include monthly information up to the end of March 1992. The DFMS model uses the same GDP data as the other models, but only monthly information up to January 1992.

We should stress that the estimation of all models are based solely on information that was available at each date, which aims to reproduce the forecasting problem of agents and Central Banks at the time the events were unfolding.

### 3.3. Real-Time Forecast Results

We examine the real-time GDP growth forecasts of the models described in section 3 and the judgmental-based forecasts from the Blue Chip indicators. As discussed earlier, our goal is to study short-term forecasts, as recent literature has shown that real-time forecasts of output growth from one-quarter and on are poor and uninformative (see, e.g., Wang,

[34] Although the DFMS model could be estimated using the only readily available information from ENAP and IP instead of waiting for the MTS data, we preferred to estimate the model with all series simultaneously, as the resulting indicator has been proved to be a reliable real-time indicator of business cycles (see Chauvet, 1998; Chauvet and Hamilton, 2006), and Chauvet and Piger (2008, 2013).

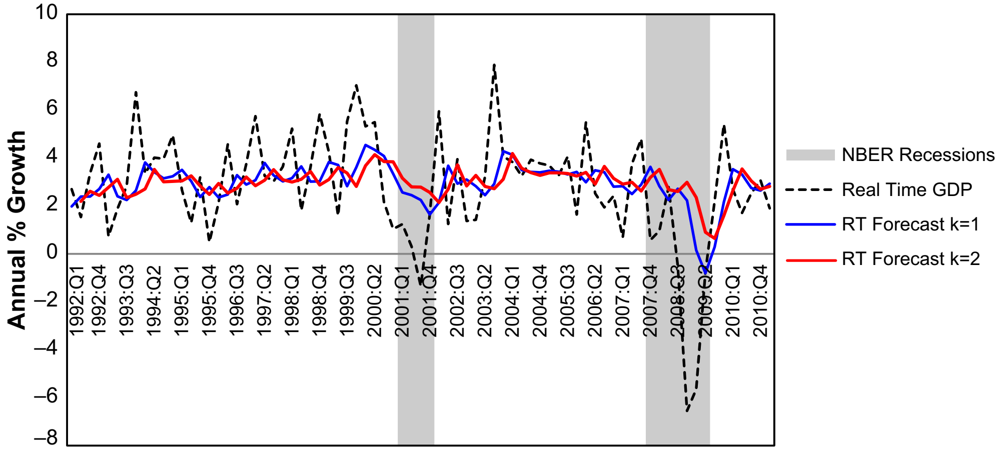


**Figure 3.1** Real-time forecasts from the benchmark AR(2) model (—), GDP growth (- - -), and NBER recessions (shaded area).

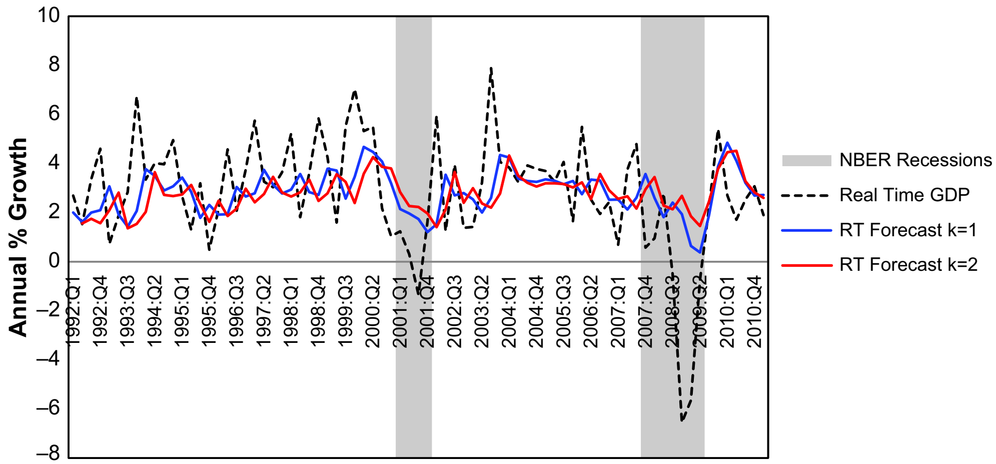


**Figure 3.2** Real-time forecasts from the CDR model (—), GDP growth (- - -), and NBER recessions (shaded area).

2009; Edge et al., 2010; Edge and Gürkaynak, 2010). Hence we focus on steps up to two quarters ahead, using quarter-over-quarter growth changes in GDP.[35]

Figures 3.1–3.13 show actual GDP growth, and real-time out-of-sample forecasts for $k = 1$ and $k = 2$ from the models and from the Blue Chip, together with shaded areas

[35] Chauvet et al. (2013) examine longer horizons and different vintages and find that the qualitative results are similar to the ones found in this paper. In particular, the ranking of the models remains roughly the same – but the long run horizon forecasts are very poor for most models. In addition, using the "advance" and the "second" releases of GDP growth leads to an overall improved performance of the nonlinear time series models over the other ones.

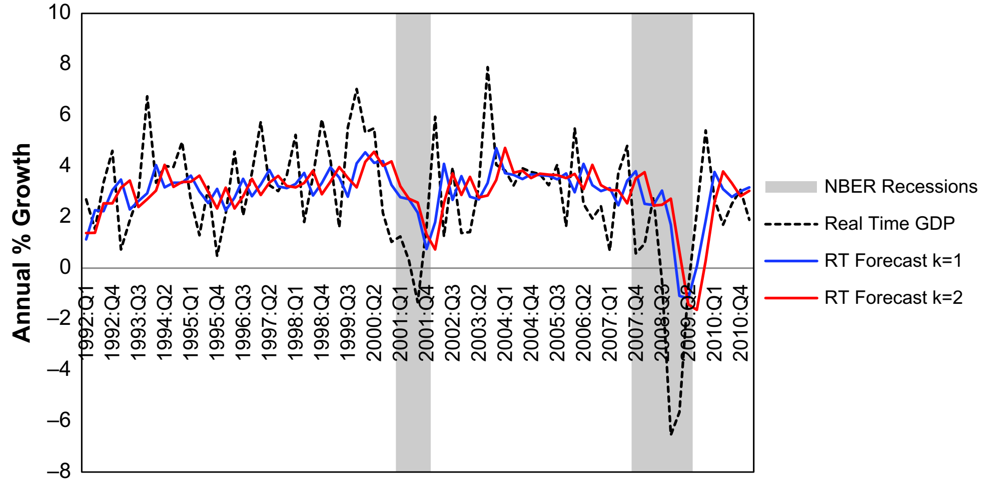


**Figure 3.3** Real-time forecasts from the MS model (—), GDP growth (- - -), and NBER recessions (shaded area).

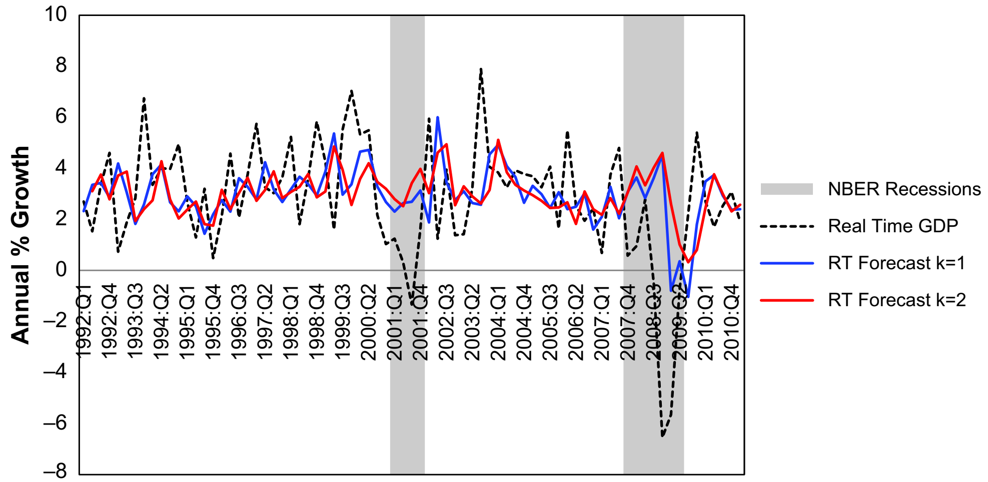


**Figure 3.4** Real-time forecasts from the VAR model (—), GDP growth (- - -), and NBER recessions (shaded area).

for NBER recessions. We compute the Theil coefficient, the RMSFE, and Pesaran and Timmermann's (1992) test for the real-time out-of-sample period. We also report the RMSFE for output growth forecasts during recession and expansion periods as determined by the NBER.[36] Tables 3.2 and 3.3 report the loss functions for the different models.

[36] There are two recessions in the period analyzed. According to the NBER, the 2001 recession started in 2001:Q1 and ended in 2001:Q4. The 2007–2009 recession started in 2007:Q4 and ended in 2009:Q2.

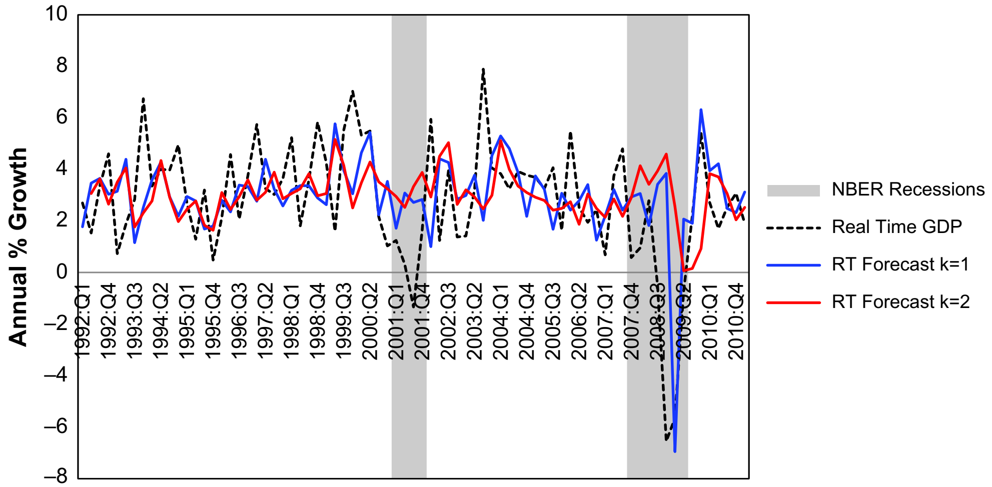


**Figure 3.5** Real-time forecasts from the VAR-Fin Aaa-Baa model (—), GDP growth (- - -), and NBER recessions (shaded area).

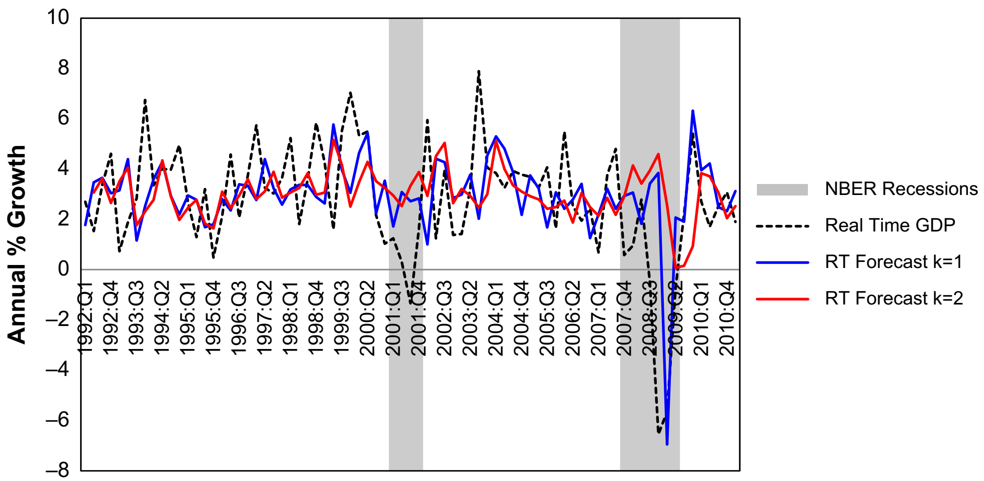


**Figure 3.6** Real-time forecasts from the VAR-Fin Aaa-T10 model (—), GDP growth (- - -), and NBER recessions (shaded area).

### *3.3.1. Real-Time Out-of-Sample Period*

The benchmark AR(2) model forecasts very well the mean of actual GDP growth, but does a poor job in predicting its volatility. This is reflected in the components of the Theil coefficient in Tables 3.1 and 3.2. The bias proportion is close to zero, but the variance proportion is around 50%, indicating that the model does not track well the variance of GDP growth.

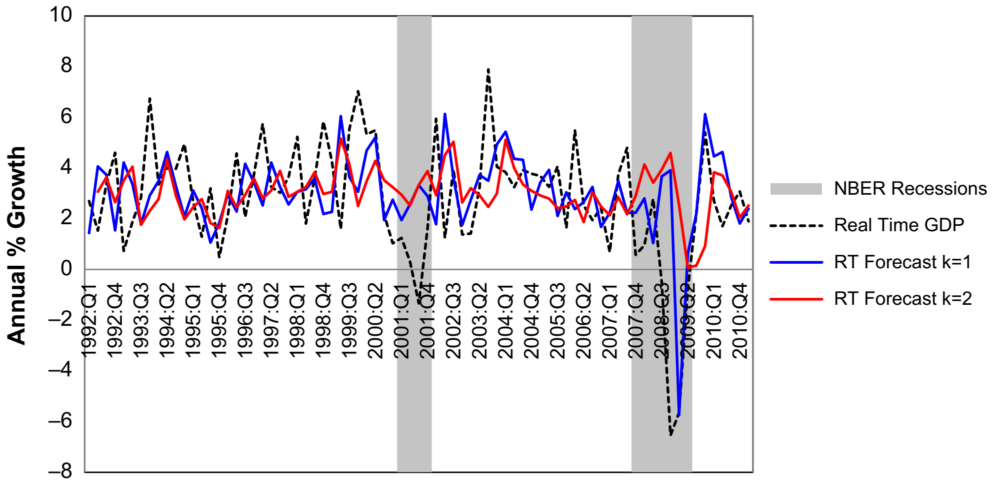


**Figure 3.7** Real-time forecasts from the VAR-Fin Baa-T10 model (—), GDP growth (- - -), and NBER recessions (shaded area).

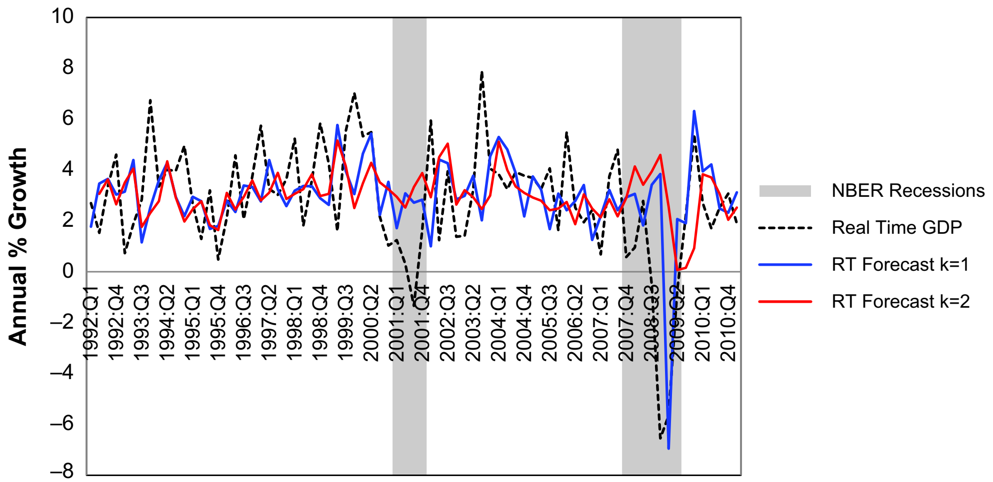


**Figure 3.8** Real-time forecasts from the VAR-Fin T10-T5 model (—), GDP growth (- - -), and NBER recessions (shaded area).

The forecasts of the CDR model are similar to the benchmark autoregressive model, with the relative RMSFE close to one. However, the CDR model displays a slight reduction in the RMSFE for forecasts of output growth during recessions, and shows a modest improvement in tracking the variance of GDP growth. The variance proportion of the Theil coefficient is 43% and 46% for $k = 1$ and $k = 2$, respectively.

Both the benchmark and the CDR models have a very good accuracy in forecasting the mean of the series, with the bias proportion close to zero. Interestingly, the RMSFE

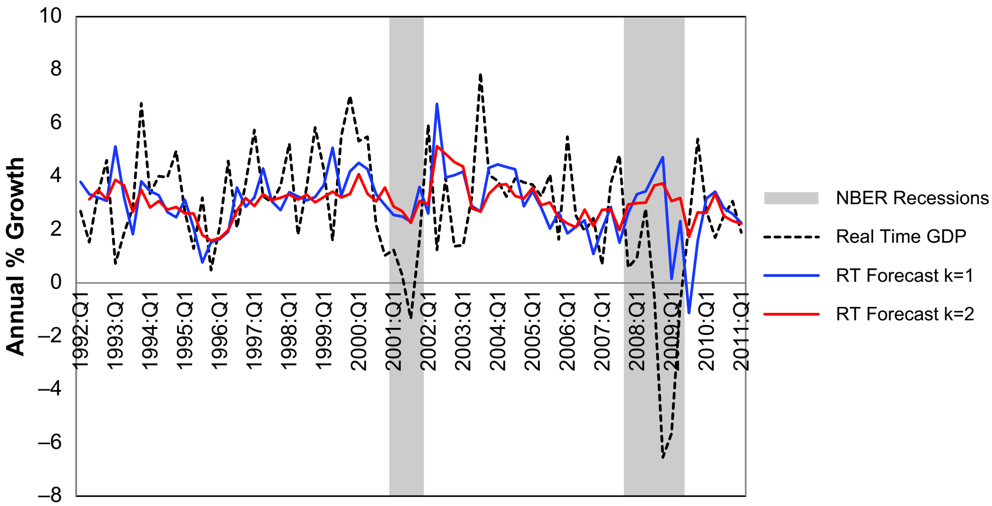


**Figure 3.9** Real-time forecasts from the BVAR model (—), GDP growth (- - -), and NBER recessions (shaded area).

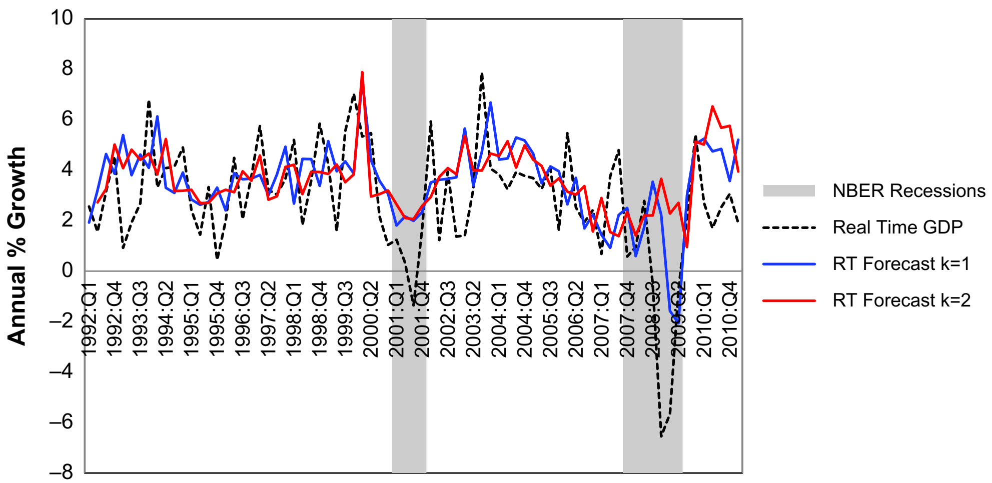


**Figure 3.10** Real-time prediction DSGE model (—), GDP growth (- - -), and NBER recessions (shaded area).

for the simple univariate AR(2) model is lower compared to most multivariate models at the one and two-quarter horizons.

The performance of the univariate MS model is somewhat similar to the CDR and DSGE models with the relative RMSFE slightly below one for $k=1$ and $k=2$. The relative Theil coefficient of the MS model to the benchmark is below one for both horizons. However, the differences between these forecasts are not significant at the

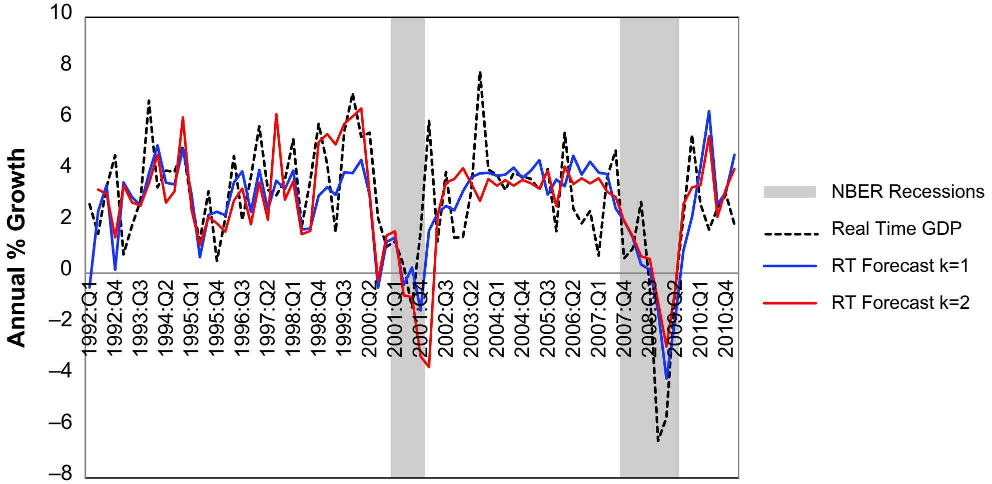


**Figure 3.11** a – Real-time Forecasts from the AR-DFMS model (—), GDP growth (- - -), and NBER recessions (shaded area).

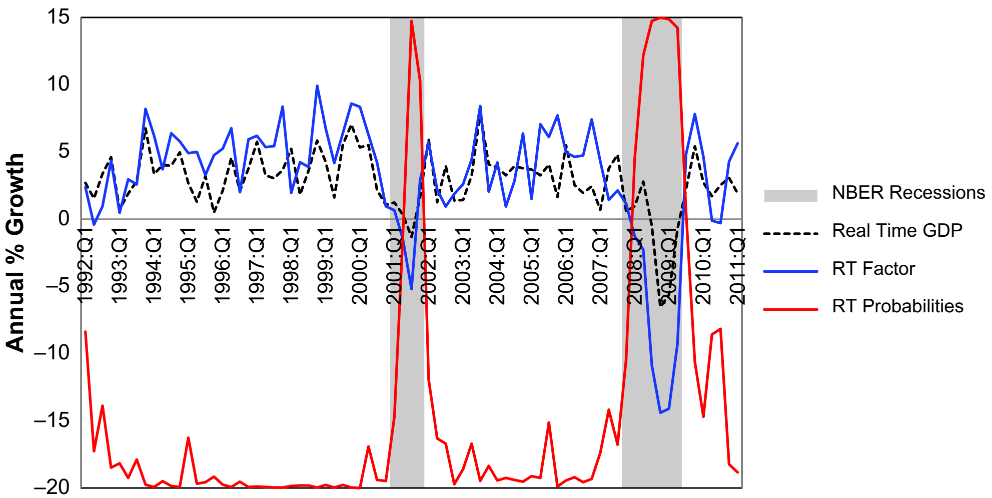


**Figure 3.11** b – Real-time Factor (—) and probabilities from the DFMS model (—), GDP growth (- - -), and NBER recessions (shaded area).

5% level using Diebold and Mariano's test (DM 1995). Nalewaik (2011) finds similar results extending Hamilton's model to consider three sates and additional series such as housing starts, GDI and unemployment rate. The forecasts are constructed similarly, based in Hamilton (1989,1994) as discussed in Section 2.7. The paper compares the results of this MS model with the Blue Chip forecasts, and finds that the performance varies across sub-samples and forecast horizons. In particular, it finds that for a longer sample starting

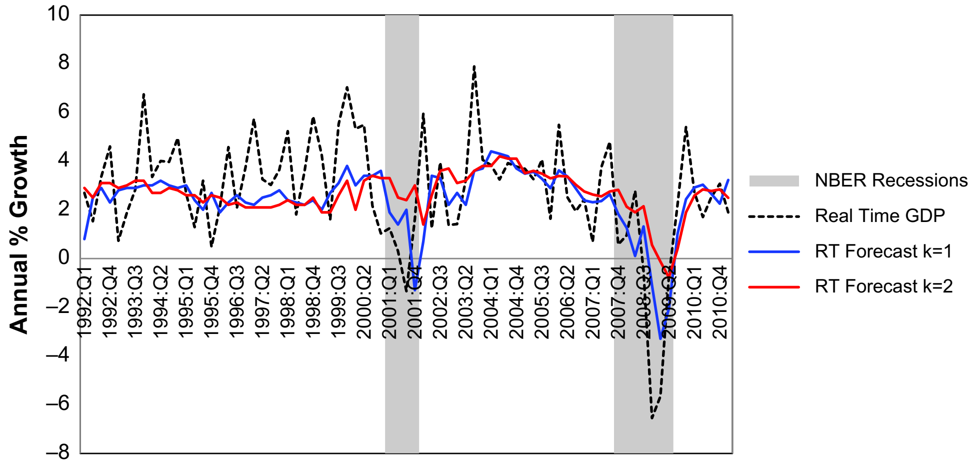


**Figure 3.12** Real-time forecasts from the Blue Chip (—), GDP growth (- - -), and NBER recessions (shaded area).

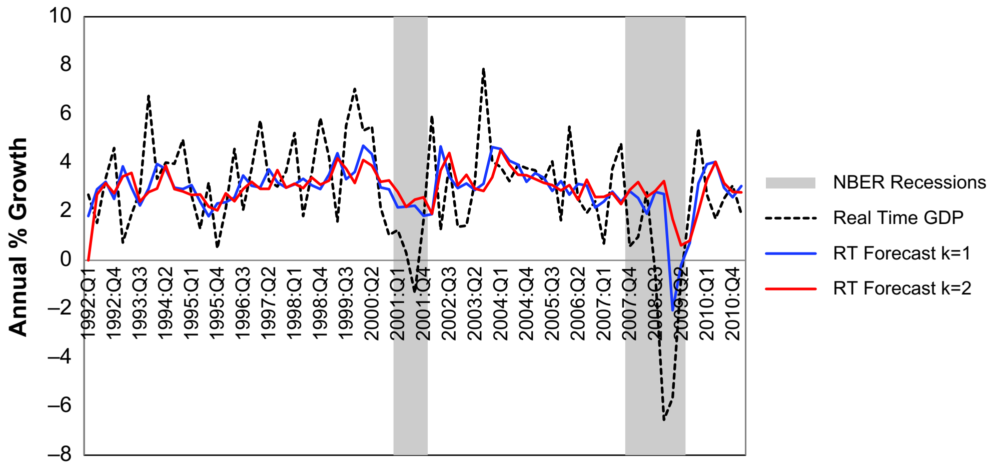


**Figure 3.13** Real-time forecasts from the forecast combination (—) GDP growth (- - -), and NBER recessions (shaded area).

in 1981, the differences in forecasts are not statistically significant, but for a sample starting in 1993 there are gains at longer horizons ($k = 4, 5$).[37] However, as discussed earlier, the literature finds that the BC forecasts are not informative for GDP growth at horizons beyond two quarters. For example, Del Negro and Schorfheide (2013) also find that the RMSFEs of the DSGE model outperforms the BC forecasts at $k => 5$, while Edge et al. (2010), Edge and Gürkaynak (2010) find similar results for DSGE, VAR, and BVAR at long horizons. Edge and Gürkaynak (2010) and Wang (2009) show that these forecasts,

[37] The difference in forecast is also statistically different from the BC forecast for recession periods at four-quarter ahead.

**Table 3.2** RMSFE, Theil Inequality, and PT Tests– Total and Relative to the Benchmark Real-Time Out-of-Sample One-Quarter Ahead Forecasts

| | RMSFE | | | Theil Inequality Coefficient | | | | |
|---|---|---|---|---|---|---|---|---|
| **Models** | **Full Sample** | **Expansion** | **Recession** | **Total** | **Bias** | **Var** | **Cov** | **PT Test** |
| AR(2) | 2.150 | 1.752 | 3.735 | 0.322 | 0.003 | 0.471 | 0.526 | 3.819** |
| CDR | 2.194 | 1.840 | 3.657 | 0.333 | 0.000 | 0.434 | 0.566 | – |
| Relative | 1.020 | 1.050 | 0.979 | 1.037 | | | | |
| DSGE | 2.205 | 1.915 | 3.466 | 0.294 | 0.111 | 0.128 | 0.667 | 5.374** |
| Relative | 1.025 | 1.093* | 0.928 | 0.916* | | | | 1.407** |
| MS | 2.111 | **1.728** | 3.644 | 0.311 | 0.010 | 0.390 | 0.600 | 5.437** |
| Relative | 0.982 | **0.986** | 0.976 | 0.967 | | | | 1.424** |
| BC | **1.923** | 1.801 | 2.534 | 0.301 | 0.039 | 0.281 | 0.680 | 5.709** |
| Relative | **0.894**** | 1.028 | 0.679** | 0.935 | | | | 1.495** |
| AR-DFMS | **1.920** | 1.879 | **2.149** | **0.281** | 0.007 | 0.075 | 0.918 | 4.500** |
| Relative | **0.893**** | 1.072 | **0.575**** | **0.874**** | | | | 1.178** |
| BVAR | 2.526 | 2.009 | 4.524 | 0.368 | 0.008 | 0.216 | 0.776 | −0.269 |
| Relative | 1.175** | 1.146** | 1.211** | 1.143** | | | | −0.07** |
| VAR | 2.452 | 1.985 | 4.296 | 0.360 | 0.005 | 0.249 | 0.746 | 2.527* |
| Relative | 1.140** | 1.133** | 1.150** | 1.120** | | | | 0.662** |
| VAR-Fin (Baa-Aaa) | 2.349 | 1.966 | 3.929 | 0.335 | 0.007 | 0.111 | 0.882 | 3.818** |
| Relative | 1.092* | 1.122** | 1.052 | 1.043 | | | | 0.999 |
| VAR-Fin (Aaa-T10) | 2.441 | 2.000 | 4.209 | 0.355 | 0.006 | 0.199 | 0.795 | 2.527* |
| Relative | 1.135** | 1.142** | 1.127** | 1.105** | | | | 0.662** |
| VAR-Fin (Baa-T10) | 2.367 | 2.002 | 3.897 | 0.338 | 0.007 | 0.110 | 0.883 | 3.818** |
| Relative | 1.101** | 1.142** | 1.043 | 1.050 | | | | 0.999 |
| VAR-Fin (T10-T5) | 2.425 | 1.879 | 4.471 | 0.357 | 0.004 | 0.274 | 0.721 | 4.312** |
| Relative | 1.128** | 1.072 | 1.197** | 1.112** | | | | 1.129** |
| Forecast | 2.101 | 1.757 | 3.517 | 0.312 | 0.004 | 0.397 | 0.598 | 5.437** |
| Combination | 0.977 | 1.003 | 0.942 | 0.969 | | | | 1.424** |

*Note*: (*) and (**) denote that the difference between the loss function from the model and the benchmark is statistically significant at the 5% and 1% level, respectively, using Diebold and Mariano's (1995) test. Full sample is from 1992:Q1 – 2011:Q1. Recession (expansion) corresponds to periods of recession (expansion) phases as dated by the NBER. RMSFE stands for root mean squared forecast error, and PT for Pesaran and Timmermann's test (1992). The loss functions are given in absolute terms, and in relative terms compared to the AR(2) model. The models are: univariate autoregressive AR(2), Cumulative Depth of Recession (CDR), Dynamic Stochastic General Equilibrium (DSGE), Univariate Markov Switching (MS), Dynamic Factor with Markov Switching (AR-DFMS), Bayesian VAR (BVAR), Baseline VAR, and VARs including Baa-Aaa (VAR-Fin Baa-Aaa), Aaa-T10 (VAR-Fin Aaa-T10), Baa-T10 (VAR-Fin Baa-T10), and T10-T5 (VAR-Fin T10-T5). FC is the forecast combination. BC stands for the Blue Chip forecasts. The (-) entry in the last column indicates that the division is undefined.

however, are outperformed by simple autoregressive processes or naïve constant growth models and conclude that the comparison is among uninformative forecasts.

We find that the VAR models do not generally perform as well as the other models. The relative RMSFE with respect to the AR(2) benchmark for the BVAR and all VAR

**Table 3.3** RMSFE, Theil Inequality, and PT Tests– Total and Relative to the Benchmark Real-Time Out-of-Sample Two-Quarter Ahead Forecasts

| | RMSFE | | | Theil Inequality Coefficient | | | | |
|---|---|---|---|---|---|---|---|---|
| **Models** | **Full Sample** | **Expansion** | **Recession** | **Total** | **Bias** | **Var** | **Cov** | **PT Test** |
| AR(2) | 2.345 | **1.796** | 4.353 | 0.351 | 0.005 | 0.562 | 0.433 | – |
| CDR | 2.373 | 1.935 | 4.098 | 0.365 | 0.000 | 0.460 | 0.540 | – |
| Relative | 1.012 | 1.077* | 0.941 | 1.039 | | | | |
| DSGE | 2.436 | 1.934 | 4.346 | 0.326 | 0.114 | 0.219 | 0.667 | – |
| Relative | 1.039 | 1.077* | 0.998 | 0.929 | | | | |
| MS | 2.373 | 1.911 | 4.163 | 0.348 | 0.009 | 0.292 | 0.699 | 2.507* |
| Relative | 1.012 | 1.064 | 0.956 | 0.990 | | | | |
| BC | **2.124** | 1.850 | 3.307 | 0.330 | 0.005 | 0.485 | 0.511 | 5.400** |
| Relative | **0.905**** | 1.030 | 0.760** | 0.938 | | | | |
| DFMS | 2.137 | 2.062 | **2.538** | **0.304** | 0.001 | 0.025 | 0.973 | 5.234** |
| Relative | 0.911* | 1.148** | **0.583**** | **0.866**** | | | | |
| BVAR | 2.472 | 1.824 | 4.748 | 0.366 | 0.008 | 0.460 | 0.532 | – |
| Relative | 1.054 | 1.016 | 1.091** | 1.043 | | | | |
| VAR | 2.594 | 1.937 | 4.933 | 0.382 | 0.007 | 0.318 | 0.675 | – |
| Relative | 1.106** | 1.079* | 1.133** | 1.087* | | | | |
| VAR-Fin (Baa-Aaa) | 2.593 | 1.955 | 4.886 | 0.381 | 0.007 | 0.292 | 0.701 | – |
| Relative | 1.105** | 1.088** | 1.122** | 1.083* | | | | |
| VAR-Fin (Aaa-T10) | 2.559 | 1.901 | 4.888 | 0.375 | 0.008 | 0.320 | 0.671 | – |
| Relative | 1.091* | 1.059 | 1.123** | 1.069 | | | | |
| VAR-Fin (Baa-T10) | 2.566 | 1.927 | 4.851 | 0.376 | 0.008 | 0.304 | 0.687 | 3.792** |
| Relative | 1.094* | 1.073* | 1.114** | 1.070 | | | | |
| VAR-Fin (T10-T5) | 2.577 | 1.897 | 4.960 | 0.379 | 0.007 | 0.325 | 0.668 | – |
| Relative | 1.099** | 1.057 | 1.139** | 1.080* | | | | |
| Forecast | 2.293 | 1.773 | 4.212 | 0.342 | 0.007 | 0.530 | 0.463 | – |
| Combination | 0.978 | 0.988 | 0.968 | 0.972 | | | | – |

*Note*: (*) and (**) denote that the difference between the loss function from the model and the benchmark is statistically significant at the 5% and 1% level, respectively, using Diebold and Mariano's (1995) test. Full sample is from 1992:Q1 – 2011:Q1. Recession (expansion) corresponds to periods of recession (expansion) phases as dated by the NBER. RMSFE stands for root mean squared forecast error, and PT for Pesaran and Timmermann's test (1992). The loss functions are given in absolute terms, and in relative terms compared to the AR(2) model. The models are: univariate autoregressive AR(2), Cumulative Depth of Recession (CDR), Dynamic Stochastic General Equilibrium (DSGE), Univariate Markov Switching (MS), Dynamic Factor with Markov Switching (AR-DFMS), Bayesian VAR (BVAR), Baseline VAR, and VARs including Baa-Aaa (VAR-Fin Baa-Aaa), Aaa-T10 (VAR-Fin Aaa-T10), Baa-T10 (VAR-Fin Baa-T10), and T10-T5 (VAR-Fin T10-T5). FC is the forecast combination. BC stands for the Blue Chip forecasts. The (-) entries in the last column indicate that the division is undefined.

models are greater than one for both horizons. The difference between the models' forecasts and the benchmark is significantly different from each other at the 1% level using DM's (1995) test for all but the BVAR model at $k = 2$. Interestingly, the BVAR does better than any other VAR at $k = 2$ and worse than any other VAR at $k = 1$.

The baseline VAR and VAR-Fin models generally have better accuracy at the two-quarter horizon than the one-quarter ahead horizon, according to the RMSFE. The inclusion of the term spread improves the forecast performance of the baseline VAR, but not as much as the default spread. For $k = 1$, the best accuracy among these models is for the VAR-Fin that includes the default risk Baa-Aaa, followed by the VAR-Fin Baa-T10, for both loss functions considered. For $k = 2$, the BVAR with no financial variable does best.

We find that the forecast accuracy of the multivariate DSGE model is similar to the univariate benchmark. For the full out-of-sample period the relative RMSFE for the DSGE model is slightly worse than the AR(2) at the one and two-quarter ahead horizon, but the difference between the models' forecasts is not significantly different from each other using DM test. This is in agreement with a vast literature on the forecast accuracy of DSGE, which finds that its forecasts are comparable or slightly superior to the ones obtained from VARs and BVAR, but not significantly different from simple benchmarks such as univariate autoregressive processes. The Theil coefficient indicates that the DSGE model forecasts relatively well the volatility of GDP fluctuations, with a variance proportion of only 13% for $k = 1$ and 22% for $k = 2$. Notice, however, that the DSGE model forecasts GDP growth with a bias, as shown in Figure 3.10 and Tables 3.1 and 3.2. The bias proportion of the Theil coefficient is 11% whereas all other models exhibit almost zero bias proportion.

Both loss functions for the AR-DFMS model are substantially lower than the one from the benchmark, and the difference is significant at the 1% or 5% level. For example, the RMSFE from the AR-DFMS model for $k = 1$ is only 1.920, considerably lower than any other specification examined, including the AR(2) (RMSFE = 2.150), the DSGE (RMSFE = 2.205), the univariate MS model (RMSFE = 2.111), and the best performing VAR for this horizon, the VAR-Fin Baa-Aaa (RMSFE = 2.349). Similar results are also found for the two-quarter ahead forecast horizon.

The AR-DFMS model also displays a very good ability in forecasting the volatility of GDP growth, outperforming all other models in this dimension. More specifically, the variance proportion of the Theil coefficient – which measures how far the forecast is from the variance of the actual series – is the smallest among all models. For $k = 1$, it is only 7.5%, whereas the variance proportion for the benchmark AR(2) is 47%. Other models that also track relatively well the volatility of GDP growth are the VAR-Fin Baa-T10 (10%), VAR-Fin Baa-Aaa (11%), and the DSGE (13%). However, for the two-quarter ahead forecast, the variance proportion for the AR-DFMS model is substantially lower than all other models (2.5%). This supports previous findings in the literature, in which the non-linearity of the Markov regime switching generates additional cyclical movements that are useful in replicating the variability of the business cycle.[38] The AR-DFMS model

[38] See, for example, Chauvet (1998, 2001), Chauvet and Hamilton (2006), Chauvet and Piger (2008), and Morley et al. (forthcoming).

also displays good forecasting accuracy of the mean growth rate of GDP, with a bias proportion almost zero.

The BC indicators are real-time judgmental forecasts made at each point at time and are not revised. The AR-DFMS model and the BC indicators have the best forecast accuracy relatively to all models to a large extent. Their performance is comparable, as their RMSFE associated with the one and two-quarter ahead forecasts are not significantly different from each other. Notice, however, that the difference between the Theil coefficient of the BC and the AR(2) model is not statistically significant for both horizons, whereas the relative forecast accuracy of the AR-DFMS to the benchmark is significant at the 1% level according to DM's test. This is in line with the evidence in Edge et al. (2010), Edge and Gürkaynak (2010), Wang (2009), Wieland and Wolters (2011), Del Negro and Schorfheide (2013), and Lundquist and Stekler (2011), who find that the judgmental-based results show modest nowcasting and short-run accuracy, but a lessen forecasting ability from one-quarter ahead and on.

The forecast combination has an overall improved accuracy compared to some models, but the AR-DFMS and the BC forecasts outperform the pooling. Notice that the forecast combination tracks relatively poorly the variance of output growth as measured by Theil coefficient.

Tables 3.2 and 3.3 also show the results for Pesaran and Timmerman's test. The null that there is a significant difference between the observed probability of a correctly signed forecast and the estimate of this probability is rejected at the 1% and 5% level for all models, except of the BVAR. This implies that the forecast from all other models have power to predict the sign of actual GDP growth.

The performance of the BVAR model, and the VAR to a lesser extent, is intriguing. However, it is somewhat in line with previous literature, which finds that the BVAR model generally yields good forecast accuracy for several variables in the system compared to other models, but not for real GDP growth. This is also the case for the VAR models (see, e.g., Christoffel et al., 2008; Edge et al., 2010; Wolters, 2010; Edge and Gürkaynak, 2010; Wieland and Wolters, 2011, etc.). Wieland and Wolters (2011) who also estimate a BVAR (4) find that this model forecasts output growth more accurately during the 1990–1991 and the 2001 recessions, but not as well when the sample includes the most recent recession. They reason that this might be related to the fact that the lag structure of the BVAR might work better during periods of lower volatility of output growth, such as during normal times and less volatile recessions as the 1990–1991 and the 2001 one.[39] This is supported in Clark (2011) who considers a BVAR with stochastic volatility. He finds that taking into account changes in the dynamics of volatility in the U.S. economy substantially improves the BVAR performance for density forecasts. A similar result is also found in a large-dimensional BVAR model with common stochastic volatility proposed in Carriero et al. (2012).

[39] Notice that our sample excludes the 1990–1991 recession.

### *3.3.2. Recession and Expansion Periods*

#### Loss Functions

Tables 3.2 and 3.3 also show the RMSFE for real-time forecasts during expansion and recession periods. Some striking results are unveiled when the real-time out-of-sample period is divided across business cycle phases, as it allows examination of the sources of differences in loss functions. Some models that perform well for the full sample display poor accuracy for expansions or recessions. Some other models that do poorly in forecasting recessions do very well in forecasting expansions and vice versa.

**Expansions.** Interestingly, the benchmark AR(2) model has a very good forecast accuracy for expansions, ranking first for $k = 2$ and ranking second only to the univariate MS model for $k = 1$. At the one-quarter ahead forecast, the difference between the performance of the benchmark forecasts and some models is large and significant, such as for the DSGE, the BVAR, the baseline VAR, and all VAR-Fin models with the exception of the one that includes the term spread (VAR-Fin T10-T5). However, the accuracy of the benchmark model is not significantly different from the non-linear time series models CDR, MS, AR-DFMS, and from the BC forecasts. At the two-period ahead forecast, the AR(2) model has the lowest RMSFE of all forecasts including the ones from the BC, but the difference is not statistically significant for most models except for the AR-DFMS, which does better, and the VAR-Fin Baa-Aaa model, which does worse. Note that the BVAR and the VAR-Fin T10-T5 also perform well during expansions, but the relative RMSFE is not statistically significant.

We also consider pooling all forecasts from the models and from the Blue Chip indicators using equal weight average. We find that it results in slight better accuracy compared to the simple AR(2) for the full sample, but the differences in forecasts are not statistically significant. During expansion periods, the MS and AR(2) models are more accurate than the forecast combination at $k = 1$ but not at $k = 2$.

Overall, these findings imply that by using a simple univariate linear (AR(2)) or nonlinear (MS) autoregressive model of GDP growth, one would have gotten in real time as good as forecasts during expansions than any other model and the professional forecasters. These simple models exhibit a very good ability to track the mean of GDP growth during normal times.

**Recessions.** The real-time performance of most models is generally poor in forecasting recessions. The RMSFEs are a lot larger during recessions compared to expansions and to the full out-of-sample period. This is the case for all models and the BC forecasts, which is consistent with several studies that show that recessions are harder to forecast than expansions.[40]

The performance of the benchmark AR(2) model is quite different for recessions compared to expansions and to the full sample. At both horizons, its forecast accuracy

[40] See, e.g., Chauvet and Guo (2003), Beaudry and Koop (1993), or Oh and Waldman (1990), among several others.

as measured by the RMSFE is worse than all but the VAR and BVAR models. The CDR model, which is designed to capture the depth of recessions, does better than the benchmark during recessions. However, their forecast accuracy is not significantly different at any significance level. This is also the case for the MS model.

The baseline VAR and the VAR-Fin models do not generally perform well in forecasting recessions, displaying the worst real-time accuracy at the one and two-quarter ahead horizons compared to all other models. The BVAR model, on one hand, has the best real-time accuracy performance of all eleven models considered and the BC forecast for $k=2$, but on the other hand is the worst of all VAR models for $k=1$.

Interestingly, the DSGE model forecasts of GDP growth at $k=1$ fare relatively well during recessions compared to the benchmark AR(2), CDR, VAR, VAR-Fin, and BVAR models. On the other hand, at $k=2$ the non-linear time series models such as the CDR, MS, and AR-DFMS do better than the benchmark, DSGE model, the VAR, VAR-Fin, and BVAR models. However, for both horizons, the difference in accuracy is only statistically significant for the VAR, VAR-Fin, and BVAR models. This is not a very informative conclusion, as the VAR models have very poor forecasting accuracy for recessions.

The best forecasting model for recessions is the AR-DFMS by a large difference compared to other models for both horizons. For example, its RMSFE (=2.149) is substantially lower compared to the AR(2), (RMSFE = 3.735), and it is about 50% to 66% lower than the RMSFE for the other models. For the two-quarter ahead horizon, the RMSFE (=2.538) for the AR-DFMS model is about only 51%–62% of the RMSFE of the other models.

The good performance of the AR-DFMS model in forecasting GDP growth during recessions is only comparable to the BC forecasts, although the model produces forecasts that are more accurate compared to the professional forecasts, with the difference in performance significant at any statistical level. The relative RMSFE during recessions for the AR-DFMS model is 0.848 of the BC forecast for $k=1$, and 0.767 for $k=2$.

We find that the forecast combination of all models and the BC is outperformed by the AR-DFMS model and the BC forecasts for the full sample and during recessions at one and two-quarter ahead. The forecast combination is also outperformed by the CDR and MS models during recessions at the two-quarter horizon.

Chauvet et al. (2013) examine the forecast ability of a linear version of the dynamic factor model, as discussed in Section 2.8. They find that this model yields forecasts that are comparable to the AR(2) model, and is outperformed by the AR-DMFS model. The comparative advantage of the AR-DFMS model to the linear DFMS and the other models considered is found to be mostly in the probability of recession terms obtained from monthly series (see Eq. (17)). Interestingly, although the forecasts of the univariate MS model are also based on probability of recessions, these are not as good forecasts as the ones from the AR-DMFS model. A possible reason is that the MS model is based only on information contained on quarterly GDP while the AR-DFMS is based on information from monthly coincident indicators of economic activity. The probabilities of recession from the DFMS model based on monthly coincident indicators timely signal recessions in real time, while the ones based solely on GDP growth yield delayed signals particularly for the last two recessions as GDP growth only mildly decreased *at their onset*.

Adding up, the accuracy of most models to output growth is relatively poor during recessions. Although the forecast ability of some models is good during expansions, most fail to forecast GDP growth during recessions. Some of the VAR-Fin models such as the one that includes the term spread T10-T5 and the BVAR do better than the DSGE model during expansions, but the gains are offset by their performance during recessions (which is the reason why their overall RMSFE for the full real-time sample is lower than the RMSFE for the DSGE model).

The best models for tracking future GDP growth during expansions are the univariate AR(2) and MS models, and the VAR-Fin T10-T5 model. For recessions, the best model is the AR-DFMS model and the BC forecasts.

The findings that the univariate autoregressive model has a good forecasting performance are in accord with Marcellino (2008). This chapter compares the accuracy of a large set of linear and non-linear models. It finds that carefully specified linear autoregressive ones (e.g., including linear trends) outperform most models.

## Graphical Analysis

An analysis of the forecast dynamics of the models in Figures 3.1 to 3.13 gives a more comprehensive picture of their performance over time and across business cycle phases. As shown in the figures, the models display a better forecast ability to GDP growth during expansions compared to recessions. In particular, all models fail to forecast negative output growth during the 2001 recession, with the exception of the AR-DFMS model and the Blue Chip forecasts. The forecast accuracy differs substantially across models during the 2007–2009 recession, especially regarding the timing and intensity of the predicted fall in GDP growth in real time, as discussed below.

Figures 3.1 and 3.2 show the real-time forecasts for the AR(2) and the CDR models. These models track closely the future mean of GDP growth for both horizons, but not as well its volatility. This is particularly accentuated during the 2001 recession and the 2007–2009 recession. Both models show a small decline relatively to the actual decrease in GDP growth during these recessions. During the 2007–2010 recession, the CDR model forecasts only a mild decline in GDP but not negative growth, including during the financial crisis and the aftermath between 2008:Q3 and 2009:Q1. Actual unrevised GDP growth had a steep fall during this period, reaching −0.5 in 2008:Q3, −6.5 in 2008:Q4, and −5.6 in 2009:Q1. The AR(2) model forecasts a decline of 0.8% in 2009:Q2 at the one-quarter ahead horizon, and only a slow, positive growth at the two-quarter ahead horizon during the worst quarters of the recession.

Figure 3.3 plots the real-time forecasts for the MS model. As the other univariate models, the MS model tracks closely future GDP growth during expansions, but it also does better in forecasting its fluctuations as well. For both horizons, the MS model forecasts a deeper fall in GDP growth during recessions compared to the benchmark and CDR models (over −1.4% for the recent recession).

Figures 3.4 to 3.9 show the forecasts for the VAR, BVAR, and VAR-Fin models. These models have good forecast performance during expansions as well. As discussed earlier, their RMSFE is slightly worse for $k = 1$ than the benchmark, but for $k = 2$, their forecasts are not statistically significant different from the benchmark, the MS, and the DSGE models. The VAR and BVAR models perform better for the two-quarter ahead horizon.

Regarding recessions, the baseline VAR, all VAR-Fin, and the BVAR models basically miss the 2001 recession, forecasting an average growth during this period. However, their performance is very different for the most recent recession. The baseline VAR and the BVAR forecast a mild negative growth, but with the wrong timing. Both forecast negative growth one quarter after the end of the recession in 2009:Q2, although the baseline VAR also forecasts a mild fall in 2009:Q1 too.

The VAR-Fin models include variables that, on hindsight, were closely associated to the financial crisis in 2008. On effect, the inclusion of the default risk Aaa-Baa, Baa-T10 or Aaa-T10 in these models lead to forecast of a deep decline in GDP growth, but after the worst quarter of the crisis in 2008:Q4. The models correctly forecast a steep fall in GDP growth in 2009:Q1 and the strong recovery in 2009:Q4. Note that the inclusion of the term spread T10-T5 does not lead to a forecast of a dramatic fall in GDP during this period.

Figure 3.10 plots the real-time forecasts from the DSGE model. The model has reasonable forecast accuracy overall, although it presents the highest forecast bias compared to the other models – the DSGE has a bias proportion of 11% while all other models have this proportion equal or below 1%. Nevertheless, it tracks oscillations in GDP growth better than most models. The DSGE model also has a reasonable forecast accuracy during recessions. As most models, it does not forecast negative growth during the 2001 recession, and the severity of this recession is less than forecasted by all but the BVAR, VAR, and VAR-Fin models. With respect to the 2007–2009 recession, at the one-quarter ahead horizon the DSGE model forecasts a stronger negative growth in GDP during the 2007–2009 recession than the univariate models, but not as intense as forecasted by the other multivariate models. For $k = 2$, however, the DSGE model completely miss the recession, predicting only a mild but positive decline in GDP growth during the worst part of the financial crisis.

Figure 3.11a shows the actual and output growth forecasts from the AR-DFMS model. The model forecasts display strong oscillations, following closely GDP growth overall, particularly during recessions. The AR-DFMS is the only model that forecasts the 2001 recession, with forecasts of negative growth matching the actual GDP data. This model also has a good performance in forecasting the 2007–2011 recession. It displays the best forecast of the timing and depth of the decline in GDP growth during this period at one and two-quarter ahead horizons. The forecasts start decreasing around the beginning of the recession, reach a trough with strong negative growth around the time of the financial crisis, and increase at around the time the recession ended.

The accuracy of the AR-DFMS forecasts is closely associated with the dynamics of the probabilities of recession and the Business Cycle Indicator obtained from this model (Figure 3.11b). The Business Cycle Indicator is highly correlated with GDP fluctuations, matching well its volatility particularly during recessions. The probabilities of recession closely match NBER expansion and recession phases. During periods in which the NBER classifies as expansions the probabilities of recession are close to zero. At around the time when the NBER recession starts, the probabilities of recessions rise substantially and remain high until around the end of the recessions as established by the NBER.[41]

For example, the model signaled in real time the onset of the Great Recession as December 2007 with information available in April 2008. The earliest possible signal, given the lag in the availability of the data, would have been in March 2008. The real-time probabilities of recession were above 50% already in April 2008, and above 80% in July 2008. The probabilities stayed close to 100% during the whole financial crisis and the most of 2009, correctly signaling the intensity of the recession. Notice, however, that for the period studied the trough dates from the model take place later compared to the NBER troughs. The model captures the "jobless recoveries" that have followed recent recessions. The accuracy of the forecasts of GDP growth from the AR-DFMS model is related to the ability of the model to forecast recessions in real time, as reflected in the probabilities of recessions and the Business Cycle Indicator shown in Figure 3.11b.

The Blue Chip indicator has a good accuracy for $k = 1$ and $k = 2$, outperforming the models except for the AR-DFMS, as discussed above. The BC indicator, as AR-DFMS model, forecasts negative growth at the one-quarter ahead horizon during the 2001 recession and 2007–2009 recession. However, the timing and intensity of the decline differ across forecast horizons to a large extent. For the 2007–2009 recession, the BC forecasts at the one-quarter horizon negative growth during and after the financial crisis, but the forecast depth is almost half of the actual decline in economic growth (Figure 3.12). At the two-quarter ahead horizon, the BC almost misses the recession – it forecasts decreasing but positive growth until 2008:Q4. When the Lehman Brothers failed, the forecasts were updated, and GDP growth was forecasted to be mildly negative in 2009:Q1 and 2009:Q2. At the two-quarter ahead horizon, the forecasting accuracy gets noticeably worse with the BC almost missing completely the recession. These results are in line with those of Edge and Gürkaynak (2010), Wieland and Wolters (2011), and Del Negro and Schorfheide

[41] Notice that this model is devised to timely signal turning points, not to forecast them. The model only includes coincident series, not leading economic variables. For an extension of this model, which uses nonlinear two dynamic factors of the yield curve and economic activity to forecast recessions, see Chauvet and Senyuz (2012).

(2013), who find that the forecasts from structural models and from the professional forecasters underpredicted recessions.

Finally, Figure 3.13 plots the forecast combination of all models and the BC forecasts. The pooling tracks relatively well output growth during expansions, but perform poorly during recessions. In particular, it misses the timing and intensity of recessions. This is partially explained by the fact that the pooling does not track well the volatility of GDP growth, which is particularly accentuated during recessions.

## 4. CONCLUSION

This chapter examines the real-time forecast accuracy of structural models and state of art reduced-form linear and non-linear time series models for U.S. output growth over time and across business cycle phases. We reproduce the forecast problem at each date that the forecast were being made in real time in the last two decades. We find that, for all models, recessions are a lot harder to forecast than expansions.

The best models for tracking future GDP growth during expansions are the AR(2) and the univariate Markov switching MS model, and the VAR model that includes the term spread. The latter and the BVAR do better than the DSGE model during expansions, but the gains are offset by their poor performance during recessions. Pooling forecasts lead to a good precision for expansions, but not as well for recessions.

In fact, we find that the accuracy of most models is relatively poor for recessions. Although the forecast ability of some models is good during expansions, most fail to forecast GDP growth during recessions. The DSGE model performance is similar to the benchmark AR(2) during recessions, but both do poorly during these periods. For recessions, the best forecast accuracy is for the non-linear multivariate AR-DFMS model and the BC forecasts. Even though the professional forecasters have information advantage over all models, the AR-DFMS model has better forecasting performance, as it is particularly advantageous by design for periods of sharp changes, such as during recessions and financial crises. The accuracy of the AR-DFMS forecasts is closely associated with the dynamics of the probabilities of recession and the business cycle indicator obtained from this model. The business cycle indicator is highly correlated with GDP fluctuations, matching well its volatility particularly during recessions. The probabilities of recession rise substantially at the beginning of recessions and remain high until around their end, as dated by the NBER. The accuracy of GDP growth forecasts from the AR-DFMS model is, thus, closely related to the ability of the model to forecast recessions in real time.

These findings imply that by using simple univariate linear autoregressive models of GDP growth, one would have gotten in real time as good as forecasts during expansions than any other model and the professional forecasters. These simple models exhibit a very good ability to track the mean of GDP growth during normal times. Although DSGE models do not score high in forecasting ability, they still appeal to policymakers as story telling tools for policy evaluation in-sample.

However, some models that display good forecast ability during normal times are not as good during periods of sharp changes, as they do not process information quickly. We find that there are large gains in using different models for recessions. Structural models and VARs are more suitable for forecasts during normal periods, although simple univariate autoregressive models do just as well. On the other hand, by using models designed to handle abrupt changes and non-linearities, such as the multivariate Markov switching model, economic agents and policymakers can hedge against those changes and obtain more reliable forecasts at times in which they are mostly needed.

## ACKNOWLEDGMENTS

Ging Cee Ng (FRB New York), Raiden Hasegawa (FRB New York), and Daniel Herbst (FRB New York) provided excellent research assistance. We thank five anonymous referees, Allan Timmerman, Graham Elliott, Frederick Joutz, Timothy Cogley, James Morley, Mi Lu, and participants of the 2012 Conference in honor to Charles Nelson at the University of Washington, and of the 87th Conference of the Western Economic Association International in San Francisco for helpful comments and suggestions. We also thank Marco Del Negro and Frank Schorfheide for sharing some of the real-time data and codes used in the paper and allowing to cross-check our results. The views expressed in this chapter do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System.

## REFERENCES

Adolfson, M., Linde, J., Villani, M., 2007a. Forecasting performance of an open economy dynamic stochastic general equilibrium model. Econometric Reviews 26 (2–4), 289–328.

Adolfson, M., Andersson, M., Linde, J., Villani, M., Vredin, A., 2007b. Modern forecasting models in action: improving macroeconomic analyses at central banks. International Journal of Central Banking 3, 111–144.

Aiolfi, M., Capistran, C., Timmermann, A., 2011. Forecast combinations. In: Clements, M., Hendry, D. (Eds.), The Oxford Handbook of Economic Forecasting. Oxford, 355–388.

Anderson, H.M., Vahid, F., 2001. Estimating the probability of a recession with nonlinear autoregressive leading indicator models. Macroeconomic Dynamics 5 (4), 482–505.

Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L., Rünstler, G., 2011. Short-term forecasts of euro area GDP growth. Econometrics Journal, Royal Economic Society 14 (1), C25–C44.

Aruoba, S.B., Diebold, F.X., 2010. Real-time macroeconomic monitoring: real activity, inflation, and interactions. American Economic Review 100, 20–24.

Aruoba, S.B., Diebold, F.X., Scotti, C., 2009. Real-time measurement of business conditions. Journal of Business and Economic Statistics 27, 417–427.

Assenmacher-Wesche, K., Pesaran, M., 2008. Forecasting the Swiss Economy Using VECX* Models: An Exercise in Forecast Combination Across Models and Observation Windows. Working Papers 2008–03, Swiss National Bank.

Banbura, M., Runstler, G., 2011. A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP. International Journal of Forecasting 27 (2), 333–346.

Banbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian VARs. Journal of Applied Econometrics 25 (1), 71–92.

Banegas, A., 2011. Predictability of Growth in Emerging Markets: Information in Financial Aggregates, Federal Reserve Board. Available at SSRN: <http://ssrn.com/abstract=1966076>or http://dx.doi.org/10.2139/ssrn.1966076.

Beaudry, P., Koop, G., 1993. Do recessions permanently change output? Journal of Monetary Economics 31, 149–163.

Bernard, H., Gerlach, S., 1998. Does the term structure predict recessions? The international evidence. International Journal of Finance and Economics 3 (3), 195–215.

Blanchard, O., Simon, J., 2000. The long and large decline in US output volatility. Brookings Papers on Economic Activity 1 (2001), 135–74.

Burns, Arthur F., Mitchell, Wesley C., 1946. Measuring Business Cycles. National Bureau of Economic Research, New York.

Camacho, M., Perez-Quiros, G., Poncela, P., 2011. Green Shoots in the Euro Area. A Real Time Measure. Working paper 1026, the Central Bank of Spain.

Camacho, M., Perez-Quiros, G., Poncela, P., 2012. Markov Switching Dynamic Factor Models in Real Time. Working Paper 1205, Bank of Spain.

Carriero, A., Clark, T., Marcellino, M., 2011. Bayesian VARs: Specification Choices and Forecasting Performance. CEPR WP 8273.

Carriero, A., Clark, T., Marcellino, M., 2012. Common Drifting Volatility in Large Bayesian VARs. CEPR WP 8894.

Chauvet, M., 1998. An econometric characterization of business cycle dynamics with factor structure and regime switches. International Economic Review 39 (4), 969–96.

Chauvet, M., 2001. A monthly indicator of Brazilian GDP. The Brazilian Review of Econometrics 21 (1), 1–15.

Chauvet, M., 2010. The Four Phases of the US Business Cycle. Working Paper. University of California Riverside.

Chauvet, M., Guo, J.T., 2003. Sunspots, animal spirits, and economic fluctuations. Macroeconomic Dynamics 7 (1), 77–103.

Chauvet, M., Hamilton, J., 2006. Dating business cycle turning points in real time. In: Van Dijk, Milas, Rothman (Eds.), Nonlinear Time Series Analysis of Business Cycles. Elsevier's Contributions to Economic Analysis series, pp. 1–54.

Chauvet, M., Piger, J., 2008. A comparison of the real-time performance of business cycle dating methods. Journal of Business Economics and Statistics 26 (1), 42–49.

Chauvet, M., Piger, J., 2013. Employment and the business cycle. Manchester School.

Chauvet, M., Popli, G., 2003. Maturing capitalism and stabilization: international evidence. Journal of Business and Economics 1 (12), 5–22.

Chauvet, M., Potter, S., 2001. Recent changes in the US business cycle. Manchester School 69 (5), 481–508.

Chauvet, M., Potter, S., 2002. Predicting recessions: evidence from the yield curve in the presence of structural breaks. Economics Letters 77 (2), 245–253.

Chauvet, M., Potter, S., 2005. Forecasting recessions using the yield curve. Journal of Forecasting 24 (2), 77–103.

Chauvet, M., Lu, M., Potter, S., 2013. Forecasting Output During Recessions and Crises. University of California Riverside. Working Paper.

Chauvet, M., Senyuz, Z., 2012. A Dynamic Factor Model of the Yield Curve as a Predictor of the Economy, Staff Working Papers in the Finance and Economics Discussion Series (FEDS), 2012-32, 1–45.

Chauvet, M., Su, Y., 2013. Robust Markov switching models. In: Wohar, M., Ma, J. (Eds.), Recent Advances in Estimating Nonlinear Models. Springer.

Chauvet, M., Tierney, H.L., 2009. Real time Changes in Monetary Policy. University of California Riverside, Working Paper.

Chib, S., 1998. Estimation and comparison of multiple change-point models. Journal of Econometrics 86 (22), 221–241.

Christiano, L.J., Eichenbaum, M., Evans, C.L., 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of Political Economy 113 (1), 1–45.

Christoffel, K., Coenen, G., Warne, A., 2008. The New Area-Wide Model of the Euro Area: Specification. Estimation Results and Properties. European Central Bank Working Paper, October 944.

Christoffel, K., Coenen, G., Warne, A., 2011. Forecasting with DSGE models. In: Clements, M., Hendry, D. (Eds.), Oxford Handbook on Economic Forecasting. Oxford University Press, pp. 89–128.

Clark, T., 2011. Real-time density forecasts from BVARs with stochastic volatility. Journal of Business and Economic Statistics 29, 327–341.

Clark, T.E., McCracken, M.W., 2010. Averaging forecasts from VARs with uncertain instabilities. Journal of Applied Econometrics 25 (1), 5–29.
Clements, M.P., Galvao, A., 2008. Macroeconomic forecasting with mixed frequency data: forecasting output growth in the United States. Journal of Business and Economic Statistics 26, 546–554.
Clements, M.P., Hendry, D.F., 2004. Pooling of forecasts. Econometrics Journal, Royal Economic Society 7 (1), 1–31.
Cogley, T., Sargent, T.J., 2001. Evolving post World War II US inflation dynamics. NBER Macroeconomics Annual 16, 331–373.
Cogley, T., Sargent, T.J., 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII US. Review of Economic Dynamics 8, 262–302.
Cooley, T.F., LeRoy, S., 1985. Atheoretical macroeconometrics. A critique. Journal of Monetary Economics 16, 283–308.
Croushore, D., Stark, T., 2001. A real-time data set for macroeconomists. Journal of Econometrics 105, 111–30.
Del Negro, M., Schorfheide, F., 2011. Bayesian macroeconometrics. The Oxford Handbook of Bayesian Econometrics, pp. 293–389.
Del Negro, M., Schorfheide, F., 2013. DSGE model-based forecasting. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 2. Elsevier.
Diebold, F.X., Mariano, R., 1995. Comparing Predictive Accuracy. Journal of Business and Economic Statistics 13, 253–263.
Diebold, F.X., Rudebusch, G.D., 1996. Measuring business cycles: a modern perspective. The Review of Economics and Statistics 78 (1), 67–77.
Edge, R., Gürkaynak, R., 2010. How useful are estimated DSGE model forecasts for central bankers? Brookings Papers on Economic Activity, 209–259 (Fall).
Edge, R.M., Kiley, M.T., Laforte, J.P., 2010. A comparison of forecast performance between federal reserve staff forecasts, simple forecasts, simple reduced-form models, and a DSGE model. Journal of Applied Econometrics 25, 720–754.
Estrella, A., Hardouvelis, G.A., 1991. The term structure as a predictor of real economic activity. Journal of Finance 46 (2), 555–576.
Estrella, A., Mishkin, F.S., 1998. Predicting US recessions: financial variables as leading indicators. Review of Economics and Statistics 80 (1), 45–61.
Evans, M.D.D., 2005. Where are we now? Real-time estimates of the macroeconomy. International Journal of Central Banking 1 (2), 129–175.
Fair, R.C., 1970. The estimation of simultaneous equation models with lagged endogenous variables and first order serially correlated errors. Econometrica 38 (3), 507–516.
Faust, J., 2012. DSGE models: I smell a rat (and it smells good). International Journal of Central Banking 8 (1), 53–64.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2003. Do financial variables help forecasting inflation and real activity in the euro area? Journal of Monetary Economics 50, 1243–1255.
Frale, C., Marcellino, M., Mazzi, G., Proietti, T., 2011. EUROMIND: a monthly indicator of the euro area economic conditions. Journal of the Royal Statistical Society Series A 174, 439–470.
Ghysels, E., Santa-Clara, P., Valkanov, R., 2004. The MIDAS Touch: Mixed Data Sampling Regression Models. CIRANO Working Papers 2004s–20. CIRANO, Montreal, Canada.
Ghysels, E., Sinko, A., Valkanov, R., 2006. MIDAS regressions: further results and new directions. Econometric Reviews 26, 53–90.
Giannone, D., Reichlin, L., 2013. Nowcasting. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 2. Elsevier.
Giannone, D., Reichlin, L., Sala, L., 2004. Monetary policy in real time. NBER Macroeconomics Annual 19, 161–224.
Giannone, D., Reichlin, L., Small, D., 2008. Nowcasting: the real-time informational content of macroeconomic data. Journal of Monetary Economics 55 (4), 665–676.
Granger, C.W.J., Teräsvirta, T., 1993. Modelling Nonlinear Economic Relationships. Oxford University Press, Oxford.
Guerin, P., Marcellino, M., 2011. Markov Switching MIDAS Models, CEPR WP 8234.

Hamilton, J.H., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–84.
Hamilton, J.H. (1994), State-space models. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North-Holland.
Kauppi, H., Saikkonen, P., 2008. Predicting US recessions with dynamic binary response models. Review of Economics and Statistics 90, 777–791.
Kim, C.J., Nelson, C.R., 1998. Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. The Review of Economics and Statistics 80 (2), 188–201 (MIT Press).
Kim, Chang-Jin, Nelson, Charles R., 1999. Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of business cycle. Review of Economics and Statistics 81 (4), 608–616.
Kim, C.J., Piger, J., Startz, R., 2008. Estimation of Markov regime-switching regression models with endogenous switching. Journal of Econometrics 143 (2), 263–273.
Klein, L.R., 1970. An Essay on Theory of Economic Prediction. Markham Publishing Company, Markham Economics Series, Chicago.
Kolasa, M., Rubaszek, M., Skrzypczynski, P., 2009. Putting the New Keynesian DSGE Model to the Real-Time Forecasting Test. European Central Bank Working Paper Series 1110, November.
Kolasa, M., Rubaszek, M., Skrzypczynski, P., 2012. Putting the new Keynesian DSGE model to the real-time forecasting test. Journal of Money, Credit and Banking 44 (7), 1301–1324.
Koop, G., Korobilis, D., 2010. Bayesian multivariate time series methods for empirical macroeconomics. Foundations and Trends in Econometrics 3 (4), 267–358.
Koop, G., Leon-Gonzalez, R., Strachan, R., 2009. On the evolution of monetary policy. Journal of Economic Dynamics and Control 33, 997–1017.
Krane, S.D., 2011. Professional forecasters' view of permanent and transitory shocks to GDP. American Economic Journal: Macroeconomics 3 (1), 184–211 (American Economic Association).
Kuzin, V., Marcellino, M., Schumacher, C., 2011. MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area. International Journal of Forecasting 27 (2), 529–542.
Kuzin, V., Marcellino, M., Schumacher, C., 2013. Pooling versus model selection for nowcasting GDP with many predictors: empirical evidence for six industrialized countries. Journal of Applied Econometrics 28 (3), 392–411.
Kydland, F.E., Prescott, E.C., 1982. Time to build and aggregate fluctuations. Econometrica 50 (6), 1345–1370.
Liu, G., Gupta, R., Schaling, E., 2009. A new-Keynesian DSGE model for forecasting the South African economy. Journal of Forecasting 28, 387–404.
Lombardi, M., Maier, P., 2011. Forecasting Economic Growth in the Euro Area during the Great Moderation and the Great Recession. European Central Bank Working Paper Series 1379, September.
Lucas, R., 1976. Econometric policy evaluation: a critique. Carnegie-Rochester Conference Series on Public Policy 1, 19–46.
Lundquist, K., Stekler, H.O., 2011. The Forecasting Performance of Business Economists During the Great Recession. Research Program on Forecasting. George Washington University Working Paper, pp. 2011–004.
Ma, J., Wohar, M. (Eds.), forthcoming. Recent Advances in Estimating Nonlinear Models. Springer Publishers.
Marcellino, M., 2008. A benchmark for models of growth and inflation. Journal of Forecasting 27, 305–340.
Marcellino, M., Schumacher, C., 2008. Factor-MIDAS for Now- and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP. CEPR Discussion Papers 6708, C.E.P.R Discussion Papers.
Marcellino, M., Stock, J., Watson, M., 2003. Macroeconomic forecasting in the euro area: country-specific versus area-wide information. European Economic Review 47, 1–18.
Mariano, R.S., Murasawa, Y., 2003. A new coincident index of business cycles based on monthly and quarterly series. Journal of Applied Econometrics 18 (427), 443.
Mazzi, G., Ozyildirim, A. (Eds.), forthcoming. Handbook on Cyclical Composite Indicators, UN/Eurostat/ Statistic Netherlands.

McConnell, M.M., Perez-Quiros, G., 2000. Output fluctuations in the United States: what has changed since the early 1980s. The American Economic Review 90 (5), 1464–1476.
Mills, Terence C., Ping, Wang, 2003. Have output growth rates stabilised? Evidence from the G-7 economies. Scottish Journal of Political Economy 50 (3), 232–46.
Morley, J., Piger, J., Tien, P.L., forthcoming. Reproducing business cycle features: are nonlinear dynamics a proxy for multivariate information? Studies in Nonlinear Dynamics and Econometrics.
Nalewaik, J.J., 2011. Forecasting Recessions Using Stall Speeds. FEDS Working Paper 2011–24.
Nalewaik, J.J., 2012. Estimating probabilities of recession in real time using GDP and GDI. Journal of Money, Credit, and Banking 44, 235–253.
Nelson, C.R., 1972. The prediction performance of the FRB-MIT-PENN model of the US economy. American Economic Review 62, 902–917.
Nyberg, H., 2010. Dynamic probit models and financial variables in recession forecasting. Journal of Forecasting 29 (1–2), 215–230.
Oh, S., Waldman, M., 1990. The macroeconomic effects of false announcements. The Quarterly Journal of Economics 105 (4), 1017–34.
Owyang, M., Piger, J.M., Wall, H.J., 2012. Forecasting National Recessions Using State Level Data. Working Paper 2012–013A, Federal Reserve Bank of Saint Louis.
Pesaran, M.H., Timmermann, A., 1992. A simple nonparametric test of predictive performance. Journal of Business and Economic Statistics 10 (4), 461–465.
Primiceri, G., 2005. Time varying structural vector autoregressions and monetary policy. The Review of Economic Studies 72, 821–852.
Proietti, T., Moauro, F., 2006. Dynamic factor analysis with non linear temporal aggregation constraints. Applied Statistics 55, 281–300.
Rotemberg, J., Woodford, M., 1997. An optimization-based econometric framework for the evaluation of monetary policy. NBER Macroeconomics Annual 12, 297–346.
Rubaszek, M., Skrzypczyński, P., 2007. Can a Simple DSGE Model Outperform Professional Forecasters? National Bank of Poland Working Paper 43.
Rubaszek, M., Skrzypczyński, P., 2008. On the forecasting performance of a small-scale DSGE model. International Journal of Forecasting 24 (3), 498–512.
Schorfheide, F., Song, D.S., 2012. Real-Time Forecasting with a Mixed-Frequency VAR. Working Papers 701. Federal Reserve Bank of Minneapolis.
Sims, C., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.
Sims, C., 2006. Comment on Del Negro, Schorfheide, Smets and Wouters. <http://sims.princeton.edu/yftp/DSSW806/DSseattleComment.pdf>.
Sims, C.A., Zha, T., 2006. Were there regime switches in US monetary policy? American Economic Review 96, 54–81.
Smets, F., Wouters, R., 2003. An estimated dynamic stochastic general equilibrium model of the euro area. Journal of the European Economic Association 1 (5), 1123–1175.
Smets, F., Wouters, R., 2007. Shocks and frictions in US business cycles: a Bayesian DSGE approach. American Economic Review 97 (3), 586–606.
Stock, J.H., Watson, M.W., 1989. New indexes of coincident and leading economic indicators. In: Blanchard, O., Fischer, S. (Eds.), NBER Macroeconomics Annual. MIT Press, Cambridge.
Stock, J.H., Watson, M.W., 1993. A procedure for predicting recessions with leading indicators: econometric issues and recent experience. In: Stock, J.H., Watson, M.W. (Eds.), Business Cycles, Indicators and Forecasting. University of Chicago Press for NBER, Chicago, pp. 255–284.
Stock, J., Watson, M., 2002. Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20, 147–162.
Stock, J.H., Watson, M.W., 2003. Forecasting output and inflation: the role of asset prices. Journal of Economic Literature 41 (3), 788–829.
Tashman, L.J., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting 16 (4), 437–450.
Timmermann, A., 2006. Forecast Combinations. Elsevier, Handbook of Economic Forecasting.
Tinbergen, J., 1939. Business Cycles in the United States, 1919–1932 – Statistical Testing of Business-Cycle Theories. League of Nations, Economic Intelligence Service, Geneva.

Tinbergen, J., 1974. The Dynamics of Business Cycles: A Study in Economic Fluctuations. University of Chicago Press.
Tong, H., 1990. Non-Linear Time Series: A Dynamical System Approach. Oxford University Press.
Tovar, C., 2009. DSGE models and central banks, economics: the open-access. Open-Assessment E-Journal 3, 16.
Trehan, B., 1989. Forecasting Growth in Current Quarter Real GNP. Federal Reserve Bank of San Francisco Economic Review, 39–51 (Winter).
van Dijk, D., Osborn, D., Sensier, M., 2002a, Changes in the Variability of the Business Cycle in the G7 Countries. CGBCR Discussion Paper 16, University of Manchester, September.
van Dijk, D., Terasvirta, T., Franses, P.H., 2002b. Smooth transition autoregressive models—a survey of recent developments. Econometric Reviews 21 (1), 1–47.
Wang, M.C., 2009. Comparing the DSGE model with the factor model: an out-of-sample forecasting experiment. Journal of Forecasting 28–2, 167–182.
Wieland, V., Wolters, M.H., 2011. The diversity of forecasts from macroeconomic models of the US economy. Economic Theory 47, 247–292.
Winter, J., 2011. Forecasting GDP Growth in Times of Crisis: Private Sector Forecasts versus Statistical Models. De Nederlandsche Bank NV Working Paper 320, November.
Wolters, M., 2010. Forecasting Under Model Uncertainty. Goethe University Frankfurt Working Paper.