# Forecasting in Nonstationary Environments:
# What Works and What Doesn't
# in Reduced-Form and Structural Models

Raffaella Giacomini and Barbara Rossi

($UCL\ and\ CeMMAP$)　　($ICREA - Univ.\ Pompeu\ Fabra,$
$Barcelona\ GSE\ and\ CREI$)

December 4, 2014

### Abstract

This review provides an overview of forecasting methods that can help researchers forecast in the presence of non-stationarities caused by instabilities. The emphasis of the review is both theoretical and applied, and provides several examples of interest to economists. We show that modeling instabilities can help, but it depends on how they are modeled. We also show how to robustify a model against instabilities.

**Keywords:** Forecasting, instabilities, structural breaks.

# 1    Introduction and Motivation

This article surveys recent developments in the estimation of forecasting models in the presence of non-stationarities caused by instabilities.[1] Instabilities are widespread in economic time series. A clear example of instabilities in macroeconomic data is the sharp reduction in the volatility of many macroeconomic aggregates around mid-1980s, a phenomenon known as "the Great Moderation", and documented by Kim & Nelson (1999) and McConnell & Perez-Quiros (2000). The Great Moderation was followed two decades later by a large decline in the growth rate of overall economic activity starting in 2007 and lasting until 2009, an especially severe financial crisis referred to as "the Great Recession". Additional examples are occasional changes in policy that may lead to changes in the transmission mechanism in the economy; examples of drastic changes in monetary policy include "the Volker disinflation" (Clarida et al. 2000) and the "Zero Lower Bound". Thus, time series are subject to occasional sharp changes in the mean growth rate of macroeconomic variables (as in the Great Recession), abrupt changes in the volatility (as in the Great Moderation) and sudden changes in the co-movements among macroeconomic variables (as when monetary policy or its transmission mechanism changes). Typically, models with fixed parameters display structural breaks and their forecasting performance changes over time and predictors that perform well at some times do not perform well at other times – see the survey chapter by Rossi (2013).

To understand why instabilities are important for forecasting, consider two motivating examples: forecasting the Great Recession, as well as forecasting inflation. Inflation is one of the key economic variables that central banks are typically concerned about. The Federal Reserve Board makes publicly available its own forecast of inflation (with a few years' lag); alternative forecasts of inflation are available via surveys collected from professional forecasters; one example is the Survey of Professional Forecasters (SPF) at the Philadelphia Fed. Figure 1 plots the relative forecasting performance of the Federal Reserve's forecasts and that of the SPF. More in detail, the figure plots the relative Mean Square Forecast Errors (MSFEs) of the two forecasts[2] over time in rolling windows (rescaled by a measure of their variability).[3] The dates in the figure indicate the mid-point in the rolling window average. Negative values indicate that the Federal Reserve Board forecasts are more precise than the SPF's; this is the case for most of the 1970s and 1980s; however, in the 1990s and 2000s, the survey forecasts became more competitive, to the point of becoming more accurate than the former. The figure illustrates how the relative performance of inflation forecasting models may change over time. This is a common phenomenon in macroeconomic data. For example, Ng & Wright (2013) find that the interest rate spread was a good predictor for output growth in the

---

[1]In the following, we will thus be using the two terms "nonstationarities" and "instabilities" interchangeably.

[2]Clearly, improvements and deteriorations in forecasting ability may depend on the choice of the loss function used to evaluate the forecsts.

[3]This is a test statistic developed by Giacomini & Rossi (2010).

US in the early 1980s but lost its predictive power since then,[4] and credit spreads became a more useful predictor, especially during the latest financial crisis. As they argue, this finding hints to the fact the Great Recession was very different from any other recession in the US: it is widely believed it was related to financial market problems rather than supply, demand or monetary policy shocks. Ng & Wright (2013, p. 1120) note that many economic data exhibit mean shifts, parameter instabilities, and stochastic volatility, which lead them to document the evolution of the properties of macroeconomic time series in post World War II data. Other examples for macroeconomic data are discussed in Stock & Watson (1996) and Rossi & Sekhposyan (2010).

As another example, consider a structural macroeconomic model such as a Dynamic Stochastic General Equilibrium (DSGE) model. Gurkaynak et al. (2013) have studied the forecasting ability of DSGEs relative to that of several reduced-form models, among others. Figure 2 (taken from Gurkaynak et al. 2013) plots rolling windows of (standardized) MSFEs of inflation forecasts of a representative DSGE model versus those of a VAR; values close to zero indicate that the two models have the same predictive ability over that period of time. Clearly, the representative DSGE is outperformed by the VAR in the last part of the sample, although it performed better in the first part.

This review provides an overview of forecasting methods that can help researchers forecast in the presence of instabilities. The emphasis of the review is both theoretical and applied, and provides several examples of interest to economists.

A few remarks are in order. In a recent survey, Rossi (2013) reviewed the literature on forecasting under instabilities; she focused on reduced-form models, in-sample Granger-causality and out-of-sample forecast evaluation in the presence of instabilities (e.g. how to perform forecast comparison or forecast rationality tests in the presence of instabilities). She also reviewed traditional model estimation methods that can be used in the presence of instabilities. We refer to her survey for these topics. In this survey, we focus instead on more recent developments in the methods for forecasting in the presence of instabilities. These include not only methods that explicitly model instabilities, but also methods that could be interpreted as robustifying devices against instabilities, such as the use of large dimensional data sets, model combination and incorporation of survey data into a model. We further include a thorough discussion of whether structural models could guard against instabilities. An important extension relative to existing surveys is that we move beyond point forecasting and discuss the role of instabilities in density forecasting. Furthermore, we focus on models and methodologies designed for frequencies of interest to macroeconomists (monthly or quarterly) rather than high frequency methods used in finance, although we discuss models with time-varying variances as they are of interest to macroeconomists.

---

[4]See also Giacomini & Rossi (2006) on the forecasting performance of the yield curve for output growth.

# 2    Methods Designed to Protect Against Instabilities in Reduced-Form Models

The examples in the previous section illustrate that rarely the best forecasting model is *the same model over time*; it is more common to observe that the best forecasting model *changes over time*. Economists have developed several techniques that help them in identifying instabilities and incorporate them into their models to improve both macroeconomic forecasting and the design of timely signals of turning points such as recessions.[5]

In the attempt of exploiting instability to improve models' forecasts, researchers have developed several tools, which we review.

## 2.1    Break Testing and Estimation with Post-break Data

As instabilities appear typically as changes in the parameters of the model, a common approach is to try to incorporate the parameter instability into the model. This section discusses several different ways to model parameter instabilities that have been considered in the literature.

The first example is to first test for the presence of structural breaks in parameters: if a break is found, then the model is estimated using dummy variables to take into account the existence of the break. For example, a simple autoregressive model with one lag and breaks in the parameters would be:

$$y_t = \alpha_t + \rho_t y_{t-1} + \varepsilon_t, \tag{1}$$

where $\varepsilon_t \sim N\left(0, \sigma_t\right)$, and:

$$
\begin{aligned}
\alpha_t &= \alpha_1 + \alpha_2 \cdot 1\left(t > \tau_\alpha\right), \\
\rho_t &= \rho_1 + \rho_2 \cdot 1\left(t > \tau_\rho\right), \\
\sigma_t &= \sigma_1 + \sigma_2 \cdot 1\left(t > \tau_\sigma\right),
\end{aligned}
\tag{2}
$$

where $\tau's$ indicate the time of the parameter change and $t = 1, 2, ...T$, $T$ being the total sample size. For example, McConnell & Perez-Quiros (2000) estimate a simplified version of eq. (1) with only a constant for the growth rate of real Gross Domestic Product (GDP) in the US between 1953:2 and 1999:2. They found that the mean growth rate of GDP $(\alpha_t)$ was constant but the variance $(\sigma_t)$ was not, and that the latter changed in 1984 (that is, $\tau_\sigma = 1984$).

Parameter instability of the kind described by eq. (2) is typically detected via structural break tests (e.g. Brown et al. 1975; Ploberger et al. 1989; Nyblom 1989; Ploberger & Krämer 1992; Andrews 1993; Andrews & Ploberger 1994; Bai & Perron 1998; and Elliott & Muller 2006, among others). Brown et al.'s (1972) test was for example used by McConnell & Perez-Quiros (2000).

---

[5] Alternative modeling methodologies involve non-linear methods (e.g. threshold models).

The reason why there are many such tests is because each one of them is designed to detect a particular type of instability. For example, Andrews (1993) considers a one-time discrete shift in the parameters, similar to the one described in eq. (2); Andrews & Ploberger (1994) consider a finite number of breaks independent of the sample size; Bai & Perron (1998) consider multiple breaks that are discrete in nature (e.g. $\alpha_t = \alpha_1 + \alpha_2 1\,(t > \tau_\alpha) + ... + \alpha_K 1\,(t > \tau_K)$); Nyblom (1989) considers breaks that follow a martingale ($\alpha_t = \alpha_{t-1} + \eta_{\alpha,t}$, where $\eta_{\alpha,t}$ is a idiosyncratic disturbance term); similarly, Elliott & Muller (2006) consider multiple, persistent breaks that are well approximated by a Wiener process (that is, that resemble random walks).

## 2.2   Time-Varying Parameter Models

Models with breaks and/or time-varying parameters can be quite general. For example, they can be adapted to Vector Autoregressive (VAR) models, which are typical models used by macroeconomists to monitor the time evolution of economic aggregates as well as forecast them. VARs provide a parsimonious representation of the complex world by summarizing it via a system of few variables that are jointly interdependent over-time:

$$Y_t = A_1 Y_{t-1} + ...A_p Y_{t-p} + \ \Omega \varepsilon_t, \tag{3}$$

where $Y_t$ is a vector of $(k \times 1)$ explanatory variables and typically $k$ is small. Traditional VARs have two important limitations. One is the fact that they are only approximations to the complex reality, even though they provide its best linear approximation (conditional on the representative variables chosen by the researcher) and a computationally convenient methodology to study interactions among variables. The second is that their parameters are typically assumed to be constant, which prevents them from adapting to a changing world. VARs with structural breaks, time varying parameters and/or stochastic volatility were designed to adapt VARs to changing environments (see Stock & Watson 2003; Cogley & Sargent 2005; Primiceri 2005; Sims & Zha 2006; Inoue & Rossi 2011, among others). This is an important issue in practice, and often economists are interested in knowing which of the models' parameters are time-varying. In fact, parameters in the conditional mean typically describe the transmission mechanism while variances are associated with the volatility of the shocks. If it is the former that change over time, then economists conclude that either policies or agents' behavior have changed, while if the latter changed then economists infer that the changes were induced by exogenous forces.

One possibility is to model the parameter path as a random walk, as in Cogley & Sargent (2005), Primiceri (2005), Stock & Watson (2003). Cogley & Sargent (2005) consider the model:

$$\begin{aligned} Y_t &= A_{1,t} Y_{t-1} + ...A_{p,t} Y_{t-p} + \ \Omega_t \varepsilon_t, \\ \Omega_t &= A_0^{-1} H_t A_0^{-1\prime}, \end{aligned} \tag{4}$$

where the coefficients $A_{j,t}$ and $H_t$ follow a random walk: $A_{j,t} = A_{j,t-1} + \eta_{A,t}$ and the $(i,j) - th$ element of the diagonal matrix $H_t$, $\sigma_{i,t}^2$, follows a random walk: $ln\left(\sigma_{i,t}^2\right) = ln\left(\sigma_{i,t-1}^2\right) + \sigma_i \eta_{i,t}$ (where $\eta_{i,t}$ is an idiosyncratic error term). Primiceri (2005) also allows $A_0$ to be time-varying to capture contemporaneous changes in the transmission mechanism.[6] Inoue & Rossi (2011) allow one-time discrete breaks as in eq. (2) in a subset of the parameters of the VAR model, eq. (4); the subset of the time-varying parameters is not known and can be estimated to identify the source of instabilities in the models. Pesaran et al. (2006) instead allow time-varying parameters derived from a meta-distribution. An alternative way to introduce time-varying parameters is via Markov switching models. For example, Sims & Zha (2006) consider Markov-switching to model changes in US monetary policy in VARs: $Y_t = A_{1,t}\left(s_t\right)Y_{t-1} + ...A_{p,t}\left(s_t\right)Y_{t-p} + \ \Omega\left(s_t\right)\varepsilon_t$, where $s_t$ is an unobserved state such that $Pr\left(s_t = b | s_{t-1} = a\right) = p_{ab}$ (that is, follows a Markov chain).

## 2.3   What Have We Learned On the Forecasting Performance of Reduced-form Models with Breaks and Time-varying Parameters?

Given the wide array of parameter instabilities that one could consider, which one should be used in practice? This is a difficult question, since it depends on the unknown process that instabilities follow. A comforting result comes from Elliott & Muller (2006), who show that most structural break tests perform similarly, independently of the exact process that instabilities follow; this is very useful for applied researchers, who can rely on their preferred test without losing much information in practice; however, the estimation, fit and forecasting performance of the model will be affected by how instabilities are modeled.

Unfortunately, detecting a break and then estimating a model that takes the break into account does not perform well in forecasting in practice. Part of the reason is that the break date is only an estimate, and that might be imprecise. Another reason is that, even if we knew the break date, there is estimation error in the size of the break (or parameter estimates); thus, a model with constant parameters may forecast more precisely than a model that correctly incorporates the instabilities via the break point. For example, when the forecasting ability is measured by the MSFE, this is due to the usual trade-off between bias and variance, as the MSFE equals the bias squared plus the variance. A mis-specified model would have a large bias, but the correctly specified model may have larger variance as parameters are estimated only in a subset of the data. To obviate this problem, Pesaran & Timmermann (2007) propose to use an estimation window that could include pre-break observations; rolling estimation may guards against instabilities, although the choice of the rolling window again will affects results (see Rossi, 2013, for a thorough discussion).[7]

---

[6]Both estimate the model with Bayesian methods. See also Muller & Petalas (2010) for a frequentist estimation of the same models.

[7]See also Pesaran & Timmermann (2002) for a reversed ordered CUSUM test to identify the time of the latest

Is it impossible then to forecast in the presence of breaks? The problem is that it is difficult to know how much to rely on past observations when estimating the models' parameters if breaks occur. In a recent paper, Pesaran et al. (2013) derive theoretical results on the optimal weight to assign to past observations in order to minimize MSFEs of one-step-ahead predictions, either when the break is large and discrete, or when the parameters are slowly changing. They propose to weigh observations such that older observations are down-weighted via an exponential function to allow the parameter to adapt to the changing nature of the true data generating process, and derive the optimal degree of down-weighting. When the breaks are discrete, the exponential function is used as an approximation, and the weights are step functions with constant weights within a regime but different weights between regimes. These results are theoretical, in the sense that they can be derived under a given parametric break process with known parameter values and known break sizes; when parameters are unknown, Pesaran et al. (2013) propose to construct robust weights that smooth over the uncertainty surrounding the dates and the size of the breaks.[8]

Regarding time-varying parameter models, the forecast accuracy of model (4) with time-varying parameters (including $A_{0,t}$) has been investigated by D'Agostino et al. (2013), who find that including stochastic volatility improves point forecast accuracy in a tri-variate VAR model with inflation, unemployment and the interest rate relative to fixed coefficient models in which the parameters are re-estimated either recursively or in rolling windows.

## 2.4 Large Dimensional Data as a Robustifying Device Against Non-stationarities

Given the empirical evidence that procedures that detect breaks may not be of much practical use to improve forecasts, coupled with the observation that models with few predictors typically do not forecast well either, the literature has turned to large dimensional data sets and models. The idea is that if models' relative performance changes over time, perhaps models that contain a large number of predictors or models that combine information across several of them might forecast better. In this section we review the literature on forecasting using large dimensional data which specifically deals with the issue of instabilities.

One such model considered in the literature is a factor model. Factor models collect information from a large dataset of predictors and express them conveniently in a small dimensional vector of "summary variables", called factors (see Forni et al. 2000, Stock & Watson 2002, Bai & Ng 2002, among others). A typical factor model is:

$$X_{i,t} = \lambda_i' F_t + e_{it}, \tag{5}$$

---

break that can be useful for forecasting.

[8]Giraitis et al. (2013) alternatively propose to use local averaging of past values of the variable in univariate models to approximate smooth weights over past observations.

where $X_{i,t}$ are the observable data, $i = 1, ..., N$, $t = 1, ..., T$, both $T$ and $N$ are large, $F_t = [F_{1,t}, ..., F_{r,t}]'$ are the unobserved components (i.e. the factors), $\lambda_i$ are the factor loadings and $r$ is small. The factors are then included as an additional explanatory variable in the model, for example an AR model (eq. 1, referred to as FAAR) or a VAR (eq. 3, referred to as a FAVAR).

Several recent developments in the literature have more directly investigated issues of instabilities in the context of factor models. The factor model, eq. 5, by averaging information across many predictors into a smaller number of factors, has the appealing feature of keeping all potentially relevant predictors in the set from which the factors are extracted. However, eq. (5) does not allow the factor loadings to change over time, whereas one would potentially give little weight to predictors at the times in which they do not forecast well and high weight at times in which they do. The idea is thus to include time variation in factor models, as in Banerjee et al. (2008), Stock & Watson (2009) and Korobilis (2013); also, Eickmeier et al. (2011) consider including time variation in FAVAR models.[9] For example, Stock & Watson (2009) consider the following model where the factor loadings are time-varying:

$$X_{i,t} = \lambda_{it}'F_t + e_{it}.$$

They find empirical instabilities in the factor loadings in a large sample of macroeconomic data series around the start of the Great Moderation sample; quite surprisingly, however, the estimated principal components are constant. One might be worried that factor model parameters might be estimated imprecisely unless the time variation in factor loadings is appropriately taken into account. Regarding this, on the one hand, Breitung & Eickmeier (2011) show that one-time structural breaks in factor loading effectively create new factors, so that ignoring instability leads to estimating too many factors; on the other hand, Bates et al. (2013) characterize the type and magnitude of parameter instability in factor loading that does not affect their consistent estimation; the intuition is that, sometimes, even large breaks in coefficients do not invalidate consistency since their effects are averaged across series, provided that these shifts have limited dependence across series.

Another possibility is to estimate large-dimensional VAR models with time-varying parameters. Large dimensional VARs with constant parameters (that is, VARs such as in eq. (3) where $k$ is large) have been considered by Banbura et al. (2010), Carriero et al. (2011), and Koop (2013), among others. Typically, since the number of the series is large and the dimension of macroeconomic datasets is small, Bayesian shrinkage is used to limit the effects of parameter proliferation; thus, these models are often referred to as Bayesian VARs. Some results are available to guide researchers in the constant parameter case.[10] While extensions that allow $k$ to be large are theoretically straightforward, the computational costs of including time-varying parameters and estimating eq.

---

[9]See Bernanke & Boivin (2003) for FAVAR models.

[10]See Carriero et al. (2011) and Giannone et al. (2012).

(4) with large $k$ are daunting, not to mention the fact that macroeconomists typically only have small samples and would face an over-parameterization problem. Koop & Korobilis (2013) suggest using forgetting factors to reduce the computational burden in the estimation; that is, they propose that the Kalman filter estimate of $\sigma^2_{i,t}$ given information at time $(t-1)$, $\sigma^2_{i,t|t-1}$, instead of being $\sigma^2_{i,t|t-1} = \sigma^2_{i,t-1|t-1} + q_t$ (like in the usual Kalman filter), be $\sigma^2_{i,t|t-1} = \lambda^{-1}\sigma^2_{i,t-1|t-1}$, where $\lambda \in (0,1]$ is the forgetting factor ($\lambda = 1$ being the case with constant coefficients).[11] Thus, the dimension of the VAR is allowed to change over time. To deal with the over-parameterization issue, they suggest shrinking.

Another technique to average information across many predictors is using forecast combinations or model averaging. Typical, forecast combinations are applied to a large dataset of predictors by averaging very simple models. For example, an approach is to estimate an Autoregressive distributed lag (ADL) model using lags of a predictor (e.g. $x_{t-1,k}$ denotes predictor "$k$", for $k = 1, 2, ..., K$) at a time in addition to the lagged dependent variable:

$$y_t = \alpha_k + \beta_k(L)x_{t-1,k} + \gamma_k(L)y_{t-1} + \varepsilon_t, \ t = 1, ..., T, \tag{6}$$

where $\beta_k(L) = \sum_{j=0}^{p} \beta_k^{(j)} L^j$ (and similarly for $\gamma_k(L)$) and $L$ is the lag operator. In practice, $p$ and $q$ are selected recursively by information criteria (BIC or AIC), potentially helping capturing instabilities in the usefulness of the predictors, first selecting the lag length for the autoregressive component only, then choosing the optimal lag length for the additional predictor. BIC provides a consistent estimate of the true number of lags while AIC does not; however, BIC penalizes model complexity more heavily. Let the forecast based on predictor $k$ be denoted by $y_{k,t+1|t}$; then the equal-weight forecast combination is $\frac{1}{K}\sum_{k=1}^{K} y_{k,t+1|t}$.

From an empirical point of view, factor models with constant factor loadings may perform well, particularly when forecasting measures of real activity (less so for inflation). Averaging information via surveys may sometimes be more successful, for example in forecasting inflation. Whether this is due to their ability to guard against instabilities is less clear.[12] Time-varying parameter models may forecast well; however, survey forecasts are typically among the best forecasters when predicting inflation (Faust & Wright 2013). Stock & Watson (2003) found that forecast combinations perform less erratically than individual forecasts, a finding emphasized in the survey by Timmermann (2006). Rossi (2013) considers comparing forecasts of several models (ADL models, equal weight forecast averaging, Bayesian model averaging (Wright 2009; Groen et al. 2013), factor models, and unobserved component models (Stock & Watson 2007), and finds that, typically, equal

---

[11] The latter can be seen by combining the latter two equations to get $q_t = (\lambda^{-1} - 1)\sigma^2_{i,t-1|t-1}$.

[12] One piece of evidence suggestive of this possibility is the finding (e.g., in Giacomini & White, 2006) that the accuracy improvements of factor forecasts over benchmarks seem to increase with the forecast horizon. Since the effects of instabilities are presumably stronger at long horizons, this finding could be evidence of the robustifying properties of factor models but whether this is the case needs to be investigated more formally.

weight forecast combinations perform the best in predicting inflation and output growth in the US. She also specifically examines time variation in the relative performance of factor models versus forecast combinations (BMA and equal weighting) and finds the latter perform better than an AR benchmark model at several points in time during the last three decades, while the performance of factors is less competitive.

## 2.5 Survey Forecasts as a Robustifying Device Against Non-stationarities

Finally, researchers have considered enlarging the forecasters' information set by including survey forecasts. The insight is that econometric models are limited in the kind and amount of data that they can consider; it is possible that forecasts might be improved by collecting and averaging information across a large set of professional forecasters, who routinely elaborate forecasts using models as well as expert judgement. Looking at Figure 1, which reports rolling averages of MSFE differences between Federal Reserve and SPF's survey forecasts, indeed it appears that the latter have improved over time. Thus, survey forecasts are typically included in the forecasters' model averaging procedures. In this section we discuss why and how survey and market-based forecasts could be incorporated into a forecasting model to improve its accuracy and potentially guard its performance against some types of non-stationarities.

A large literature has shown that forecasts from surveys of professional forecasters or market-based forecasts such as those extracted from futures contracts are difficult to beat benchmarks when forecasting key macroeconomic variables such as inflation, interest rates and output growth. Even though the accuracy of survey forecasts in absolute terms varies across variables and forecast horizons (e.g., survey forecasts of inflation are accurate at all horizons, as illustrated by Faust & Wright (2013), whereas Del Negro & Schorfheide (2012) show that for output growth the performance of surveys deteriorates rapidly as the horizon increases), a robust finding is that survey- and market participants appear to have a superior ability than models at incorporating into their forecasts the vast amount of information available to them about the current state of the economy. Moreover, notwithstanding possible frictions in the degree of attention or the different quality of information that forecasters may receive (e.g., Coibion & Gorodnichenko 2012), there are several examples showing that surveys incorporate new information that can be relevant for forecasting in a timely fashion. This point is well illustrated by the following picture from Altavilla et al. (2013), which shows the consensus forecast of the three-month yield from the Blue Chip Analysts survey relative to the forecast of the same variable implied by the popular Diebold & Li (2006) model (with the information sets for the model and the survey aligned). In Figure 3, the curves are the forecasts made at the beginning of every month between June 2011 and November 2011 and points on the curves are 1- to 4-quarters ahead forecasts. We chose this time period to illustrate the effect of the monetary policy announcement made by the FOMC on August 9th, 2011 which contained

an instance of "forward guidance" that has been recently used as an instrument of monetary policy by several monetary authorities around the world. In this specific instance the announcement declared that the committee anticipated "[...] exceptionally low levels for the federal funds rate at least through mid 2013".

Figure 3 shows that before the announcement both the survey and the model predicted interest rates to increase over the following year, but survey participants immediately incorporated the information that the monetary authority intended to keep interest rates low over the following year, whereas the model continued to predict an increase in interest rates that did not in fact materialize. This example illustrates how, as long as information about the current and future state of the economy is available, survey forecasts can incorporate this information in a timely manner and thus provide a forecast that is more accurate (and potentially robust to predictable changes, even though strictly speaking in the example the information was about keeping interest rates unchanged) than that implied by a model that is not able to incorporate such information in a timely manner.

There are two important caveats to discuss before drawing the implication from the above discussion that forecasting models are useless, and that one should solely rely on survey forecasts. First, survey forecasts are only available for some variables and at given frequencies, and thus one still needs a model to be able to forecast at other times and all other variables. Second, even when survey forecasts are available, they are not always more accurate than model-based forecasts. This point is well illustrated for Blue Chip Analysts forecasts of interest rates, which are more accurate than models at short forecast horizons but less accurate at long horizons (Chun 2008 and Altavilla et al. 2013). Rossi & Sekhposyan (2014), among others, have also found time-varying systematic biases in survey forecasts. We are thus left with two open questions: 1) when to use survey data and 2) how to best incorporate survey information into a model in order to obtain accuracy gains for all variables in the model? Most of the literature on incorporating survey data into models has ignored the first question. For example, Monti (2010) proposes a way to incorporate survey forecasts into a structural model (expressed in state-space form) but makes specific assumptions about how the survey forecasts are produced. Del Negro & Schorfheide (2012) also incorporate survey forecasts into Smets & Wouters' (2007) DSGE model but do not discuss how the survey forecasts were selected. To our knowledge, the only method that gives guidance on both how to choose which surveys to use and how to incorporate them into a forecast is the method proposed by Altavilla et al. (2013) which we illustrate in the next section.

## 2.6   When and How to Use Survey Data

This section illustrates how to incorporate survey forecasts into a model using the exponential tilting method illustrated by Altavilla et al. (2013), which builds on Robertson et al. (2005) and Giacomini

& Ragusa (2014). The method is generally applicable in that it can be applied for incorporating surveys into any given model-based multivariate density forecast (see Giacomini & Ragusa 2014 for the general approach). Here we illustrate the case in which the model-based density forecast is a multivariate normal distribution, as in this case the method reduces to a convenient analytical expression. The basic ingredients are an $h-$step ahead forecast made at time $t$ for a vector $y_{t+h}$, and survey based forecasts of the conditional mean of a subset of the components of $y_{t+h}$, which, without loss of generality, we collect in $y_{1t+h}$. The model-based forecast for $y_{t+h} = \binom{y_{1t+h}}{y_{2t+h}}$ is thus $\widehat{f}_t(y_{t+h}) \sim N(\widehat{\mu}_t, \Sigma) = N\left( \binom{\widehat{\mu}_{1t}}{\widehat{\mu}_{2t}}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$ and the survey forecast for $y_{1t+h}$ is $\widetilde{\mu}_{1t}^{survey}$. The method solves the constrained optimization problem of finding a new "tilted" density forecast $\widetilde{f}_t$ for the whole vector $y_{t+h}$ such that, out of all the densities that have conditional mean $\widetilde{\mu}_{1t}^{survey}$ for $y_{1t+h}$, it is the closest to the original model-based density forecast $\widehat{f}_t$ according to a Kullback-Leibler measure of divergence. The solution to this constrained optimization problem is:

$$
\begin{aligned}
\widetilde{f}_t(y_{t+h}) &\sim & N(\widetilde{\mu}_t, \Sigma), \\
\widetilde{\mu}_t &=& \begin{pmatrix} \widetilde{\mu}_{1t}^{survey} \\ \widetilde{\mu}_{2t} \end{pmatrix} \\
\widetilde{\mu}_{2t} &=& \widehat{\mu}_{2t} - \Sigma_{21}\Sigma_{11}^{-1}\left( \widehat{\mu}_{1t} - \widetilde{\mu}_{1t}^{survey} \right).
\end{aligned}
$$

Note that the tilting results in a new forecast for all the variables in the system: the forecast for $y_{1t+h}$ by construction equals the survey forecast and the forecast for all remaining variables $y_{2t+h}$ is also changed in a way that is mediated by the covariance between the two blocks of variables and the difference between the model-based and the survey forecasts for $y_{1t+h}$.

In terms of the answer to the questions posed at the end of the previous section (when to use survey data and how to incorporate them in a way that improves the forecast of the whole system), Altavilla et al. (2013) prove that, if the survey forecast for a given yield is informationally efficient relative to (i.e., it encompasses) the model-based forecast, the whole tilted density forecast is more accurate than the original density forecast, where accuracy is measured by the logarithmic scoring rule of Amisano & Giacomini (2007). In particular, they found that, if $\widehat{e}_{t+h,i}$ is the model-based forecast error for the $i - th$ element of $y_{1t+h}$ and $e_{t+h,i}^{survey}$ is the survey forecast error for the same variable, then, if $E\left[ e_{t+h,i}^{survey}(e_{t+h,i}^{survey} - \widehat{e}_{t+h,i}) \right] \leq 0$ then $E\left[ \log \widetilde{f}_t(y_{t+h}) \right] \geq E\left[ \log \widehat{f}_t(y_{t+h}) \right]$. The encompassing condition $E\left[ e_{t+h,i}^{survey}(e_{t+h,i}^{survey} - \widehat{e}_{t+h,i}) \right] \leq 0$ involves future observables so in principle it cannot be verified directly at time $t$. If one however assumes some measure of persistence in informational efficiency, one can expect that if the survey forecast encompassed the model-based forecast in the recent past, it might reasonably continue to do so in the immediate future. This consideration prompts Altavilla et al. (2013) to suggest testing the encompassing condition over historical data. They propose a modification of the fluctuation test of Giacomini & Rossi (2010),

which allows for the possibility that the encompassing condition may not have held at all dates and, importantly, gives an indication of whether it was satisfied in the recent past. The test involves first obtaining two sequences of historical (out-of-sample) forecast errors for the model and the corresponding survey forecast errors: $\{\widehat{e}_{j+h,i}\}_{j=1}^{t-h}$ and $\left\{e_{j+h,i}^{survey}\right\}_{j=1}^{t-h}$ and it is implemented by choosing a fraction $\delta$ of the total sample $t$ and computing a sequence of standardized rolling means:

$$F_{j,\delta} = \widehat{\sigma}^{-1}\left(\delta t\right)^{-1/2} \sum_{s=j-\delta t+1}^{j} e_{s+h,i}^{survey}(e_{s+h,i}^{survey} - \widehat{e}_{s+h,i}), \ j = \delta t, ..., t,$$

where $\widehat{\sigma}$ is a (HAC) estimator of the standard deviation of $e_{s+h,i}^{survey}(e_{s+h,i}^{survey} - \widehat{e}_{s+h,i})$ computed over the rolling window (e.g., a HAC estimator with truncation lag $h-1$). The null hypothesis that the survey forecast was informationally efficient at all points in the period $1, ..., t$ against the one-sided alternative that the survey did not encompass the model forecast at least at one point in the sample can be rejected at the significance level $\alpha$ as long as

$$\max_{j \leq t} F_{j,\delta} > k_{\delta,\alpha},$$

where the critical value $k_{\delta,\alpha}$ is in Table 1.

Altavilla et al. (2013) apply the method to the problem of improving the forecasting ability of Diebold & Li's (2006) yield curve model. The example is relevant for the problem considered in this survey because Diebold & Li's (2006) model suffered from instabilities in its forecast performance: whereas the model performed well on the data considered by the authors, Altavilla et al. (2013) show that the accuracy of the model substantially deteriorated over the years 2000-2013, when the model was uniformly outperformed by the random walk. During this time period, a user interested in forecasting yields would have had access to survey forecasts for a subset of the yields considered by the model and Altavilla et al. (2013) investigate whether incorporating this information would have improved the forecasting performance of Diebold & Li's (2006) model. The former apply the method to, first, conclude that only the survey forecast for the 3-month yield would have been useful during this time period and, second, that incorporating it into the model would have resulted in large and significant improvements, with typical accuracy gains around 30% and up to 52% relative to Diebold & Li's (2006) model.

## 3    Are Structural Models a Safeguard Against Non-stationarities?

From the perspective of the topic of this article, there might be good reasons for why structural models may be less sensitive to non-stationarities than reduced form models. Going back to one interpretation of Hurwicz's (1962) classic definition, if one defines structural relationships as those that are invariant to interventions (by man or nature), then one may expect such relationships to

13

be more stable over time than those based on empirical regularities observed in historical data. Insofar as a model is able to capture some of these deep and stable relationships, we should hope such a model to provide forecasts that are more robust to structural instabilities than those based on a-theoretical relationships. This point is made in Giacomini (2014), who suggests a role for economic theory in guiding the search for stable relationships, and discusses econometric methods for incorporating some common types of theoretical restrictions into forecasting models. She discusses some contributions in the literature that have shown how simple theory-based restrictions on the parameters of a model (e.g., exclusion restrictions), or moment restrictions involving future observables (e.g., Euler equations) could be usefully exploited for forecasting.

Whether the complex set of structural relationships and restrictions explicitly and implicitly embedded in Smets & Wouters's (2007) model can offer an antidote to structural instability is still up for debate. One subtle point made by Fernandez-Villaverde & Rubio-Ramirez (2008, p. 238) is that a strict interpretation of Hurwicz's (1962) definition is that a structure "represents not a property of the material system under observation, but rather a property of the anticipations of those asking for predictions concerning the state of the system", and there is thus no reason to believe that the relationships embedded in a structural model such as Smets & Wouters' (2007) DSGE model should be stable over time. Building on this argument, Fernandez-Villaverde & Rubio-Ramirez (2008) estimate a DSGE model with time-varying parameters and indeed find widespread evidence of instability.

## 3.1   Smets & Wouters (2007)

From the forecasting point of view, a sizable literature has emerged in recent years that investigates the forecasting performance of DSGE models, with the vast majority confining attention to (small variations of) Smets & Wouters' (2007) model. Recent examples include Adolfson et al. (2007), Edge & Gurkaynak (2010) and the literature reviews by Del Negro & Schorfheide (2012) and Gurkaynak et al. (2013). The general conclusions from this literature suggest that this class of DSGE models may not be a safeguard against nonstationarities.

First, the forecasting performance of Smets & Wouters' (2007) DSGE model has been unstable over time (Gurkaynak et al., 2013). The performance has in particular deteriorated in the years after the publication of the article, which may either be caused by the onset of the crisis (and the model being inadequate during recessions) or it may indicate that the good performance of the model documented using pre-crisis data could be due to modelling choices that reflected the properties of such data. A piece of evidence potentially supporting the latter conjecture is the finding by Chauvet & Potter (2012) that separately considering the forecasting performance during booms and recessions does not change the conclusion (also in Del Negro & Schorfheide 2012 and Gurkaynak et al. 2013) that DSGE models are generally outperformed by reduced-form models.

A second conclusion from the literature which could at first sight be interpreted as evidence in favour of the DSGE model's ability to guard against non-stationarities is the finding by Del Negro & Schorfheide (2012) that DSGE models are outperformed by reduced-form models and survey forecasts at short forecast horizons, but appear to be successful at forecasting output growth at long forecast horizons. This is because short horizon forecasting has presumably more to do with efficient extraction of information from large datasets (as is the case for factor models (Stock & Watson 2006) or Bayesian shrinkage estimation (De Mol et al. 2008)) whereas at long forecast horizons nonstationarities may play a more prominent role. However, a competing factor besides robustness to nonstationarities that may differentially affect the forecasting performance of models at short and long forecast horizons is the modelling of trends within the model. There is already some evidence in the literature documenting the sensitivity of the performance of DSGE models to the trend specification (e.g., Canova & Ferroni 2011). The latter note that, while typically DSGE models are stationary, the data typically are not; thus, the data need first to be filtered to eliminate stochastic trends and breaks; for example, in the presence of breaks, practitioners select sub-samples that contain only stationary data. They propose to avoid an arbitrary filter and, instead, propose an ideal filter for DSGE models. Furthermore, Giacomini & Rossi (2014) show how the de-trending method can affect the relative performance of DSGE models and reduced-form models, lending empirical support to a conjecture originally made by Sims (2003). In particular, Giacomini & Rossi (2014) show that, if one were to estimate a DSGE model de-trending over the full sample, its forecasting performance would be comparable to that of BVARs up to the early 1990s, after which the DSGE model would be significantly better than the BVAR; however, if one were to estimate the same DSGE model recursively de-trending the data up to the forecasting point, the conclusion would be reversed and the BVAR would forecast better. Whether the good performance of DSGE models for long-horizon output forecasting is indeed due to their robustness to instabilities or to their modelling of trends is an unanswered question that seems to deserve further investigation.

## 3.2 Beyond Smets & Wouters (2007)

Since the finding that Smets & Wouters' (2007) model failed to predict the Great Recession in 2008Q4-2009 that followed the financial crisis, a number of contributions in the literature have tried to improve on the model in various directions. Many of them have focused on the mis-specification of the original model, notably its missing financial channel. As a result, a number of DSGE models with financial frictions have appeared in the literature and there is some evidence (Del Negro et al. 2013) that adding such frictions could have improved the forecasting performance of Smets & Wouters' (2007) model, to the point that the model would have been able to predict the recession as early as 2008Q3.

With regards to non-stationarities, some authors have incorporated into DSGE models some of the devices that have been considered in the reduced-form literature. One example is time-varying parameters, which, as we discussed in Section 2, have proven an important modelling device in the VAR literature (e.g., Cogley & Sargent 2005; Primiceri 2005). Notwithstanding the point made by Cogley & Sbordone (2006) that a fixed-parameter DSGE could be compatible with time-varying parameters VAR, several contributions have showed the usefulness of incorporating time-varying parameters into structural models. Besides the already discussed Fernandez-Villaverde & Rubio-Ramirez (2008) DSGE model with time-varying parameters, these include Canova (2006) and Justiniano & Primiceri (2008).

A different approach to modelling nonstationarities in the context of structural models is to postulate Markov-switching processes for different aspects of the model. Examples are Sims & Zha (2006), Davig & Leeper (2007), Farmer et al. (2009) and Bianchi (2013). The empirical performance of this class of models has been investigated to a lesser extent than their time-varying parameter counterparts, possibly due to the greater computational difficulties that they present. For example, the estimation of these models via perturbation methods has until recently faced the challenge of obtaining a well-defined steady state around which one can obtain higher-order approximations to the solutions of the model. Foerster et al. (2013) have recently proposed a method for overcoming this challenge, which hopefully will pave the way for future investigation of the empirical performance of Markov-switching DSGE models.

Finally, as we discussed in Section 2, the reduced-from literature has shown the potential benefits of model combination as a device for robustifying forecasts against nonstationarities. In a recent paper, Wieland et al. (2012) advocate adopting a similar approach to policy analysis and forecasting in the presence of model uncertainty and develop a database collecting several examples of macroeconomic models for the US economy, the Eurozone and multi-country models that are used at policy institutions like the IMF, the ECB, the Fed, and in academia. From a more theoretical point of view, the literature has made notable advances towards discussing model uncertainty and incorporating ambiguity about aspects of the economy into models of agents' behavior. For example, Ilut & Schneider (2012) propose a DSGE model in which agents face time-varying ambiguity about the distribution of technology over time. Even though this line of enquiry could eventually lead to the development of structural models which are robust to non-stationarities, it is probably safe to say that the literature is not there just yet.

## 4  Density Forecasts and Instabilities

So far, the discussion has centered on point forecasts. However, a growing literature is moving towards considering density forecasts. A density forecasts is the probability distribution of the

forecast, and assigns a probability at each possible future value of the target variable. Therefore, a density forecast completely characterizes the forecast distribution, while point forecasts are typically the mean of the distribution. Working with density forecasts has several advantages, among which: (i) it provides the researcher with information on the possible outcomes and the probability that the forecasting model assigns to them; (ii) automatically provides a measure of uncertainty around point forecasts (e.g. the 2.5 and 97.5 quantiles of the density forecast provide a 95% confidence interval for the point forecast), thus providing a valuable tool to quantify risk in forecast-based decisions; (iii) allows researchers more complete descriptions of the models' outcomes and the possibility to evaluate several aspects at the same time (e.g. Value-at-Risk measures); (iv) allow researchers to obtain "fan charts", which summarize the uncertainty in the forecast density simultaneously at various forecast horizons. For example, central banks and policy-makers routinely use density forecasts to communicate to the public the uncertainty around their point forecasts using fan charts – a pioneer in the use of fan charts was the Bank of England.

A density forecast is typically obtained by making distributional assumptions on the error term of a forecasting model. For example, in the simple autoregressive model considered in eq. (1), the researcher may assume that the error term follows a Normal distribution with zero mean and variance $\sigma^2$. Thus, the one-step ahead conditional forecast density given the information set at time $t$, $\mathcal{F}_t$, is a Normal, with mean $\alpha + \rho y_t$ and variance $\sigma^2$: $\phi(y_{t+1}|\mathcal{F}_t) \sim N(\alpha + \rho y_t, \sigma^2)$. The unknown parameters in the conditional mean of the forecast density (in this example $\alpha$ and $\rho$) are typically estimated via either rolling or recursive procedures, and the variance is proxied by the in-sample variance of the fitted errors. Forecast densities in the ADL model, eq. (6), can be obtained similarly; under Normality, the one-step-ahead forecast density is $\phi(y_{t+1}|\mathcal{F}_{t,k}) \sim N(\beta_{k,0} + \beta_{k,1}(L)x_{t-1,k} + \beta_{k,2}(L)y_{t-1}, \sigma^2)$. In the case of factor models, the explanatory variables $x_{t-1,k}$ in eq. (6) are the factors extracted from a large dataset of predictors, $F_t$. Clearly, many other ways to obtain forecast densities are available, e.g. directly from surveys or via Bayesian methods.

## 4.1 Structural Breaks in Density Forecasts

It is typical to evaluate correct calibration of density forecasts using traditional tests such as Diebold et al. (1998, 1999), Berkowitz (2001), Corradi & Swanson (2006) or log-scores, and to evaluate forecast density comparisons via Amisano & Giacomini (2007) and Diks et al. (2011).

Regarding correct calibration of density forecasts, under correct calibration their Probability Integral Transform (PIT) should be uniform. Corradi & Swanson (2006) propose to measure the largest distance between the cumulative distribution of the PITs (denoted by $\Phi(y_{t+h}|\mathcal{F}_t)$) and the 45-degree line (indexed by $r \in (0,1)$) over the whole out-of-sample period, $t = R, ..., T$, where $R$ and $T$ denote the time of the first and last estimation periods. Corradi & Swanson (2006) propose

17

a Kolmogorov test: $\sup_{r \in [0,1]} \left| (T-R)^{-1/2} \sum_{t=R}^{T} (1\{\Phi(y_{t+h}|\mathcal{F}_t) \leq r\} - r) \right|$. However, clearly, in the presence of time variation, the models could be correctly calibrated at some points in time and mis-specified in some other periods. Thus, lack of correct calibration might be difficult to detect with traditional techniques when there are instabilities in the data. To address lack of correct calibration that could be evolving over time, Rossi & Sekhposyan (2013) design a test of the correct specification of density forecasts that can be applied in unstable environments. Their test evaluates whether the PIT is uniform at each point in time, where the point in time is indexed by $s(T-R)$, $s \in (0,1)$. Their Kolmogorov test procedure is as follows: $\kappa_P \equiv \sup_s \sup_r Q_P(\tau, r)$, where

$$
\begin{aligned}
Q_P(\tau, r) \equiv P^{-1} &\left( \sum_{t=R}^{R+[s(T-R)]} (1\{\Phi(y_{t+h}|\mathcal{F}_t) \leq r\} - r) - s\sum_{t=R}^{T} (1\{\Phi(y_{t+h}|\mathcal{F}_t) \leq r\} - r) \right)^2 \\
&+ P^{-1} \left( \sum_{t=R}^{T} (1\{\Phi(y_{t+h}|\mathcal{F}_t) \leq r\} - r) \right)^2
\end{aligned} \tag{7}
$$

Their correct calibration test combines Corradi & Swanson's (2006) test (captured by the second component in eq. 7) with a component that tests time variation in the misspecification (the first component in eq. 7). Rossi & Sekhposyan (2013) find that SPF-based predictive densities of both output growth and inflation are mis-specified; furthermore, the mis-specification is time-varying. The instability affects nowcasts (i.e. current year forecasts) and one-year-ahead forecasts of inflation and output growth. The exact break dates differ across horizons as well as across variables. However, it appears that overall the correct specification of the nowcasts breaks around the beginning of Volker's chairmanship, suggesting that the major change in monetary policy at that time resulted in a significant change in the way forecasters formed inflation expectations. For one-year-ahead forecasts, on the other hand, the break happens around mid-1990, which may reflect changes in productivity growth or its measurement.

Regarding forecast density comparisons, Amisano & Giacomini (2007) instead propose to compare two competing predictive densities $\phi_1\left(y_{t+1}|\mathcal{F}_t^{(1)}\right)$ and $\phi_2\left(y_{t+1}|\mathcal{F}_t^{(2)}\right)$ via average log-scores differences. To compare density forecasts in a way robust to instabilities, one could simply implement Amisano & Giacomini's (2007) test in rolling window over the out-of-sample period, as in Manzan & Zerom (2013), and use the critical values in Giacomini & Rossi (2010) to assess significance. This involves the following test statistic:

$$
\sup_j m^{-1/2} \sum_{t=j-m+1}^{j} w_t \left[ \log \phi\left(y_{t+1}|\mathcal{F}_t^{(1)}\right) - \log \phi\left(y_{t+1}|\mathcal{F}_t^{(2)}\right) \right], \ j = R+m, ..., T.
$$

Note that the procedure would work for models whose parameters are estimated with a rolling scheme.

Finally, parallel to the literature on structural breaks on parameters of a model, there exist a literature on detecting instabilities in densities that could be applied to forecast densities. Indeed, forecast densities could be unstable. For example, during the Great Moderation period, forecast errors were less volatile than pre-1984, thus the forecast error distribution was tighter after 1984 than before; on the other hand, during the financial crisis of 2007-2009 the forecast error distribution became more spread out. More in general, all the features of the forecast error distribution could change over time. Inoue (2001) proposed a test to evaluate whether a density changed over time. Inoue (2001) proposes to test the difference between two density functions using the difference in their non-parametric estimates. Let $\tau$ denote the unknown candidate structural break date and let $x_t$ be a p-dimensional random vector for which the researcher would like to test the stability in the distribution function, that is, whether there is a distribution function such that $Pr\left(x_t \leq i\right) = F\left(t\right)$ for all $i \in \Re^p$. Inoue's (2001) test is based on comparing the distribution function before and after the potential break-date, for all possible break dates:

$$sup_{1 \leq \tau \leq T} sup_i \left| \left(\frac{\tau}{T}\right) \left(1 - \frac{\tau}{T}\right) T^{1/2} \left\{ \tau^{-1} \sum_{t=1}^{\tau} 1\left(x_t \leq i\right) - (T-\tau)^{-1} \sum_{t=\tau+1}^{T} 1\left(x_t \leq i\right) \right\} \right|.$$

Forecast densities, however, are based on estimated parameters. Thus, to test whether forecast densities have changed over time requires taking into account the consequences of parameter estimation error, and Rossi & Sekhposyan (2013) discuss the appropriate test statistic and critical values. Empirically, instabilities in density forecasts are important, as shown by Andrade et al. (2012) and Rossi & Sekhposyan (2013). For example, the former document time-variation in SPF's inter-quantile ranges and skewness measures while the latter find strong evidence of instabilities in the Survey of Professional Forecasters' density forecasts of both inflation and output growth nowcasts and one-quarter-ahead forecasts. Also, the latter find that instabilities in nowcasts are prominent at the time of the Great Moderation, while in one-year ahead forecasts the date of the structural break is closer to the late 1990s.

## 4.2   Do Model Combinations and Time-Varying Parameters Help When Forecasting Densities?

To hedge against time variation, it might be useful to combine information across various predictors/models simultaneously, as previously discussed. This can be done via combining density forecasts too. One way to do so is by using equal weights, see Mitchell & Wallis (2011). More in detail, this is implemented simply as follows: $\phi\left(y_{t+1}|\mathcal{F}_t\right) = \frac{1}{K} \sum_{k=1}^{K} \phi\left(y_{t+1}|\mathcal{F}_{t,k}\right)$; if the densities are Normal, their combination is a finite mixture of Normal distributions.

A second way to pool information across several models is Bayesian Model Averaging (BMA). In this case, $\phi\left(y_{t+1}|\mathcal{F}_t\right) = \frac{1}{K} \sum_{k=1}^{K} w_k \phi\left(y_{t+1}|\mathcal{F}_{t,k}\right)$, where the weights are proportional to the

models' posterior probabilities. By following this procedure, BMA assigns a higher weight to models that have a higher likelihood, according to past data; however, estimating such weights may not necessarily improve the forecasting performance of the models, as the weights will contain parameter estimation error. Two commonly used variants of BMA models are described in Wright (2009). The first, which we will refer to as BMA-OLS, has time-varying weights $P_t(M_k|D_t)$, which represent the posterior probability of model $k$ denoted by $M_k$, given the observed data $D_t$, while the parameters of the models are estimated by OLS. The second is fully Bayesian, where the estimated parameters are posterior estimates. Recursive logarithmic score weights are another option to pool densities, where the model densities are combined using Bayes' rule using equal prior weight on each model (e.g. Jore et al., 2010).

A finding in the literature on forecast densities parallels that on point forecasts: equal weight model averaging has been shown to perform well for point forecasts by Stock & Watson (2003, 2004); a similar result has been found for density forecasts by Rossi & Sekhposyan (2014). The latter find that forecast density specification tests are favorable to equal weight forecast combinations when predicting US output growth and inflation, while most of the other models fail to pass the tests in some dimension.

A typical issue encountered in practice is that the variance is time-varying as well as the parameters in the conditional mean of the models. For example, during the Great Moderation, the volatility of output growth decreased substantially; similarly, the volatility increased during the latest financial crisis and the recent increases in variation in energy prices. Thus, estimating models with constant variance would clearly lead to misspecification in the uncertainty in the forecast density. Clark (2011) considers the performance of BVARs with stochastic volatility. He considers a VAR where $A(L) Y_t = \Omega_t \varepsilon_t$ and $\Omega_t = A^{-1} \Lambda_t^{1/2}$, where $\varepsilon_t$ is a multivariate standard Normal and $\Lambda_t$ is a diagonal matrix of time-varying variances, each one of which evolves according to a random walk: $\Lambda_t = diag([\lambda_{1,t}, ... \lambda_{k,t}, ..., \lambda_{K,t}])$, $log(\lambda_{k,t}) = \log(\lambda_{k,t-1}) + v_{k,t}$, where $v_{k,t}$ is a normal with zero mean and constant variance. The random walk in the volatility is similar to the modeling framework used in Cogley & Sargent (2005) to fit empirical models for post-World War II data in the US. Perhaps not surprisingly, Clark (2011) finds that including stochastic volatility in BVARs does improve the accuracy of density forecasts. Billio et al. (2013) propose a Bayesian combination of multivariate forecast densities that takes into account instabilities. Their idea is to combine predictive densities from a set of models using weights that are time varying, and are derived from a distribution such that the dynamic process of the weights is guided both by the densities' past empirical performance and by learning. This allows them to model time varying patterns as well as possible breaks in the series of interest.

### 4.3 Do Surveys Help When Forecasting Densities?

Finally, in the previous section, we saw that a possible way to guard against instabilities is to use structural models. How useful is structural information when the target is to produce a well-calibrated forecast density (as opposed to point forecasts with low MSFE)? Wolden et al. (2011) combine predictive densities of several VARs as well as a DSGE model to forecast inflation. They find that the DSGE model receives a large weight at short horizons only when the VARs do not take into account structural breaks; however, the resulting density forecast is poorly calibrated. When VARs are allowed to have breaks, the weight on the DSGE model decreases considerably and the resulting forecast density is well-calibrated.

## 5 Conclusions

Forecasting models are subject to instabilities, both in their parameters and in their performance over time. This is a common phenomenon in macroeconomic data. For example, a variable could be a good predictor during certain times and a useless predictor at other times. Thus, taking into account instabilities when forecasting is very important. The question is how, and which methods work empirically. This is what we examined in this survey.

Our analysis leads to several broad conclusions. First, modeling instabilities can help, but it depends on how they are modeled. Relying on structural break tests and then estimating a model that takes the break into account does not perform well when forecasting in practice. Similarly, the evidence in favor of Markov-switching models in the context of forecasting is scant. There is instead some evidence supporting the use of models with time-varying parameters.

Second, there exist approaches that could be interpreted as ways to robustify a model against non-stationarities. Among these, we found that a forecasting method that has been shown time and again to improve performance is forecast combination, for both point and density forecasting. Whether this is because combining models is a way to robustify the forecast against instabilities - as opposed to being, say, a guard against model misspecification or an effective way to pool different information sets - remains to be shown incontrovertibly. Furthermore, survey forecasts could be used to improve the performance of a forecast model and robustify it against instabilities.
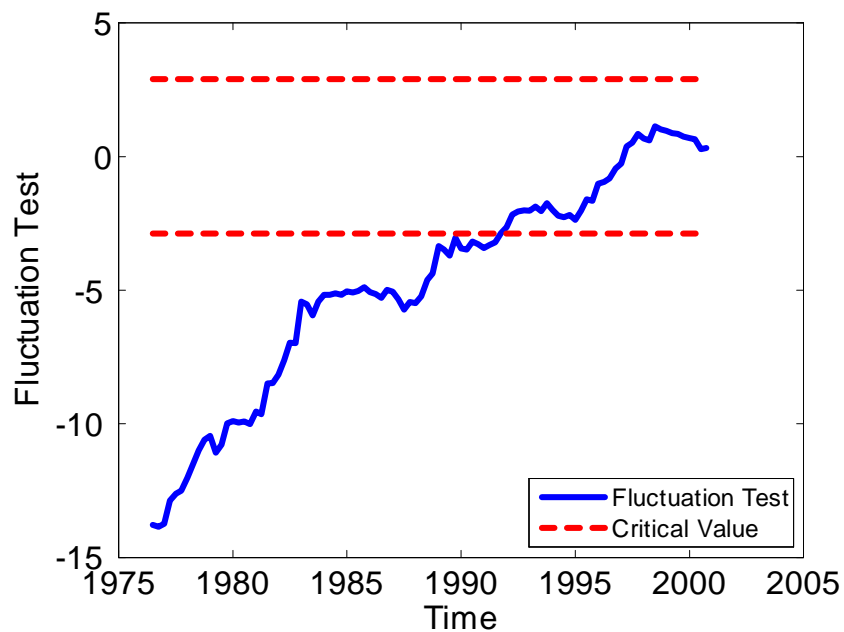
Finally, whether structural models can provide a guard against instabilities in not clear. The conclusion from the literature is that there is yet little evidence that they do, at least for the case of the popular DSGE model by Smets & Wouters (2007). It is possible that economic theory could provide insights and ultimately be a helpful guide in the search for stabile relationships that can be exploited for forecasting, but that search is not over yet. On the other hand, there is also the possibility that structural models may be misspecified, and the misspecification could present itself as parameter changes. Inoue et al. (2014) provide methodologies designed to help researchers

identify which parts of their models are mis-specified and could potentially help limit the effects of instabilities.

Overall, instability has been recognized to be one of the major causes of forecast failure, which prevents the use of in-sample methodologies to select forecasting models. This survey has discussed recent avenues of research that could potentially improve traditional forecasts, several of which are still only partially explored. Constructing methodologies that improve forecasts in the presence of instabilities has been and continues to be an interesting and lively area of research, and one in which payoffs are potentially big. After all, forecasts are made and used every day by policy-makers, researchers in Central banks and academia, as well as firms and consumers.
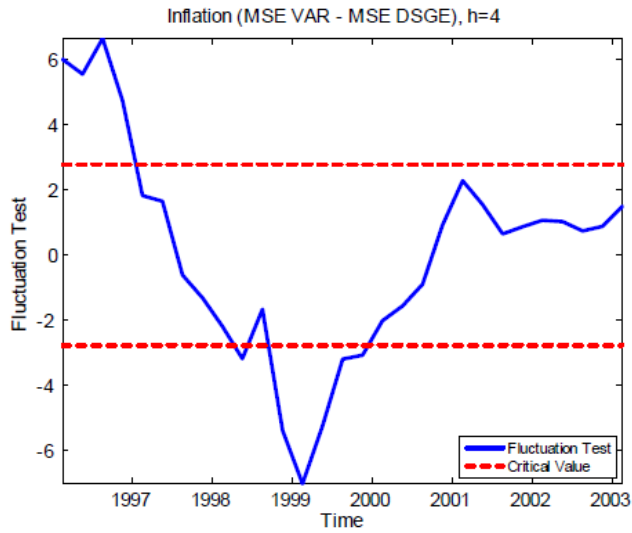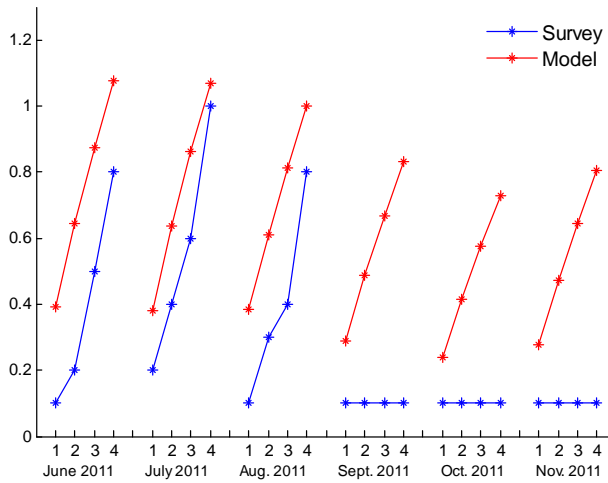
# Figures

## Figure 1



The figure plots Giacomini & Rossi's (2010) Fluctuation test statistic for testing equal forecasting performance, with 10% boundary lines. Negative values favor the Federal Reserve's forecast performance over the Survey of Professional Forecasters.

**Figure 2**



Inflation (MSE VAR - MSE DSGE), h=4

The figure plots Giacomini & Rossi's (2014) test statistic for equal
performance of the DSGE and the VAR, with 10% boundary lines for
testing equal fit. Negative values favor the VAR model's performance.

**Figure 3**



The figure shows 1- to 4-quarters ahead forecasts of the 3-month yield
from the Blue Chip Analyst survey and from Diebold & Li's (2006)
model made at the beginning of each month on the horizontal axis.

**Table 1. Critical values $(k_{\delta,\alpha})$**

| | | $\alpha$ | |
|---|---|---|---|
| $\delta$ | | 0.05 | 0.10 |
| | | | |
| 10% | | 3.176 | 2.928 |
| 20% | | 2.938 | 2.676 |
| 30% | | 2.770 | 2.482 |
| 40% | | 2.624 | 2.334 |
| 50% | | 2.475 | 2.168 |
| 60% | | 2.352 | 2.030 |
| 70% | | 2.248 | 1.904 |
| 80% | | 2.080 | 1.740 |
| 90% | | 1.975 | 1.600 |

Notes. The table shows the critical value for the informational efficiency test described in Section 2.

# References

Adolfson M, Linde J, Villani M. 2007. Forecasting performance of an open economy Dynamic Stochastic General Equilibrium model. *Econometric Reviews* 26: 289–328.

Altavilla C, Giacomini R, Ragusa G. 2013. Anchoring the yield curve using survey expectations. Centre for Microdata Methods and Practice. *CeMMAP Working Paper* 52/13. London: The Institute for Fiscal Studies.

Amisano G, Giacomini R. 2007. Comparing density forecasts via weighted Likelihood ratio tests. *Journal of Business and Economic Statistics* 25(2): 177–190.

Andrade P, Ghysels E, Idier J. 2012. Tails of inflation forecasts and tales of monetary policy. *Banque of France Working Paper* 407. Paris: Banque of France.

Andrews DWK. 1993. Tests for parameter instability and atructural change with unknown xhange point. *Econometrica* 61(4): 821–856.

Andrews DWK, Ploberger W. 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62(6): 1383–1414.

Bai J, Ng S. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70(1): 191–221.

Bai J, Perron P. 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* 66: 47–78.

Banbura M, Giannone D, Reichlin L. 2010. Large Bayesian vector auto regressions. *Journal of Applied Econometrics* 25: 71–92.

Banerjee A, Marcellino M, Masten I. 2008. Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. In: *Forecasting in the Presence of Structural Breaks and Model Uncertainty (Frontiers of Economics and Globalization),* Vol. 3, ed. DE Rapach, ME Wohar, pp. 149–194. Bingley (UK): Emerald Group Publishing Limited.

Bates B, Plagborg-Moller M, Stock J, Watson MW. 2013. Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics* 177: 289–304.

Berkowitz, J. 2001. Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics* 19(4): 465–474.

Bianchi, F. 2013. Regime switches, agents' beliefs, and post-World War II U.S. macroeconomic dynamics. *Review of Economic Studies* 80(2): 463–490.

Billio M, Casarin R, Ravazzolo F, Van Dijk H. 2013. Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics* 177(2): 153–170.

Breitung J, Eickmeier S. 2011. Testing for structural breaks in dynamic factor models. *Journal of Econometrics* 163(1): 71–84.

Brown R, Durbin J, Evans J. 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society Series B* 37: 149–163.

Canova, F. 2006. Monetary policy and the evolution of the U.S. economy. *Centre for Economics Policy Research Working Paper* 5467. London: CEPR.

Canova F, Ferroni F. 2011. Multiple filtering devices for the estimation of cyclical DSGE models. *Journal of Quantitative Economics* 2: 73–98.

Carriero A, Clark T, Marcellino M. 2011. Bayesian VARs: specification choices and forecast accuracy. *FRB of Cleveland Working Paper* 11–12. Cleveland (OH): Federal Reserve Bank of Cleveland.

Chauvet M, Potter S. 2012. Forecasting output. In *Handbook of Economic Forecasting*, Vol. 2, ed. G Elliott, A Timmermann, pp. 1–56. Amsterdam: Elsevier-North Holland.

Chun, A. 2008. Forecasting interest rates and inflation: Blue Chip clairvoyants or econometrics? *Mimeo.*

Clarida R, Galí J, Gertler M. 2000. Monetary policy rules and macroeconomic stability: evidence and some theory. *The Quarterly Journal of Economics* 115(1): 147–180.

Clark, TE. 2011. Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics* 29(3): 327–341.

Cogley T, Sargent TJ. 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics* 8(2): 262–302.

Cogley T, Sbordone A. 2006. Trend inflation and inflation persistence in the new Keynesian Phillips curve. *FRB of New York Staff Report* 270. New York: Federal Reserve Bank of New York.

Coibion O, Gorodnichenko Y. 2012. What can survey forecasts tell us about informational rigidities? *Journal of Political Economy* 120: 116–159.

Corradi V, Swanson NR. 2006. Predictive density evaluation. In *Handbook of Economic Forecasting*, Vol. 1, ed. G Elliott, CWJ Granger, A Timmermann, pp. 197–284. Amsterdam: Elsevier-North Holland.

D'Agostino A, Gambetti L, Giannone D. 2013. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* 28(1): 82-101.

Davig T, Leeper EM. 2007. Generalizing the Taylor principle. *American Economic Review* 97: 607–635.

De Mol C, Giannone D, Reichlin L. 2008. Forecasting using a large number of predictors: is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics* 146: 318–328.

Del Negro M, Schorfheide F. 2012. DSGE model-based forecasting. *Federal Reserve Bank of New York Staff Report* 554. New York: Federal Reserve Bank of New York.

Del Negro M, Giannoni M, chorfheide F. 2013. Inflation in the Great Recession and New

Keynesian models. *Federal Reserve Bank of New York Staff Report* 618. New York: Federal Reserve Bank of New York.

Diebold FX, Gunther TA, Tay AS. 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39(4): 863–883.

Diebold FX, Li C. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130: 337–364.

Diebold FX, Tay AS, Wallis KF. 1999. Evaluating density forecasts of inflation: the survey of professional forecasters. In *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger,* ed. RF Engle, H White H, pp. 76–90. Oxford: Oxford University Press.

Diks C, Panchenko V, Van Dijk D. 2011. Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics* 163(2): 215–230.

Edge RM, Gürkaynak RS. 2010. How useful are estimated DSGE model forecasts for central bankers? *Brookings Papers on Economic Activity* 41: 209–259.

Eickmeier S, Lemke W, Marcellino M. 2011. Classical time-varying FAVAR models - Estimation, forecasting and structural analysis. *Deutsche Bundesbank Discussion Paper* 04-2011. Frankfurt: Deutsche Bundesbank.

Elliott G, Müller U. 2006. Efficient tests for general persistent time variation in regression coefficients. *Review of Economic Studies* 73: 907–940.

Farmer R, Waggoner D, Zha T. 2009. Indeterminacy in a forward looking regime switching model. *International Journal of Economic Theory* 5: 69–84.

Faust J, Wright JH. 2013. Forecasting inflation. In *Handbook of Economic Forecasting*, Vol. 2, ed. G Elliott, A Timmermann, pp. 3-56. Amsterdam: Elsevier-North Holland.

Fernández-Villaverde J, Rubio-Ramírez JF. 2008. How structural are structural parameters? In *NBER Macroeconomics Annual 2007*, Vol. 22, ed. D Acemoglu, K Rogoff, M Woodford, pp. 83–137. Chicago: University of Chicago Press.

Foerster A, Rubio-Ramirez J, Waggoner D, Zha T. 2013. Perturbation methods for Markov-switching DSGE model. *Federal Reserve Bank of Kansas City Research Working Paper* 13-01. Kansas City (Mo.): Federal Reserve Bank of Kansas City.

Forni M, Hallin M, Lippi M, Reichlin L. 2000. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics* 82(4): 540–554.

Giacomini, R. 2014. Economic theory and forecasting: lessons from the literature. *Mimeo*.

Giacomini R, Ragusa G. 2014. Theory-coherent forecasting. *Journal of Econometrics* 182: 145–155.

Giacomini R, Rossi B. 2009. Detecting and predicting forecast breakdown. *Review of Economic Studies* 76(2): 669–705.

Giacomini R, Rossi B. 2010. Forecast comparisons in unstable environments. *Journal of Applied*

*Econometrics* 25(4): 595–620.

Giacomini R, Rossi B. 2014. Model comparisons in unstable environments. *Barcelona Graduate School of Economics Working Paper* 784. Barcelona: Barcelona GSE.

Giacomini R, Rossi B. 2006. How stable is the forecasting performance of the yield curve for output growth? *Oxford Bulletin of Economics and Statistics* 68(s1): 783-795.

Giannone D, Lenza M, Primiceri G. 2012. Prior selection for Vector Autoregressions. *Review of Economics and Statistics*, In press.

Giraitis L, Kapetanios G, Price S. 2013. Adaptive forecasting in the presence of recent and ongoing structural change. *Journal of Econometrics* 177(2): 153–170.

Groen JJJ, Paap R, Ravazzolo F. 2013. Real-time inflation forecasting in a changing world. *Journal of Business and Economic Statistics* 31(1): 29–44.

Gurkaynak R, Kısacıkoğlu B, Rossi B. 2013. Do DSGE models forecast more accurately out-of-sample than reduced-form models?. In *VAR Models in Macroeconomics, Financial Econometrics, and Forecasting - New Developments and Applications Essays in Honor of Christopher A. Sims, Advances in Econometrics,* Vol. 32, ed. T Fomby, L Kilian, A Murphy, pp. 27–80. Bingley UK: Emerald Insight.

Hurwicz, L. 1962. On the structural form of interdependent systems. In *Methodology and Philosophy of Science. Proceedings of the 1960 International Congress*, ed. L. Nagel, P Suppes, A Tarski, pp. 232–239. Stanford (CA): Stanford University Press.

Ilut C, Schneider M. 2012. Ambiguous business cycles. *National Bureau of Economic Research Working Paper* 17900. Cambridge (MA): NBER.

Inoue A. 2001. Testing for distributional change in time series. *Econometric Theory* 17(1): 156–187.

Inoue A, Kuo CH, Rossi B. 2014. Identifying the sources of model misspecification. *Centre for Economic Policy Research Working Paper* 10140. London: CEPR.

Inoue A, Rossi B. 2012. Out-of-sample forecast tests robust to the choice of window size. *Journal of Business and Economic Statistics* 30(3): 432–453.

Inoue A, Rossi B. 2011. Identifying the sources of instabilities in macroeconomic fluctuations. *The Review of Economics and Statistics* 93(4): 1186–1204.

Jore AS, Mitchell J, Vahey SP. 2010. Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics* 25(4): 621–634.

Justiniano A, Primiceri G. 2008. The time-varying volatility of macroeconomic fluctuations. *American Economic Review* 98: 604–41.

Kim, CJ, Nelson CR. 1999. Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics and Statistics* 81(4): 608–616.

Koop, G. 2013. Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics* 28(2): 177–203.

Koop G, Korobilis D. 2013. Large time-varying parameter VARs. *Journal of Econometrics* 177(2): 185–198.

Korobilis D, 2013. Assessing the transmission of monetary policy using time-varying parameter dynamic factor models. *Oxford Bulletin of Economics and Statistics* 75(2): 157–179.

Manzan S, Zerom D. 2013. Are macroeconomic variables useful for forecasting the distribution of U.S. inflation? *International Journal of Forecasting* 29(3): 469–478.

McConnell MM, Pérez-Quirós G. 2000. Output fluctuations in the United States: What has changed since the early 1980's? *American Economic Review* 90(5): 1464–1476.

Mitchell J, Wallis KF. 2011. Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics* 26(6): 1023–1040.

Monti F. 2010. Combining judgment and models. *Journal of Money, Credit and Banking* 42: 1641-1662.

Müller U, Petalas PE. 2010. Efficient estimation of the parameter path in unstable time series models. *Review of Economic Studies* 77: 1508–1539.

Ng S, Wright JH. 2013. Facts and challenges from the Great Recession for forecasting and macroeconomic modeling. *Journal of Economic Literature* 51(4): 1120–1154.

Nyblom, J. 1989. Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 84: 223–230.

Pesaran MH, Pettenuzzo D, Timmermann A. 2006. Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies* 73(4): 1057–1084.

Pesaran MH, Pick A, Pranovich M. 2013. Optimal forecasts in the presence of structural breaks. *Journal of Econometrics* 177(2): 134–152.

Pesaran MH, Timmermann A. 2007. Selection of estimation window in the presence of breaks. *Journal of Econometrics* 137: 134–161.

Pesaran MH, Timmermann A. 2002. Market timing and return prediction under model instability. *Journal of Empirical Finance* 9(5): 495–510.

Ploberger W, Krämer W. 1992. The CUSUM Test with OLS residuals. *Econometrica* 60: 271–285.

Ploberger W, Krämer W, Kontrus K. 1989. A new test for structural stability in the linear regression model. *Journal of Econometrics* 40: 307–318.

Primiceri G. 2005. Time varying structural Vector Autoregressions and monetary policy. *Review of Economic Studies* 72(3): 821–852.

Robertson JC, Tallman EW, Whiteman CH. 2005. Forecasting using relative entropy. *Journal of Money, Credit and Banking* 37: 383–401.

Rossi, B. 2013. Advances in forecasting under instabilities. In *Handbook of Economic Forecasting*, Vol. 2, ed. G Elliott, A Timmermann, pp. 1203-1324. Amsterdam: Elsevier-North Holland.

Rossi B, Sekhposyan T. 2014. Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set. *International Journal of Forecasting* 30(3): 662–682.

Rossi B, Sekhposyan T. 2014. Forecast rationality tests in the presence of instabilities, with applications to Federal Reserve and survey forecasts. *Journal of Applied Econometrics*: In press.

Rossi B, Sekhposyan. 2013. Conditional predictive density evaluation in the presence of instabilities. *Journal of Econometrics* 177(2): 199–212.

Rossi B, Sekhposyan T. 2010. Have economic models' forecasting performance for US output growth and inflation changed over time, and when? *International Journal of Forecasting* 26(4): 808–835.

Sims, C. 2003. *Comment on Smets and Wouters*. Princeton University. https://sims.princeton.edu/yftp/Ottawa/SWcommentSlides.pdf

Sims C, Zha T. 2006. Were there regime switches in U.S. monetary policy? *American Economic Review* 96(1): 54–81.

Smets F, Wouters R. 2007. Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review* 97: 586–606.

Stock JH, Watson MW. 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics* 14(1): 11–30.

Stock JH, Watson MW. 2002. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97(460): 1167–1179.

Stock JH, Watson MW. 2003. Has the business cycle changed and why? In *NBER Macroeconomics Annual 2002*, Vol. 17, ed. M Gertler, K Rogoff, pp. 159–230. Cambridge (MA): National Bureau of Economic Research.

Stock JH, Watson MW. 2003. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature* 41(3): 788–829.

Stock JH, Watson MW. 2004. Combination forecasts of output growth in a seven country data set. *Journal of Forecasting* 23(6): 405–430.

Stock JH, Watson MW. 2006. Forecasting with many predictors. In *Handbook of Economic Forecasting*, Vol. 1, ed. G Elliott, CWJ Granger, A Timmermann, pp. 515–554. Amsterdam: Elsevier-North Holland.

Stock JH, Watson MW. 2007. Has inflation become harder to forecast? *Journal of Money, Credit and Banking* 39(1): 3–33.

Stock JH, Watson MW. 2009. Forecasting in dynamic factor models subject to structural instability. In *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, ed. DF Hendry, J Castle, N Shephard, pp. 173–205. Oxford: Oxford University Press.

Timmermann A. 2006. Forecast combinations. In *Handbook of Economic Forecasting*, Vol. 1, ed. G Elliott, CWJ Granger, A Timmermann, pp. 135–196. Amsterdam: Elsevier-North Holland.

Wieland V, Cwik T, Müller GJ, Schmidt S, Wolters MH. 2012. A new comparative approach to macroeconomic modelling and policy analysis. Center for Financial Studies. Mimeo. Frankfurt: CFS.

Wright, JH. 2009. Forecasting US inflation by Bayesian model averaging. *Journal of Forecasting* 28(2): 131–144.

Wolden Bache I, Sofie Jore A, Mitchell J, Vahey SP. 2011. Combining VAR and DSGE forecast densities. *Journal of Economic Dynamics and Control* 35(10): 1659-1670.