Incorporating theoretical restrictions into forecasting by projection methods

Raffaella Giacomini
UCL, Cemmap, and CEPR

Giuseppe Ragusa
Luiss University

October 3, 2011

Abstract

We propose a method for modifying a given density forecast in a way that incorporates the information contained in theory-based moment conditions. An example is "improving" the forecasts from atheoretical econometric models, such as factor models or Bayesian VARs, by ensuring that they satisfy theoretical restrictions given for example by Euler equations or Taylor rules. The method yields a new density (and thus point-) forecast which has a simple and convenient analytical expression and which by construction satisfies the theoretical restrictions. The method is flexible and can be used in the realistic situation in which economic theory does not specify a likelihood for the variables of interest, and thus cannot be readily used for forecasting.

1 Introduction

A survey of the literature on macroeconomic forecasting over the past couple of decades reveals two general and opposite trends. On the one hand, the vast majority of the literature has focused on "atheoretical" econometric models that are able to capture both the dynamics of a variable and the interdependencies among variables in a way that results in accurate forecasts. One of the lessons learned from this strand of the literature is that accurate forecasting relies in great part on being able to extract the information contained in large datasets while at the same time avoiding the curse of dimensionality, which is the common theme of various econometric methods that have now become standard tools in the forecaster toolbox, such as factor models (Stock and Watson, 2002, Forni, Hallin, Lippi, and Reichlin, 2000), Bayesian VARs (BVARs, e.g., Litterman, 1986; Giannone, Lenza, and Primiceri, 2010) and forecast combination such as Bayesian Model Averaging (Raftery, Madigan, and Hoeting, 1997; Aiolfi, Capistrán Carmona, and Timmermann, 2010), bagging (Inoue and Kilian, 2008), and other similar techniques. On the other hand, some forecasters, in particular at central banks and policy institutions, have become increasingly dissatisfied with the inability of these econometric models to "tell a story" (Edge, Kiley and Laforte, 2008). A small "theoretical" literature has thus emerged that tries to address these concerns and suggests forecasting using theory-based models directly, such as Dynamic Stochastic General Equilibrium models (DSGEs, e.g., Smets and Wouters, 2003; Edge, Kiley, and Laforte, 2010; Christoffel, Coenen, and Warne, 2010) or no-arbitrage affine term structure models (Ang and Piazzesi, 2003). One major limitation of this approach, which the methods that we propose in this paper overcome, is that economic theory typically only provides a set of partial- or general-equilibrium restrictions that are moreover nonlinear, which implies that in general a likelihood for the variables of interest is not known and the model must be approximated before it can be used for forecasting. A final branch of the literature has considered "hybrid" approaches which for example use the theoretical model to form priors to impose on the parameters of the econometric model (An and Schorfheide, 2007; Schorfheide, 2000) or construct an optimal combination of the theoretical and econometric models (Del Negro and Schorfheide, 2004; Carriero and Giacomini, 2011).

The motivation for this paper is similar to that of the "hybrid" approach, since our premise is that it is legitimate to ask whether economic theory can be useful for forecasting, but we also argue that one should not discard what has been learned from decades of forecasting with atheoretical econometric models. We depart from the "hybrid" literature by considering an environment in which a forecaster has a way to generate density forecasts for a vector of economic variables (for example using BVARs or factor models) and also has a set of theoretical restrictions expressed as (nonlinear) moment conditions which are typically not satisfied by the initial density forecasts. These moment conditions could for example be Euler equations or moment conditions implied by Taylor rules. One key feature of our approach is that it does not require the theoretical restrictions to specify a likelihood for the variables of interest. In fact, the typical situation that a forecaster encounters is that the econometric model uses a large number of predictors (often hundreds for factor models and BVARs) but the theoretical restrictions only involve a small number of variables. Further, economic theory often only suggests partial equilibrium restrictions or, even in the context of general equilibrium models, one might want to focus on only a subset of the equations in the model. In both the "theoretical" and the "hybrid" literatures, instead, the key assumption is that the theoretical model specifies a likelihood for all the variables in the system, which is problematic since for the vast majority of economic models, such as DSGEs, the exact likelihood is not known due to the model being highly nonlinear. The necessity to obtain a likelihood in the existing literature means that in practice one needs to approximate the model by linearizing it around the steady-state and adding distributional assumptions on the errors. One undesirable consequence of this focus on likelihoods is that the theoretical models which are currently considered in the literature are only approximations of the truth which moreover do not necessarily satisfy the restrictions implied by the theory.

In a nutshell, we propose the following procedure for forecasting the vector Y_{t+h} based on the information set Ω_t, when there are a set of theory-based moment restrictions that are assumed to be valid for all t :

$$E_t[g(Y_{t+h}, \theta_0)] = 0. \tag{1}$$

The subscript t indicates conditioning on Ω_t and θ_0 is assumed to be known, calibrated, or estimated on a different data set than the one used for forecasting. The moment conditions could involve only a subset of the variables in Y_{t+h}.

1. For each $t = 1, ..., T-h$ (where for example $t = 1, ..., T-h$ is the out-of-sample portion of the available sample), start with an econometric model that does not in general satisfy the moment conditions but is known to give accurate forecasts (e.g., a BVAR or a factor model), and let $f_t(y_{t+h})$ be the density forecast implied by the model at time t.

2. Project $f_t(y_{t+h})$ onto the space of distributions that satisfy the moment conditions. Proposition 1 shows that this projection step yields a new density $\widetilde{f}_t(y_{t+h})$ which is known in closed form and is given by

$$\widetilde{f}_t(y_{t+h}) = f_t(y_{t+h}) \exp\left\{\eta_t + \lambda_t' g(y_{t+h}, \theta_0)\right\}. \tag{2}$$

   The new density by construction satisfies the moment conditions (1).

3. Estimate η_t and λ_t by (numerically) solving the following equations:

$$\begin{aligned} \lambda_t &= \min_{\lambda} \int f_t(x) \exp\left\{\lambda' g(x, \theta_0)\right\} dx \\ \eta_t &= \log\left\{\int f_t(x) \exp\left\{\lambda_t' g(x, \theta_0)\right\} dx\right\}^{-1} \end{aligned} \tag{3}$$

   and use this new density for (point- or density-) forecasting.

The "projected" or "tilted" density $\widetilde{f}_t(y_{t+h})$ can be interpreted as the density which, out of all the densities that satisfy the moment conditions $E_t[g(Y_{t+h}, \theta_0)] = 0$, is the closest to the initial density $f_t(y_{t+h})$ according to a Kullback-Leibler measure of divergence (see, e.g., White (1996)). Note that the theoretical restrictions will in typical applications involve only a subset of the components of Y_{t+h}, but, because of the likely dependence among the variables in the system, imposing the moment conditions on the joint density of Y_{t+h} potentially affects all the conditional densities and thus the forecast for each variable. Moreover, the projected density will by construction reflect any nonlinearity in the dependence among the component of Y_{t+h} that is specified by the moment conditions. Such nonlinearity is difficult to capture using existing methods due to the fact that the moment conditions only give partial information about the joint distribution of the variables.

The parameter η_t is an integration constant that ensures that (2) is a well-defined density. As we show in Section 2, η_t can also be related to the relative Kullback Leibler Information Criterion (KLIC), which measures the relative divergence of the original and the projected densities from the true density. In the context of out-of-sample forecast evaluation, the relative KLIC equals the relative logarithmic scoring rule which forms the basis for the test of equal density forecast accuracy proposed by Amisano and Giacomini (2007).

The vector λ_t represents the weights assigned to each moment condition, and, loosely speaking, it can be interpreted as a measure of the importance of each theoretical restriction, since a λ_t close to zero indicates that the original model almost satisfies the moment conditions, whereas large values of λ_t suggest that the initial density requires a large amount of tilting in order to satisfy the theoretical restrictions.

It is important to note that the procedure yields a new density which has a known analytical form but in general is not a member of a known family (for example, even though the initial density is normal in typical applications to macroeconomic forecasting, the projected density will not be normal in all but a few special cases). The fact that our method gives a density that is not in the same family as the original density forecast is a useful feature of our approach, as it allows one to understand the effects of imposing theoretical restrictions on the entire shape of the distribution, rather than just focusing on the first and possibly second moment, as is typically done in the literature.

To gain some insight into how the projection step modifies a density forecast and why this could result in an improvement in accuracy, we show in Figure 1 an actual example from our empirical application. The figure shows two out-of-sample density forecasts for CPI at one particular point in time: the histogram is the one-step-ahead density forecast of CPI implied by an "atheoretical" BVAR for CPI, GDP, nondurable and service real consumption (C_t), the federal funds rate and the real yield of a one year bond R_t. The dashed line is the density forecast obtained by incorporating into the BVAR the Euler condition $E_t\left[0.96\left(\frac{C_{t+1}}{C_t}\right)^{-2}(1+R_{t+1})-1\right]=0$, using our projection method. The vertical line represents the realization of the variable.

[Figure 1 about here.]

As one can see from the figure, in this instance the incorporation of the Euler equation restrictions does not seem to modify the shape of the BVAR density but rather it shifts it and almost centers it around the actual realization of the variable, thus yielding a substantially more accurate point- and density forecast.

2 Motivating example

This section shows a simple example where an analytical expression for the projected density can be easily obtained, which provides some intuition for our method. Suppose that the true conditional density $h_t(y_{t+1})$ of the variable of interest Y_{t+1} is unknown apart from its conditional mean μ_t, which implies the moment condition $E_t[Y_{t+1}-\mu_t] = \int (y_{t+1}-\mu_t)\, h_t(y_{t+1})dy_{t+1} = 0$. In this example μ_t can be time-varying or constant, and the only simplifying assumption is that μ_t is known or estimated using a different dataset than the one on which the density forecasts are evaluated. For example, μ_t could be pinned down by economic theory or depend on structural parameters that are calibrated or estimated using microeconometric data, or in this simple example it could represent, e.g., the mean forecast from a survey of professional forecasters.

Suppose that one has available a sequence of one-step-ahead density forecasts $\{f_t(y_{t+1})\}_{t=1}^{T-1}$ $\sim N(\widehat{\mu}_t, 1)$, which do not necessarily satisfy the moment conditions in that $\widehat{\mu}_t$ may be different from μ_t. In order to obtain a new sequence of density forecasts which by construction have the correct mean, our procedure suggests projecting $f_t(y_{t+1})$ onto the space of densities with mean μ_t at each $t = 1, ..., T-1$, which yields a sequence of new density forecasts $\widetilde{f}_t(y_{t+1})$:

$$\begin{aligned} \widetilde{f}_t(y_{t+1}) &= f_t(y_{t+1}) \exp\{\eta_t + \lambda_t(y_{t+1}-\mu_t)\} \\ &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y_{t+1}-\widehat{\mu}_t)^2 + \eta_t + \lambda_t(y_{t+1}-\mu_t)\right\}. \end{aligned} \tag{4}$$

In this simple case one can find analytical expressions for λ_t and η_t :

$$\begin{aligned} \lambda_t &= \mu_t - \widehat{\mu}_t \\ \eta_t &= \frac{1}{2}(\mu_t - \widehat{\mu}_t)^2, \end{aligned} \tag{5}$$

which shows that λ_t and η_t are related to the degree of misspecification in the conditional mean of the initial forecast, with the limiting case $\eta_t = \lambda_t = 0$ occurring when the initial density satisfies the moment conditions and thus no projection is necessary. In this simple example of a mean restriction on a single variable the projection step simply recenters the density forecast at the theoretical mean. It also happens that the projected density (4) is a normal with mean μ_t and variance 1, but the projected density is no longer normal as soon as one considers a different type of (e.g., nonlinear) moment condition or a non-normal initial density. In more realistic scenarios in which the initial density is multivariate the projection step will in general affect both the marginal densities and the dependence among the variables. For example, in the empirical application in Section 4 we show how to impose the restrictions implied by the Euler equation on a multivariate density forecast implied by a BVAR.

Since the Euler equation imposes a nonlinear restriction on the joint density of consumption and interest rates, the projected density forecast will by construction reflect the nonlinear dependence structure implied by the Euler equation as well as its effects on the conditional density of each variable in the VAR. Note that, because of the nonlinearity of the Euler equation and the fact that it only gives partial information about the distribution of the variables involved, it is otherwise challenging to construct a multivariate conditional density that satisfies these restrictions.

Besides representing the integration constant, η_t can also be related to the "local relative Kullback Leibler information criterion" $\Delta KLIC_t$ considered by Giacomini and Rossi (2010), which in this case measures the relative divergence of the initial and the projected densities from the true density:

$$\begin{aligned} \Delta KLIC_t &= E\left[\log h_t\left(Y_{t+1}\right)/f_t\left(Y_{t+1}\right)\right] - E\left[\log h_t\left(Y_{t+1}\right)/\widetilde{f}_t\left(Y_{t+1}\right)\right] \\ &= E\left[\log \widetilde{f}_t\left(Y_{t+1}\right) - \log f_t\left(Y_{t+1}\right)\right] = E\left[\eta_t + \lambda_t(Y_{t+1} - \mu_t)\right] \\ &= E\left[E_t\left[\eta_t + \lambda_t(Y_{t+1} - \mu_t)\right]\right] = E\left[\eta_t\right]. \end{aligned} \tag{6}$$

In the context of an out-of-sample evaluation exercise where f_t is the density forecast and Y_{t+1} the realization of the variable, one can further interpret (6) as a measure of the relative accuracy of the density forecasts $\widetilde{f}_t\left(y_{t+1}\right)$ and $f_t\left(y_{t+1}\right)$, according to a logarithmic scoring rule, as proposed by Amisano and Giacomini (2007). One can see from (5) that $E\left[\eta_t\right] \geq 0$, which implies that the projected density is always weakly closer to the truth and the projected density forecast is weakly more accurate than the initial density forecast. We prove in Proposition 2 below that this result is valid in general.

3 Projected density forecasts

In this section we show how to incorporate theoretical restrictions expressed in the form of moment conditions into an existing density forecast, such as that implied by an theoretical econometric model.

Let Y_t be a $n \times 1$ vector of variables of interest, and suppose the user has available a sequence of $h-$step ahead density forecasts $\{f_t(y_{t+h})\}_{t=1}^{T-h}$, where the index t reflects the fact that the density forecasts are based on the information set at time t and are thus typically time-varying, either in their parameters, or, less commonly, in their functional form. The forecasts can be, but are not necessarily, obtained by estimating some econometric model in an out-of-sample fashion, in which case $t = 1$ is the first out-of-sample observation and $T - h$ is the size of the out-of-sample portion of the sample. Even though we don't make it explicit for notational convenience, the forecasts in this case will depend on parameters estimated using in-sample data by any of the three schemes usually employed in the literature: a fixed scheme (where all forecasts depend on the same in-sample parameters estimated using data up to time $t = 1$), rolling scheme (the time-t forecast depends on parameters estimated using the most recent R observations, for some arbitrary R) or recursive scheme (the time-t forecast depends on parameters estimated using all observations in the sample up to time t).

We assume that the user wishes to incorporate in the forecasts a set of theoretical restrictions given by k moment conditions:

$$E_t[g(Y_{t+h}, \theta_0)] = 0, \tag{7}$$

which could for example be (a subset of) the equilibrium conditions from a structural economic model, such as Euler conditions.

We allow for the possibility that the moment conditions only involve a subset of the components of Y_t.

A key simplifying assumption that we make in the paper is that θ_0 is either known (e.g., $g(\cdot)$ are moments of the distribution from surveys of forecasters data), calibrated (e.g., $g(\cdot)$ are equilibrium conditions from a calibrated structural model) or estimated on a different data set than the one on which the forecasts are evaluated (e.g., θ_0 are deep parameters characterizing tastes and technology that can be estimated using microeconometric data). In addition, for example when some of the parameters do not have an obvious structural interpretation, θ_0 could be estimated using in-sample data, in which case the moment conditions should be interpreted as being conditional on the in-sample parameter estimate $\widehat{\theta}_t$, so that the theoretical restriction (7) becomes

$$E_t[g(Y_{t+h}, \widehat{\theta}_t)] = 0. \tag{8}$$

The starting point and motivation for our procedure is the acknowledgment that the initial density forecasts f_t do not necessarily satisfy the moment conditions (7), in the sense that $\int g(x, \theta_0) f_t(x) dx$ may not be zero at time t. In order to incorporate the information contained in the theoretical restrictions, we propose projecting each density forecast onto the space of densities that satisfy the moment conditions, which yields a new sequence of density forecasts $\left\{ \widetilde{f}_t(y_{t+h}) \right\}_{t=1}^{T-h}$ that by construction satisfy the moment conditions. The projected density forecast at time t, $\widetilde{f}_t(y_{t+h})$, is the (unique) density which, out of all the densities that satisfy the moment conditions, is the closest to the initial density $f_t(y_{t+h})$ according to a Kullback-Leibler measure of divergence. The following proposition shows how to construct the projected density forecasts.

Proposition 1. *If a a solution $\widetilde{f}_t\,(y_{t+h})$ to the constrained minimization*

$$\begin{aligned} \min_{h_t \in \mathcal{H}} \int \log \frac{h_t(x)}{f_t(x)} h_t(x) dx \\ s.t. \int g(x, \theta_0) h_t(x) dx &= 0, \end{aligned} \tag{9}$$

exists, then it is unique and it is given by

$$\widetilde{f}_t\left(y_{t+h}\right) = f_t(y_{t+h}) \exp\left\{\eta_t + \lambda_t' g(y_{t+h}, \theta_0)\right\}, \tag{10}$$

where η_t *and* λ_t *solve*

$$\begin{aligned} \lambda_t &= \min_{\lambda} \int f_t(x) \exp\left\{\lambda' g(x, \theta_0)\right\} dx \\ \eta_t &= \log\left\{\int f_t(x) \exp\left\{\lambda_t' g(x, \theta_0)\right\} dx\right\}^{-1}. \end{aligned} \tag{11}$$

Proof. The optimization problem can be restated as

$$\begin{aligned} \min_{h_t \in \mathcal{H}} \int \gamma\left(\frac{h_t(x)}{f_t(x)}\right) f_t(x) dx \\ s.t. \int g(x, \theta_0) h_t(x) dx &= 0, \end{aligned} \tag{12}$$

where $\gamma(u) = u \log u$, is a convex function. By convexity, we have that $\gamma(u) \leqslant \gamma(v) + \gamma'(v)(u - v)$. Let $h_t(x)$ be any feasible density. Feasible means that $\int g(x, \theta_0)\tilde{f}_t(x)dx = 0$. Then, when evaluated at $u = h_t(x)/f_t(x)$ and $v = \tilde{f}_t(x)/f_t(x) = \exp(\eta_t + \lambda_t g(x, \theta_0))$, the inequality above can be rewritten as

$$\gamma\left(\frac{h_t(x)}{f_t(x)}\right) \leqslant \gamma\left(\frac{\tilde{f}_t(x)}{f_t(x)}\right) + \left(\eta_t + \lambda_t' g(x, \theta_0) + 1\right)\left(h_t(x)/f_t(x) - \tilde{f}_t(x)/f_t(x)\right). \tag{13}$$

Integrating both sides with respect to $f_t(x)$, and using the feasibility of $h_t(x)$ and the optimality of $\tilde{f}_t(x)$, we obtain

$$\int \gamma\left(\frac{h_t(x)}{f_t(x)}\right) f_t(x)dx \leqslant \int \gamma\left(\frac{\tilde{f}_t(x)}{f_t(x)}\right) f_t(x)dx, \tag{14}$$

from which the result follows. □

Proposition 1 does not give conditions for the existence of the density. More simply it says that, if a solution exists, it must be of the exponential form given in (10). In facts, giving conditions under which the solution to the constrained optimization exists is a nontrivial task. A sufficient condition is that the set of densities satisfying the moment conditions is closed. The problem in (10) is often referred to as a (constrained) maximum entropy problem, see, among other, Jaynes (1968); Csiszár (1975); Golan et al. (1996); Golan (2002, 2008); Maasoumi (1993, 2007).

The construction of the projected density at each step t in practice can be carried out by numerically approximating the integrals in (11) and thus implementing the following steps:

1. For each $t = 1, ..., T - h$, generate S draws $\left\{y_{t+h}^s\right\}_{s=1}^S$ from $f_t(y_{t+h})$
2. Solve $\lambda_t = \min_\lambda \frac{1}{S}\sum_{s=1}^S f_t(y_{t+h}^s) \exp\left\{\lambda' g(y_{t+h}^s, \theta_0)\right\}$

3. Let $\eta_t = \log \left\{ \frac{1}{S} \sum_{s=1}^{S} f_t(y^s_{t+h}) \exp\left\{\lambda_t' g(y^s_{t+h}, \theta_0)\right\}\right\}^{-1}$

The following proposition shows that the projected density forecast obtained by the method just described is weakly more accurate than the initial forecast, when accuracy is measured by the logarithmic scoring rule (e.g., Amisano and Giacomini, 2007).

Proposition 2. *Consider the logarithmic scoring rule for the $h-$step ahead density forecast f_t :*

$$L(f_t, Y_{t+h}) = \log f_t(Y_{t+h}),$$

where Y_{t+h} is the realization of the variable at time $t+h$. A density forecast f_t is more accurate the larger the expected value of $L(f_t, Y_{t+h})$. If $E_t[g(Y_{t+h}, \theta_0)] = 0$ for all t, then

$$E\left[L\left(\widetilde{f}_t, Y_{t+h}\right) - L\left(f_t, Y_{t+h}\right)\right] \geq 0 \text{ for all } t.$$

Proof. Note that

$$\begin{aligned} E\left[L\left(\widetilde{f}_t, Y_{t+h}\right) - L\left(f_t, Y_{t+h}\right)\right] &= E\left[\log f_t(Y_{t+h}) + \eta_t + \lambda_t' g(Y_{t+h}, \theta_0) - \log f_t(Y_{t+h})\right] \\ &= E\left[E_t\left(\eta_t + \lambda_t' g(Y_{t+h}, \theta_0)\right)\right] = E\left[\eta_t\right]. \end{aligned}$$

We show that $\eta_t \geqslant 0$, for each t, which in turn implies that $E[\eta_t] \geq 0$. By the information inequality (e.g., Theorem 2.3 of White (1996)) we have that

$$\int \log \frac{\widetilde{f}_t(x)}{f_t(x)} \widetilde{f}_t(x) dx \geqslant 0,$$

with equality if and only $f_t = \widetilde{f}_t$, almost surely. Since $\widetilde{f}_t(x) = \exp(\eta_t + \lambda_t g(x, \theta_0)) f_t(x)$, we have that

$$\begin{aligned} 0 &\leqslant \int \log \frac{\widetilde{f}_t(x)}{f_t(x)} \widetilde{f}_t(x) dx = \int \log \frac{\exp(\eta_t + \lambda_t g(x, \theta_0)) f_t(x)}{f_t(x)} \widetilde{f}_t(x) dx \\ &= \int \log \exp(\eta_t + \lambda_t g(x, \theta_0)) \widetilde{f}_t(x) dx = \int \eta_t \widetilde{f}_t(x) dx + \lambda_t \int g(x, \theta_0) \widetilde{f}_t(x) dx \\ &= \eta_t, \end{aligned}$$

where the last equality follows from the fact that $\widetilde{f}_t(x)$ satisfies the moment conditions by construction. □

4 Empirical illustration: incorporating Euler equation restrictions into BVARs

In this section we show an application of our techniques to macroeconomic forecasting using the US dataset described by Stock and Watson (2008). This dataset has been widely investigated, and a number of econometric methods have been shown to produce accurate forecasts, including forecast combinations, factor models and Bayesian shrinkage. All of these methods tend to perform equally well in applications (e.g., Giacomini and White, 2006 and De Mol, Giannone, and Reichlin, 2008), so we choose one method, a BVAR, as representing the best currently known method for forecasting key macroeconomic variables in the Stock and Watson (2008) dataset. The goal of this application is to ask whether the already accurate but "atheoretical" forecasts based on the BVAR can be further improved by incorporating the economic restrictions embedded in a simple Euler equation.

Our starting point is the sequence of $h-$step ahead density forecasts based on the model considered by Giannone, Lenza and Primiceri (2010), which is a BVAR(4):

$$\begin{aligned} Y_t &= C + B_1 Y_{t-1} + \ldots + B_p Y_{t-4} + \varepsilon_t \qquad (15) \\ \varepsilon_t &\sim N(0, \Sigma) \end{aligned}$$

with a prior that is a natural conjugate variant of the Minnesota prior of Doan et al. (1984) and Litterman (1986), also considered by Banbura et al. (2010). This variant shrinks all VAR coefficients towards zero except for coefficients on the own lags of each component of Y_t, which are either set to one (for variables that exhibit persistence) or to zero (for close to stationary variables). The degree of shrinkage is controlled by a single hyperparameter. See Giannone et al. (2010) for details on the model and the priors. Note that the BVAR predictive density for the one-step ahead forecast is a multivariate normal, whereas for longer horizons the predictive density is obtained by simulation.

We will present results for various forecast horizons h. The BVAR point forecasts are the expected value of the predictive density

$$E_t^f Y_{t+h} = \int y_{t+h} f_t(y_{t+h}) dy_{t+h}, \qquad (16)$$

which can be approximated by the average of the draws:

$$\widehat{Y}_{t+h}^f = \frac{1}{S} \sum_{s=1}^{S} y_{t+h}^s. \qquad (17)$$

The theoretical moment restriction that we seek to incorporate into the BVAR density forecasts is the Euler equation:

$$E_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\alpha} (1 + R_{t+1}) - 1 \right] = 0, \qquad (18)$$

which for the $h-$step forecast horizon implies

$$E_t \left[\beta \left(\frac{C_{t+h}}{C_{t+h-1}} \right)^{-\alpha} (1 + R_{t+h}) - 1 \right] = 0, \qquad (19)$$

where C_t is nondurable and service real consumption, and R_{t+1} is the real yield of a one year bond. The bond yield is deflated by using the Consumer Price Index.

When imposing the Euler equation restrictions on the BVAR density we have the choice of interpreting (19) as a restriction on the levels of consumption or on consumption's growth rates. Note that for the levels case, equation (19) is a restriction on the joint density of Y_{t+h} and Y_{t+h-1}, which can be obtained as $f(y_{t+h}, y_{t+h-1}) = f(y_{t+h-1})f_{t+h-1}(y_{t+h})$. To construct the projected density forecast at time t, $\tilde{f}_t(y_{t+h})$, we first obtain the projected joint density as

$$\tilde{f}(y_{t+h}, y_{t+h-1}) = f(y_{t+h}, y_{t+h-1}) \exp(\eta_t + \lambda_t g(y_{t+h}, y_{t+h-1})), \tag{20}$$

where

$$g(Y_{t+h}, Y_{t+h-1}) = \beta \left(\frac{C_{t+h}}{C_{t+h-1}} \right)^{-\alpha} (1 + R_{t+h}) - 1 \tag{21}$$

and (η_t, λ_t) are constructed by drawing $\{y^s_{t+h}, y^s_{t+h-1}\}^S_{s=1}$ from $f(y_{t+h}, y_{t+h-1}) = f(y_{t+h-1})f_{t+h-1}(y_{t+h})$ and solving

$$\arg\min_{(\eta,\lambda)} \sum_{s=1}^{S} \exp(\eta + \lambda g(y^s_{t+h}, y^s_{t+h-1})). \tag{22}$$

The $h-$step ahead point forecast at time t based on the projected density is then given by

$$E^{\tilde{f}}_t Y_{t+h} = \int y_{t+h} \tilde{f}(y_{t+h}, y_{t+h-1}) dy_{t+h} dy_{t+h-1}, \tag{23}$$

which can be approximated by

$$\widehat{Y}^{\tilde{f}}_{t+h} = \sum_{s=1}^{S} w_{ts} y^s_{t+h}, \tag{24}$$

with $w_{ts} = \exp(\eta_t + \lambda_t g(y^s_{t+h}, y^s_{t+h-1}))/S$.

The next section reports the results of an out-of-sample exercise comparing the performance of the BVAR to that of the projected BVAR that incorporates the Euler equation restrictions. We will conduct an out-of-sample evaluation exercise that compares the point forecast accuracy of the two models in terms of the mean square forecast error(MSFE) of the point forecast $\widehat{Y}_{t+h}$:

$$MSFE = \frac{1}{T-h} \sum_{t=1}^{T-h} \left[(Y_{t+h} - \widehat{Y}_{t+h})^2 \right], \tag{25}$$

and the density forecast accuracy in terms of the average logarithmic scoring rule of the density forecast f_t:

$$\frac{1}{T-h} \sum_{t=1}^{T-h} \log f_t(Y_{t+h}). \tag{26}$$

4.1 Results

Similarly to Giannone et al (2010), we consider a subset of Stock and Watson's (2008) dataset which originally contained 149 U.S. quarterly variables covering a broad range of categories including income, industrial production, capacity, employment and unemployment, consumer prices, producer prices, wages, housing starts, inventories and orders, stock prices, interest rates for different maturities, exchange rates, and money aggregates. The time span is from the first quarter of 1959 through the last quarter of 2008.

We present results for both a "small" model with 5 variables (GDP, CPI, Federal Fund Rate, the yield on the one year bond, Consumption) and a "medium" model which contains 25 variables including all the variables considered by Smets and Wouters (2007) as well as investment, hours worked, wages and other labor market, financial and monetary variables.

Both small and medium BVARs are estimated in an out-of-sample fashion using a rolling window of length $R = 80$ and for forecast horizons $h = 1, 4, 8$. The procedure yields a total of $T - h = 114 - h$ BVAR forecast densities $f_t(y_{t+h})$. We have tried different value for the hyperparameters that regulates the amount of shrinkage and, although the analysis is quantitatively sensitive to this choice, the qualitative results are rather robust. For this reason, we only present the results for a single hyperparameter.

We then construct the projected density forecasts that incorporate the Euler equation (19), for $\beta = 0.96$ and $\alpha = \{1, 1.3, 1.6, 2, 2.3, 2.6\}$. For these combinations of parameters, we fail to reject at the 5% significance level the null hypothesis that the moment conditions hold.

To motivate our method and give a sense of the gains in forecast performance that one could expect, it is useful to ask whether the density forecasts implied by the BVAR already satisfy the moment conditions, in which case the scope for improvement when using our method would be minimal. For this purpose, Figure 2 reports the Euler equation errors implied by the BVAR and computed as

$$\frac{1}{S}\sum_{s=1}^{S} g(y^s_{t+h}, y^s_{t+h-1}),$$

where $g(\cdot)$ is defined in (21) for $\beta = 0.96$, $\alpha = 2$ and (y^s_{t+h}, y^s_{t+h-1}) are draws from the BVAR as in (22). The dotted lines represent the associated 95% predictive interval.

[Figure 2 about here.]

Figure 2 confirms that there are periods in which the densities implied by the BVAR do not satisfy the Euler equation. This is particularly noticeable for the small BVAR and for forecast horizon $h = 1$, in which case one would expect to observe gains from imposing the Euler restrictions, whereas the small BVAR densities for $h = 8$ appear to already satisfy the Euler equations and one would not therefore expect major gains in forecasting performance from the projection step. These predictions are borne out in the MSFE comparisons presented in Table 1 and discussed in the next section. In the following sections we investigate whether the extent to which the BVAR densities fail to satisfy the theoretical restrictions is sufficient to induce significant improvements in forecasting performance when forecasting with Euler-projected densities.

4.1.1 Point forecast performance

Table 1 reports the MSFE ratios of the point forecasts implied by the small BVAR relative to those implied by the Euler-projected BVAR. An entry greater than one indicates that the projected forecasts are more accurate and stars signal rejection at the 5equal MSFE according to the Giacomini and White (2006) test. Table 2 contains the results for the same selection of five variables but when the initial forecast is that implied by the medium BVAR.

[Table 1 about here.]

[Table 2 about here.]

From Table 1 we see that in the small BVAR the incorporation of the Euler restrictions in the vast majority of cases results in an improvement in accuracy, but the improvement is only statistically significant for the one-step-ahead forecast of CPI, for which the accuracy gains are moreover sizable. Table 2 reveals that in the context of the medium BVAR there are generally gains from imposing the Euler condition, but the gains are not significant except in the case of the four-step-ahead forecast of consumption. The results in Tables 1 and 2 are robust to a number of different choices for the Euler equation parameter α and for the hyperparameter that controls the amount of shrinkage in the BVAR.

To gain some insight into how and why the projection step results in more accurate forecasts, we focus on the one-step-ahead forecast of CPI in the small model, which is the case in which we observed the largest gains from imposing the Euler equation restrictions. To understand whether the superior performance is due to a few isolated episodes as opposed to being observed throughout the sample, Figure 3 plots the difference in absolute forecast errors for the BVAR and the Euler-projected BVAR.

[Figure 3 about here.]

The figure shows that the Euler-projected BVAR is more accurate than the BVAR in the vast majority of the sample, and the magnitude of the improvement is particularly large in the first part of the sample.

4.1.2 Density forecast performance

Here we present results about the relative performance of the density forecast implied by the BVAR relative to that of the Euler-projected BVAR, which considers the multivariate density as a whole. Tables 3 and 4 report the relative accuracy of the two forecasts in the small and the medium BVAR, respectively, and for different forecast horizons and choices of the Euler equation parameter α. As we can see from Table 3, in the case of the small BVAR there appears to be no improvement in the density forecast accuracy of the model as a whole when imposing the Euler condition, as indicated by values of the Amisano and Giacomini (2007) test statistic which are all insignificant. The situation is radically different for the medium BVAR, for which Table 4 reveals that there are significant improvements in the density forecast performance deriving from the incorporation of the Euler equation restrictions, at medium and long forecast horizons. The improvements are moreover robust to a number of different choices for the Euler equation parameters and the forecast horizons.

[Table 3 about here.]

[Table 4 about here.]

4.1.3 Summary of empirical results and comments

Our results generally point to the conclusion that there can be improvements in accuracy from incorporating theoretical restrictions into an already accurate atheoretical forecast. In terms of density forecast performance, the gains in accuracy were large and significant when the initial forecast was based on the medium BVAR and for medium and long forecast horizons, whereas there were no significant gains when the initial forecast was from the small BVAR. In terms of point forecast performance, the improvements were only significant in two cases (the one-step-ahead forecast of CPI implied by the small BVAR and the four-step-ahead forecast of real nondurable consumption implied by the medium BVAR). When they did occur, the gains in accuracy were of sizable magnitude. The fact that imposing the Euler condition improved the CPI forecast in the small BVAR but not in the medium BVAR can be reconciled with the findings in Stock and Watson (1999), who showed that the best model for forecasting U.S. inflation is an econometric model that incorporates an index of aggregate economic activity based on a large number of variables. This suggests that one possible reason for the lack of gains from projection that we observed in the medium BVAR is that this model is large enough to already capture something akin to the index of aggregate activity, whereas the small BVAR is too small to do so and therefore its forecasts still have scope for improvement. There are several directions in which our empirical analysis could be extended, for example by incorporating additional economic restrictions and/or investigating whether the low gains from the procedure could be due to misspecification of the Euler condition. We do not pursue these extensions in this paper and leave them for future research.

5 Conclusion

Economic theory often implies restrictions on the joint distribution of variables that are expressed as (nonlinear) moment conditions, such as Euler equations, which do not generally provide conditional densities that can be used for forecasting. On the other hand, there are several methods in the literature that have been shown to provide accurate forecasts, but they are usually based on atheoretical econometric models that do not satisfy the theory-based restrictions. We bridge this gap by proposing a method that takes as the starting point a density forecast, such as an accurate forecast implied by an econometric model, and modifies it in a way that ensures that the new density forecast satisfies the theoretical restrictions. The incorporation of the theoretical restrictions is achieved by a projection step which involves solving a relatively simple numerical optimization problem, whose complexity grows with the number of restrictions one wishes to impose.

We illustrate our method with an application to the Stock and Watson (2008) dataset that asks whether imposing the restrictions implied by an Euler condition can improve the point- and density forecasts implied by a BVAR - currently known as one of the best methods for forecasting a number of key macroeconomic variables. We find that even imposing such a simple restriction can sometimes give sizable gains in forecast accuracy, which suggests that it is worthwhile to use our methods to further investigate whether economic theory does indeed "tell a story" that forecasters should listen to.

References

Aiolfi, M., C. Capistrán Carmona, and A. Timmermann (2010): "Forecast combinations," *CREATES Research Paper No. 2010-21.*

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

An, S. and F. Schorfheide (2007): "Bayesian analysis of DSGE models," *Econometric Reviews*, 26, 113–172.

Ang, A. and M. Piazzesi (2003): "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables* 1," *Journal of Monetary economics*, 50, 745–787.

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

Carriero, A. and R. Giacomini (2011): "How useful are no-arbitrage restrictions for forecasting the term structure of interest rates?" *Journal of Econometrics*.

Christoffel, K., G. Coenen, and A. Warne (2010): "Forecasting with DSGE models," *ECB Working Paper No. 1185.*

Csiszár, I. (1975): "I-Divergence Geometry of Probability Distributions and Minimization Problems," *The Annals of Probability*, 3, 146–158.

De Mol, C., D. Giannone, and L. Reichlin (2008): "Forecasting using a large number of predictors: Is Bayesian regression a valid alternative to principal components?" *Journal of Econometrics*, 146, 318.

Del Negro, M. and F. Schorfheide (2004): "Priors from general equilibrium models for VARs," *International Economic Review*, 643–673.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3, 1–100.

Edge, R., M. Kiley, and J. Laforte (2010): "A comparison of forecast performance between federal reserve staff forecasts, simple reduced-form models, and a DSGE model," *Journal of Applied Econometrics*, 25, 720–754.

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

Giacomini, R. and B. Rossi (2010): "Forecast comparisons in unstable environments," *Journal of Applied Econometrics*, 25, 595–620.

Giacomini, R. and H. White (2006): "Tests of conditional predictive ability," *Econometrica*, 74, 1545–1578.

Giannone, D., M. Lenza, and G. Primiceri (2010): "Prior selection for vector autoregressions," Tech. rep., mimeo.

Golan, A. (2002): "Information and entropy econometrics-editor's view," *Journal of Econometrics*, 107, 1–16.

——— (2008): *Information and Entropy Econometrics: A Review and Synthesis*, Now Publisher, Hanover, MA, USA.

Golan, A., G. Judge, and D. Miller (1996): *Maximum entropy econometrics: Robust estimation with limited data*, John Wiley & Sons Inc, New York, USA.

Inoue, A. and L. Kilian (2008): "How useful is bagging in forecasting economic time series? A case study of US consumer price inflation," *Journal of the American Statistical Association*, 103, 511–522.

Jaynes, E. (1968): "Prior probabilities," *IEEE Transactions on Systems Science and Cybernetics*, 4, 227–241.

Litterman, R. (1986): "A statistical approach to economic forecasting," *Journal of Business & Economic Statistics*, 1–4.

Maasoumi, E. (1993): "A compendium to information theory in economics and econometrics," *Econometric reviews*, 12, 137–181.

——— (2007): "5 On Econometric Methodology," *Economic Record*, 64, 340–343.

Raftery, A., D. Madigan, and J. Hoeting (1997): "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, 179–191.

Schorfheide, F. (2000): "Loss function-based evaluation of DSGE models," *Journal of Applied Econometrics*, 15, 645–670.

Smets, F. and R. Wouters (2003): "An estimated dynamic stochastic general equilibrium model of the euro area," *Journal of the European Economic Association*, 1, 1123–1175.

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

——— (2008): "Forecasting in dynamic factor models subject to structural instability," *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173–205.

White, H. (1996): *Estimation, inference and specification analysis*, vol. 22, Cambridge Univ Pr.

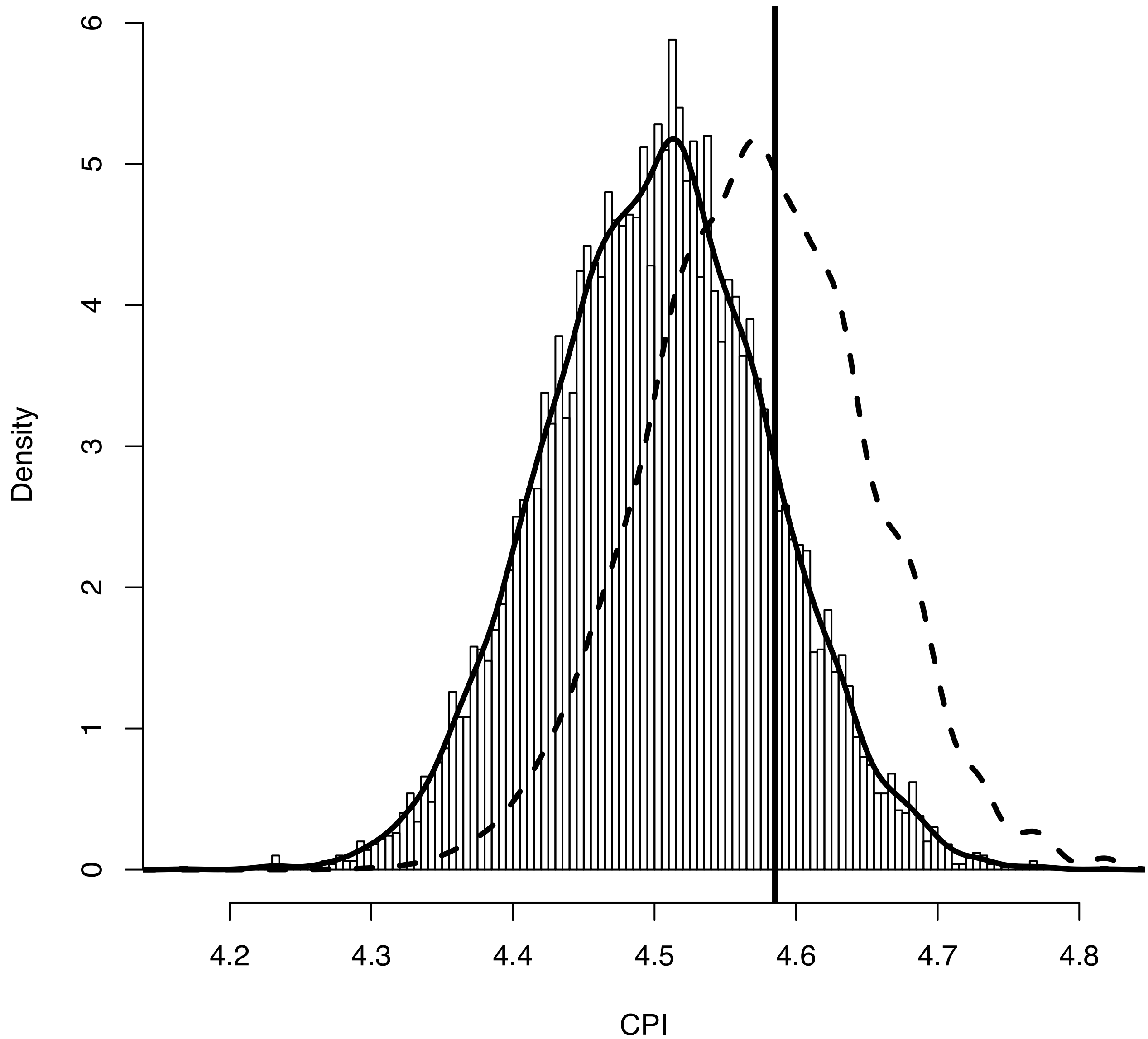

Figure 1: The figure shows two out-of-sample density forecasts for CPI at one particular point in time: the histogram is the one-step-ahead density forecast of CPI implied by a BVAR for CPI, GDP, nondurable and service real consumption, the federal funds rate and the real yield of a one year bond; the dashed line is the projected density forecast that incorporates the Euler equation restrictions

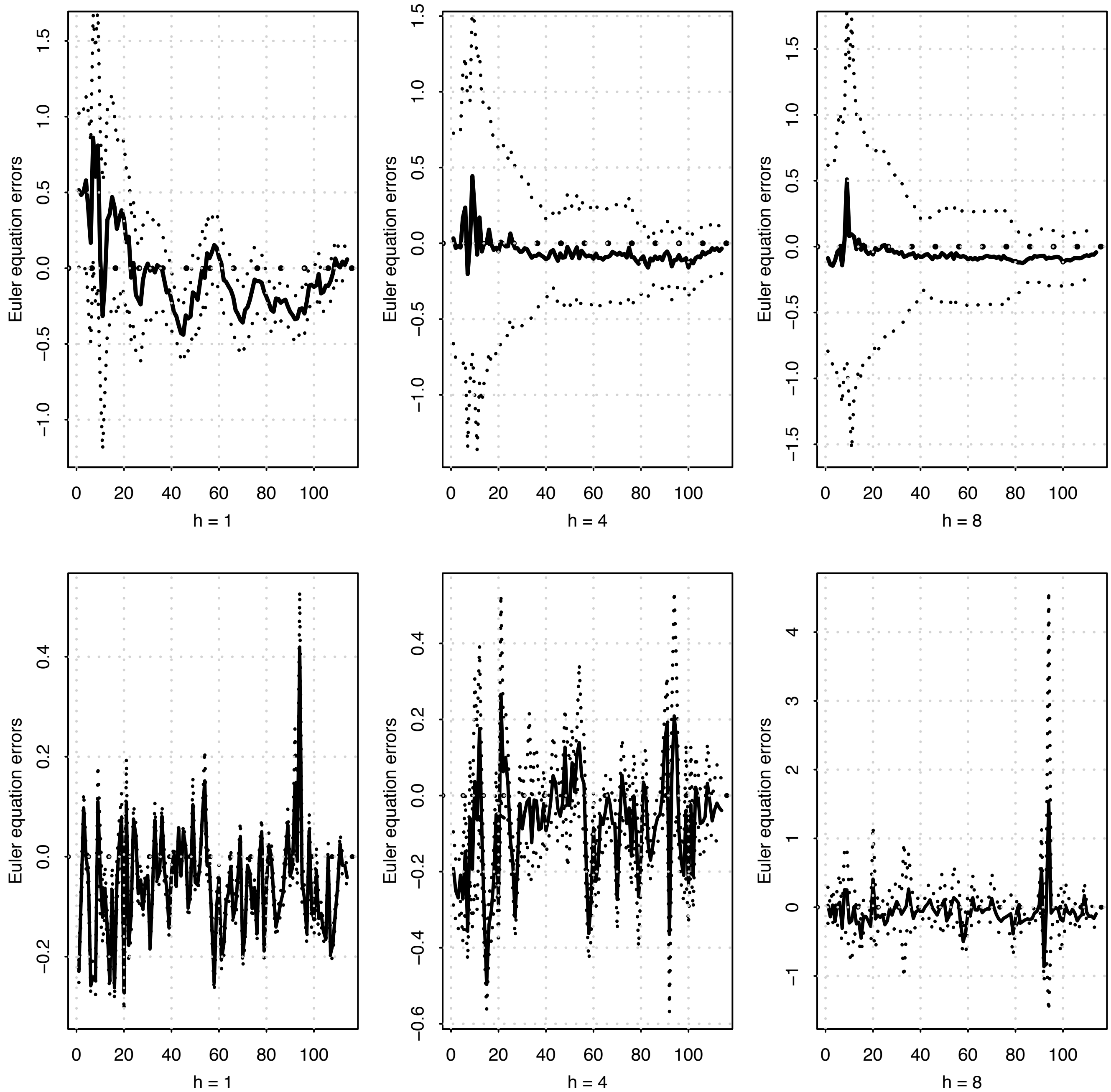

Figure 2: Euler equation error for forecast horizons $h = 1, 4$, and 8. Each panel plots the Euler equation ($\beta = .96$, and $\alpha = 2$) error under the approximated density implied by the small BVAR (top panels) and medium BVAR (bottom panels) and the 95% predictive bands.

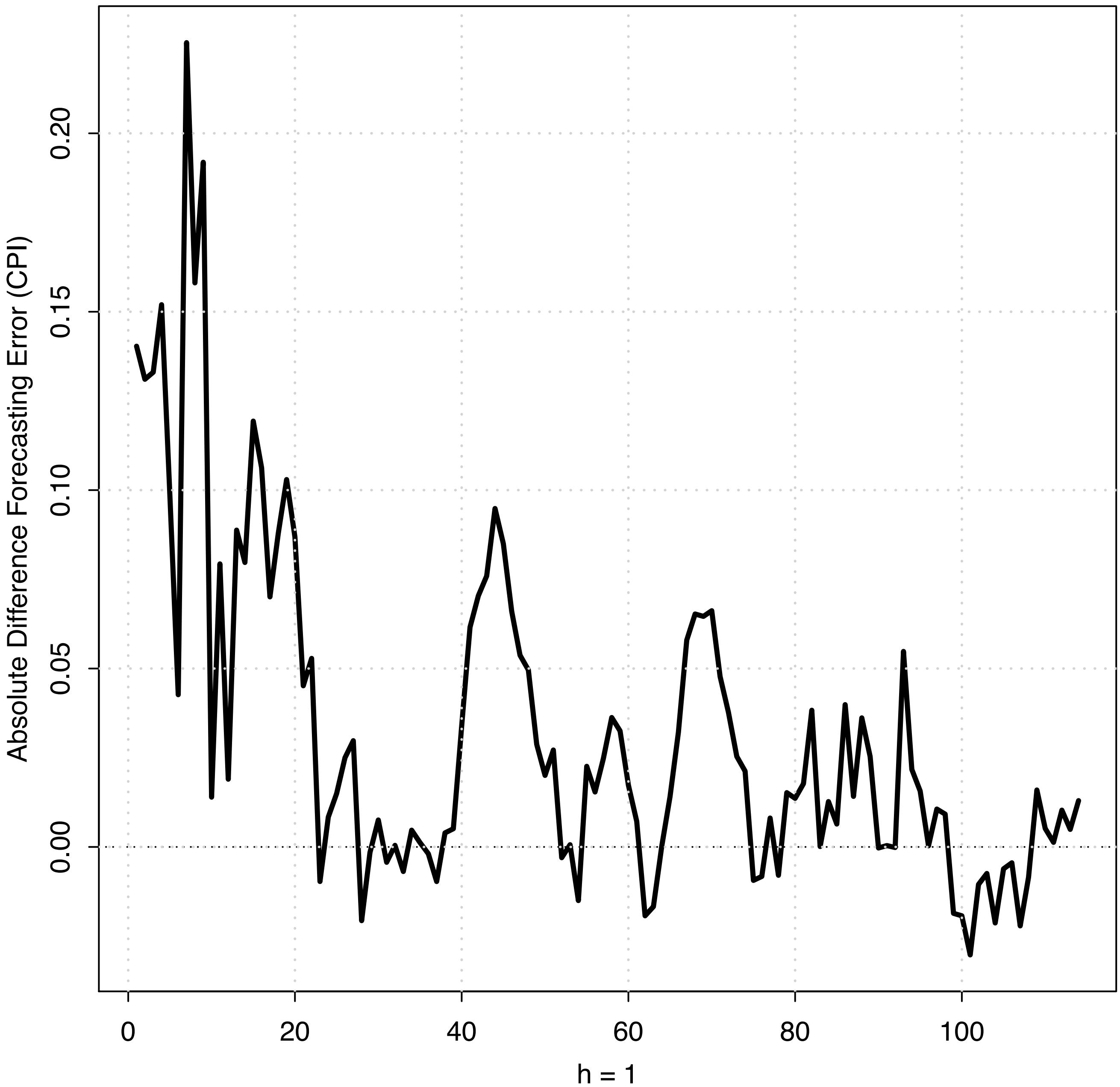

Figure 3: Absolute forecasting error of the CPI series. The black line is the difference between the absolute forecasting error of the Small BVAR and the absolute forecasting error of the Euler-projected forecast with parameters $\beta = .96$ and $\alpha = 2$.

	GDP	*CPI*	*FFR*	*Tbill*	*Cons*
			$\alpha = 1$		
h = 1	1.06	9.95*	0.87	1.02	0.95
h = 4	1.00	1.10	1.00	1.01	1.00
h = 8	1.00	1.04	1.00	1.00	0.99
			$\alpha = 1.6$		
h = 1	1.03	9.40*	0.85	1.03	0.80
h = 4	1.00	1.09	1.00	1.00	0.99
h = 8	1.00	1.04	0.99	1.00	0.99
			$\alpha = 2$		
h = 1	0.97	9.44*	0.85	1.03	0.71
h = 4	1.00	1.09	0.99	1.00	0.98
h = 8	1.00	1.03	0.99	1.00	0.98
			$\alpha = 2.3$		
h = 1	0.99	8.12*	0.85	1.04	0.68
h = 4	1.00	1.08	0.99	1.00	0.98
h = 8	1.00	1.01	0.99	1.00	0.98

Table 1: Point forecast performance. Each entry consists of the ratio between the MSFE of the Small BVAR and the MSFE of the Euler-projected BVAR. A star indicates that the difference is significant at the 5% level according to the Giacomini and White (2006) test.

	GDP	*CPI*	*FFR*	*Tbill*	*Cons*
			$\alpha = 1$		
h = 1	1.02	1.09	1.04	1.02	0.99
h = 4	0.98	1.02	1.06	0.99	1.33**
h = 8	1.02	1.03	1.06	1.04	1.08
			$\alpha = 1.6$		
h = 1	1.01	1.04	1.02	0.99	0.99
h = 4	1.06	1.00	1.08	1.00	1.53**
h = 8	1.01	1.03	1.04	1.03	1.07
			$\alpha = 2$		
h = 1	1.01	0.99	1.01	1.00	1.04
h = 4	1.08	1.01	1.08	0.99	1.58**
h = 8	1.02	1.02	1.02	1.00	1.09
			$\alpha = 2.3$		
h = 1	1.01	0.97	1.01	1.00	1.07
h = 4	1.08	1.02	1.08	0.98	1.60**
h = 8	1.01	1.02	1.03	1.02	1.05

Table 2: Point forecast performance. Each entry consists of the ratio between the MSFE of the Medium BVAR and the MSFE of the Euler-projected BVAR. Two stars indicates that the difference is significant at the 1% level according to the Giacomini and White (2006) test.

	$\alpha = 1$	$\alpha = 1.3$	$\alpha = 1.6$	$\alpha = 2$	$\alpha = 2.3$
$h = 1$	0.9764	0.8693	0.8143	0.7247	0.5439
$h = 2$	-0.0251	-0.0520	-0.0817	-0.1222	-0.1582
$h = 3$	-0.0555	-0.0808	-0.1084	-0.1457	-0.1785
$h = 4$	-0.0567	-0.0797	-0.1034	-0.1393	-0.1664
$h = 5$	-0.0659	-0.0872	-0.1064	-0.1418	-0.1635
$h = 6$	-0.0612	-0.0809	-0.1025	-0.1337	-0.1554
$h = 7$	-0.0539	-0.0695	-0.0907	-0.1172	-0.1363
$h = 8$	-0.0503	-0.0665	-0.0732	-0.1081	-0.1302

Table 3: Density forecast performance. Entries are the test statistics of the Amisano and Giacomini (2007) test comparing the small BVAR and the Euler-projected density forecasts. Positive (negative) values indicate that the projected density forecast is more accurate (less accurate) than the BVAR forecasts. Values outside the $(-1.96, 1.96)$ interval denote rejection of the null hypothesis of equal performance at the 5% significance level.

	$\alpha = 1$	$\alpha = 1.3$	$\alpha = 1.6$	$\alpha = 2$	$\alpha = 2.3$
h = 1	-0.824	-0.943	-2.021	-1.629	-0.598
h = 2	-0.208	-0.103	-0.748	0.407	0.530
h = 3	1.114	1.018	1.920	2.240	2.231
h = 4	3.508	3.520	3.223	3.267	3.292
h = 5	3.087	3.296	3.503	3.687	3.842
h = 6	2.758	3.144	3.570	4.366	4.205
h = 7	3.086	3.203	3.337	3.539	3.865
h = 8	3.137	2.734	3.039	4.330	3.615

Table 4: Density forecast performance. Entries are the test statistics of the Amisano and Giacomini (2007) test comparing the medium BVAR and the Euler-projected density forecasts. Positive (negative) values indicate that the projected density forecast is more accurate (less accurate) than the BVAR forecasts. Values outside the $(-1.96, 1.96)$ interval denote rejection of the null hypothesis of equal performance at the 5% significance level.