Theory-coherent forecasting*

Raffaella Giacomini
UCL, Cemmap, and CEPR

Giuseppe Ragusa
Luiss University

May 20, 2013

Abstract

We consider a method for producing multivariate density forecasts that satisfy moment restrictions implied by economic theory, such as Euler conditions. The method starts from a base forecast that might not satisfy the theoretical restrictions and forces it to satisfy the moment conditions using exponential tilting. Although exponential tilting has been considered before in a Bayesian context (Robertson et al., 2005), our main contributions are: 1) to adapt the method to a classical inferential context with out-of-sample evaluation objectives and parameter estimation uncertainty; and 2) to formally discuss the conditions under which the method delivers improvements in forecast accuracy. An empirical illustration which incorporates Euler conditions into forecasts produced by Bayesian vector autoregressions shows that the improvements in accuracy can be sizable and significant.

1 Introduction

Economic theory often provides moment conditions which restrict the dynamic behavior of key macroeconomic variables, but which cannot be used directly to produce forecasts that are coherent with theory. An expectational Euler condition, for example, imposes a nonlinear restriction on the joint density of future consumption and real interest rates, conditional on current observables, and thus potentially provides valuable information for forecasting the future path of both variables. This paper considers a method for constructing multivariate density forecasts that are coherent with theory. The idea is to start from a base multivariate density forecast which does not necessarily satisfy the theoretical restrictions and then use exponential tilting to obtain a new density forecast that by construction satisfies the moment conditions. In practice,

*We thank the Domenico Giannone, Norman Swanson, three anonymous referees and seminar participants at Princeton, Columbia, Institute for Advanced Studies at Vienna, University of Bath, Université libre de Bruxelles for useful comments.

the method consists of an importance sampling procedure where draws from the base density forecasts are reweighted using weights obtained by numerical optimization. The computational cost of the procedure is low, only depending on the number of theoretical restrictions one wishes to incorporate.

Even though exponential tilting has been considered before in microeconometrics and Bayesian econometrics (in particular by Robertson et al., 2005), the contribution of this paper is to formalize the method in a classical inferential context with out-of-sample evaluation objectives, address the issue of parameter estimation uncertainty and present formal conditions under which the incorporation of theoretical restrictions leads to forecast accuracy gains. Our main result shows that a tilted density forecast which incorporates moment conditions that are true in population but may depend on consistently estimated parameters is more accurate than the base density forecast, provided that accuracy is measured by the logarithmic scoring rule of Amisano and Giacomini (2007). One practical implication of this result is the recommendation to separately estimate the parameters in the base density and the parameters in the moment condition.

The paper offers a way to incorporate economic theory into forecasting without resorting to estimation of full-fledged Dynamic Stochastic General Equilibrium (DSGE) models, and is part of a small literature which proposes hybrid approaches that combine elements of economic theory and reduced-form modelling. There are several reasons why a hybrid approach might be an appealing alternative to forecasting with fully specified theoretical models. First, a DSGE model in general does not directly provide a conditional density that can be used for forecasting, which means that one typically would end up forecasting with an approximated density (either from the linearized model or, less frequently, from higher order approximations or numerical solutions of the model). A few articles have investigated the forecast performance of linearized DSGE models compared to reduced-form models (e.g., Smets and Wouters, 2003; Edge, Kiley, and Laforte, 2010; Christoffel, Coenen, and Warne, 2010), but it remains to be seen whether the results are robust to different choices of DSGE priors, different time periods and different specifications for the reduced form model, to name a few. Second, the user has to take a stand on many aspects of the model for which theory provides no guidance, necessitating ad hoc modeling choices. Third, DSGE models are not capable of incorporating the rich data sets that have proved helpful in reduced-form forecasting of variables such as inflation and real output. Examples of "hybrid" approaches have been considered in the context of linearized DSGE models are Schorfheide, 2000, who proposes priors based on the DSGE model to perform Bayesian inference in the VAR model, and Del Negro and Schorfheide, 2004 and Carriero and Giacomini, 2011, who consider combinations of the DSGE and the VAR model. Even though the focus of these methods is on estimation, they could in principle be used for forecasting. The method considered in this paper has several advantages over these existing hybrid approaches as it allows full flexibility in the choice of the base model, which can be driven by considerations about its empirical performance instead of the requirement that the base model contain the same variables as the DSGE model. Further, it does not require one to put equal faith in all of the restrictions embedded in a DSGE model, but to choose which restrictions to impose. Finally, the restrictions can be nonlinear, whereas the approaches mentioned above necessitate restricting attention to linearized models.

A noteworthy feature of the tilting procedure is that it yields a new multivariate forecast which has a known analytical form but in general is not a member of a known family of distributions (for example, even if the base density is a multivariate normal, the tilted density will not be normal in all but a few special cases). The fact that the method gives a density that is not in the same family as the original density forecast is a useful feature of the approach, as it allows one to understand the effects of imposing theoretical restrictions on the entire shape of the distribution, including the marginal densities and the dependence structure. To illustrate how the tilting method can modify the marginal density forecast and why this could result in accuracy improvements for the individual variables, we show in Figure 1 an actual example from our empirical application. The application considers as a base model a Bayesian VAR with 22 variables including real consumption (C_t), the real return on the Fama-French (1993) portfolio R_t and real GDP, and incorporates the Euler condition $E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\alpha}R_{t+1}-1\right]=0$, with $\beta=0.999$ and $\alpha=0.6$.

[Figure 1 about here]

The left panel of Figure 1 shows the effect of tilting at a particular point in time on the density forecast for R_t, a variable which is directly restricted by the Euler equation, whereas the right panel shows that the tilting also affects the density forecast for real GDP, which does not enter the Euler condition directly but is nonetheless indirectly modified by the procedure along with all other variables in the base model. The histograms represent the density forecasts implied by the base model and the dashed lines are the tilted density forecasts. The vertical lines are the realizations of the variables. In both cases the incorporation of the Euler equation restrictions modifies the shape of the density forecasts implied by the base model by making them left skewed and by shifting them towards the actual realizations of the variables, thus yielding more accurate point- and density forecasts for both variables. Even though the figure only illustrates the benefits of tilting for asset returns at one point in time, we show in Section 4 that this tends to be true on average.

2 Motivating example

This section shows a simple example where an analytical expression for the tilted density can be easily obtained, which provides some intuition for the 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:

$$E_t[Y_{t+1} - \mu_t] = \int (y_{t+1} - \mu_t) h_t(y_{t+1}) dy_{t+1} = 0.$$

Suppose that one has available a one-step-ahead density forecast $f_t(y) \sim N(\widehat{\mu}_t, 1)$, which does not necessarily satisfy the moment conditions in that $\widehat{\mu}_t$ may be different from μ_t. In order to obtain a new density forecast which by construction have the correct mean, the tilting procedure finds a new density forecast $\widetilde{f}_t(y)$ with mean μ_t and which is closest to $f_t(y)$ according to the Kullback-Leibler measure of divergence. The solution of this constrained optimization can be shown to be

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

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

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

which shows that η_t and τ_t are related to the degree of misspecification in the conditional mean of the initial forecast, with the limiting case $\eta_t = \tau_t = 0$ occurring when the initial density satisfies the moment conditions and thus no tilting is necessary. In this simple example of a mean restriction on a single variable the tilting simply recenters the density forecast at the theoretical mean. It also happens that the projected density (1) is a normal with mean μ_t and variance 1, but the tilted density would no longer be 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 an Euler equation on a multivariate density forecast implied by a Bayesian VAR. 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.

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 base and the tilted densities from the true density. We will show in the next section that $\Delta KLIC_t = E\left[\eta_t\right]$, which in our simple example (2) cannot be negative, implying that the tilted density is weakly closer to the true density than the base density. Our main result is to prove that this is true under more general conditions.

3 The tilting method

This section illustrates the exponential tilting method for producing multivariate density forecasts that satisfy theoretical moment conditions.

Suppose the objective at a given time t is to construct an $h-$step ahead density forecast for the $n \times 1$ random vector Y_{t+h}, based on the information set Ω_t. This could represent a step in an out-of-sample evaluation procedure, in which case t is a point in the out-of-sample portion of the sample of size P and the tilting is thus implemented P times. We assume that the forecaster has available a base model that can be used to produce a density forecast $f_{t,h}\left(\cdot\right)$. It is not necessary for the base forecast to have a closed-form analytical expression, but it must be possible to generate draws from it. Even though we don't make it explicit for notational convenience, the base density forecast can depend on parameters estimated using in-sample data. Any estimation procedure and any of the three schemes usually employed in the literature are allowed: a fixed scheme (where all forecasts depend on the same in-sample parameters estimated using data dated 1 to R, for some arbitrary R), a rolling scheme (where the time-t forecast depends on parameters estimated using the most recent R observations) or a recursive scheme (where 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 into the forecast at time t a set of theoretical restrictions expressed as k moment conditions:

$$E_t[g(Y_{t+h}, \theta_0)] = 0, \tag{3}$$

which could for example be (a subset of) the equilibrium conditions from a structural economic model, such as Euler conditions. The subscript t indicates conditioning on the information set Ω_t. We allow for the possibility that the moment conditions only involve a subset of the components of Y_{t+h}.

We start by considering the case in which θ_0 is either known or calibrated (e.g., $g\left(\cdot\right)$ are equilibrium conditions from a calibrated structural model). We then relax this assumption and consider the situation in which the moment conditions depend on parameters estimated in-sample.

The starting point and motivation for the procedure is the acknowledgment that the base density forecast $f_{t,h}$ does not necessarily satisfy the moment conditions (3), in the sense that $\int g(y,\theta_0)f_{t,h}(y)dy$ may not be zero at time t. The tilting procedure yields a new density forecast $\widetilde{f}_{t,h}$ which is the (unique) density which, out of all the densities that satisfy the moment conditions, is the closest to the base density $f_{t,h}$ according to a Kullback-Leibler measure of divergence. The following proposition shows how to construct the tilted density forecast.

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

$$\min_{h_{t,h}\in\mathcal{H}} \int \log \frac{h_{t,h}(y)}{f_{t,h}(y)} h_{t,h}(y)dy \tag{4}$$

$$s.t. \int g(y,\theta_0)h_{t,h}(y)dy = 0, \tag{5}$$

exists, then it is unique and it is given by

$$\widetilde{f}_{t,h}(y) = f_{t,h}(y)\exp\left\{\eta_{t,h}(\theta_0) + \tau'_{t,h}(\theta_0)g(y,\theta_0)\right\}, \tag{6}$$

where $\eta_{t,h}(\theta_0)$ and $\tau_{t,h}(\theta_0)$ are given by

$$\begin{aligned}
\tau_{t,h}(\theta_0) &= \arg\min_{\tau} \int f_{t,h}(y)\exp\left\{\tau' g(y,\theta_0)\right\}dy \qquad (7)\\
\eta_{t,h}(\theta_0) &= \log\left\{\int f_{t,h}(y)\exp\left\{\tau_{t,h}(\theta_0)' g(y,\theta_0)\right\}dy\right\}^{-1}.
\end{aligned}$$

Proof. The optimization problem can be restated, after indexes are dropped, as

$$\begin{aligned}
\min_{h\in\mathcal{H}} \int \gamma\left(\frac{h(y)}{f(y)}\right) f(y)dy & \qquad (8)\\
s.t. \int g(y,\theta_0)h(y)dy &= 0,
\end{aligned}$$

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 $h(y)$ be any feasible density. Feasible means that $\int g(y,\theta_0)h(y)dy = 0$. Then, when evaluated at $u = h(y)/f(y)$ and $v = \tilde{f}(y)/f(y) = \exp(\eta + \tau' g(y,\theta_0))$, the inequality above can be rewritten as

$$\gamma\left(\frac{h(x)}{f(x)}\right) \leqslant \gamma\left(\frac{\tilde{f}(x)}{f(x)}\right) + \left(\eta + \tau' g(x,\theta_0) + 1\right)\left(h(x)/f(x) - \tilde{f}(x)/f(x)\right). \tag{9}$$

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

$$\int \gamma\left(\frac{h(x)}{f(x)}\right) f(x)dx \leqslant \int \gamma\left(\frac{\tilde{f}(x)}{f(x)}\right) f(x)dx, \tag{10}$$

from which the result follows. □

The optimization problem described by (5) plays a basic role in the information-theoretic approach to statistics (see its origins in the information theoretic literature (Kullback, 1968; Jaynes, 1968; Csiszár, 1975)). It has also been studied in econometrics, where it has been applied to the problem of parameter estimation in moment condition models (Kitamura and Stutzer, 1997; Newey and Smith, 2004; Ragusa, 2011) and in other related settings (Golan et al., 1996; Golan, 2002, 2008; Maasoumi, 1993, 2007; Komunjer and Ragusa, 2009). Robertson et al. (2005) consider the procedure in a Bayesian setting.

The choice of the Kullback-Leibler measure of divergence is in principle not unique, and other measures of divergence could in principle be chosen. The two main reasons for choosing the Kullback-Leibler measure in our context are: 1) it provides a convenient analytical expression for the tilted density 2) unlike other measures of distance, it has a direct counterpart in the logarithmic scoring rule, which is a common and well-studied measure for evaluating density forecasts (Amisano and Giacomini, 2007).

It is important to note that 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 (6). In fact, giving conditions under which the solution to the constrained optimization exists is a nontrivial task. A sufficient condition for the existence of the projection density is given in Csiszár (1975) and it requires that the following hold

$$\int \sup_{\tau \in \mathcal{E}} \exp\left\{\tau' g(y, \theta_0)\right\} f_{t,h}(y) dy < \infty,$$

where $\mathcal{E} \subset \mathbb{R}^k$ is an open set containing 0. This condition, sometimes referred to as *weak Cramér condition*, can be interpreted as requiring that all exponential moments of g are finite. Whether this conditions is satisfied in practice depends both on the density $f_{t,h}$ and the function g.

The construction of the projected density at each step t in practice can be carried out by numerically approximating the integrals in (7) and thus implementing the following steps:

1. Generate D draws $\left\{y^d\right\}_{d=1}^{D}$ from the base density forecast $f_{t,h}(y)$
2. Numerically solve $\tau_{t,h} = \arg\min_{\tau} \frac{1}{D}\sum_{d=1}^{D} f_{t,h}(y^d)\exp\left\{\tau' g(y^d, \theta_0)\right\}$
3. Obtain $\eta_{t,h}$ as $\eta_{t,h} = \log\left\{\frac{1}{D}\sum_{d=1}^{D} f_{t,h}(y^d)\exp\left[\tau'_{t,h} g(y^d, \theta_0)\right]\right\}^{-1}$.

3.1 Main result

In this section we provide a theoretical justification for the tilting procedure, which was missing in Robertson et al. (2005). Specifically, we ask whether and in what sense incorporating a moment restriction into an existing density forecast can give improvements in forecast accuracy. The main results are Propositions 2 and 3, which show that, when the moment condition is true in population, incorporating it using the tilting method results in a weakly more accurate density 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 $f_{t,h}$:*

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

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

$$E\left[L(\widetilde{f}_{t,h}, Y_{t+h}) - L(f_{t,h}, Y_{t+h})\right] \geq 0.$$

Proof. We have:

$$\begin{aligned} E\left[L(\widetilde{f}_{t,h}, Y_{t+h}) - L(\widetilde{f}_{t,h}, Y_{t+h})\right] &= E\left[\log f_{t,h}(Y_{t+h}) + \eta_{t,h} + \tau'_{t,h} g(Y_{t+h}, \theta_0) - \log f_{t,h}(Y_{t+h})\right] \\ &= E[\eta_{t,h} + \tau'_{t,h} g(Y_{t+h}, \theta_0)] \\ &= \eta_{t,h}, \end{aligned}$$

where the last equality follows from the law of iterated expectations, the fact that $\eta_{t,h}$ and $\tau_{t,h}$ depend only on variables that are in the information set at time t and from $E_t[g(Y_{t+h}, \theta_0)] = 0$. We now show that $\eta_t \geqslant 0$. By the information inequality (e.g., Theorem 2.3 of White (1996)) we have that

$$\int \log \frac{\widetilde{f}_{t,h}(y)}{f_{t,h}(y)} \widetilde{f}_{t,h}(y) dy \geqslant 0,$$

with equality if and only $f_{t,h} = \widetilde{f}_{t,h}$, almost surely. Since $\widetilde{f}_{t,h}(y) = \exp(\eta_{t,h} + \tau'_{t,h} g(x, \theta_0)) f_{t,h}(y)$, we have that

$$\begin{aligned} 0 &\leqslant \int \log \frac{\widetilde{f}_{t,h}(y)}{f_{t,h}(y)} \widetilde{f}_{t,h}(y) dy = \int \log \frac{\exp(\eta_{t,h} + \tau'_{t,h} g(x, \theta_0)) f_{t,h}(y)}{f_{t,h}(y)} \widetilde{f}_{t,h}(y) dy \\ &= \int \log \exp(\eta_{t,h} + \tau'_{t,h} g(x, \theta_0)) \widetilde{f}_{t,h}(y) dy = \int \eta_{t,h} \widetilde{f}_{t,h}(y) dy + \tau'_{t,h} \int g(x, \theta_0) \widetilde{f}_{t,h}(y) dy \\ &= \eta_{t,h}, \end{aligned}$$

where, again, the last equality follows from the fact that $\widetilde{f}_{t,h}(y)$ satisfies the moment conditions by construction. The result then follows from the law of iterated expectations. We have that

$$\begin{aligned} E\left[L(\widetilde{f}_{t,h}, Y_{t+h}) - L(\widetilde{f}_{t,h}, Y_{t+h})\right] &= E\left\{E_t\left[L(\widetilde{f}_{t,h}, Y_{t+h}) - L(\widetilde{f}_{t,h}, Y_{t+h})\right]\right\} \\ &= E[\eta_{t,h}] \geq 0, \end{aligned}$$

as required. $\square$

Proposition 2 assumes knowledge of the true parameter vector θ_0. The next result shows that, at least asymptotically, the tilted forecast is weakly more accurate than the base forecast when the parameter in the moment condition can be consistently estimated using past observations. Let $\hat{\theta}_t$ denote an estimator of θ_0, based on Ω_t, and let $\widetilde{f}_{t,h}(Y_{t+h}, \hat{\theta}_t)$ denote the tilted density when the moment restriction is imposed with θ_0 estimated by $\hat{\theta}_t$.

Proposition 3. *Suppose that* $\hat{\theta}_t \xrightarrow{p} \theta_0$ *as* $t \to \infty$, *where* $\theta_0 \in \overline{\Theta}$, $\overline{\Theta}$ *a compact set; and (i)* $\eta_{t,h}(\theta)$, $\tau_{t,h}(\theta)$, *and* $g(\cdot, \theta)$ *are continuously differentiable over* $\theta \in \overline{\Theta}$ *for all* t; *(ii)* $\sup_t E\| \sup_{\theta\in\overline{\Theta}} \partial\eta_{t,h}(\theta)/\partial\theta'\| = O_p(1)$ *and* $\sup_t E\| \sup_{\theta\in\overline{\Theta}} \partial[\tau_{t,h}(\theta)' g(Y_{t+h}, \theta)]/\partial\theta'\| = O_p(1)$; *(iii)* $\sup_t E \sup_{\theta\in\overline{\Theta}} |\eta_{t,h}(\theta)|^2 < \infty$ *and* $\sup_t E[\sup_{\theta\in\overline{\Theta}} |\tau_{t,h}(\theta)' g(Y_{t+h}, \theta)|^2] < \infty$. *Then*

$$\lim_{t\to\infty} E\left[\log \frac{\widetilde{f}_{t,h}(Y_{t+h}, \hat{\theta}_t)}{f_{t,h}(Y_{t+h})}\right] = \Delta \geq 0.$$

Proof. Notice that

$$\log \frac{\widetilde{f}_{t,h}(Y_{t+h}, \hat{\theta}_t)}{f_{t,h}(Y_{t+h})} = \eta_{t,h}(\hat{\theta}_t) + \tau_{t,h}(\hat{\theta}_t)' g(Y_{t+h}, \hat{\theta}_t).$$

Conditions (i), a mean value expansion around θ_0, and the triangular inequality imply that we may write

$$\begin{aligned}
&|\eta_{t,h}(\hat{\theta}_t) + \tau_{t,h}(\hat{\theta}_t)' g(Y_{t+h}, \hat{\theta}_t) - \eta_{t,h}(\theta_0) + \tau_{t,h}(\theta_0)' g(Y_{t+h}, \theta_0)| \\
&\quad \leq |\partial\eta_{t,h}(\bar{\theta})/\partial\theta'(\hat{\theta}_t - \theta_0) + \partial[\tau_{t,h}(\bar{\theta})' g(Y_{t+h}, \bar{\theta})]/\partial\theta(\hat{\theta}_t - \theta_0),
\end{aligned}$$

where $\bar{\theta}$ is the mean value. Consistency of $\hat{\theta}_t$ and condition (ii) imply

$$\begin{aligned}
&|\eta_{t,h}(\hat{\theta}_t) + \tau_{t,h}(\hat{\theta}_{t,h})\prime g(Y_{t+h}, \hat{\theta}_t) - \eta_{t,h}(\theta_0) + \tau_{t,h}(\theta_0)\prime g(Y_{t+h}, \theta_0)| \\
&\quad \leq \|\sup_{\theta\in\overline{\Theta}} \partial\eta_{t,h}(\theta)/\partial\theta'\| \|\hat{\theta}_t - \theta_0\| + \|\sup_{\theta\in\overline{\Theta}} \partial[\tau_{t,h}(\theta)\prime g(Y_{t+h}, \theta)]/\partial\theta\| \|\hat{\theta}_t - \theta_0\| \\
&\quad = O_P(1) o_p(1) + O_p(1) o_p(1) \\
&\quad = o_p(1),
\end{aligned}$$

which shows that

$$\log\left[\frac{\widetilde{f}_{t,h}(Y_{t+h}, \hat{\theta}_t)}{f_{t,h}(Y_{t+h})}\right] - \log\left[\frac{\widetilde{f}_{t,h}(Y_{t+h})}{f_{t,h}(Y_{t+h})}\right] \xrightarrow{p} 0.$$

Condition (iii) implies that $\eta_{t,h}(\hat{\theta}_t) + \tau_{t,h}(\hat{\theta}_t)\prime g(Y_{t+h}, \hat{\theta}_t)$ is uniformly integrable. Then, convergence in probability can be turned into convergence in L_1, and the first result of the proposition obtains with

$$\Delta = E \log\left[\frac{\widetilde{f}_{t,h}(Y_{t+h})}{f_{t,h}(Y_{t+h})}\right].$$

Non-negativity of Δ follows from Proposition 2. $\square$

Continuous differentiability of $\eta_{t,h}(\theta)$ and $\tau_{t,h}(\theta)$ can be shown to hold when additional smoothness conditions are imposed on the function g (see Komunjer and Ragusa, 2009). Condition (ii) requires that these derivatives are bounded in probability; condition (iii) is needed in order to guarantee that $\log \frac{\widetilde{f}_{t,h}(Y_{t+h},\theta_0)}{f_{t,h}(Y_{t+h})}$ is uniformly integrable. These are high level assumptions and it is difficult to ascertain whether they hold in general situations. When the function g is a linear function, these conditions hold since, as shown in Section 2, both $\eta_{t,h}$ and $\tau_{t,h}$ can be expressed as smooth functions of means. For general nonlinear functions g, obtaining a closed-form solution for $\eta_{t,h}$ and $\tau_{t,h}$ is not possible and thus the plausibility of the assumptions in Proposition 3 must be assessed on a case by case basis.

One implication of the result in Proposition 3 is the practical recommendation to separately estimate the parameters in the base density and the parameters in the moment condition. If one were to estimate $\widetilde{f}$ directly by maximum likelihood, one would be estimating jointly the parameters in the base density $f_{t,h}$ and θ_0, and there would be no guarantee that, in general, the maximum likelihood estimator of θ_0 is consistent for θ_0. If the moment condition is estimated separately, instead, one could simply verify the standard conditions for consistency of GMM estimators.

4 Empirical illustration: incorporating Euler equation restrictions into BVARs

In this section we show an application of the tilting method 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 VARs (BVARs). All of these methods tend to perform equally well in applications (e.g., Giacomini and White, 2006 and De Mol, Giannone, and Reichlin, 2008a), so we choose one method, a BVAR, as representing one of the best currently known methods 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 an Euler equation. The specific BVAR we consider is based on Bańbura et al. (2009) to which we refer the reader for the details.

Let Y_t be a vector containing the 22 variables listed in Table 1 and let

$$Y = XB + U,$$

where $Y = (Y_1, \ldots, Y_T)'$, $X = (X_1, \ldots, X_T)'$ with $X_t = (1, Y'_{t-1}, \ldots, Y'_{t-p})'$, $B = (C, B_1, , \ldots, B_p)'$,

$U = (u_1, \dots, u_T)'$. The matrix B contains all autoregressive coefficients and constants. The elements of U are mean zero Gaussian with variance $E[u_t u_t'] = \Sigma$.

We consider the following prior

$$vec(B)|\Sigma \sim N(vec(B_0), \Sigma \otimes \Omega_0), \quad \text{and} \quad \Sigma \sim iW(\Sigma_0, \alpha_0),$$

where the prior parameters B_0, Ω_0, Σ_0, and α_0 are chosen so that the VAR is centered around a random walk with drift:

$$E[(B_k)_{ij}] = \begin{cases} 1 & j = i, k = 1 \\ 0, & \text{otherwise} \end{cases}, \quad \text{Var}[(B_k)_{ij}] = \begin{cases} \frac{\lambda^2}{k^2}, & j = i \\ \frac{\lambda^2}{k^2}\frac{\sigma_i^2}{\sigma_j^2}, & \text{otherwise,} \end{cases}$$

and the prior on C is centered at zero with a diffuse variance. As is common in the BVAR literature, the scale parameter σ_i is set equal to the residual standard error of a p-lag univariate autoregression for the variable Y_{it}. Due to the conjugate nature of the prior, the posterior distribution of B (conditional on Σ) is normal and the posterior of Σ is inverted Wishart.

The hyperparameter λ controls the overall tightness of the prior. As $\lambda \to 0$, the prior becomes tighter around the random walk with drift and when $\lambda \to \infty$ the prior becomes loose and the posteriors coincide with the distributions of the classical linear model. The amount of shrinkage should depend on the size of the VAR: larger VARs should correspond to smaller values of λ to avoid over-fitting (De Mol et al., 2008b).

Bańbura et al. (2009) show that, when applied to the macroeconomic dataset we consider in this section, the forecasting performances of the BVAR can be improved by imposing a prior on a linear combination of the VAR coefficient as initially proposed by Doan et al. (1984). This additional prior restricts $\Pi = (I_m - B_1 - \ldots - B_m)$ to zero and can be thought of as "inexact differencing". This prior, first proposed by Doan et al. (1984), can be easily implemented using Theil's mixed estimation device by adding a set of "dummy" observations to the data.

To apply our method, we need to be able to draw from $h-$step ahead density forecast implied by the BVAR. For the one-step ahead forecast the density forecast is a multivariate $\mathcal{T}$ and thus it is easy to simulate from it. For longer horizons, draws from the predictive density are obtained by simulation. Let $\{y_{t+h}^{(d)}\}_{d=1}^{D}$ be D draws from the predictive distribution of Y_{t+h}. The BVAR point forecasts we consider are the expected values of the predictive distribution

$$E_t^f Y_{t+h} = \int y_{t+h} f_{t,h}(y) dy, \tag{11}$$

which are approximated by averaging the predictive draws:

$$\widehat{Y}_{t+h}^f = \frac{1}{D} \sum_{d=1}^{D} y_{t+h}^{(d)}. \tag{12}$$

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

$$E_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\alpha} R_{j,t+1} - 1 \right] = 0, \quad j = 1, \ldots, J, \tag{13}$$

which for the $h-$step forecast horizon implies

$$E_t \left[\beta \left(\frac{C_{t+h}}{C_{t+h-1}} \right)^{-\alpha} R_{j,t+h} - 1 \right] = 0, \quad j = 1, \ldots, J. \tag{14}$$

where C_t is real consumption, $\beta \in (0, 1)$ is a discount factor, γ is the elasticity of inter-temporal substitution, and $R_{j,t+1}$ is any traded asset real return indexed by j. Equation (14) arises from the assumption that a representative agent with preferences of the constant relative risk aversion type chooses consumption by solving

$$\max_{\{C\}_{t=1}^{\infty}} E_t \sum_{s=0}^{\infty} \beta^j \left(\frac{C_{t+s}^{1-\gamma}}{1-\gamma} \right),$$

subject to a budget constraint

$$W_{t+1} = \sum_{j=1}^{J} R_{j,t+1} B_{j,t}, \quad W_t - C_t = \sum_{j=1}^{J} B_{j,t}$$

where W_t is the stock of wealth and $B_{j,t}$ is the amount of aggregate wealth devoted to asset j at time t (Cf. Hansen and Singleton, 1982).

In the empirical application, the measure of consumption we use is non-durables plus services. The quarterly, seasonally adjusted observations on aggregate real consumption of non-durables and services are from the Bureau of Economic Analysis. The measure of consumption is paired with two sets of stock returns: (a) the returns on all stocks of the S&P 500 index and (b) the returns on one of the six Fama-French (1993) portfolios sorted by size and book-to-market equity. In particular, we use the portfolio formed by stocks of small market equity firms with high book value relative to market equity. Both returns are deflated by using the price index of non-durable and service consumption.

We impose the Euler equation restrictions on the BVAR predictive density interpreting (14) as a restriction on the levels of consumption. In this case, equation (14) is a restriction on the joint density of $Y_{t+h} = (C_{t+h}, R_{t+h})$ and $Y_{t+h-1} = (C_{t+h-1}, R_{t+h-1})$ which is the output of the Gibbs sampler when the out-of-sample forecasts are obtained recursively. To construct the tilted density forecast at time t, we first obtain the tilted joint density as

$$\tilde{f}(y_{t+h}, y_{t+h-1}) = f(y_{t+h}, y_{t+h-1}) \exp(\eta_t + \tau_t g(y_{t+h}, y_{t+h-1})), \tag{15}$$

where

$$g(Y_{t+h}, Y_{t+h-1}) = \beta \left(\frac{C_{t+h}}{C_{t+h-1}} \right)^{-\alpha} R_{j,t+h} - 1, \tag{16}$$

and (η_t, τ_t') are constructed by drawing $\{y^d_{t+h}, y^d_{t+h-1}\}_{d=1}^{D}$ from $f(y_{t+h}, y_{t+h-1})$ and solving

$$\arg\min_{(\eta,\tau)} \sum_{d=1}^{D} \exp(\eta + \tau' g(y^h_{t+h}, y^h_{t+h-1})). \tag{17}$$

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

$$E_t^{\tilde{f}} Y_{t+h} = \int y_{t+h} \tilde{f}(y_{t+h}, y_{t+h-1}) dy_{t+h} dy_{t+h-1}, \tag{18}$$

which can be approximated by

$$\widehat{Y}^{\tilde{f}}_{t+h} = \sum_{d=1}^{D} w_{th} y^d_{t+h}, \tag{19}$$

with

$$w_{th} = \exp(\eta_t + \tau_t' g(y^d_{t+h}, y^d_{t+h-1}))/D.$$

The next section reports the results of an out-of-sample exercise comparing the performance of the BVAR to that of the tilted BVAR that incorporates the Euler equation restrictions. The exercise compares the point forecast accuracy of the two models in terms of the mean square forecast error (MSFE) of the point forecast $\widehat{Y}_{t+h}$ over an out-of-sample period including data from period R to period T:

$$MSFE = \frac{1}{T-R-h} \sum_{t=R}^{T-h} \left[(Y_{t+h} - \widehat{Y}_{t+h})^2 \right], \tag{20}$$

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

$$\frac{1}{T-R-h} \sum_{t=R}^{T-h} \log f_t(Y_{t+h}) \tag{21}$$

We stress here that our theoretical results on the tilted forecasts concern the logarithmic scoring rule but are silent about the effect that our procedures has on the MSFE and other measures of point-forecast accuracy

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 first quarter of 2009.

We present results for a BVAR model with $p = 5$ lags and contains 22 variables. The variables are similar to those used by Banbura et al. (2010) in their "medium" model. The differences are due to the fact that we have to add the variables necessary to evaluate the Euler condition. We report forecasting results for the following variables: real Gross Domestic Product, Federal Fund Rate, non-durables plus services consumption, price index of non-durables plus services, S&P 500 price index (SP500), and returns on the Fama-French portfolio (FFP). The model also includes variables on investment, hours worked, wages and other labor market, financial and monetary indicators. The full list of variables is given in Table 1.

[Table 1 about here]

The hyperparameter λ that regulates the amount of shrinkage is chosen to maximize the performance of the BVAR. We produce out-of-sample forecasts from BVARs that use different λ and we pick the value of λ that gives the best forecasting accuracy in terms of average MSFE for three of the key six variable. We use this procedure consistently with our goal of verifying whether adding the theoretical restrictions improves the performance of a BVAR with already good forecasting accuracy.[1]

The BVAR model is estimated in an out-of-sample fashion using a rolling window of length $R = 80$ and for forecast horizons $h = \{1, 2, 3, 4, 5, 6, 7, 8\}$. The predictive densities are evaluated for $t = R, \ldots, T$, where R is 1979:Q2. The procedure yields a total of $T - R - h = 114$ BVAR forecast densities $f_{t,h}(y)$.

We then construct the tilted density forecasts that incorporate the Euler equation (14). The value of β is fixed across all simulation to $\beta = 0.999$. This value if consistent with estimates found in the literature (cf. Hansen and Singleton, 1982; Ludvigson, 2011). For the value of the elasticity of intertemporal substitution we follow two approaches. In the first, α is kept fixed. In the second approach, at each $t = R, \ldots, T - h$, we estimate the parameter α using the same data on which the BVAR is estimated.

When following the first approach, instead of choosing a unique value for α, we consider a grid of values:

$$\alpha = \{0.1, 0.6, 1.1, 1.62.1, 2.6, 3.1, 3.6, 4.1, 4.6\}.$$

The grid is chosen in such a way to contain values that are deemed theoretically reasonable (Cf. Mehra and Prescott, 1985) and to be consistent with those obtained by estimating α on

[1] Experimentation with different values of λ shows that when the BVAR has poorer predictive ability the effect of imposing the theoretical restrictions is larger.

a rolling window scheme having the same length of the one used for the out-of-sample exercise. Figure 2 plots the rolling estimates of α when (14) uses the returns on the S&P 500 (left panel) or the returns on the Fama-French portfolio (right panel). The graphs also report the full sample estimates which are in both case very close to 0.6. The rolling estimates behave instead very differently: the estimates based on the S&P 500 go from negative values to positive and large values; the ones based on the Fama-French portfolio are instead closer to the full sample estimate. The estimates are obtained by (exactly identified) GMM and the point-wise 95% confidence intervals shown in Figure 2 are based on the normal asymptotic approximation of the sampling distribution of estimator. The asymptotic variance is estimated by using the inverse of the Hessian of the GMM objective functions.

[Figure 2 about here]

To 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

$$\frac{1}{D}\sum_{d=1}^{d} g(y^d_{t+h}, y^d_{t+h-1}),$$

where $g(\cdot)$ is as defined in (16) for $\beta = 0.999$, $\alpha = 0.6$ and (y^d_{t+h}, y^d_{t+h-1}) are draws from the predictive density of the BVAR. The dotted lines represent the associated 95% predictive interval. Figure 3 confirms that there are periods in which the densities implied by the BVAR do not satisfy the Euler equation. In the next 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 tilted densities that are coherent with an Euler condition.

[Figure 3 about here]

4.1.1 Point forecast performance

Table 2 and Table 3 report the MSFE ratios of the point forecasts implied by the BVAR and its tilted version when the return is based on the S&P 500 and the Fama French portfolio, respectively. An entry smaller than one indicates that the tilted forecasts are more accurate. Entries marked with an asterisk signal rejection at the 5% significance level of the null hypothesis of equal MSFE according to the Giacomini and White (2006) test. The tables show both the results when α is calibrated and the case in which α is recursively estimated using in-sample data.

[Tables 2 and 3 about here]

From Table 2 we see that 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 forecasts of the returns on the S&P 500, for which the accuracy gains are relatively sizable. The results in Table 3 show a similar patterns: there are generally gains from imposing the Euler condition, but the gains are not significant except in the case of the returns on the Fama-French portfolio. In both cases, the improvement in MSFE are similar to those we obtain when α is calibrated. The results in Tables 1 and 2 are robust to a number of different choices for the Euler equation parameter β.

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 the return of the Fama-French portfolio, 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 consistently throughout the sample, Figure 4 plots the difference in absolute forecast errors for the BVAR and the tilted BVAR.

[Figure 4 about here]

The figure shows that the tilted BVAR is more accurate than the BVAR in the vast majority of the sample, and the magnitude of the improvement appears to be particularly large during recessions.

4.1.2 Density forecast performance

Here we consider the multivariate density forecast as a whole and evaluate the relative performance of the density forecast implied by the BVAR relative to that of the tilted BVAR. Tables 4 and 5 report the relative accuracy of the two forecasts for different forecast horizons and choices of the Euler equation parameter α. As we can see from these tables, there are large improvements 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 generally all significant.

[Tables 4 and 5 about here]

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 an exponential tilting method 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 and show that imposing the restrictions implied by a simple Euler condition can improve the point- and density forecast performance of a Bayesian VAR - currently known as one of the best methods for forecasting a number of key macroeconomic variables.

Mnemonic	Description	Transformation
GDP251	Real GDP, quantity index	log
FYFF	Interest rate: federal funds (pct per annum)	level
PCEPI	Consumption price index non-durables plus ervices	log
RPCNDSV	Real consumption non-durables plus services	log
FFP	Return on fama french portfolio	level
FSPCOM	S&P 500 index	log
CPIAUCSL	Consumer Price Index All Items	log
PCEPILFE	Real spot market price index: all commodities	log
FMRRA	Depository inst reserves: nonborrowed (mil USD)	log
FMFBA	Depository inst reserves: total (mil USD)	log
FSDJ	Money stock: M1 (bil USD)	log
MZMSL	Money stock: M2 (bil USD)	log
IPS10	Industrial production index: total	log
LBOUT	Capacity utilization: manufacturing (SIC)	log
HSBR	Housing starts: Total (thousands)	log
LBMNU	Employees, nonfarm: total private	log
GDP288A	Real avg hrly earnings, non-farm prod. workers	log
LHNAG	Unemp. rate: All workers, 16 and over (%)	log
FYGM3	Interest rate: US T-bills, sec mkt, 3-month	level
FYGT5	Interest rate: US treasury const. mat., 5-yr	level
FYGT10	Interest rate: US treasury const. mat., 10-yr	level
sFYBAAC	US effective exchange rate: index number	level

Table 1: Description of the variables entering the Bayesian VAR.

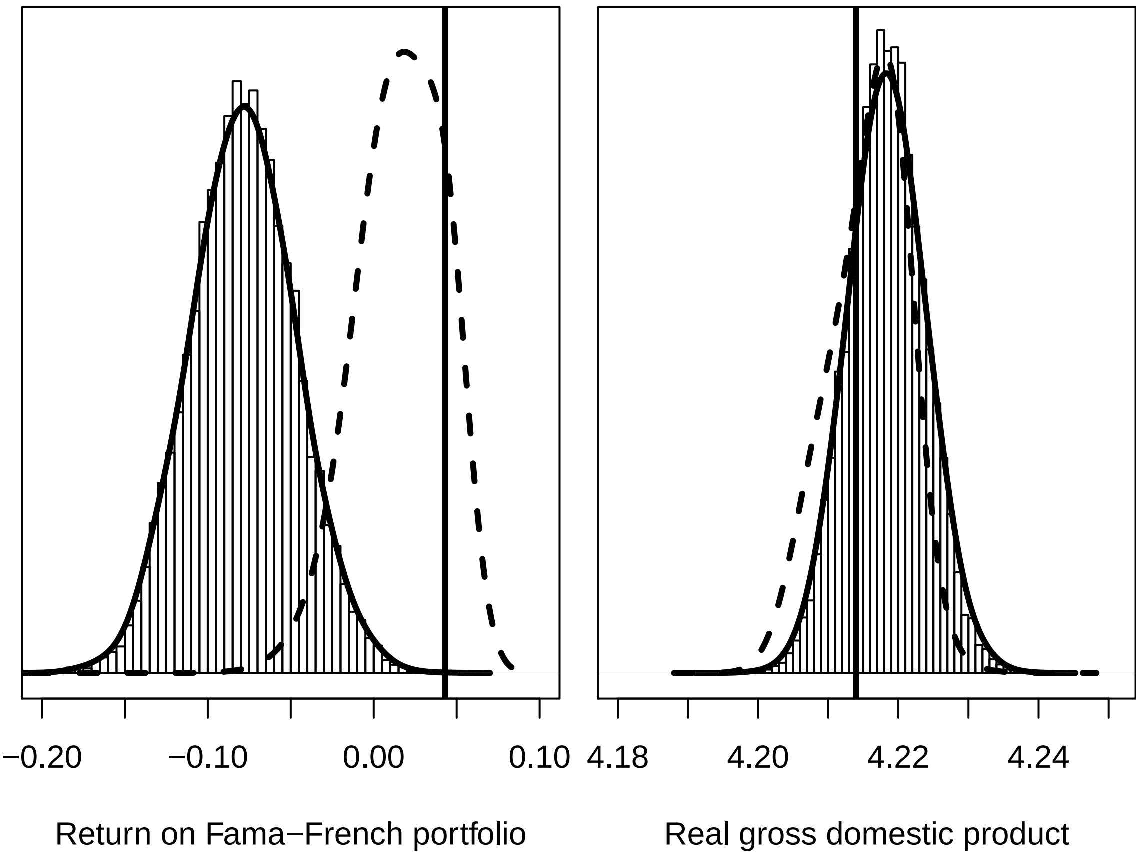

Figure 1: The figure shows two out-of-sample density forecasts for the real GDP and for the return on a Fama-French portfolio at one particular point in time (1988:Q1): the histogram is the one-step-ahead density forecast implied by a BVAR with 22 variables which include real GDP, non-durables and services real consumption, the federal funds rate and the return on the Fama-French portfolio. In each graph, the dashed line is the projected density forecast that incorporates the Euler equation; the solid vertical line is the realization of the variable.

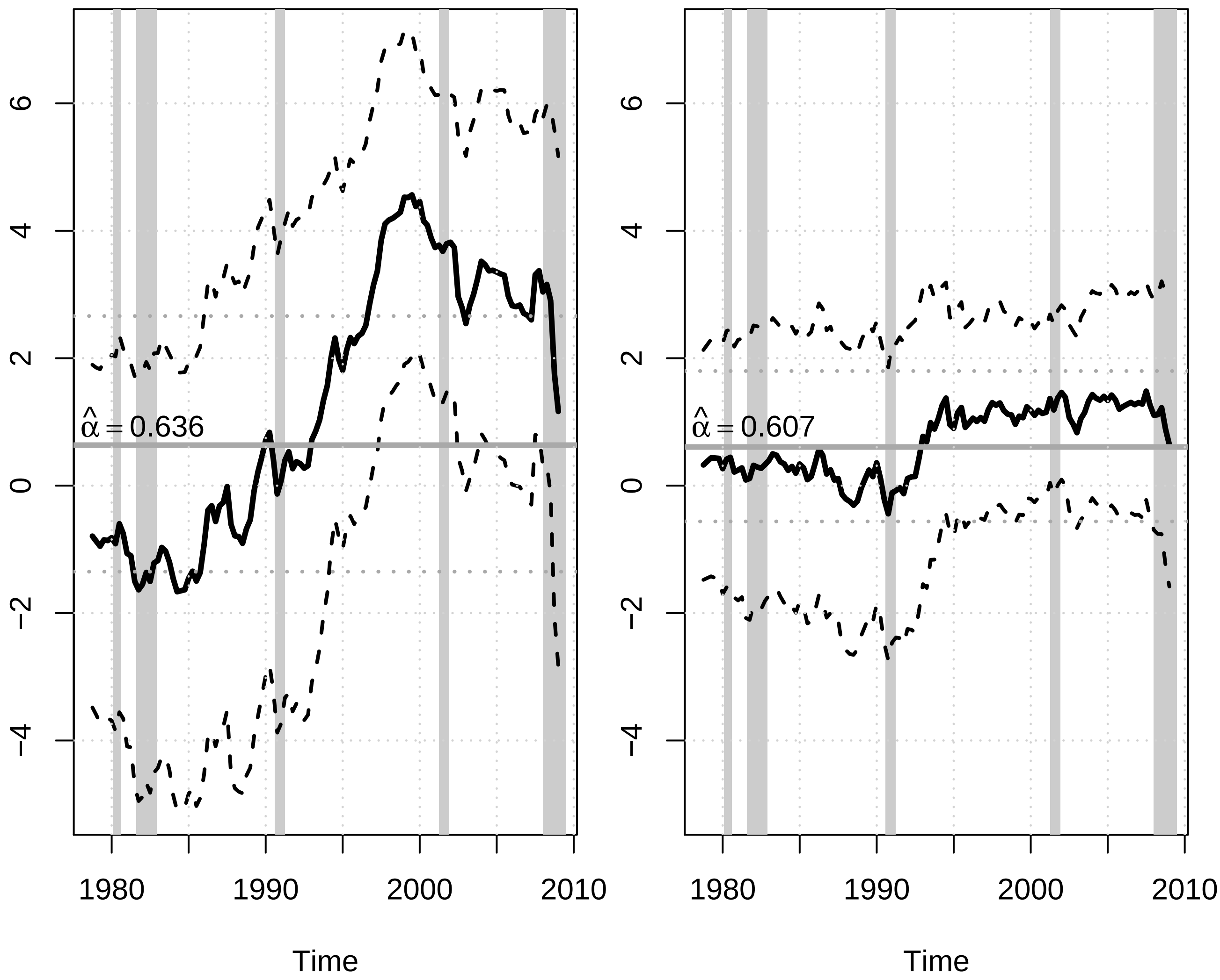

Figure 2: Rolling (darker continuous line) and full sample estimates (lighter continuous line) of the parameter α. The parameter is estimated using a rolling window of length $R = 80$ starting from the period 1959:1. Dotted lines represent the 95% (point-wise) confidence interval. The results plotted in the Left panel use the S%P 500 composite index. The ones plotted on the right panel use the Fama-French portfolio.

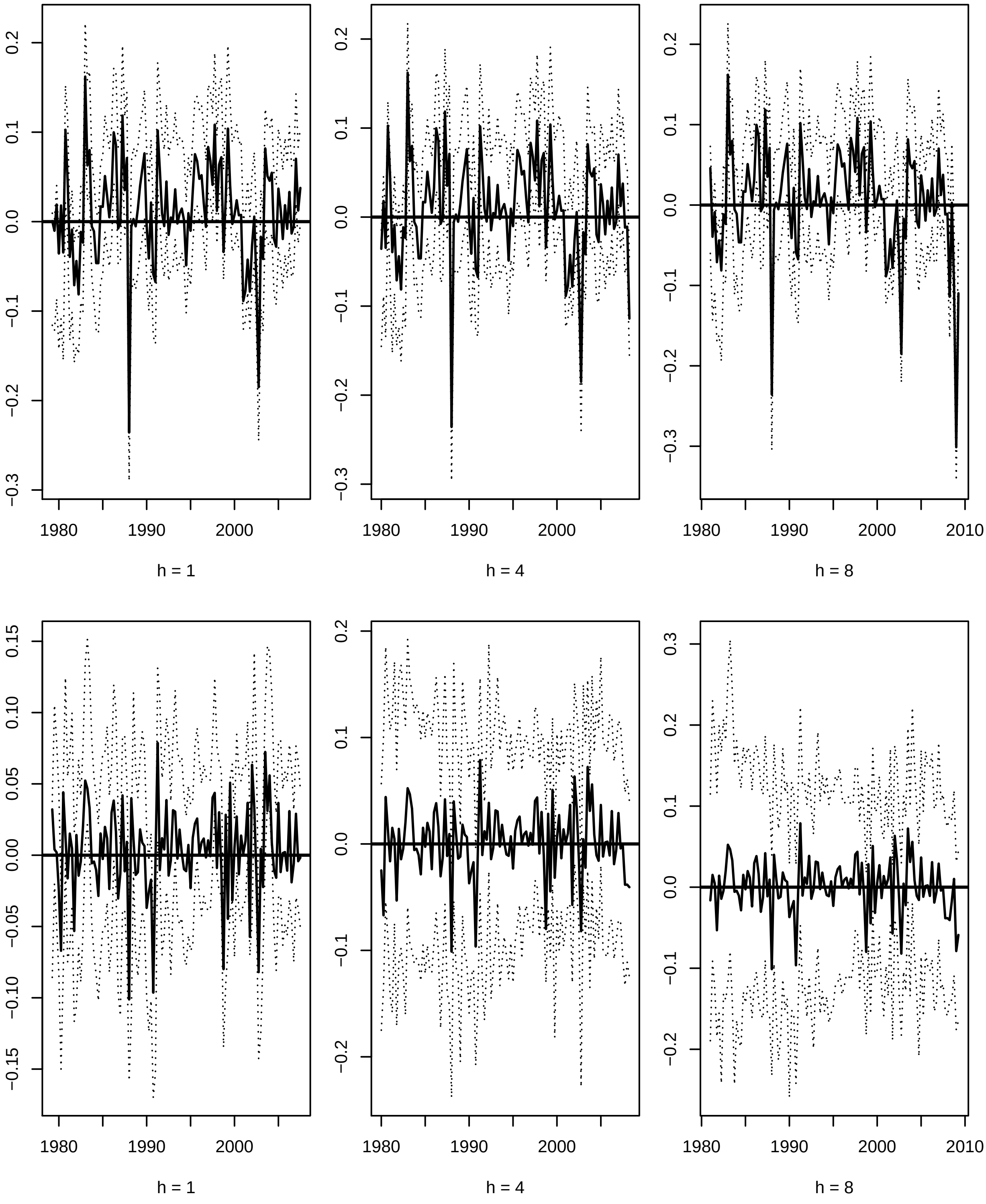

Figure 3: Euler equation error for forecast horizons $h = 1, 4$, and 8. Each panel plots the Euler equation ($\beta = .999$, and $\alpha = 0.6$) error under the approximated density implied by the medium BVAR. The top panel plots the Euler equation that uses the S&P 500 composite index, the bottom panel the one that uses the Fama-French portfolio. The lighter lines are the 95% predictive bands.

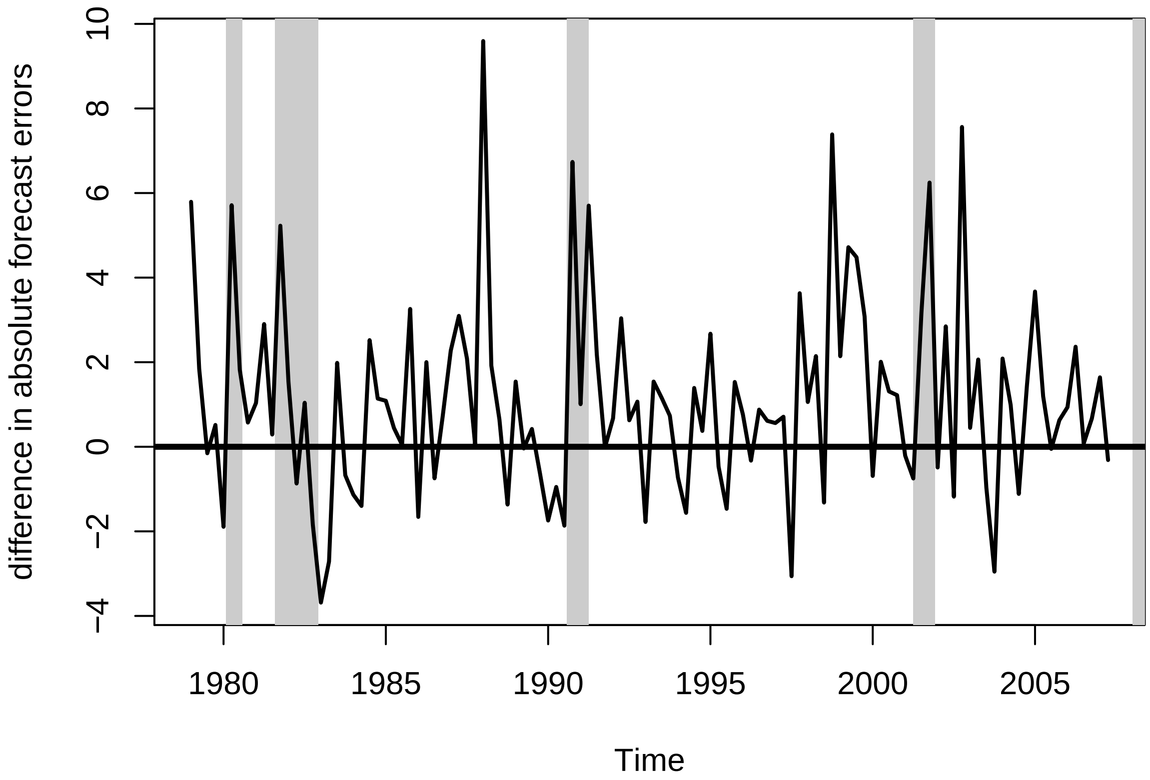

Figure 4: Absolute forecasting one-step-ahead error of the return on the Fama-French portfolio. The black line is the difference between the absolute forecasting error of the BVAR and the absolute forecasting error of the Euler-projected forecast with parameters $\beta = .999$ and $\alpha = 0.6$.

	Euler equation parameter										
	$\alpha = 0.1$	$\alpha = 0.6$	$\alpha = 1.1$	$\alpha = 1.6$	$\alpha = 2.1$	$\alpha = 2.6$	$\alpha = 3.1$	$\alpha = 3.6$	$\alpha = 4.1$	$\alpha = 4.6$	$\hat{\alpha}_s$
RGDP											
h=1	1.006	1.007	1.007	1.007	1.007	1.006	1.003	1.002	1.003	1.000	1.014
h=2	1.005	1.005	1.004	1.003	0.997	1.001	0.995	0.994	0.996	0.994	1.012
h=3	1.001	1.001	1.001	1.001	1.000	0.999	0.998	0.997	0.996	0.995	1.003
h=4	0.996	0.996	0.996	0.996	0.996	0.996	0.995	0.995	0.994	0.993*	0.994
h=5	0.995	0.995	0.996	0.996	0.996	0.995	0.995	0.995	0.994	0.993	0.991
h=6	0.993	0.994	0.994	0.995	0.995	0.995	0.995	0.995	0.994	0.994*	0.986
h=7	0.988	0.989	0.990	0.991	0.992	0.993	0.993	0.994	0.994	0.994*	0.978
h=8	0.986	0.987	0.989	0.990	0.991	0.992	0.993	0.994	0.994	0.994	0.973
FFR											
h=1	1.011	1.011	1.011	1.011	1.010	1.010	1.007	1.007	1.009	1.006	1.011
h=2	1.006	1.004	1.003	1.001	1.011	0.999	1.008	1.007	0.996	0.995	1.009
h=3	1.007	1.005	1.004	1.003	1.002	1.000	0.999	0.999	0.998	0.997	1.003
h=4	1.010	1.009	1.008	1.008	1.007	1.006	1.006	1.005	1.005	1.005	1.002
h=5	1.012	1.011	1.010	1.009	1.008	1.007	1.007	1.006	1.006	1.006	0.999
h=6	1.006	1.005	1.005	1.005	1.005	1.004	1.004	1.004	1.005	1.005	0.985
h=7	1.006	1.007	1.007	1.008	1.009	1.010	1.011	1.011	1.012	1.013	0.980
h=8	1.014	1.015	1.016	1.016	1.017	1.018	1.019	1.020	1.020	1.021	0.986
PCEPI											
h=1	1.029	1.021	1.013	1.006	1.000	0.995	0.991	0.987	0.984	0.984	1.036
h=2	1.003	0.998	0.994	0.990	0.993	0.984	0.988	0.986	0.981	0.981	0.998
h=3	0.993	0.989	0.987	0.985	0.983	0.982	0.981	0.981	0.982	0.984	0.980
h=4	0.986	0.984	0.982	0.981	0.981	0.980	0.981	0.982	0.983	0.985	0.969
h=5	0.984	0.983	0.982	0.982	0.982	0.983	0.984	0.985	0.987	0.989	0.964
h=6	0.980	0.980	0.980	0.980	0.981	0.982	0.983	0.985	0.987	0.989	0.957
h=7	0.979	0.979	0.980	0.981	0.982	0.984	0.985	0.987	0.990	0.992	0.952
h=8	0.976	0.977	0.978	0.980	0.981	0.983	0.985	0.987	0.990*	0.992	0.948
Cons.											
h=1	1.022	1.017	1.014	1.010	1.007	1.005	1.002	0.999	0.996	0.993	1.051
h=2	1.003	1.003	1.002	1.001	0.999	1.000	0.998	0.997	0.996	0.994	1.016
h=3	0.991	0.992	0.994	0.995	0.996	0.997	0.998	0.998	0.998	0.998	0.988
h=4	0.984	0.987	0.989	0.991	0.993	0.995	0.996	0.997	0.998	0.998	0.975
h=5	0.982	0.984	0.987	0.989	0.991	0.993	0.995	0.996	0.997	0.998	0.966
h=6	0.980	0.983	0.985	0.988	0.990	0.992	0.994	0.995	0.996	0.997	0.961
h=7	0.978	0.981	0.984	0.987	0.990	0.992	0.994	0.996	0.997	0.998	0.957
h=8	0.977	0.981	0.984	0.987	0.989	0.992	0.994	0.995	0.996	0.997	0.954
S&P											
h=1	1.009	0.993	0.979	0.968	0.960	0.955	0.961	0.964	0.962	0.978	1.036
h=2	0.936	0.929	0.923	0.918	0.929*	0.914*	0.927*	0.928*	0.919*	0.924	0.968
h=3	0.916	0.912	0.908*	0.906*	0.904*	0.903*	0.904*	0.906*	0.909*	0.913*	0.948
h=4	0.912	0.910	0.907*	0.906*	0.905*	0.906*	0.907*	0.908*	0.911*	0.915*	0.944
h=5	0.906	0.904*	0.903*	0.902*	0.902*	0.903*	0.905*	0.907*	0.910*	0.914*	0.934
h=6	0.904*	0.902*	0.902*	0.902*	0.902*	0.904*	0.906*	0.908*	0.912*	0.916*	0.931
h=7	0.905*	0.904*	0.905*	0.905*	0.907*	0.909*	0.912*	0.915*	0.919*	0.923*	0.931*
h=8	0.908*	0.909*	0.909*	0.911*	0.913*	0.916*	0.919*	0.922*	0.926*	0.931*	0.935*

Table 2: Point forecast performance: S&P returns. Each entry consists of the ratio between the medium MSFE of the BVAR and its Euler-projected version when the returns are based on S&P 500 index. Entries marked with an asterisk indicate that the difference in MSFE is significant at the 5% level according to the Giacomini and White (2006) test.

	Euler equation parameter										
	$\alpha = 0.1$	$\alpha = 0.6$	$\alpha = 1.1$	$\alpha = 1.6$	$\alpha = 2.1$	$\alpha = 2.6$	$\alpha = 3.1$	$\alpha = 3.6$	$\alpha = 4.1$	$\alpha = 4.6$	$\hat{\alpha}_s$
RGDP											
h=1	0.988	0.998	0.994	0.989	0.987	0.984	0.980	0.978	0.968	0.982	0.987
h=2	0.984	0.980	0.976	0.972	0.968	0.964	0.960	0.956	0.972	0.969	0.989
h=3	0.988	0.985	0.982	0.979	0.976	0.973	0.970	0.967	0.965	0.957	0.992
h=4	0.992	0.990	0.988	0.986	0.984	0.982	0.980	0.978	0.976	0.974	0.996
h=5	0.995	0.994	0.993	0.991	0.990	0.988	0.987	0.985	0.983	0.982	1.001
h=6	0.998	0.997	0.996	0.995	0.994	0.992	0.991	0.990	0.988	0.987	1.001
h=7	1.000	1.000	0.999	0.998	0.997	0.996	0.995	0.994	0.993	0.992	1.003
h=8	1.002	1.002	1.001	1.001	1.000	0.999	0.999	0.998	0.997	0.996	1.003
FFR											
h=1	1.003	0.999	0.998	0.998	0.994	0.995	1.004	0.999	1.004	1.004	1.004
h=2	0.969	0.972	0.971	0.971	0.970	0.970	0.970	0.972	0.978	0.979	0.971
h=3	0.991	0.990	0.989	0.989	0.988	0.988	0.988	0.988	0.980	0.989	0.984
h=4	0.994	0.993	0.992	0.991	0.991	0.991	0.990	0.990	0.990	0.990	0.983
h=5	0.998	0.997	0.996	0.995	0.995	0.994	0.994	0.993	0.993	0.993	0.986
h=6	1.005	1.005	1.004	1.003	1.003	1.002	1.002	1.002	1.001	1.001	0.993
h=7	1.007	1.007	1.006	1.006	1.006	1.005	1.005	1.005	1.005	1.004	0.995
h=8	1.012	1.011	1.011	1.010	1.010	1.009	1.009	1.009	1.009	1.008	0.999
PCEPI											
h=1	1.010	1.006	1.004	1.003	1.004	1.005	1.002	1.010	1.015	1.018	1.008
h=2	1.004	1.003	1.002	1.002	1.003	1.003	1.005	1.005	1.006	1.008*	1.001
h=3	1.002	1.002	1.002	1.002	1.002	1.003	1.004	1.006	1.007	1.009	1.000
h=4	1.003	1.003	1.003	1.003	1.004	1.004	1.005	1.006	1.007	1.008	1.000
h=5	1.004	1.004	1.004	1.004	1.005	1.005	1.006	1.006	1.007	1.008	1.001
h=6	1.005	1.005	1.005	1.005	1.006	1.006	1.006	1.007	1.008	1.008	1.001
h=7	1.006	1.006	1.006	1.007	1.007	1.007	1.008	1.008	1.009	1.009	1.001
h=8	1.008	1.008	1.008	1.008	1.008	1.009	1.009	1.009	1.010	1.010	1.003
Cons.											
h=1	0.996	0.990	0.985	0.980	0.972	0.969	0.978	0.966	0.969	0.981	1.016
h=2	1.006	1.001	0.997	0.993	0.989	0.984	0.980	0.978	0.988	0.986	1.010*
h=3	1.002	1.001	1.000	0.998	0.997	0.995	0.993	0.992	0.988	0.983	1.008
h=4	1.002	1.002	1.001	1.000	0.999	0.998	0.997	0.996	0.994	0.993	1.006
h=5	1.001	1.001	1.001	1.001	1.000	1.000	0.999	0.998	0.998	0.997	1.003
h=6	1.002	1.002	1.002	1.001	1.001	1.000	1.000	0.999	0.998	0.998	1.002
h=7	1.002	1.002	1.002	1.002	1.002	1.001	1.001	1.001	1.000	0.999	1.002
h=8	1.003	1.003	1.003	1.003	1.003	1.003	1.003	1.002	1.002	1.001	1.001
FFP											
h=1	0.568*	0.493*	0.492*	0.498*	0.583*	0.600*	0.621*	0.601*	0.629*	0.649*	0.540*
h=2	0.607*	0.553*	0.553*	0.559*	0.570*	0.587*	0.610*	0.674*	0.763*	0.797	0.557*
h=3	0.557*	0.553*	0.555*	0.562*	0.576*	0.594*	0.618*	0.647*	0.702*	0.781	0.554*
h=4	0.617*	0.616*	0.621*	0.632*	0.649*	0.672*	0.700*	0.734*	0.772	0.815	0.663*
h=5	0.624*	0.623*	0.628*	0.639*	0.656*	0.679*	0.707*	0.741	0.779	0.822	0.622*
h=6	0.615*	0.613*	0.617*	0.627*	0.644*	0.666*	0.693*	0.726*	0.764*	0.807	0.608*
h=7	0.671*	0.674*	0.683*	0.698*	0.719*	0.746*	0.778*	0.815	0.858	0.905	0.669*
h=8	0.812*	0.817*	0.829*	0.847*	0.872	0.904	0.942	0.986	1.036	1.091	0.811*

Table 3: Point forecast performance: Fama-French returns. Each entry consists of the ratio between the medium MSFE of the BVAR and its Euler-projected version when the returns are those from the Fama French portfolio. Entries marked with an asterisk indicate that the difference in MSFE is significant at the 5% level according to the Giacomini and White (2006) test.

	Euler equation parameter										
	0.1	0.6	1.1	1.6	2.1	2.6	3.1	3.6	4.1	4.6	$\hat{\alpha}_s$
h=1	0.545	0.681	0.689	0.688	0.611	0.593	0.512	0.564	0.526	0.481	0.593
	(0.15)	(0.17)	(0.17)	(0.17)	(0.16)	(0.16)	(0.14)	(0.16)	(0.16)	(0.16)	(0.15)
h=2	0.224	0.257	0.262	0.264	0.263	0.257	0.248	0.207	0.157	0.137	0.243
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.07)	(0.07)	(0.07)
h=3	0.197	0.197	0.195	0.190	0.183	0.174	0.163	0.149	0.133	0.083	0.193
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
h=4	0.128	0.132	0.135	0.135	0.134	0.131	0.126	0.119	0.111	0.101	0.113
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.03)
h=5	0.096	0.096	0.094	0.091	0.086	0.081	0.074	0.066	0.056	0.046	0.092
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
h=6	0.093	0.097	0.099	0.101	0.101	0.100	0.098	0.095	0.092	0.087	0.099
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
h=7	0.069	0.069	0.069	0.068	0.066	0.064	0.060	0.056	0.050	0.044	0.070
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
h=8	0.031	0.031	0.030	0.028	0.025	0.022	0.018	0.014	0.008	0.003	0.031
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)

Table 4: Density forecast performance: Fama-French portfolio. Entries are the average logarithmic scoring rule (e.g., Amisano and Giacomini, 2007). Positive (negative) values indicate that the projected density forecast is more accurate (less accurate) than the BVAR forecasts. Values in parenthesis are heteroskedasticity- and autocorrelation-consistent (HAC) standard errors.

	Euler equation parameter										
	0.1	0.6	1.1	1.6	2.1	2.6	3.1	3.6	4.1	4.6	$\hat{\alpha}_s$
h=1	−0.024	0.005	0.032	0.055	0.074	0.089	0.087	0.092	0.103	0.087	−0.073
	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)	(0.08)	(0.08)
h=2	0.065	0.090	0.112	0.131	0.103	0.156	0.123	0.126	0.159	0.150	0.021
	(0.11)	(0.10)	(0.09)	(0.08)	(0.07)	(0.08)	(0.07)	(0.07)	(0.09)	(0.09)	(0.09)
h=3	0.134	0.155	0.173	0.187	0.197	0.203	0.205	0.202	0.194	0.181	0.103
	(0.12)	(0.11)	(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)	(0.11)
h=4	0.151	0.166	0.178	0.187	0.192	0.193	0.190	0.183	0.171	0.155	0.117
	(0.12)	(0.11)	(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.10)	(0.10)
h=5	0.180	0.194	0.204	0.212	0.216	0.216	0.213	0.206	0.194	0.178	0.160
	(0.13)	(0.12)	(0.10)	(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.09)	(0.10)	(0.12)
h=6	0.172	0.181	0.187	0.190	0.190	0.187	0.181	0.171	0.158	0.141	0.145
	(0.13)	(0.12)	(0.11)	(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.11)
h=7	0.209	0.208	0.204	0.198	0.188	0.175	0.159	0.139	0.116	0.090	0.158
	(0.16)	(0.14)	(0.13)	(0.11)	(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.09)	(0.11)
h=8	0.172	0.167	0.160	0.150	0.137	0.121	0.102	0.081	0.057	0.030	0.110
	(0.17)	(0.15)	(0.13)	(0.12)	(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)

Table 5: Density forecast performance: S&P 500 composite index. Entries are the average logarithmic scoring rule (e.g., Amisano and Giacomini, 2007). Positive (negative) values indicate that the projected density forecast is more accurate (less accurate) than the BVAR forecasts. Values in parenthesis are heteroskedasticity- and autocorrelation-consistent (HAC) standard errors.

References

Bańbura, M., D. Giannone, and L. Reichlin (2009): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25, 71–92.

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 (2008a): "Forecasting using a large number of predictors: Is Bayesian regression a valid alternative to principal components?" *Journal of Econometrics*, 146, 318.

——— (2008b): "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. 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.

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.

HANSEN, L. P. AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–86.

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

KITAMURA, Y. AND M. STUTZER (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–74.

KOMUNJER, I. AND G. RAGUSA (2009): "Existence and Uniqueness of Semiparametric Projections," University of California at San Diego.

KULLBACK, S. (1968): *Information theory and statistics*, Courier Dover Publications.

LUDVIGSON, S. (2011): "Advances in Consumption-Based Asset Pricing: Empirical Tests," *NBER Working Paper Series*, 16810.

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.

MEHRA, R. AND E. C. PRESCOTT (1985): "The equity premium: A puzzle," *Journal of Monetary Economics*, 15, 145–161.

NEWEY, W. AND R. J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–256.

RAGUSA, G. (2011): "Minimum Divergence, Generalized Empirical Likelihoods, and Higher Order Expansions," *Econometric Reviews*, 30, 406–456.

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 (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.