

Department of Economics

A Bayesian Framework for the Expectations Hypothesis. How to Extract Additional Information from the Term Structure of Interest Rates

Andrea Carriero

Working Paper No. 591

March 2007

ISSN 1473-0278



A Bayesian Framework for the Expectations Hypothesis. How to Extract Additional Information From the Term Structure of Interest Rates.

Andrea Carriero*
Queen Mary, University of London

This version January 2007. First draft September 2005

Abstract

Even if there is a fairly large evidence against the Expectations Hypothesis (EH) of the term structure of interest rates, there still seems to be an element of truth in the theory which may be exploited for forecasting and simulation. This paper formalizes this idea by proposing a way to use the EH without imposing it dogmatically. It does so by using a Bayesian framework such that the extent to which the EH is imposed on the data is under the control of the researcher. This allows to study a continuum of models ranging from one in which the EH holds exactly to one in which it does not hold at all. In between these two extremes, the EH features transitory deviations which may be explained by time varying (but stationary) term premia and errors in expectations. Once cast in this framework, the EH holds on average (i.e. after integrating out the effect of the transitory deviations) and can be safely and effectively used for forecasting and simulation.

JEL Classification: C11, E43, E44, E47.

Keywords: Bayesian VARs, Expectations Theory, Term Structure.

*Department of Economics, Queen Mary University of London, Mile End Rd., London E1 4NS. Email: a.carriero@qmul.ac.uk. I thank Fabio Canova, Carlo Favero, Massimiliano Marcellino, Nicola Scalzo, Ulf Söderström, seminar participants at Bocconi University, ECB, and University of Brescia, conference participants at the VI Workshop in Quantitative Finance, Milan, at the First Italian Congress of Econometric and Empirical Economics, Venice, at the First Symposium on Econometric Theory and Applications, Academia SINICA, Taipei, and at the Frontiers in Time Series Analysis conference, Olbia. Errors are mine. Previously circulated as: "Validating the Expectations Hypothesis as a Set of Uncertain Restrictions".

1 Introduction

The Expectations Hypothesis of the term structure of interest rates (EH) states that actual long-term interest rates are determined by the market's expectation of the future short-term rates. Popularized by Fisher (1930), Keynes (1930), and Hicks (1953), this theory continues to be a way that many economists think about the determination of long-term interest rates.

Most central banks use the EH in their economic forecasting and financial models. Central bank researchers routinely impose the EH to forecast short-term rates, to assess how monetary policy affects long-term rates, and to measure market expectations about interest rates and inflation (Clews 2002, Scholtes 2002, Söderlind and Svensson, 1997, European Central Bank "Monthly Bulletin"). Also the monetary VAR literature (Rudebusch, 1998, Krueger and Kuttner, 1996) often imposes the EH to disentangle expected from unexpected movements in interest rates.

However, the empirical evidence casts serious doubts on the appropriateness of using the EH for these forecasting and simulation exercises. Indeed, the EH has been widely tested, and almost invariably rejected. A nonexhaustive list of studies rejecting the EH comprises Fama (1984), Fama and Bliss (1987), Stambaugh (1988), Campbell and Shiller (1991), Campbell (1995), Backus et al. (2001), Bekaert et al. (1997), Cochrane and Piazzesi (2005). The failure of the EH may be due both to the presence of irrational agents and to time variation in term premia. This latter explanation seems to be the most relevant, as many studies have presented evidence that term premia in bond returns are time varying. In particular, Fama and Bliss (1987) show that term premia do vary through time and are forecastable via the forward rates. Campbell and Shiller (1991) find similar results using yield spreads to predict yield changes. Recently, Cochrane and Piazzesi (2005) have strengthened these results showing that the same linear combination of forward rates predicts bonds returns at all maturities.

On the other side, Campbell and Shiller (1987) found an anomaly, i.e. the EH is statistically rejected but the theoretical yield spread between the long-term and the short-term interest rate based on its validity has a very high correlation with the actual yield spread. This leads them to conclude that "...deviations from the present value model for bonds are transitory...". Building on their framework, Carriero et al. (2006) show that the difference between the actual yield spread and the theoretical yield spread is not statistically significant. In a different setup, previous studies report empirical evidence contradicting the EH, but still potentially valuable for forecasting short-term interest rates, see, e.g. Fama (1984), Fama and Bliss (1987) and Mishkin (1988).

Thus, even if there is a fairly large evidence against the EH, there still seems to be an element of truth in the theory which may be exploited for forecasting and simulation.

This paper formalizes this idea by proposing a way to use the EH without imposing it dogmatically. It does so by using a Bayesian framework such that the extent to which the EH is imposed on the data is under the control of the researcher. This allows to study a continuum of models ranging from one in which the EH hold exactly to one in which it does not hold at all. In this respect, our approach closely parallels that of Ingram and Whiteman (1994) and Del Negro and Schorfheide (2004), who used DSGE models to derive priors for VARs.

In particular, the EH is used to derive a prior on a VAR in the yield spread and the variation in the short-term rates. A tightness hyperparameter controls the noise around the restrictions implied by the EH on the VAR. When the tightness is set to zero, the EH is imposed exactly, while as the tightness goes to infinity the VAR becomes entirely unrestricted. For intermediate values of the tightness there is a whole range of models in which the EH restrictions hold with some degree of uncertainty.

From a statistical point of view, modeling the EH with uncertainty may improve on the traditional way of imposing it exactly. As stressed by Sims (2003), even in simple situations model comparison methods will misbehave when a discrete collection of models is serving as a proxy for a more realistic continuous parameter space. If in the true data generating process a given set of restrictions holds up to some noise, then imposing the restrictions exactly would be suboptimal.

The rationale of modelling the EH as a noisy relation is not merely statistical. The uncertainty around the restrictions has a neat economic interpretation. The EH may be affected by a time varying but stationary term premium and expectations errors. These deviations may be thought of as stationary disturbances around the EH relation. When restrictions are derived from the EH, the stochasticity of the disturbances is transferred to the restrictions which then become inherently fuzzy.

Our analysis leads to two main conclusions.

First, we confirm that the EH does not hold exactly, but we show that it may hold on average, i.e. after integrating out the effect of the deviations which may affect it in the short run. When the EH restrictions are imposed exactly they are rejected but if we allow for some noise around them they are supported by the data. Indeed, the model fitting the data better is neither the one in which the EH is exactly imposed, nor the unrestricted VAR, but a model in which the EH restrictions hold with noise. This finding suggests that the EH prior can be safely imposed to perform forecasting and simulation exercises. It also explains both the common result of rejection, and

the anomalous high correlation between actual and EH consistent spread documented in Campbell and Shiller (1987).

Second, the use of the EH prior provides significant gains in forecast accuracy. The EH prior clearly dominates the unrestricted VAR in predicting both the yield spread and the change in short-term rates. Therefore, using the EH restrictions as priors allows to extract additional information from the term structure of interest rates. The information about the short-term rate contained in long-term rate is extracted and exploited to improve the forecasts of the change in the short-term rate. Then, having a better forecast of the short-term rate allows to improve the forecasts of the yield spread as well. Depending on the estimation window and the forecast horizon, the gains in terms of mean square error can be up to 4 percent in predicting the change in short-term rates and up to 10 percent in predicting the yield spread. These results also explain why previous results contradicting the EH are still potentially valuable for forecasting short-term interest rates (see, e.g. Fama, 1984, Fama and Bliss, 1987 and Mishkin, 1988).

To check whether this good performance is merely due to the use of a shrinkage estimator, the EH prior is also compared to a more competitive forecast model such as a VAR with a Minnesota prior (Doan et al., 1984). This prior shrinks the VAR coefficients to univariate root representations and it has proved to be empirically successful (Litterman, 1986, Todd, 1984) but has the important limitation that it lacks economic justification. As a result, the EH prior does also significantly better than the Minnesota prior in predicting changes in the short-term rates, while the Minnesota prior produces the best forecasts of the yield spread. However, when interpreting this latter result, one should bear in mind that random walk assumption has only a statistical but not an economic justification, while the EH prior is based on economic theory.

The paper is organized as follows: Section 2 introduces the basic framework, Section 3 describes our proposed Bayesian framework, Section 4 provides empirical evidence, Section 5 deals with forecast accuracy, Section 6 concludes. Section 7 contains appendices detailing on the derivation of some results used in the paper.

2 Basic Framework

To make the paper self-contained, this section briefly states the EH and derives a set of restrictions implied by its validity on a bivariate VAR. This section draws heavily on Shiller (1979) and Campbell and Shiller (1987), to which the interested reader may refer for more detailed derivations.

2.1 A linearized expectations model

The EH states that actual long-term interest rates are determined by the market's expectation of future short-term rates. Most simple linear term structure models relate long-term interest rates to an unweighted simple average of expected short rates. Those models are appropriate for pure discount bonds. For coupon-carrying bonds Shiller (1979) proposes a linearized model relating the T -period interest rate (the yield to maturity on T -period bonds) R_t to a weighted average of expected future one-period (short-term) interest rates r_t, r_{t+1}, \dots :

$$R_t = \frac{1 - \gamma}{1 - \gamma^T} \sum_{i=0}^{T-1} \gamma^i E_t(r_{t+i}) + TP_T. \quad (1)$$

Here t denotes the time period, γ is a constant of linearization $0 < \gamma < 1$, TP_T is a constant (i.e. dependent on maturity only) term premium and E_t denotes expectations given information at time t . The parameter γ is set equal to $\gamma = 1/(1 + \bar{R})$, where \bar{R} is the average of R_t . Then (1) relates R_t to the present value of future short-term interest rates discounted by \bar{R} .

Rearranging (1) gives an expression involving the spread $S_t = R_t - r_t$ and the change in short-term rate $\Delta r_t = r_t - r_{t-1}$:

$$S_t = \sum_{i=1}^{T-1} \gamma^i E_t(\Delta r_{t+i}) + \gamma^T (R_t - TP_T) + TP_T. \quad (2)$$

As $T \rightarrow \infty$ this simplifies to

$$S_t = \sum_{i=1}^{\infty} \gamma^i E_t(\Delta r_{t+i}) + TP_{\infty}. \quad (3)$$

where TP_{∞} is the term premium for a bond with an infinite maturity.

2.2 Expectations Hypothesis restrictions

If r_t is stationary in first differences, a necessary condition for the EH to hold is that the vector $x_t = [r_t \ R_t]'$ should be cointegrated with cointegrating vector $\alpha = [-1 \ 1]'$ (up to a scalar normalization). Indeed according to equation (3) the spread $S_t = \alpha' x_t$ has to be stationary, as it equals a constant TP_{∞} plus the discounted sum of the stationary variables $E_t \Delta r_{t+i}$. To impose cointegration Campbell and Shiller (1987) use the following

VAR representation for Δr_t and S_t :

$$\begin{bmatrix} \Delta r_t \\ S_t \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} + \begin{bmatrix} a(L) & b(L) \\ c(L) & d(L) \end{bmatrix} \begin{bmatrix} \Delta r_{t-1} \\ S_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \quad (4)$$

where the polynomials in the lag operator $a(L)$, $b(L)$, $c(L)$, and $d(L)$ are all of order p , and the disturbances are a vector white noise. From equation (3) it is possible to derive the following set of p restrictions implied on the VAR in equation (4):

$$a_j + c_j = 0, \forall j = 1 \dots p; \quad b_1 + d_1 = 1/\gamma; \quad b_j + d_j = 0, \forall j = 2 \dots p. \quad (5)$$

See Appendix A or Campbell and Shiller (1987) for details. To simplify notation, define α as a $2(2p+1)$ vector collecting all the VAR coefficients:

$$\alpha = \left[a_1 \quad b_1 \quad \dots \quad a_p \quad b_p \quad k_1 \quad c_1 \quad d_1 \quad \dots \quad c_p \quad d_p \quad k_2 \right]', \quad (6)$$

the $2p$ restrictions in (5) can be compactly written as:

$$H\alpha = \mu_{EH}, \quad (7)$$

where

$$H = \begin{bmatrix} I_{2p} & \mathbf{0}_{2p \times 1} & I_{2p} & \mathbf{0}_{2p \times 1} \end{bmatrix}, \quad (8)$$

$$\mu_{EH} = \begin{bmatrix} 0 & \frac{1}{\gamma} & \mathbf{0}_{1 \times (2p-2)} \end{bmatrix}'. \quad (9)$$

Notice that the validity of the EH implies that the $2p$ couples of coefficients attached to a given variable in the two equations must be perfectly negatively correlated.

Campbell and Shiller (1987) test the restrictions in (7) via a Wald test, and find strong rejection. As documented below in Section 4, the rejection result is confirmed also using more recent data. Bekaert and Hodrick (2001) argue that a Wald test has bad small sample properties and propose some alternative tests (Lagrange multiplier and distance metric tests) still rejecting the EH for the US. On the other side, Campbell and Shiller (1987) show that imposing the rejected restrictions on the VAR in equation (4) does not yield a significant loss of fit, as the implied fitted value for the spread is very highly correlated with the actual spread. This leads them to the conclusion that there is "some element of truth" in the theory, which seems to hold up to transitory deviations. In the following section we formalize this idea by allowing for some noise around the restrictions.

3 A Bayesian Framework for the Expectations Hypothesis

In this section we develop an extended version of the EH which allows transitory deviations from the theory to occur. This extension leads to a more general framework which comprises the traditional one as a special case. In particular, the EH is used to derive a prior on a VAR in the yield spread and the variation in the short-term rates. A tightness hyperparameter controls the noise around the restrictions implied by the EH on the VAR. When the tightness is set to zero, the EH is imposed exactly, while as the tightness goes to infinity the VAR becomes entirely unrestricted. For intermediate values of the tightness there is a whole range of models in which the EH restrictions hold with an (increasing) degree of uncertainty.

The Section is organized as follows. In Subsection 3.1 and 3.2 we introduce the noise around the EH restrictions and discuss about its economic interpretation. Subsection 3.3 derives the prior from the EH, Subsection 3.4 derives posteriors and the marginal likelihood. Subsection 3.5 extends the framework to priors on second order parameters. Subsection 3.6 describes Bayesian inference.

3.1 Adding uncertainty

By definition any economic theory is a simplification of reality, and as such it can not hold exactly even if the theory is “true”. Suppose the EH does hold, but only on average, i.e. some noise causes temporary departures from the EH restrictions in (7). Formally, let the uncertainty introduced by this noise be measured by the parameter σ . The resulting set of stochastic constraints is:

$$H\alpha \sim N(\mu_{EH}, \sigma I_p). \quad (10)$$

The hyperparameter σ can be interpreted as the tightness of the restrictions. By controlling this hyperparameter we can study a continuum of models ranging from one in which the EH restrictions hold exactly to one in which they do not hold at all. When $\sigma = 0$ the EH restrictions are imposed without noise, which provides a model equivalent to the VAR in equation (4) restricted according equation (7). When $\sigma \rightarrow \infty$ the uncertainty about the restrictions is so high that they are not binding, and this provides a model equivalent to the VAR in equation (4) without any restrictions. For intermediate values of σ the EH is imposed on the model in a non-dogmatic way, i.e. it features some noise. As we shall discuss below, this noise can be interpreted as the effect of time varying (but stationary) term premia and errors in expectations.

3.2 Interpreting noise

In this subsection we discuss why the set of restrictions in equation (7) may hold only up to some noise, as shown in equation (10). Some noise directly affects the EH relation, while some additional noise arises when EH restrictions are derived within the VAR framework.

First, the EH may not hold due to deviations from full market rationality caused by irrational behaviour of some agents or by market frictions. A number of researchers have identified possible sources for this, as the presence of "chartist" or "technical" analysts (Frankel and Froot, 1990, Taylor and Allen, 1992) and/or because of learning of some traders (Lewis 1989) and/or because of the presence of noisy traders (DeLong et al 1990). Second, it may well be the case that the term premium is not constant but features some movement around its mean. Many studies have presented evidence that term premia in bond returns are time varying. In particular, Fama and Bliss (1987) show that term premia do vary through time and are forecastable via the forward rates. Campbell and Shiller (1991) find similar results using yield spreads to predict yield changes. Recently, Cochrane and Piazzesi (2005) have strengthened these results showing that the same linear combination of forward rates predicts bond returns at all maturities. Finally, the EH relation in equation (1) cannot be considered an exact relation, as it comes from a linearization which ignores the Jensen inequality term.

These sources of noise may be thought of as a stationary error term appended to equation (1) as in Clarida and Taylor (1997), who develop a framework featuring the testable implication that deviations from a rational expectations model, whether due to risk aversion or to nonrational expectations, are realizations of a stationary stochastic process. In this light our framework shall allow us to assess the relevance of these deviations and to see how far we are from a world in which in any period term premiums are constant and all agents are fully rational. Clearly, with a stationary disturbance, equation (1) would hold only on average (i.e. after integrating out the effect of the noise) and when restrictions are derived from it the stochasticity of the disturbance is transferred to the restrictions which then become inherently fuzzy.

Additional noise arises when the restrictions are derived within the VAR framework. Indeed, a second approximation is used in order to get linear restrictions when the infinite sum in equation (1) is computed. Moreover, agents' expectations are not observable. Even in presence of market rationality, the econometrician proxies the unobserved expectations of the agents by using a linear projection of a model which is not necessarily the same used by the agents. Again, these sources of noise make restrictions fuzzy.

3.3 A prior from the EH

The set of restriction (10) can be thought of in a Bayesian perspective as a prior on the coefficients of the VAR in equation (4). Define $y = \text{vec}([\Delta r \ S])$, $\Xi = I_2 \otimes [\Delta r_{-1} \ S_{-1} \ \dots \ \Delta r_{-p} \ S_{-p} \ 1]$, and $\varepsilon = \text{vec}([u_1 \ u_2])$. The subscript t has been removed as we are considering the vector of data for each variable. We can now rewrite the VAR in the data-matrix notation:

$$y = \Xi\alpha + \varepsilon. \quad (11)$$

Given a sample size T , y and ε are $2T \times 1$ vectors, and Ξ is the $2T \times 2(2p+1)$ matrix of regressors. Defining Σ_u as the variance matrix of the disturbances in equation (4), the vector ε of disturbances of the vectorized model has variance $\Omega = \Sigma_u \otimes I_T$.

We will refer to the system consisting of the VAR in (11) and the restrictions in (10) as the $EH(\sigma)$ model:

$$\begin{cases} y = \Xi\alpha + \varepsilon \\ H\alpha \sim N(\mu_{EH}, \sigma I_p) \end{cases}. \quad (12)$$

In this model the prior is expressed in terms of linear combinations of the coefficients. In Appendix B we derive from it the following representation which specifies a prior directly on the vector of coefficients:

$$\begin{cases} y = \Xi\alpha + \varepsilon \\ \alpha \sim N(\alpha_{EH}, \Sigma_{EH}) \end{cases}, \quad (13)$$

where

$$\alpha_{EH} = \begin{bmatrix} 0 & 0 & 1/\gamma & 0 \\ 1 \times (2p+1) & & & 1 \times (2p-1) \end{bmatrix}', \quad (14)$$

and

$$\Sigma_{EH} = \begin{bmatrix} \delta I_{2p} & 0 & -\delta I_{2p} & 0 \\ 0 & \delta & 0 & 0 \\ -\delta I_{2p} & 0 & (\sigma + \delta) I_{2p} & 0 \\ 0 & 0 & 0 & \delta \\ 1 \times 2p & & 1 \times 2p & \end{bmatrix}. \quad (15)$$

The parameter δ is the prior variance of the unrestricted coefficients: as there are $2(2p+1)$ coefficients and $2p$ restrictions, $2p+2$ coefficients are unrestricted. To these coefficients is assigned a variance of δ which is set to an arbitrary high number to get uninformative.¹

¹In the application δ is set to 10^6 following Doan et al. (1984).

To clarify the role played by the tightness parameter σ it is worth to look at the correlation matrix of the coefficients under the EH-prior, which is easily derived from (15):

$$Corr(\alpha) = \begin{bmatrix} I_{2p} & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} I_{2p} & 0 \\ 0 & 1 & 0 & 0 \\ \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} I_{2p} & 0 & I_{2p} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (16)$$

Notice that depending on the value of the tightness parameter σ we move from the exact restrictions case ($\sigma = 0$) to the unrestricted VAR ($\sigma \rightarrow \infty$). If $\sigma = 0$, the EH is in the traditional form and involves perfect negative correlation between the relevant $2p$ couples of coefficients of the VAR. Letting $\sigma > 0$ we allow this correlation to be imperfect. As $\sigma \rightarrow \infty$, the correlation across the relevant couples of coefficients goes to zero and the correlation matrix approaches that of a VAR without cross equation restrictions.

3.4 Posteriors and marginal likelihood

The $EH(\sigma)$ model described in equation (13) features exact closed form solutions for the coefficients posterior densities and the marginal likelihood. The posterior density is normal with variance:

$$\Sigma_{\bar{\alpha}} = (\Sigma_{EH}^{-1} + \Xi' \Omega^{-1} \Xi)^{-1}, \quad (17)$$

and mean

$$\bar{\alpha} = \Sigma_{\bar{\alpha}} (\Sigma_{EH}^{-1} \alpha_{EH} + \Xi' \Omega^{-1} y). \quad (18)$$

The marginal likelihood is:

$$p(y) = (2\pi)^{-T} |\Omega|^{-1/2} |\Sigma_{\bar{\alpha}}|^{1/2} |\Sigma_{EH}|^{-1/2} \exp(-Q/2), \quad (19)$$

where

$$Q = y' \Omega^{-1} y - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH}. \quad (20)$$

Derivations of the posterior and marginal likelihood are contained in Appendix C.

3.5 Second order parameters

The priors used so far feature a fixed covariance matrix of the errors ($\Omega = \Sigma_u \otimes I_N$) as in Theil and Goldberger (1961) and Litterman (1986), and imply the existence of the closed form solutions for posteriors and marginal likelihoods described above in equations

(17)-(20). More generally, we could specify a prior also on the variance matrix of the disturbances. A natural choice would be a Wishart prior:

$$\Sigma_u^{-1} \sim Wi(\nu_0, \Sigma_{u0}^{-1}). \quad (21)$$

Uninformativeness is achieved by setting $\nu_0 = 0$ and $\Sigma_{u0} = 0_{2 \times 2}$. If we assume independency between α and Σ_u equations (21) and (13) constitute the so called independent Normal-Wishart prior. Posteriors and marginal likelihoods for this prior are computed by simulation, see appendix C for details.

3.6 Bayesian inference

Bayesian inference is drawn by means of the Bayes factor. The Bayes factor is a summary of the evidence provided by the data in favour of one theory, represented by a statistical model, as opposed to another.

Following Kass and Raftery (1995), consider some data \mathbf{D} assumed to have arisen under one of the two theories H_1 and H_2 according to a probability density $pr(\mathbf{D}|H_1)$ or $pr(\mathbf{D}|H_2)$. Given a priori probabilities $pr(H_1)$ and $pr(H_2) = 1 - pr(H_1)$, the data produce a posteriori probabilities $pr(H_1|\mathbf{D})$ and $pr(H_2|\mathbf{D}) = 1 - pr(H_1|\mathbf{D})$. Since any prior opinion gets transformed to a posterior opinion through consideration of the data, the transformation itself represents the evidence provided by the data. Once we convert to the odds scale ($odds = probability/(1 - probability)$) the transformation takes a simple form. Using Bayes theorem, we obtain $\frac{pr(H_2|\mathbf{D})}{pr(H_1|\mathbf{D})} = \frac{pr(\mathbf{D}|H_2) pr(H_2)}{pr(\mathbf{D}|H_1) pr(H_1)}$, so that the transformation is simply multiplication by

$$B_{21} = \frac{pr(\mathbf{D}|H_2)}{pr(\mathbf{D}|H_1)}. \quad (22)$$

The factor B_{21} is the Bayes factor of theory H_2 as opposed to theory H_1 . In our example, the data are in y and each model H_i corresponds to a different value of the tightness σ .

Kass and Raftery (1995) extensively discuss the use of Bayes factors and propose a scale to interpret it. Their suggested interpretation appears in Table 1. Notice that Bayes factors can equally well provide evidence in favour of a null hypothesis. For example, a $2 \ln B_{21}$ between 6 and 10 provides both evidence against H_1 and in favour of H_2 , while a $2 \ln B_{21}$ between -10 and -6 provides both evidence against H_2 and in favour of H_1 .

4 Empirical Evidence

What is usually done in a classical framework is a comparison of a restricted model against an unrestricted model. This strategy may result to be ineffective whenever in the true data generating process a given set of restrictions holds on average, i.e. it holds up to some noise. As stressed by Sims (2003), even in simple situations model comparison methods will misbehave when the discrete collection of models is serving as a proxy for a more realistic continuous parameter space.

In this section we estimate the VAR in equation (4) under the stochastic constraints given by equation (10) for several different values of the tightness of such constraints. This amounts in exploring a continuum of models ranging from one in which the EH holds exactly to one in which it does not hold at all.

The rationale for allowing for noise around the EH is not merely statistical. As discussed in Section 3.2, the uncertainty around the restrictions has a neat economic interpretation. The EH may be affected from a time varying but stationary term premium and expectations errors.

We apply our framework to US data. First, we gauge the appropriateness of the VAR in equation (4) in describing the data. Then we proceed to the Bayesian estimation of the model for different values of the tightness parameter and we draw inference by mean of Bayes factors. Finally, we discuss our results.

4.1 Data and preliminary results

Our data set is at monthly frequency and consists of the 1-month certificate of deposit rate in the U.S. secondary market and the 10-year U.S. Treasury bond yield, at a constant maturity rate. Both series are provided by the Federal Reserve of St.Louis.

First, we check whether the VAR in equation (4) provides a good representation of the data. In order to avoid problems with parameter instability, we leave out of the sample the so called Volcker reserve-targeting period and use the sample 1983:1 to 2006:12. In this subsample the VAR is stationary and recursive *OLS* estimates and Chow tests do not detect any structural break. The lag length of the VAR is set to 3 by using the Schwartz criterion (with a maximum lag length of 13). The residuals feature some outliers which are easily removed by means of five dummy variables. Diagnostic tests provide evidence in favor of nonautocorrelation and homoscedasticity of the disturbances. In particular, the LM test statistic reported in Johansen (1995) for the null of no autocorrelation up to order 4 is well below the critical value, and the White (1980) test does not reject the null of homoscedasticity at the 5% confidence level.

Then we turn to the classical estimation of VAR in equation (4), both unrestricted and restricted according to equation (7). Estimation results for these cases are reported in the first and third columns of Table 2. The remaining columns on Table 2 refer to the Bayesian estimation and we shall discuss them in next subsection. The Wald test strongly rejects the restrictions in (7), consistently with Campbell and Shiller (1987).

4.2 Bayesian estimation

Now we turn to the Bayesian estimation of the $EH(\sigma)$ model described by equation (13).

We start from the two extreme cases. The first case is the entirely restricted model, i.e. the $EH(\sigma)$ model with $\sigma = 0$. The second extreme is the entirely unrestricted VAR, which is the $EH(\sigma)$ model with $\sigma = \infty$. For future reference, we name these models respectively $EH(0)$ and $EH(\infty)$.² As shown in Table 2, the $EH(0)$ is equivalent to the VAR in equation (4) restricted according to equation (3). Indeed the Bayesian posterior estimates of $EH(0)$ are virtually the same as those obtained by *FIML* estimation of the restricted VAR. Similarly, the $EH(\infty)$ model is equivalent to the VAR in equation (4) without any restrictions, as Bayesian posterior estimates of $EH(\infty)$ are virtually the same as *OLS* estimates of the unrestricted VAR.

Then, we estimate the $EH(\sigma)$ model for a grid of values of σ ranging from 0 to ∞ . The marginal likelihood of each of these models is graphed in Figure 1. As is clear from the graph, the marginal likelihood is hump-shaped and features a sharp peak in the point $\sigma = \sigma^* = 0.085$. This means that the best model (i.e. the one with the higher marginal likelihood) is neither the one which imposes the EH exactly, nor the unrestricted VAR, but a model in which the EH restrictions hold on average, i.e. up to some noise. We call this model $EH(\sigma^*)$. Posterior estimates of this model are reported in the 5th column of Table 2, while the last two columns of the table report respectively the deviation from the estimates of the $EH(0)$ and $EH(\infty)$ models.

Notice that if we would have used the classical framework to test restrictions, we would have focused only on the two extreme cases $\sigma = 0$ and $\sigma = \infty$. This means that we would end up ignoring the model featuring the highest marginal likelihood, which is in between the two extremes.

All the above results are based on a fixed covariance matrix of the errors as in Theil and Goldberger (1961) and Litterman (1986). As discussed in Section 3.5 we could specify a prior also on the matrix of second order parameters. In that case we do not have closed form solutions for posteriors and marginal likelihoods and we need to estimate

²In the application, neither the case $\sigma = 0$ nor the case $\sigma = \infty$ are computationally feasible, therefore we use respectively $\sigma = \textit{eps}$ and $\sigma = 10^6$, where *eps* is the precision of the used software.

them by simulation.³ For details see appendix C. The marginal likelihood for this case is also reported in Figure 1. As is clear, it virtually coincides with that computed under the simpler Normal prior with fixed variance. Recomputing the estimates contained in Table 2 yields the same results as in the previous case. As this case is much more demanding from a computational perspective, in the remainder of the paper we use the prior with fixed covariance matrix of the errors.

4.3 Bayes factors

The sharp peak featured by the marginal likelihood in the point σ^* provides neat evidence that the model with restrictions imposed with uncertainty is the best one. To check how strong is the evidence in favour of that model one may use the Bayes factor. The $(2ln)$ Bayes factor of the $EH(\sigma^*)$ model versus the $EH(0)$ is 19.95, while the $(2ln)$ Bayes factor of the $EH(\sigma^*)$ model versus the $EH(\infty)$ is 90.59. These figures signal very strong evidence in favor of the $EH(\sigma^*)$ model with respect to the entirely restricted and the entirely unrestricted VAR.

There are still two issues to be investigated. First, an important concern about the above results is related to the so called Lindley paradox. It is well known that a prior with a very high variance is likely to be significantly disadvantaged respect to a tighter one. In our example the $EH(\infty)$ model features a much higher variance with respect to both the $EH(0)$ and the $EH(\sigma^*)$ models. Therefore, the rejection of the $EH(\infty)$ model may be the spurious result of the fact that it features a too high variance.⁴ Second, it would be interesting to see the shape of the Bayes factor for the EH prior against the unrestricted VAR for all the possible values of the tightness.

To address these issues we specify an alternative competing model, call it *UVAR*, which also does not impose the EH restrictions, but which features a prior with much less variance than the $EH(\infty)$ model. In particular we use pre-sample data (from 1966:1 to 1982:12) to estimate the VAR by *OLS*, and then we use these *OLS* estimates to elicitate a prior for the competing model.⁵ On the other side, the $EH(\sigma)$ model is left completely unaltered. Then we compute the Bayes factors for all the possible values of the tightness parameter σ .

³To compute posteriors and marginal likelihoods we use an algorithm implementing Gibbs sampling . We use the BACC algorithm for MATLAB available at www2.cirano.qc.ca/~bacc/index.html

⁴Importantly, the Lindley paradox effect does not affect the other result, namely that the $EH(\sigma^*)$ model is better than $EH(0)$. Indeed, in that case the model featuring more variance is the $EH(\sigma^*)$ model, so if there is any Lindley paradox effect this would work against and not in favour of it.

⁵The prior mean is fixed to the OLS estimates, while the variance of the coefficients is a diagonal homoskedastic matrix featuring the same determinant as OLS estimates.

The results from this analysis are reported in Figure 2, which plots the Bayes factor (B_{21}) as a function of the EH prior tightness σ , together with the inconclusive region.⁶ If we allow for very little noise, letting $\sigma \rightarrow 0$, the Bayes factor supports the *UVAR*. This is the common result of rejection, as letting the tightness go to zero amounts to imposing the EH without noise. On the other hand, allowing for large departures from the EH leads the Bayes factor to the inconclusive region. Intuitively, the noise on the constraints becomes so large that the $EH(\sigma)$ becomes virtually equivalent to the *UVAR*, the two models end up having the same marginal likelihood, and the Bayes factor converges to 1. For intermediate values of σ the Bayes factor strongly supports the $EH(\sigma)$ model. In the point $\sigma = \sigma^* = 0.085$ the $(2ln)$ Bayes factor reaches the value of 11.99, which is lower than the value of 90.59 previously found. This signals that there was a relevant Lindley paradox effect, but it still provides strong evidence in favour of the $EH(\sigma)$ model and against the unrestricted VAR.

4.4 Discussion

In this section we estimated a continuum of models ranging from one in which the EH restrictions hold exactly to one in which they do not hold at all. As a result, the best model is neither the one in which the EH is exactly imposed, nor the unrestricted one, but the model in which the EH restrictions hold with uncertainty.

We have drawn inference by means of the Bayes factors. When the EH restrictions are imposed exactly, they are rejected. If we allow for some noise around the EH restrictions, they are supported by the data. Therefore, the EH holds on average, i.e. after integrating out the effect of the deviations which may affect it in the short run. As stressed in section 3.2 these deviations have a neat economic interpretation, as the EH may be affected from a time varying but stationary term premium and expectations errors.

The fact that the EH holds on average suggests that the EH prior can be safely imposed on the data to perform simulation exercises. Moreover, it explains both the common result of rejection, and the anomalous high correlation between actual and EH consistent spread documented in Campbell and Shiller (1987).

As the data indicate that the best model is that in which the EH is imposed in a non-dogmatic way it is natural to ask whether using the EH prior may improve the accuracy of forecasts. This issue is addressed in the next section.

⁶Notice that the shape of the Bayes factor is the same of that of the Marginal likelihood, as the *UVAR* does not depend on σ . This can not be immediately seen from figures 1 and 2 as the Bayes factor is graphed in logs.

5 How to Extract Additional Information From the Term Structure

In this subsection we show that the VAR with the EH prior, i.e. the $EH(\sigma)$ model, produces significant improvements in forecast accuracy with respect to the VAR estimated without imposing any prior information. Moreover, the EH prior has a significantly better performance with respect to the Minnesota prior in forecasting variations in the short-term rate. This means that using the EH as a prior allows to extract additional information from the term structure of interest rates. Paralleling Del Negro and Schorfheide (2004) we assess forecast accuracy for several values of the prior tightness σ , and we distinguish between the case in which σ is chosen ex post or ex ante.

5.1 Preliminaries

We start with evaluating the overall forecasting performance of the VAR with the EH prior for different values of the tightness parameter σ and for different forecast horizons. The results are based on a forecasting experiment performed on the whole sample with a rolling estimation window of 12 years, which is roughly 1/2 of the available sample (i.e. 144 observations). Figure 3 plots the percentage gain in forecasting (z axis) for different values of the EH prior tightness σ (x axis) and for different forecast horizons (y axis). The percentage gain is defined as the difference between the *in – det* statistic proposed by Doan et al. (1984) obtained using the EH prior and that obtained using the simple unrestricted VAR described in equation (4).⁷ Positive values of the difference between the *in – det* statistic imply a positive gain from using the EH as prior information. Several things can be seen from the picture.

First, overall the forecasting performance of the $EH(\sigma)$ model is a hump-shaped function of σ , and is maximized by small but positive values of the tightness parameter σ . For the one-step ahead case the value of σ providing the best forecasts is 0.03. The optimal value for the tightness decreases as the forecast horizon increases.

Second, when $\sigma = \infty$ the gain is 0 by definition, as the prior is so loose that it is practically ineffective. Recall from Table 2 that the Bayesian estimates of the model $EH(\sigma)$ with $\sigma = \infty$ are virtually identical to the *OLS* estimates of the unrestricted VAR in equation (4).

⁷Specifically, the statistic is given by the log determinant of the forecast error covariance matrix divided by the number of forecasted variables. Aside from covariance terms, this number is the average mean square error made in forecasting each variable in the VAR. As the statistic is in logs, the percentage gain/loss is obtained simply by taking the differences and multiplying by 100.

Third, for longer horizons, imposing the restrictions exactly is suboptimal respect to using them as a prior, but it still provides pretty good forecasts. Indeed, for horizons longer than two-step ahead, the gain in using $\sigma = 0$ is smaller than using the optimal σ , but it is still high. This is consistent with those results contradicting the EH but still potentially valuable for forecasting short-term interest rates, see, e.g. Fama (1984), Fama and Bliss (1987) and Mishkin (1988).

Finally, the forecasting performance of the $EH(\sigma)$ is higher at intermediate horizons. While the gain at the 1-month horizon is about 2%, and at the 12-month horizon about 4%, for intermediate horizons it goes up to 7%.

5.2 Forecasts comparisons

The analysis performed so far is intended mostly as an initial inspection of the forecasting performance of our $EH(\sigma)$ model. There are still several issues to investigate. First, the analysis was based on ex post values for σ , so it is unclear whether the forecast gains can still be obtained when the optimal tightness has to be chosen ex-ante, i.e. before the actual forecasts errors become available. Second, one may want to look at the forecasting performance for each variable under analysis. Third, we may want to assess formally the statistical significance (if any) of the gain in using the EH prior. Finally, we may want to compare the EH prior to a more competitive opponent than a simple unrestricted VAR, e.g. the Minnesota prior.

To see whether the forecast gains can still be obtained when the optimal tightness for the EH prior is chosen ex-ante, at each point in time, before estimating the $EH(\sigma)$ model, we estimate the optimal tightness σ as the value which maximizes the Marginal likelihood. This is simply done by using a grid search over some values of σ ranging from 0 to ∞ .

We look at the forecasting performance for each variable under analysis, i.e. the variation in short-term rates, and the spread between long-term and short-term rates. As loss function we choose the mean squared forecast error. To assess whether the difference in the forecasts is significant we use the test for predictive accuracy recently developed by Giacomini and White (2006). This is a test for the null of equal forecasting method accuracy and as such can handle forecasts based on both nested and non-nested models, regardless from the estimation procedures used in the derivation of the forecasts, including Bayesian and semi- and non-parametric estimation methods.

To check whether the good forecasting performance is merely due to the use of a shrinkage estimator, we compare the EH prior also to a more competitive forecast

model such as a VAR with a Minnesota prior. This prior shrinks the VAR coefficients to univariate root representations and it has proved to be empirically successful (Litterman, 1986 Todd, 1984) but has the important limitation that it lacks economic justification. In particular, the Minnesota prior we use is that described in Doan et al. (1984), the only difference being that the prior mean on the first lag of the variables is set to 0 rather than to 1. This is due to the fact that the variables assumed (a-priori) to follow the random walk are R_t and r_t . If the coefficient on their first lag is shrunk to 1, the coefficient on the first lag of their transformations S_t and Δr_t , those entering equation (4), has to be shrunk to 0 to be consistent with the random walk hypothesis. We optimize the choice for the hyperparameters of the Minnesota prior as well, by doing at each point in time a grid search over the two hyperparameters controlling the tightness of the prior.⁸

Results of this analysis are displayed in Tables 3 and 4. Table 3 compares the forecasting performance of the EH prior against the unrestricted VAR, while Table 4 compares the EH prior to the Minnesota prior. The tables contain results for different lengths of the estimation window used for the forecasting exercise. In particular, we use our baseline estimation window of 12 years (1/2 of the sample, 144 obs.) as well as a shorter window of 10 (2/5 of the sample, 120 obs.) and 8 years (1/3 of the sample, 96 obs.). We do so to check for robustness, as there is a trade off between the precision of model estimates and the precision of mean squared forecast error estimates. Indeed, using a wider estimation window improves the precision in estimating the forecasting model but yields fewer observations for the estimation of the mean squared forecast error. Several conclusions can be drawn.

First, the EH prior does significantly better than the unrestricted VAR in predicting both the change in short-term rates and the yield spread. For horizons longer than 2-step ahead, the (significant) gain in using the EH prior ranges from 1.39 to 3.95 percent when forecasting changes in the short-term rates and from 3.49 to 10.26 percent when forecasting the spread. For shorter horizons the forecasts of the two models are not statistically different.

Second, the EH prior outperforms the Minnesota prior in predicting the change in the short-term rates. Again, for shorter horizons there is no significant difference between the two models, but for horizons longer than 2-step ahead the EH prior produces significant gains in forecasting, ranging from 1.15 to 5.25 percent depending on the forecast horizon and the estimation window. The only exception to this is the significantly

⁸Regarding the remaining two hyperparameters of the Minnesota prior, the decay parameter is held fix at a linear rate, while the variance of the constants is set equal to that used for the unrestricted coefficients in the EH prior.

worse performance at the very long horizons when a 8-year estimation window is used.

Third, the Minnesota prior is on average the best model for predicting the spread. When using an 8-year estimation window, the difference between the EH prior and the Minnesota prior forecasts is insignificant, but the results obtained with longer estimation windows show that using the Minnesota prior produces significant and high gains, which may be up to 13.38 percent. Also in this case, there are not significant differences at the very short horizons.

5.3 Discussion

In this section we investigated whether using the EH as prior information yields significant improvements in forecasting.

We found that the EH prior improves significantly over the unrestricted VAR in predicting both the variation in short-term rates and the yield spread. The EH prior extracts the information about the short-term rate contained in the long-term rate, and exploit this information to improve the forecasts of the short-term rate. Then, having a better forecast of the short-term rate allows to improve the forecasts of the long-term rate as well. Depending on the estimation window and the forecast horizon, the gains in terms of mean squared error can be up to 4 percent when predicting the change in short-term rates and up to 10 percent when predicting the yield spread.

When the EH prior is compared to a more competitive benchmark, the Minnesota prior, it still produces the best forecasts for the variation in short-term rates. However, the Minnesota prior produces the best forecasts of the spread. The fact that the EH improves forecasts of the short-term rates but is beaten by the Minnesota prior in predicting the spread is a bit puzzling. A similar paradoxical result has been found by Campbell and Shiller (1991) who show that the slope of the term structure almost always gives a forecast in the wrong direction for the short-term change in the yield on the longer bond, but gives a forecast in the right direction for long-term changes in short rates.

The point is that even if using the EH as a prior improves the accuracy in forecasting the short-term rates, the size of such improvement decays as the forecast horizon increases. Forecasting the 10-year rate requires forecasting the short-term rate up to the 10-year horizon, and for very long horizons using the EH prior does not yield any advantage. This explains why a model imposing the random walk assumption may do better. However, one should bear in mind that the random walk assumption has only a statistical justification, while the EH prior has been derived from economic theory.

6 Conclusions

Central bank researchers routinely impose the EH to forecast short-term rates, to assess how monetary policy affects long-term rates, and to measure market expectations. Also the monetary VAR literature often imposes the EH to disentangle expected from unexpected movements in interest rates.

However, the fairly large evidence against the EH casts serious doubts on the appropriateness of using the EH for these forecasting and simulation exercises.

This paper has proposed a way to use the EH without imposing it dogmatically on the data. In particular, rather than being used to derive a set of exact restrictions, the EH has been used to derive a prior on a VAR in the yield spread and the variation in the short-term rates. A hyperparameter controls the tightness of the EH prior. When the tightness is set to zero, the EH is imposed exactly on the VAR, while as the tightness goes to infinity the VAR becomes entirely unrestricted. For intermediate values of the tightness there is a whole range of models in which the EH restrictions hold with an (increasing) degree of uncertainty. In this respect our approach is very close to that of Ingram and Whiteman (1994) and Del Negro and Schorfheide (2004), who used respectively RBC and DSGE models to derive priors for VARs.

As a result, the best model is neither the one in which the EH is exactly imposed, nor the unrestricted VAR, but the model in which the EH restrictions hold with noise. This result explains both the common result of rejection, and the surprisingly high correlation between actual and EH consistent spread documented in Campbell and Shiller (1987). The Bayes factor provides evidence that the EH holds on average, i.e. after integrating out the effect of the deviations (time varying stationary risk premia and errors in expectations) which may affect it in the short run. This suggests that the EH prior can be safely imposed on the data to perform simulation exercises.

Using the EH as a prior also allows to extract additional information from the term structure of interest rates. Our forecasting exercise provides evidence that the EH prior clearly dominates the unrestricted VAR in predicting both the yield spread and the change in short-term rates. The gains in terms of mean square error can be up to 4 percent in predicting the change in short-term rates and up to 10 percent in predicting the yield spread. These results may also explain why previous results contradicting the EH are still potentially valuable for forecasting short-term interest rates (see, e.g. Fama, 1984, Fama and Bliss, 1987 and Mishkin, 1988). The EH prior also does significantly better than the Minnesota prior in predicting changes in the short-term rates, while the Minnesota prior produces the best forecasts of the yield spread.

7 Appendices

A. Derivation of the EH restrictions

Here we sketch the derivation of the EH restrictions. For a complete derivation see Campbell and Shiller (1987). Demean the variables and stack the VAR in equation (4) as:

$$\begin{bmatrix} \Delta \tilde{r}_t \\ \Delta \tilde{r}_{t-1} \\ \dots \\ \Delta \tilde{r}_{t-p+2} \\ \Delta \tilde{r}_{t-p+1} \\ \tilde{S}_t \\ \tilde{S}_{t-1} \\ \dots \\ \tilde{S}_{t-p+2} \\ \tilde{S}_{t-p+1} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p & b_1 & b_2 & \dots & b_{p-1} & b_p \\ 1 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ c_1 & c_2 & \dots & c_{p-1} & c_p & d_1 & d_2 & \dots & d_{p-1} & d_p \\ 0 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta \tilde{r}_{t-1} \\ \Delta \tilde{r}_{t-2} \\ \dots \\ \Delta \tilde{r}_{t-p+1} \\ \Delta \tilde{r}_{t-p} \\ \tilde{S}_{t-1} \\ \tilde{S}_{t-2} \\ \dots \\ \tilde{S}_{t-p+1} \\ \tilde{S}_{t-p} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ 0 \\ \dots \\ 0 \\ 0 \\ u_{2t} \\ 0 \\ \dots \\ 0 \\ 0 \end{bmatrix},$$

where the $\tilde{\cdot}$ indicates that the variables are taken in deviations from their mean. Define A the coefficient matrix, v_t the vector of disturbances, and z_t as the vector containing Δr_t , S_t and their lags. The VAR can be compactly written as:

$$z_t = Az_{t-1} + v_t.$$

The EH (in deviation from the means) states that:

$$\tilde{S}_t = \sum_{i=1}^{\infty} \gamma^i E_t(\Delta \tilde{r}_{t+i}),$$

Define now two selector vectors g' and h' , both composed by $2p$ elements, all of which are zero except for the $(p+1)$ th element of g' and the 1st element of h' which are unity. In this notation, the above equation is:

$$g' z_t = \sum_{i=1}^{\infty} \gamma^i E_t(h' z_{t+i}).$$

Using the VAR projection to proxy for expectations yields:

$$g'z_t = \sum_{i=1}^{\infty} \gamma^i A^i z_t.$$

Since the above expression has to hold in general, it holds for any z_t :

$$g' = \sum_{i=1}^{\infty} \gamma^i h' A^i.$$

Then, exploiting the properties of geometric series we have:

$$g' = h' \gamma A (I - \gamma A)^{-1},$$

and postmultiplying by $(I - \gamma A)$ provides the following set of $2p$ linear restrictions:

$$g'(I - \gamma A) = h' \gamma A,$$

i.e.:

$$\begin{cases} a_j + c_j = 0, \forall j = 1 \dots p \\ b_1 + d_1 = 1/\gamma \\ b_j + d_j = 0, \forall j = 2 \dots p \end{cases}.$$

B. Derivation of the EH prior

Define $y = \text{vec}([\Delta r \ S])$, $\Xi = I_2 \otimes [\Delta r_{-1} \ S_{-1} \ \dots \ \Delta r_{-p} \ S_{-p} \ 1]$, and $\varepsilon = \text{vec}([u_1 \ u_2])$. The subscript t has been removed as we are considering the vector of data for each variable.

We can now rewrite the VAR in the data-matrix notation:

$$y = \Xi \alpha + \varepsilon,$$

i.e.:

$$\begin{aligned} \underbrace{\begin{bmatrix} \Delta r \\ S \end{bmatrix}}_{2T \times 1} &= \underbrace{\begin{bmatrix} I_2 \otimes X \\ \end{bmatrix}}_{2T \times 2(2p+1)} * \underbrace{\alpha}_{2(2p+1) \times 1} + \underbrace{\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}}_{2T \times 1}, \\ X &= \begin{bmatrix} \Delta r_{-1} & S_{-1} & \Delta r_{-2} & S_{-2} & \Delta r_{-3} & S_{-3} & 1 \end{bmatrix}, \\ \alpha &= \begin{bmatrix} a_1 & b_1 & a_2 & b_2 & a_3 & b_3 & k_1 & c_1 & d_1 & c_2 & d_2 & c_3 & d_3 & k_2 \end{bmatrix}', \\ \varepsilon &\sim N\left(0, \Omega = \Sigma_u \otimes I_T\right), \end{aligned}$$

Given a sample size T , y and ε are $2T \times 1$ vectors, and Ξ is the $2T \times 2(2p+1)$ matrix of regressors, and the vector α is the $2(2p+1)$ vector collecting all the VAR coefficients. Defining Σ_u as the variance matrix of the disturbances, the vector ε of disturbances of the vectorized model has variance $\Omega = \Sigma_u \otimes I_T$. The EH restrictions are:

$$\begin{cases} a_j + c_j = 0, \forall j = 1 \dots p \\ b_1 + d_1 = 1/\gamma \\ b_j + d_j = 0, \forall j = 2 \dots p \end{cases}.$$

These $2p$ restrictions can be compactly written as:

$$H\alpha = \mu_{EH},$$

where

$$H = \begin{bmatrix} I_{2p} & 0_{2p \times 1} & I_{2p} & 0_{2p \times 1} \end{bmatrix},$$

and

$$\mu_{EH} = \begin{bmatrix} 0 & \frac{1}{\gamma} & 0_{1 \times (2p-2)} \end{bmatrix}'.$$

Adding the noise to each of the restrictions yields:

$$H\alpha \sim N(\mu_{EH}, \sigma I_p).$$

The generic form of a normal prior satisfying the EH restrictions would be:

$$\alpha \sim N(\alpha_{EH}, \Sigma_{EH}),$$

multiplying by H :

$$H\alpha \sim N(H\alpha_{EH}, H\Sigma_{EH}H'),$$

so there is the following relation between the prior moments of the vector of restrictions and those of the vector of coefficients:

$$\begin{aligned} \mu_{EH} &= H\alpha_{EH} \\ \sigma I_p &= H\Sigma_{EH}H' \end{aligned}.$$

The above system has no unique solution for α_{EH} and Σ_{EH} : as there are $2(2p + 1)$ coefficients and only $2p$ restrictions, $2(p + 1)$ coefficients are not restricted and H is not square. To solve this problem we simply set a prior with arbitrarily high variance δ on the unrestricted coefficients.

$$H_2\alpha_{EH} \sim N(\mu_{2EH}, \Sigma_2),$$

with:

$$H_2 = \begin{bmatrix} I_{2p} & \mathbf{0}_{2p \times 1} & \mathbf{0}_{2p \times 2p} & \mathbf{0}_{2p \times 1} \\ \mathbf{0}_{1 \times 2p} & 1 & \mathbf{0}_{1 \times 2p} & \mathbf{0} \\ I_{2p} & \mathbf{0}_{2p \times 1} & I_{2p} & \mathbf{0}_{2p \times 1} \\ \mathbf{0}_{1 \times 2p} & \mathbf{0} & \mathbf{0}_{1 \times 2p} & 1 \end{bmatrix}, \mu_{2EH} = \begin{bmatrix} \mathbf{0}_{2p \times 1} \\ \mathbf{0} \\ \mu_{EH} \\ \mathbf{0} \end{bmatrix}, \Sigma_2 = \begin{bmatrix} \delta I_{2p} & \mathbf{0}_{2p \times 1} & \mathbf{0}_{2p \times 2p} & \mathbf{0}_{2p \times 1} \\ \mathbf{0}_{1 \times 2p} & \delta & \mathbf{0}_{1 \times 2p} & \mathbf{0} \\ \mathbf{0}_{2p \times 2p} & \mathbf{0}_{2p \times 1} & \sigma I_{2p} & \mathbf{0}_{2p \times 1} \\ \mathbf{0}_{1 \times 2p} & \mathbf{0} & \mathbf{0}_{1 \times 2p} & \delta \end{bmatrix}.$$

The third block of this matrices produces the EH set of restrictions, while the remaining blocks specify a uninformative prior on the unrestricted coefficients. Now we can invert the matrix H_2 and get a prior explicitly specified in terms of the vector of coefficients, rather than in terms of the restrictions.

$$\alpha \sim N(\alpha_{EH}, \Sigma_{EH}),$$

where

$$\alpha_{EH} = H_2^{-1} \mu_{2EH} = \begin{bmatrix} 0 & 0 & 1/\gamma & 0 \\ 1 \times (2p+1) & & & 1 \times (2p-1) \end{bmatrix}',$$

and

$$\Sigma_{EH} = H_2^{-1} \Sigma_2 H_2'^{-1} = \begin{bmatrix} \delta I_{2p} & 0 & -\delta I_{2p} & 0 \\ 0 & \delta & 0 & 0 \\ -\delta I_{2p} & 0 & (\sigma + \delta) I_{2p} & 0 \\ 0 & 0 & 0 & \delta \\ 1 \times 2p & & 1 \times 2p & \end{bmatrix}.$$

The $EH(\sigma)$ model consists of the VAR plus the EH-restrictions:

$$\begin{cases} y = \Xi \alpha + \varepsilon \\ \alpha \sim N(\alpha_{EH}, \Sigma_{EH}) \end{cases}$$

The correlation matrix is:

$$Corr(\alpha) = \begin{bmatrix} I_{2p} & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} I_{2p} & 0 \\ 0 & 1 & 0 & 0 \\ \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} I_{2p} & 0 & I_{2p} & 0 \\ 0 & 0 & 0 & 1 \\ 1 \times 2p & & 1 \times 2p & \end{bmatrix}$$

C. Derivation of Posterior Densities and Marginal Likelihood.

Consider the VAR in equation (13):

$$\begin{aligned} y_{2T \times 1} &= \Xi_{2T \times Mk} \alpha_{2k \times 1} + \varepsilon_{2T \times 1}, \\ \alpha &\sim N(\alpha_{EH}, \Sigma_{EH}), \\ e &\sim N(0, \Omega = \Sigma_u \otimes I_T), \end{aligned}$$

Here 2 is the number of equations, p is the number of lags included, $k = 2p + 1$ is the number of regressors and T is the sample size, while α_{EH} and Σ_{EH} are the coefficients prior moments. The prior density is:

$$p(\alpha) = (2\pi)^{-2k/2} |\Sigma_{EH}|^{-1/2} \exp \left\{ -\frac{1}{2} (\alpha - \alpha_{EH})' \Sigma_{EH}^{-1} (\alpha - \alpha_{EH}) \right\},$$

the likelihood is⁹:

$$p(y|\alpha) = (2\pi)^{-2T/2} |\Omega|^{-1/2} \exp \left\{ -\frac{1}{2} (y - \Xi\alpha)' \Omega^{-1} (y - \Xi\alpha) \right\},$$

a posterior density kernel is:

$$\begin{aligned} p(y|\alpha)p(\alpha) &= (2\pi)^{-2(T+k)/2} |\Omega|^{-1/2} |\Sigma_{EH}|^{-1/2} \\ &\exp \left\{ -\frac{1}{2} [(y - \Xi\alpha)' \Omega^{-1} (y - \Xi\alpha) + (\alpha - \alpha_{EH})' \Sigma_{EH}^{-1} (\alpha - \alpha_{EH})] \right\}. \end{aligned} \quad (23)$$

Now define¹⁰:

$$\Sigma_{\bar{\alpha}} = (\Sigma_{EH}^{-1} + \Xi' \Omega^{-1} \Xi)^{-1}, \quad (24)$$

$$\bar{\alpha} = \Sigma_{\bar{\alpha}} (\Sigma_{EH}^{-1} \alpha_{EH} + \Xi' \Omega^{-1} y). \quad (25)$$

⁹Notice that: $|\Omega|^{-1/2} = |\Sigma_u \otimes I_T|^{-1/2} = (|\Sigma_u|^T |I_T|^2)^{-1/2} = |\Sigma_u|^{-T/2}$.

¹⁰Notice that:

$$\begin{aligned} \Sigma_{\bar{\alpha}} &= [\Sigma_{EH}^{-1} + \Xi' \Omega^{-1} \Xi]^{-1} = [\Sigma_{EH}^{-1} + (I_2 \otimes X)' (\Sigma_u \otimes I_T) (I_2 \otimes X)]^{-1} \\ &= [\Sigma_{EH}^{-1} + I_2' \Sigma_u I_2 \otimes X' I_T X]^{-1} = [\Sigma_{EH}^{-1} + \Sigma_u \otimes X' X]^{-1}, \\ \bar{\alpha} &= \Sigma_{\bar{\alpha}} [\Sigma_{EH}^{-1} \alpha_{EH} + \Xi' \Omega^{-1} y] = \Sigma_{\bar{\alpha}} [\Sigma_{EH}^{-1} \alpha_{EH} + (I_2 \otimes X)' (\Sigma_u \otimes I_T) y] \\ &= \Sigma_{\bar{\alpha}} [\Sigma_{EH}^{-1} \alpha_{EH} + (I_2' \Sigma_u \otimes X' I_T) y] = \Sigma_{\bar{\alpha}} [\Sigma_{EH}^{-1} \alpha_{EH} + (\Sigma_u \otimes X') y]. \end{aligned}$$

Using the above definitions and completing the square yields the following expression for the term in square brackets on the right-hand side of equation (23):

$$\begin{aligned}
& (y - \Xi\alpha)' \Omega^{-1} (y - \Xi\alpha) + (\alpha - \alpha_{EH})' \Sigma_{EH}^{-1} (\alpha - \alpha_{EH}) \\
= & y' \Omega^{-1} y - y' \Omega^{-1} \Xi \alpha - \alpha' \Xi' \Omega^{-1} y + \alpha' \Xi' \Omega^{-1} \Xi \alpha \\
& + \alpha' \Sigma_{EH}^{-1} a - \alpha' \Sigma_{EH}^{-1} a_{EH} - \alpha_{EH}' \Sigma_{EH}^{-1} \alpha + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH} \\
= & y' \Omega^{-1} y - (y' \Omega^{-1} \Xi + \alpha_{EH}' \Sigma_{EH}^{-1}) \alpha - \alpha' (\Xi' \Omega^{-1} y + \Sigma_{EH}^{-1} a_{EH}) \\
& + \alpha' (\Xi' \Omega^{-1} \Xi + \Sigma_{EH}^{-1}) \alpha + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH} \\
= & y' \Omega^{-1} y - (\Sigma_{\bar{\alpha}}^{-1} \bar{\alpha})' \alpha - \alpha' (\Sigma_{\bar{\alpha}}^{-1} \bar{\alpha}) + \alpha' (\Sigma_{\bar{\alpha}}^{-1}) \alpha + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH} \\
= & y' \Omega^{-1} y - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \alpha - \alpha' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + \alpha' \Sigma_{\bar{\alpha}}^{-1} \alpha + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH} \\
= & y' \Omega^{-1} y - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH},
\end{aligned}$$

where the last line follows using some algebra.¹¹ Substituting this result in equation (23), the posterior density kernel can be written as follows:

$$\begin{aligned}
p(y|\alpha)p(\alpha) &= (2\pi)^{-2(T+k)/2} |\Omega|^{-1/2} |\Sigma_{EH}|^{-1/2} \\
& \exp \left\{ -\frac{1}{2} \left[y' \Omega^{-1} y - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH} \right] \right\}.
\end{aligned}$$

Defining

$$Q = y' \Omega^{-1} y - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH},$$

¹¹Since:

$$\begin{aligned}
\alpha' \Sigma_{\bar{\alpha}}^{-1} \alpha &= (-\alpha + \bar{\alpha} - \bar{\alpha})' \Sigma_{\bar{\alpha}}^{-1} (-\alpha + \bar{\alpha} - \bar{\alpha}) = (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) \\
& + (-\bar{\alpha})' \Sigma_{\bar{\alpha}}^{-1} (-\bar{\alpha}) + (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (-\bar{\alpha}) + (-\bar{\alpha})' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha),
\end{aligned}$$

we have that:

$$\begin{aligned}
& y' \Omega^{-1} y - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \alpha - \alpha' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + [\alpha' \Sigma_{\bar{\alpha}}^{-1} \alpha] + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH} \\
= & y' \Omega^{-1} y - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \alpha - \alpha' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + \left[\begin{array}{l} (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) + \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + \\ -(\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) \end{array} \right] + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH} \\
= & y' \Omega^{-1} y - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \alpha - \alpha' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) \\
& + \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} - (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH} \\
= & y' \Omega^{-1} y - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} [\alpha + (\bar{\alpha} - \alpha)] - [(\bar{\alpha} - \alpha)' + \bar{\alpha}'] \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} \\
& + (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) + \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH} \\
= & y' \Omega^{-1} y - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) + \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH} \\
= & y' \Omega^{-1} y - \bar{\alpha}' \Sigma_{\bar{\alpha}}^{-1} \bar{\alpha} + (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) + \alpha_{EH}' \Sigma_{EH}^{-1} \alpha_{EH}.
\end{aligned}$$

we have:

$$p(y|\alpha)p(\alpha) = (2\pi)^{-2(T+k)/2} |\Omega|^{-1/2} |\Sigma_{EH}|^{-1/2} \exp \left\{ -\frac{1}{2} [(\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) + Q] \right\}. \quad (26)$$

Forgetting constants:

$$p(y|\alpha)p(\alpha) \propto \exp \left\{ -1/2 [(\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha)] \right\} \implies p(\alpha|y) \sim N(\bar{\alpha}, \Sigma_{\bar{\alpha}}),$$

which shows that $\bar{\alpha}, \Sigma_{\bar{\alpha}}$ in equations (24) and (25) are the moments of the posterior. The posterior properly normalized density is:

$$p(\alpha|y) = (2\pi)^{-2k/2} |\Sigma_{\bar{\alpha}}|^{-1/2} \exp \left\{ -\frac{1}{2} (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) \right\}. \quad (27)$$

The marginal likelihood is given by the integral over the $2 \times k$ dimensional space of the product of the properly normalized prior and data densities:

$$p(y) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p(y|\alpha)p(\alpha) d\alpha_1 \dots d\alpha_{2k} = \int_{\mathfrak{R}^{2k}} p(y|\alpha)p(\alpha) d\alpha.$$

Using equation (26):

$$\begin{aligned} p(y) &= \int_{\mathfrak{R}^{2k}} (2\pi)^{-2(T+k)/2} |\Omega|^{-1/2} |\Sigma_{EH}|^{-1/2} \exp \left\{ -\frac{1}{2} [(\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) + Q] \right\} d\alpha, \\ &= (2\pi)^{-2(T+k)/2} |\Omega|^{-1/2} |\Sigma_{EH}|^{-1/2} \exp(-Q/2) \int_{\mathfrak{R}^{2k}} \exp \left\{ -\frac{1}{2} [(\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha)] \right\} d\alpha. \end{aligned} \quad (28)$$

Notice it is important that the properly normalized prior and properly normalized likelihood, and not arbitrary kernels of these densities, be used in forming the marginal likelihood. Now exploit the fact that the posterior properly normalized density in equation (27) integrates to one:

$$\begin{aligned} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p(\alpha|y) d\alpha_1 \dots d\alpha_{2k} &= \int_{\mathfrak{R}^{2k}} p(\alpha|y) d\alpha = 1 \implies \\ \int_{\mathfrak{R}^{2k}} (2\pi)^{-2k/2} |\Sigma_{\bar{\alpha}}|^{-1/2} \exp \left\{ -\frac{1}{2} (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) \right\} d\alpha &= 1 \implies \\ \int_{\mathfrak{R}^{2k}} \exp \left\{ -\frac{1}{2} (\bar{\alpha} - \alpha)' \Sigma_{\bar{\alpha}}^{-1} (\bar{\alpha} - \alpha) \right\} d\alpha &= \frac{1}{(2\pi)^{-2k/2} |\Sigma_{\bar{\alpha}}|^{-1/2}}. \end{aligned} \quad (29)$$

Substituting (29) in the expression on the right-hand side of equation (28) gives the following expression for the marginal likelihood:

$$p(y) = (2\pi)^{-2(T+k)/2} |\Omega|^{-1/2} |\Sigma_{EH}|^{-1/2} \exp(-Q/2) \frac{1}{(2\pi)^{-2k/2} |\Sigma_{\bar{\alpha}}|^{-1/2}},$$

which simplifies to:

$$p(y) = (2\pi)^{-T} |\Omega|^{-1/2} |\Sigma_{\bar{\alpha}}|^{1/2} |\Sigma_{EH}|^{-1/2} \exp(-Q/2).$$

The above results apply to the case with fixed variance matrix of the errors ($\Omega = \Sigma_u \otimes I_T$) as in Theil (1961) and Litterman (1986). Alternatively, we could specify a prior also on this matrix, as a Wishart:

$$\Sigma_u^{-1} \sim Wi(\nu_0, \Sigma_{u0}^{-1}).$$

Uninformativeness for this prior is achieved by setting $\nu_0 = 0$ and $\Sigma_{u0} = 0_{2 \times 2}$. If we assume independency between α and Σ_u , the above prior, coupled with the normal prior on coefficients constitutes the so called independent Normal-Wishart prior:

$$\begin{aligned} p(\alpha, \Sigma_u) &= p(\alpha)p(\Sigma_u), \\ \alpha &\sim N(\alpha_{EH}, \Sigma_{EH}), \\ \Sigma_u^{-1} &\sim Wi(\nu_0, \Sigma_{u0}^{-1}). \end{aligned}$$

This prior implies the following conditional posterior distributions (for a derivation see Geweke 2005):

$$\begin{aligned} \alpha|y, \Sigma_u^{-1} &\sim N(\bar{\alpha}, \Sigma_{\bar{\alpha}}), \\ \Sigma_u^{-1}|y, \alpha &\sim Wi(\nu_0 + T, (\Sigma_{u0} + S)^{-1}), \end{aligned}$$

where the generic element of the matrix S is $s_{ij} = (y_i - \Xi_i \alpha)(y_j - \Xi_j \alpha)$, and $i, j = 1, 2$ signal the subvector or submatrix composed by the T rows associated with the i -th and j -th equation (so for example in our case $y_1 = \Delta r_t$ and $y_2 = S_t$). These conditional posterior distributions are the foundation of a Gibbs sampling algorithm which successively draws from $p(\alpha|y, \Sigma_u^{-1})$ and $p(\Sigma_u^{-1}|y, \alpha)$ to simulate draws from the unconditional posteriors. Marginal likelihoods are then computed numerically. For details see Geweke (2005) p. 165 or Koop (2003) p.137.

References

- [1] Backus D., Foresi S., Mozumdar A., Wu L., 2001, "Predictable Changes in Yields and Forward Rates", *Journal of Financial Economics*, Vol. 59, pp. 281-311.
- [2] Bekaert G., Hodrick R.J., Marshall D., 1997, "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates", *Journal of Financial Economics*, Vol. 44, pp. 309-48.
- [3] Bekaert, G., Hodrick R.J., 2001, Expectations Hypotheses Tests, *The Journal of Finance*, Vol. 56, pp 1357–1394.
- [4] Campbell J., 1995, "Some Lessons from the Yield Curve", *Journal of Economic Perspectives*, Vol. 9, pp. 129-152.
- [5] Campbell J., Shiller R., 1987, "Cointegration and Tests of Present Value Models", *Journal of Political Economy*, Vol. 95, pp.1062-1088.
- [6] Campbell J., Shiller R., 1991, "Yield Spreads and Interest Rate Movements: A Bird's Eye View", *Review of Economic Studies*, Vol. 58, pp. 495-514.
- [7] Carriero A., Favero C., Kaminska I., 2006, "Financial factors, Macroeconomic Information and the Expectations Theory of the Term Structure of interest Rates", *Journal of Econometrics*, Vol. 131, pp. 339-358.
- [8] Clarida R., Taylor M.P., 1997, "The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors", *The Review of Economics and Statistics*, Vol. 79, pp. 353-361.
- [9] Clews R., 2002. "Asset Prices and Inflation" *Quarterly Bulletin Bank of England*, Summer 2002, 178-85.
- [10] Cochrane J., Piazzesi M., 2005, "Bond Risk Premia", *American Economic Review*, Vol. 95, pp. 138-160.
- [11] De Long J. B., Shleifer A., Summers L.H., Waldmann R., 1990, "Noise Trader Risk in Financial Markets", *Journal of Political Economy*, Vol. 98, pp. 703-738.
- [12] Del Negro, M., Schorfheide, F., 2004, "Priors from General Equilibrium Models for VARs," *International Economic Review*, Vol. 45, 643-673.

- [13] Doan T., Litterman R., Sims C., 1984, "Forecasting and Conditional Projection Using Realistic Prior Distributions", *Econometric Reviews*, Vol. 3, pp. 1-100.
- [14] *ECB Monthly Bulletin*. <http://www.ecb.int/pub/mb/html/index.en.html>
- [15] Fama E.F., 1984, "The Information in the Term Structure", *Journal of Financial Economics*, Vol.13, pp. 509-528.
- [16] Fama E., Bliss R.R., 1987, "The Information in Long-Maturity Forward Rates", *American Economic Review*, Vol. 77, 680-692.
- [17] Fisher I., 1930, *The Theory of Interest*, Macmillan, New York.
- [18] Frankel J.A., Froot K.A., 1990, "Chartists, Fundamentalists, and the Demand for Dollars", in Courakis A. S. and Taylor M. P. (eds), *Private Behaviour and Government Policy in Interdependent Economies*, Oxford University Press, Oxford, pp. 73-128.
- [19] Geweke, J. (2005) *Contemporary Bayesian Econometrics and Statistics*. Wiley.
- [20] Giacomini R., White H., 2006, "Tests of Conditional Forecast Accuracy", *Econometrica*, Vol. 74, pp. 1545-1578.
- [21] Hicks J., 1953, *Value and Capital*, Oxford University Press, Oxford.
- [22] Ingram B., Whitemann C., 1994, "Supplanting the Minnesota prior. Forecasting Macroeconomic Time Series Using Real Business Cycle Priors", *Journal of Monetary Economics*, Vol. 34, pp. 497-510.
- [23] Johansen S., 1995, *Likelihood-based Inference in Cointegrated Vector Auto-regressive Models*, Oxford University Press, Oxford.
- [24] Kass R. E., Raftery A., 1995, "Bayes Factors", *Journal of the American Statistical Association*, Vol. 90, pp. 773-795.
- [25] Keynes J.M., 1930, *Treatise on Money*, Harcourt, Brace and Howe Inc., New York.
- [26] Koop, G., 2003, *Bayesian Econometrics*, Wiley, Chicester.
- [27] Krueger J.T., Kuttner, K.N, 1996, "The Fed Funds Futures Rate as a Predictor of Federal Reserve Policy", *Journal of Futures Markets*, Vol. 16, pp. 965-979.

- [28] Lewis K.K., 1989, "Changing beliefs and systematic rational forecast errors with evidence from foreign exchange", *American Economic Review*, Vol. 79, pp. 269-276.
- [29] Litterman R.B., 1986, "Forecasting with Bayesian Vector Autoregressions-Five Years of Experience", *Journal of Business & Economic Statistics*, Vol. 4, pp. 25-38.
- [30] Mishkin F.S., 1988, "The Information in the Term Structure: Some Further Results", *Journal of Applied Econometrics*, Vol. 3, pp. 307-14.
- [31] Rudebusch G.D., 1998, "Do Measures of Monetary Policy in a VAR Make Sense?", *International Economic Review*, Vol. 39, pp. 907-31.
- [32] Scholtes C., 2002, "On Market-Based Measures of Inflation Expectations", *Quarterly Bulletin of Bank of England*, Spring 2002, pp. 67-77.
- [33] Shiller R., 1979, "The Volatility of Long Term Interest Rates and Expectations Models of the Term Structure", *Journal of Political Economy*, Vol. 87, pp. 1190-1219.
- [34] Sims C.A., 2003, "Probability Models for Monetary Policy Decisions", manuscript, Princeton University.
- [35] Söderlind P., Svensson L.E.O., 1997, "New Techniques to Extract Market Expectations from Financial Instruments", *Journal of Monetary Economics*, Vol. 40, pp. 373-429.
- [36] Stambaugh R.F., 1988, "The Information in Forward Rates: Implications for Models of the Term Structure", *Journal of Financial Economics*, Vol. 21, pp. 41-70,
- [37] Taylor M.P., Allen H.L., 1992, "The use of technical analysis in the foreign exchange market", *Journal of International Money and Finance*, Vol. 11, 304-314.
- [38] Theil H., Goldberger A.S., 1961, "On Pure and Mixed Estimation in Economics", *International Economic Review*, Vol. 2, pp. 65-78.
- [39] Todd R.M., 1984, "Improving Economic Forecasting with Bayesian Vector Autoregression", *Federal Reserve Bank of Minneapolis Quarterly Review*, pp.18-29.
- [40] White H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity", *Econometrica*, Vol. 48, pp. 817-838.

Tables and Figures

Table 1: Interpreting Bayes factors

$2 \ln B_{21}$	B_{21}	Evidence Against H_1
0 to 2	1 to 3	Bare Mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
> 10	> 150	Very Strong

Source: Kass and Raftery (1995).

Table 2: Estimation results

	Unrestricted VAR		Restricted VAR		VAR with EH prior		
	<i>OLS</i>	<i>EH</i> (∞)	<i>FIML</i>	<i>EH</i> (0)	<i>EH</i> (σ^*)	% <i>d</i> _{<i>EH</i>(0)}	% <i>d</i> _{<i>EH</i>(∞)}
<i>a</i> ₁	0.4528	0.4528	0.2901	0.2901	0.4293	48.0%	-5.18%
<i>b</i> ₁	0.2591	0.2591	0.1058	0.1057	0.2378	124%	-8.23%
<i>a</i> ₂	-0.0036	-0.0036	0.0914	0.0914	0.0146	-84.0%	-509%
<i>b</i> ₂	-0.2781	-0.2781	-0.0610	-0.0610	-0.2372	288%	-14.6%
<i>a</i> ₃	0.2016	0.2016	0.2054	0.2054	0.1987	-3.28%	-1.46%
<i>b</i> ₃	0.0409	0.0409	-0.0232	-0.0232	0.0212	-191%	-48.2%
<i>k</i> ₁	-0.0219	-0.0219	-0.0234	-0.0234	-0.0218	-6.73%	-0.45%
<i>c</i> ₁	-0.0298	-0.0298	-0.2901	-0.2901	-0.0674	-76.7%	125%
<i>d</i> ₁	1.1338	1.1338	0.8884	0.8885	1.0997	23.7%	-3.00%
<i>c</i> ₂	-0.2433	-0.2433	-0.0914	-0.0914	-0.2142	134%	-11.9%
<i>d</i> ₂	-0.2861	-0.2861	0.0610	0.0610	-0.2208	-461%	-22.8%
<i>c</i> ₃	-0.2115	-0.2115	-0.2054	-0.2054	-0.2162	5.25%	2.22%
<i>d</i> ₃	0.1256	0.1256	0.0232	0.0232	0.0940	305%	-25.1%
<i>k</i> ₂	0.0117	0.0117	0.0093	0.0094	0.0118	27.0%	1.37%

The first group of columns provides estimates of the VAR without restrictions. The first column reports the OLS estimates, the second column the Bayesian estimates. The second group of columns provides estimates of the VAR with the EH restrictions exactly imposed. The first column reports the FIML estimates, the second the Bayesian estimates. The third group of columns provides the estimates of the VAR with the EH prior. The first column reports the estimates, the second and the third the percentage deviations from the estimates of the unrestricted and the restricted VAR.

Table 3: Forecasting performance of EH prior against the unrestricted VAR

h	96-obs		120-obs		144-obs	
	Δr	S	Δr	S	Δr	S
1	-1.065	-0.105	0.755	-1.31	1.469	-2.769
2	1.335	5.727	0.853	2.387	1.501	-0.483
3	2.639**	10.26***	1.537	5.568**	1.563	3.390
4	3.957***	9.824***	2.867***	5.441**	3.397***	4.474**
5	3.379***	8.589***	2.282***	4.654**	3.438***	5.142***
6	3.028***	7.288***	2.302***	3.498*	3.227***	5.089***
7	3.066***	6.124***	2.518***	2.557	3.149***	4.790***
8	2.674***	5.381***	2.186***	1.906	2.684***	4.650***
9	2.463***	4.862***	2.208***	1.333	2.365***	4.379***
10	2.176**	4.511***	2.039***	0.980	2.057***	4.106***
11	1.747**	4.297***	1.654***	0.773	1.670***	3.887***
12	1.454**	4.167***	1.452***	0.679	1.386***	3.655***

The table reports the percentage gain in forecasting using the VAR with the EH prior rather than the unrestricted VAR. The column h reports the forecast horizon. The percentage gain is defined as the percentage decrease in Mean Squared Error. The symbols *, **, ***, mean significancy at the 10, 5, 1 percent confidence level according to the Giacomini and White (2006) test. Results are computed for three different rolling estimation windows. The first group of columns provides results based on a rolling estimation window of 1/3 of the sample (96-obs), the second group provides results based on a rolling estimation window of 2/5 of the sample (120-obs), the third group provides results based on a rolling estimation window of 1/2 of the sample (144-obs).

Table 4: Forecasting performance of EH prior against the Minnesota prior

h	96-obs		120-obs		144-obs	
	Δr	S	Δr	S	Δr	S
1	2.149	2.329	-0.148	-0.767	-2.063	-2.5610
2	3.405	1.658	0.773	-4.369	0.642	-5.4880
3	3.967*	0.260	1.649	-7.266**	1.958	-7.222**
4	5.255***	-2.242	4.503***	-10.85***	4.446***	-10.18***
5	3.867***	-3.191	3.252**	-12.86***	4.063***	-11.23***
6	2.834**	-2.801	2.974**	-13.38***	3.994***	-11.39***
7	1.82**	-2.148	3.173***	-13.08***	3.294***	-11.12***
8	0.522	-1.23	2.174***	-12.25***	2.466***	-10.20***
9	-0.177	-0.564	1.631***	-11.16***	2.026***	-9.131***
10	-0.953	-0.262	1.231***	-10.06***	1.152***	-8.082***
11	-1.767***	-0.136	0.423	-9.042***	0.562	-7.051***
12	-2.075***	-0.291	-0.065	-8.148***	0.167	-6.174***

The table reports the percentage gain in forecasting using the VAR with the EH prior rather than the a VAR with the Minnesota prior. The column h reports the forecast horizon. The percentage gain is defined as the percentage decrease in Mean Squared Error. The symbols *, **, ***, mean significance at the 10, 5, 1 percent confidence level according to the Giacomini and White (2006) test. Results are computed for three different rolling estimation windows. The first group of columns provides results based on a rolling estimation window of 1/3 of the sample (96-obs), the second group provides results based on a rolling estimation window of 2/5 of the sample (120-obs), the third group provides results based on a rolling estimation window of 1/2 of the sample (144-obs).

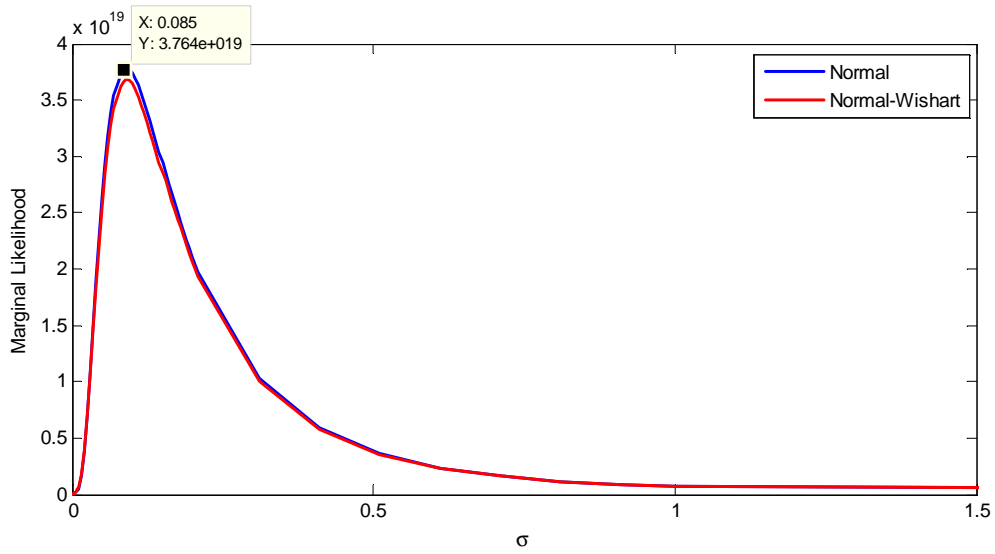


Figure 1: Marginal likelihood of the $EH(\sigma)$ model, as a function of σ . The blue line is based on the Normal prior with fixed variance matrix for the disturbances, while the red line is based on the independent Normal-Wishart prior.

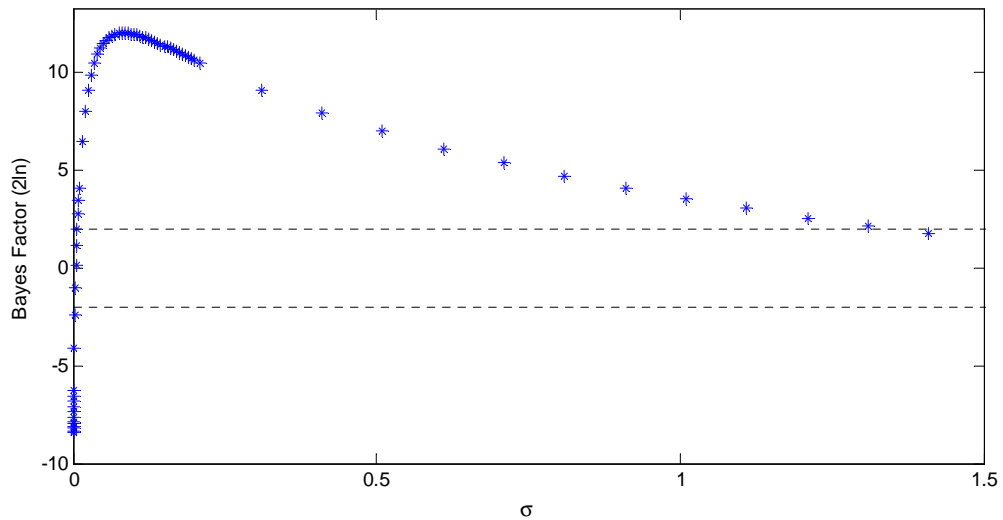


Figure 2: Twice the log of the Bayes factor of the $EH(\sigma)$ model against the $UVAR$, as a function of σ . The two dotted lines represent the so-called inconclusive region. Below the dotted lines the data provide evidence in favour of the $UVAR$. Above the dotted lines the data provide evidence in favour of the $EH(\sigma)$ model. Within the dotted lines the data do not provide evidence in favour of any of the two models.

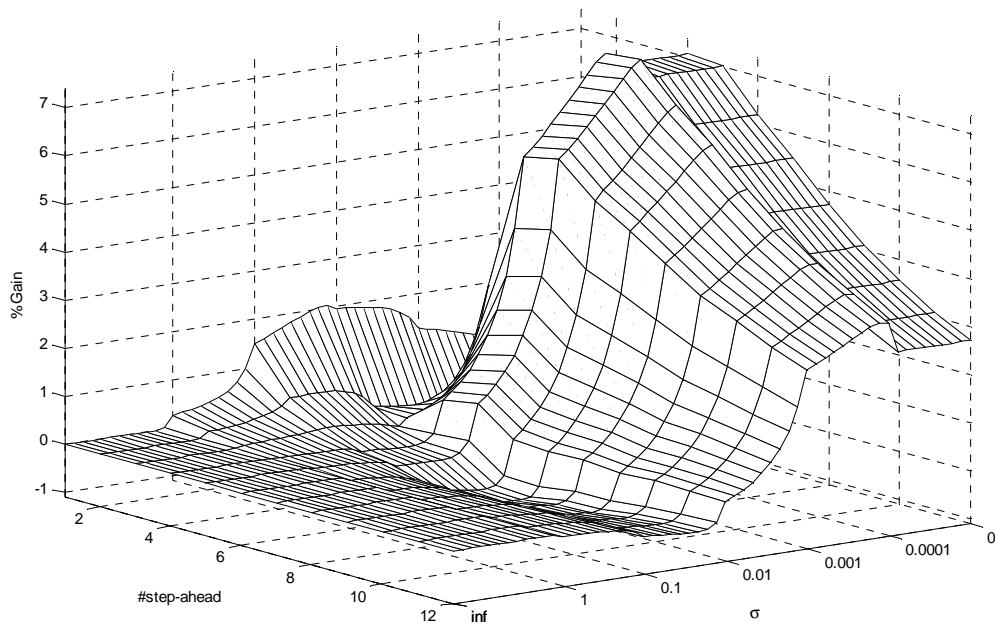


Figure 3: Percentage gain in forecasting using the VAR with the EH prior rather than the unrestricted VAR. The x axis reports the EH prior tightness, the y axis the forecast horizon. The percentage gain is the difference in the multivariate in-det statistic. The in-det statistic is given by the log determinant of the forecast error covariance matrix divided by the number of forecasted variables..

**This working paper has been produced by
the Department of Economics at
Queen Mary, University of London**

**Copyright © 2007 Andrea Carriero
All rights reserved**

**Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm**