Resurrecting the Expectations Hypothesis: How to Extract Additional Information From the Term Structure of Interest Rates

Andrea Carriero*
Università Bocconi

First draft: March 2004
This version: July 2005

Abstract

In this paper we propose a new way of modelling the Expectations Hypothesis (EH) of the term structure of interest rates and provide striking evidence validating it on both statistical and economic grounds. The idea is to model the EH as a noisy relation, allowing for temporary departures from it. We do so using a Bayesian framework in which the EH can be viewed as a prior on a gaussian VAR. Importantly, our approach is very general and comprises the traditional framework as a special case. Once the EH is modeled as a noisy relation it is strongly supported by the data and is entirely consistent with the behavior of the U.S. 10-year rate from the seventies onwards. Moreover, our evidence explains the common result of rejection and the anomaly found by Campbell and Shiller (1987). Finally, our approach allows to extract additional information from the term structure and then to significantly increase the accuracy of a Taylor-rule based model in predicting future short term rates.

JEL Classification: C11, E43, E44, E47.

Keywords: Bayesian VARs, Expectations Theory, Macroeconomic Information in Finance, Term Structure, Uncertain Restrictions

*Address: IEP Università Bocconi, Via Gobbi 5, 20136 Milano, Italy. E-mail: andrea.carriero@unibocconi.it. Homepage: www.igier.unibocconi.it/carriero. I am indebted to Carlo Favero. I thank, without implicating them, Marco Aiolfi, Gianni Amisano, Pierluigi Balduzzi, Fabio Canova, Marco Del Negro, Andrea Ghiringhelli, Alfred Haug, Bruce N. Lehmann, Tae-Hwy Lee, Massimiliano Marcellino, Linlin Niu, Giorgio Primiceri, Barbara Rossi, Luca Sala, Nicola Scalzo, George Tauchen, Ulf Söderström, and seminar participants at Bocconi University and University of Brescia. This paper has been presented 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 Restictions".

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 the writings of 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. Notwithstanding its important role in macroeconomics and finance, the EH has been widely criticized on theoretical grounds and has received little empirical support.

From a theoretical perspective the EH was not viewed as a viable model of the term structure for several years, due to the result of Cox, Ingersoll and Ross (1981) that the EH is not consistent with the absence of arbitrage. More recently, however, McCulloch (1993) and Fisher and Gilles (1998) have presented counterexamples to the Cox, Ingersoll and Ross (1981) proof, and Longstaff (2000) shows that all traditional forms of the EH can be consistent with the absence of arbitrage if markets are incomplete. In particular he demonstrates that the Cox, Ingersoll and Ross (1981) proof hinges on the complete market hypothesis, which does not necessarily hold, as shown by several studies (see, for example, Daves and Ehrhardt 1993, Amihud and Mendelson 1991, Kamara 1994, Duffee 1996, Longstaff 1992, Boudoukh and Whitelaw 1991, Cornell and Shapiro 1990, Elton and Green 1998). Therefore, the EH cannot be ruled out on theoretical grounds and its validity is purely an empirical issue.

Turning to the empirical validation, the EH has been widely tested, and almost invariably rejected. The bulk of the available literature (see, for example, Fama and Bliss 1987, Campbell 1995, and Cochrane 2001) testing the EH has two common features. First, it uses a single-equation, limited information approach. Second, it uses ex-post realized returns as a proxy for ex-ante expected returns. There are several problems related to this testing strategy. First, as noted by McCallum (1994), the limited information approach might cause a bias in the estimates due to simultaneity. Moreover, Elton (1999) asserts that there is ample evidence against the belief that information surprises tend to cancel out over time. Hence realized returns cannot be considered as an appropriate proxy for expected returns. Finally, Campbell (1995) finds strong effects of expectation errors on the single-equation tests, which are confirmed by a number of papers relating expectation errors to peso problems. Therefore, having a good approximation of expected returns is crucial when testing the theory. Keeping these latter points in mind, it is not surprising that a single equation approach proxying expected returns with ex-post realized returns rejects the EH. In many cases expectations that were entirely reasonable ex-ante may turn out to be completely wrong ex-post. In all these instances using ex-post realized returns amounts to using irrational forecasts and this biases the test of the EH.

Campbell and Shiller (1987) circumvent these problems by using a bivariate framework. Their approach provides model-based proxies for ex-ante expected returns. Still, as noted by Carriero, Favero and Kaminska (2005), their results are biased by the fact that they use information from the whole sample to simulate investors' expectations while investors can only use historically available information to generate predictions of short-term rates. Campbell and Shiller (1987) implement a Wald test which still rejects the EH but their analysis of the data leads them to conclude that there is an important element of truth in the EH. In particular, they find an anomaly: 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. Building on a similar framework, Carriero, Favero and Kaminska (2005) show that these two spreads are not statistically different, while the high correlation between them leads Campbell and Shiller (1987) to conclude that "...deviations from the present value model for bonds are transitory...".

We take this last point and develop an extended version of the EH in which transitory deviations from the theory may occur. This is the only key assumption made throughout the paper, and it seems reasonable. By definition any economic theory is a simplification of reality, and as such it cannot hold exactly even if the theory is "true". Indeed, in the real world there is always some kind of noise which blurs any equilibrium. The proposed extension leads to a more general framework which comprises the traditional one as a special case.

Many studies have presented forward rate regressions providing 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 and show that the same linear combination of forward rates predicts bonds returns at all maturities. These empirical results have been interpreted in the literature as strong evidence against the EH, as the EH is viewed to be consistent only with constant term premia, and equivalent to the statement that excess returns should not be predictable.

Two comments are in order. First, even if the EH involves a constant (i.e. dependent on maturity only) term premium, our framework does not rule out the possibility of transitory variations in the term premium. Indeed once we allow for transitory deviations from the theory, the constancy of the term premium has a slightly different interpretation: if (as we will show) the EH holds "on average", then the term premium can be a time varying process, with the only requirement that it has to revert to its constant mean. Our results provide evidence that the variations in the term premium are not large enough to reject the theory. Second, the existence of a time varying term premium is not necessarily at odds with the EH: Longstaff (1990) has shown that the EH can actually imply time varying term premia if the time frame for which the EH holds differs from the return measurement period.

We adopt the Bayesian approach developed by Jeffreys (1935, 1961) as a major part of his program for scientific inference. In this approach, statistical models are introduced to represent the probability of the data according to several competing theories, and Bayes's theorem is used to compute the posterior probability that a theory is correct. Then the theories can be compared using the Bayes factor, which is a summary of the evidence provided by the data in favour of one theory as opposed to another. In particular, we have two competing theories, represented by two different priors on the coefficients of a gaussian bivariate VAR. The first theory does not impose any restriction, and it is shaped into a loose, proper prior. The second theory imposes the restrictions derived from the EH, and is shaped into a prior on some linear combinations of the coefficients.

To ensure the robustness of our results, we extend the testing framework in two dimensions, by providing statistical results recursively through time, and by adding macroeconomic information to the picture. The EH implies that monetary policy affects long-term rates by influencing expectations about future short-term rates, but as central banks also look at bond markets to extract information about inflation expectations, monetary policy (i.e. the short-term rate) is likely to respond to bond market conditions.

Our main results are: i) The EH is strongly supported by the data both on statistical and economic grounds (see Section 4); ii) The EH is entirely consistent with U.S. 10-year rate dynamics from the beginning of the seventies onwards (see Section 5); iii) The EH allows to extract additional information from the term structure and therefore to significantly increase the accuracy of a Taylor-rule based model in predicting future short-term rates (see Section 6). Moreover, our evidence comprises the common result of rejection of the EH as a special case and explains the anomaly found by Campbell and Shiller (1987).

The paper is organized as follows: Section 2 introduces the basic framework, Section 3 derives our extended framework, and Section 4 provides statistical evidence. Section 5 discusses the ability of our model to explain the dynamics of the 10-year rate, Section 6 evaluates forecast accuracy, Section 7 concludes.

2 The Basic Framework

In this section we introduce our basic framework, developed by Shiller (1979) and Campbell and Shiller (1987). First, we formally state the EH, then we derive a set of restrictions implied by its validity on a bivariate VAR.

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) $R_{t,T}$ to a weighted average of expected future one-period (short-term) interest rates $r_t, r_{t+1}, \ldots$:

$$R_{t,T} = \frac{1-\gamma}{1-\gamma^T} \sum_{i=0}^{T-1} \gamma^i E_t r_{t+i} + TP_T. \tag{1}$$

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

$$S_{t,T} = \sum_{i=1}^{T-1} \gamma^i E_t \Delta r_{t+i} + \gamma^T \left(R_{t,T} - TP_T\right) + TP_T. \tag{2}$$

As $T \to \infty$ this simplifies to

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

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

2.2 Expectations Hypothesis restrictions

Our data set consists of the 1-month certificate of deposit in the U.S. secondary market rate and the 10-year U.S. Treasury bond yield, at a constant maturity rate. Data are monthly and go from 1966:1 to 2004:1. Both series are provided by the Federal Reserve of St.Louis. Following Campbell and Shiller (1987) we consider a VAR for S_t and Δr_{t+i}:

$$\begin{aligned}\Delta r_t &= k_1 + a_1\Delta r_{t-1} + a_2\Delta r_{t-2} + a_3\Delta r_{t-3} \\ &\quad + b_1 S_{t-1} + b_2 S_{t-2} + b_3 S_{t-3} + u_{1t}, \\ S_t &= k_2 + c_1\Delta r_{t-1} + c_2\Delta r_{t-2} + c_3\Delta r_{t-3} \\ &\quad + d_1 S_{t-1} + d_2 S_{t-2} + d_3 S_{t-3} + u_{2t},\end{aligned} \tag{4}$$

where the lag length has been chosen via the Bayesian information criterion performed over the whole sample with a maximum lag length of 13.

From equation (3) it is possible to derive a set of restrictions implied by the EH on the VAR in equation (4). In appendix A we show that, provided the VAR is stable, these restrictions are given by:

$$\begin{bmatrix} a_1 + c_1 \\ b_1 + d_1 \\ a_2 + c_2 \\ b_2 + d_2 \\ a_3 + c_3 \\ b_3 + d_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/\gamma \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{5}$$

which can be compactly written as:

$$H\alpha_{EH} = \mu_{EH_0}, \tag{6}$$

where:

$$\begin{aligned} H &= \begin{bmatrix} I_6 & \underset{6\times 1}{\mathbf{0}} & I_6 & \underset{6\times 1}{\mathbf{0}} \end{bmatrix}, \\ \alpha_{EH} &= \begin{bmatrix} a_1 & b_1 & a_2 & b_2 & a_3 & b_3 & k_1 & c_1 & d_1 & c_2 & d_2 & c_3 & d_3 & k_2 \end{bmatrix}', \\ \mu_{EH_0} &= \begin{bmatrix} 0 & \frac{1}{\gamma} & 0 & 0 & 0 & 0 \end{bmatrix}'. \end{aligned}$$

The subscript EH denotes the fact that the vector of coefficients α satisfies the restrictions implied by the EH.

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

3 The Expectations Hypothesis as a Set of Uncertain Restrictions

By definition any economic theory is a simplification of reality, and as such it cannot hold exactly even if the theory is "true". 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.

3.1 Adding uncertainty

Suppose the EH does hold, but only on average, i.e. some noise causes temporary departures from the EH restrictions in (6). Formally, let the uncertainty introduced by this noise be measured by the parameter σ. The resulting set of stochastic constraints is:

$$\begin{bmatrix} a_1 + c_1 \\ b_1 + d_1 \\ a_2 + c_2 \\ b_2 + d_2 \\ a_3 + c_3 \\ b_3 + d_3 \end{bmatrix} \sim N \left(\mu_{EH_0} = \begin{bmatrix} 0 \\ 1/\gamma \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma_{EH_0} = \begin{bmatrix} \sigma & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma \end{bmatrix} \right), \tag{7}$$

which can be compactly written as:

$$H\alpha_{EH} \sim N\left(\mu_{EH_0}, \Sigma_{EH_0}\right). \tag{8}$$

It is straightforward to interpret the parameter σ as the tightness of the restrictions: a large value of σ implies that the EH restrictions hold with a lot of uncertainty, while as σ decreases the EH restrictions become more binding and eventually become certain. Therefore it is crucial to calibrate σ in an appropriate way. Of course, if we allow for very little variation (or no variation) this essentially implies ruling out any noise and imposing the EH restrictions to hold exactly, while allowing for very large deviations from the restrictions would lead to an insignificant version of the EH. Indeed, any theory is likely to be supported by the data if we allow its restrictions to hold with a sufficiently large amount of noise. We will show in subsection 3.3 that our estimate for the parameter σ provides a sensible set of uncertain restrictions, i.e. the implied degree of uncertainty is sufficiently high to avoid imposing the theory to hold exactly, and sufficiently low to be effectively binding.

The diagonal, homoskedastic structure of the matrix Σ_{EH_0} may not seem very general. Regarding diagonality, there is no a priori economic rationale to think that the restrictions may be correlated, and the independence between restrictions minimizes the distance from the certainty case. Some rationale could be used instead to question the homoskedasticity of the restrictions. For example, it could be argued that restrictions which are linked to variables further away in time could bear more uncertainty than those which are nearer in time. We have tried alternative specifications consistent with this latter argument, each specifying different proportions among the variances of each single restriction and the overall statistical results were very robust to all such modifications.

3.2 The EH in a Bayesian framework

The set of restriction (8) can be thought of in a Bayesian perspective as a prior on the coefficients of the VAR in equation (4). Therefore, we can test the EH by using the approach developed by Jeffreys (1935, 1961). In this approach, statistical models are introduced to represent the probability of the data according to several competing theories, and Bayes's theorem is used to compute the posterior probability that a theory is correct. Then the theories can be compared using the Bayes factor, which is a summary of the evidence provided by the data in favour of one theory as opposed to another.

In particular, we shall compare two competing theories: the first theory does not impose any restrictions on the coefficients of the VAR in equation (4), while the second theory imposes the restrictions derived from the EH. Rewrite the VAR in equation (4) in the following way:

$$y = \Xi\alpha + \varepsilon, \tag{9}$$

with:

$$\begin{aligned}
y &= \begin{bmatrix} \Delta r_t & S_t \end{bmatrix}', \\
\Xi &= \left[I_2 \otimes X \right], \\
X &= \begin{bmatrix} \Delta r_{t-1} & S_{t-1} & \Delta r_{t-2} & S_{t-2} & \Delta r_{t-3} & S_{t-3} & 1 \end{bmatrix}, \\
\alpha &= \begin{bmatrix} a_1 & b_1 & a_2 & b_2 & a_3 & b_3 & k_1 & c_1 & d_1 & c_2 & d_2 & c_3 & d_3 & k_2 \end{bmatrix}', \\
\varepsilon &= \begin{bmatrix} u_{1t} & u_{2t} \end{bmatrix}' \sim N\left(0,\ \Omega\right),\ \Omega = \Sigma_u\ \otimes I_T,
\end{aligned}$$

where T is the sample size and where y, ε and α are $2T \times 1$, $2T \times 1$, and 14×1 vectors, and X, Ξ, Σ_u are $T \times 7$, $2T \times 14$, and 2×2 matrices.

The first theory does not impose any restriction on the coefficients and so it is easily shaped into a loose prior:

$$\alpha \sim N(\alpha_0 = \underset{14\times 1}{\mathbf{0}}, \Sigma_{\alpha_0} = \delta I_{14}). \tag{10}$$

We will refer to the VAR with this loose prior as the unrestricted VAR (UVAR). For a sufficiently large δ the prior does not add any information to that of the likelihood, and the posterior mean of α is numerically identical to the OLS estimate. With our data, a value of $\delta = 10$ is large enough to ensure that this is the case, and it is innocuous in terms of our results, as documented in section 4.3.1.

The second theory imposes the restriction scheme (8), implied by the EH:

$$H\alpha_{EH} \sim N\left(\mu_{EH_0}, \Sigma_{EH_0}\right). \tag{11}$$

We will refer to (11) as the EH prior and to the system consisting of (9) and (11) as the EH-restricted VAR (RVAR):

$$\begin{aligned} y &= \Xi\alpha_{EH} + \varepsilon, \\ H\alpha_{EH} &\sim N\left(\mu_{EH_0}, \Sigma_{EH_0}\right). \end{aligned} \tag{12}$$

In Appendix B we derive the following alternative representation of the EH prior, expressed in terms of the vector of coefficients of the VAR rather than in terms of the vector of restrictions:

$$\alpha_{EH} \sim N\left(\alpha_{EH0}, \Sigma_{\alpha_{EH0}}\right), \tag{13}$$

where

$$\alpha_{EH0} = \left[\begin{array}{ccc} \underset{1\times 8}{\mathbf{0}} & 1/\gamma & \underset{1\times 5}{\mathbf{0}} \end{array}\right]', \tag{14}$$

and

$$\Sigma_{\alpha_{EH0}} = \left[\begin{array}{cccc} \delta I_6 & \underset{6\times 1}{\mathbf{0}} & -\delta I_6 & \underset{6\times 1}{\mathbf{0}} \\ \underset{1\times 6}{\mathbf{0}} & \delta & \underset{1\times 6}{\mathbf{0}} & 0 \\ -\delta I_6 & \underset{6\times 1}{\mathbf{0}} & (\sigma+\delta) I_6 & \underset{6\times 1}{\mathbf{0}} \\ \underset{1\times 6}{\mathbf{0}} & 0 & \underset{1\times 6}{\mathbf{0}} & \delta \end{array}\right]. \tag{15}$$

The parameter δ is the prior variance of the coefficients. Provided that δ is sufficiently large this specification is perfectly equivalent to (11). Again, with our data, a value of $\delta = 10$ is large enough to ensure this is the case. Equations (9) and (13) lead to the following alternative representation of the RVAR in equation (12):

$$\begin{aligned} y &= \Xi \alpha_{EH} + \varepsilon, \\ \alpha_{EH} &\sim N\left(\alpha_{_{EH0}}, \Sigma_{\alpha_{EH0}}\right). \end{aligned} \tag{16}$$

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):

$$Corr(\alpha_{EH0}) = \begin{bmatrix} I_6 & \underset{6\times1}{\mathbf{0}} & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} I_6 & \underset{6\times1}{\mathbf{0}} \\ \underset{1\times6}{\mathbf{0}} & 1 & \underset{1\times6}{\mathbf{0}} & 0 \\ \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} I_6 & \underset{6\times1}{\mathbf{0}} & I_6 & \underset{6\times1}{\mathbf{0}} \\ \underset{1\times6}{\mathbf{0}} & 0 & \underset{1\times6}{\mathbf{0}} & 1 \end{bmatrix}. \tag{17}$$

Notice that depending on the value of the tightness parameter σ we move from the exact restrictions case ($\sigma = 0$) to the unrestricted VAR (as $\sigma \to \infty$). If $\sigma = 0$, the EH is in the traditional form and involves perfect negative correlation between six couples of coefficients of the VAR. Letting $\sigma > 0$ we allow this correlation to be imperfect. As $\sigma \to \infty$, the correlation across the relevant couples of coefficients goes to zero and the correlation matrix approaches that of the loose prior (the identity matrix), thus the two priors become virtually identical.

3.3 Estimation

Both the UVAR and the RVAR are linear regression models subject to a set of stochastic linear restrictions on the regression coefficients. To estimate such models, Theil (1971) proposed the method of mixed estimation, which involves using the uncertain restrictions to supplement data. The added restrictions act as prior information on the coefficients and GLS is numerically equivalent to Bayesian estimation. Derivations of the posterior and marginal likelihood are contained in Appendix C. The parameter σ is estimated to be 0.13 by maximizing the marginal likelihood of the model (see Figure 1).

Before moving on, it is crucial to check that the parametrization $\sigma = 0.13,\ \delta = 10$ is a sensible one. This is ensured by three features. First, neither the loose prior nor the EH prior should impose any restriction on the individual coefficients. Second, only the EH prior should impose restrictions on linear combinations of the coefficients. Indeed the EH does not say anything about individual coefficients, it does say something about some linear combinations of them. Third, the EH restrictions should be uncertain but effective, namely they should be binding.

The RVAR has all these features, as shown in Figures 2 and 3. Figure 2 plots prior and posterior distributions of the coefficients for the RVAR and the UVAR. Notice that both priors are very loose with respect to the individual coefficients. However, Figure 2 hides the key feature of the EH prior, namely to put restrictions on some linear combinations of the coefficients. To see this feature Figure 3 shows the prior and posterior distributions of the restrictions imposed by the EH: it is clear that while the loose prior does not bind, the EH prior is much tighter. Figure 3 also shows that the EH prior is effectively binding since the posterior estimates are shrunk toward the prior mean.

Turning to the correlation matrix of the coefficients, the parametrization $\sigma = 0.13$, $\delta = 10$ implies a correlation between the relevant pairs of coefficients of -0.99, which is very close to the value of -1 implied by the validity of the exact restrictions and very far from the value of 0 implied by the UVAR. Thus, the correlation decreases with respect to the exact restrictions case, but still remains very high.

4 Statistical Evidence

In this section we provide evidence clearly supporting the EH. We do so by computing Bayes factors of the RVAR versus the UVAR as a function of the EH prior tightness σ. For derivations of all the formulas used in this section see Appendix C.

4.1 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

$$\frac{pr(H_2|\mathbf{D})}{pr(H_1|\mathbf{D})} = \frac{pr(\mathbf{D}|H_2)}{pr(\mathbf{D}|H_1)} \frac{pr(H_2)}{pr(H_1)}, \tag{18}$$

so that the transformation is simply multiplication by

$$B_{21} = \frac{pr(\mathbf{D}|H_2)}{pr(\mathbf{D}|H_1)}, \tag{19}$$

which is the Bayes factor of theory H_2 as opposed to theory H_1.

Kass and Raftery (1995) suggested the following interpretation for the value of B_{21} and twice its natural logarithm $2\ln B_{21}$:

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

We speak here in terms of B_{21}, because weighting evidence against a null hypothesis is more familiar, but 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.

From (19) it is clear that we need $pr(\mathbf{D}|H_2)$ and $pr(\mathbf{D}|H_1)$ in order to compute the Bayes factor. As shown in Appendix C, when H_i is a gaussian VAR with fixed variance, $pr(\mathbf{D}|H_i)$ is given by:

$$pr(\mathbf{D}|H_i) = (2\pi)^{-MT/2} |\Omega|^{-1/2} \frac{|\Sigma_{\alpha_{i0}}|^{-1/2}}{|\Sigma_{\bar{\alpha}_i}|^{-1/2}} \exp\left\{-Q_i/2\right\}, \tag{20}$$

for $i = 1, 2$, where

$$Q_i = y'\Omega^{-1}y - \bar{\alpha}_i'\Sigma_{\bar{\alpha}i}^{-1}\bar{\alpha}_i + \alpha_{0i}'\Sigma_{\alpha_{0i}}^{-1}\alpha_{0i}, \tag{21}$$

and where α_{i0}, $\bar{\alpha}_i$ and $\Sigma_{\alpha_{i0}}$, $\Sigma_{\bar{\alpha}_i}$ are prior and posterior means and variances of the vector of coefficients, and Ω is the variance-covariance matrix of the residuals.

In our case, theory H_1 is the UVAR, with $\alpha_{10} = \alpha_0$, $\bar{\alpha}_1 = \bar{\alpha}$, $\Sigma_{\alpha_{10}} = \Sigma_{\alpha_0}$, $\Sigma_{\bar{\alpha}_1} = \Sigma_{\bar{\alpha}}$, while theory H_2 is the RVAR, with $\alpha_{20} = \alpha_{EH0}$, $\bar{\alpha}_2 = \bar{\alpha}_{EH}$, $\Sigma_{\alpha_{20}} = \Sigma_{\alpha_{EH0}}$, $\Sigma_{\bar{\alpha}_2} = \Sigma_{\bar{\alpha}_{EH}}$. Thus the Bayes factor for the RVAR versus the UVAR is:

$$B_{21} = \left[\frac{\frac{|\Sigma_{\alpha_{EH0}}|}{|\Sigma_{\bar{\alpha}_{EH}}|}}{\frac{|\Sigma_{\alpha_0}|}{|\Sigma_{\bar{\alpha}}|}}\right]^{-1/2} \exp\left\{\frac{Q_{UVAR} - Q_{RVAR}}{2}\right\}. \tag{22}$$

4.2 Results

Figure 4 plots the Bayes factor (B_{21}) and twice its natural logarithm ($2\ln B_{21}$) as a function of the EH prior tightness σ together with their inconclusive regions. The inconclusive region for B_{21} ranges from $\frac{1}{3}$ to 3: in this region neither $B_{21} > 3$ nor $B_{12} = B_{21}^{-1} > 3 \Rightarrow B_{21} < \frac{1}{3}$. The inconclusive region for $2\ln B_{21}$ ranges from -2 to 2: in this region neither $2\ln B_{21} > 2$ nor $2\ln B_{12} = -2\ln B_{21} > 2 \Rightarrow 2\ln B_{21} < -2$. In these regions the evidence in favour of theory H_2 as opposed to theory H_1 and of H_1 as opposed to theory H_2 are not worth more than a bare mention.

If we allow for very little noise, letting $\sigma \to 0$, the Bayes factor supports the UVAR ($B_{21} = 4.96e - 007$ and $2\ln B_{21} = -29.031$). This is the common result of rejection. Indeed, letting the tightness go to zero amounts to imposing the EH without noise. Therefore our general framework nests the traditional one as a special case, and is also consistent with the empirical findings rejecting the EH.

On the other hand, allowing for very large departures from the EH restrictions, letting $\sigma \to \delta$, leads the Bayes factor to the inconclusive region. Intuitively, the noise on the constraints becomes too large and data cannot distinguish between the restricted and the unrestricted VAR. Indeed, if we allow for too large departures from the EH, the RVAR becomes virtually equivalent to the UVAR, so the Bayes factor B_{21} converges to 1 (i.e. the two models end up having the same marginal likelihood) and twice its natural logarithm $2\ln B_{21}$ converges to 0.

For intermediate values of σ the Bayes factor strongly supports the RVAR. Importantly, at the estimated EH prior tightness $\sigma =$ 0.13 the value of $2\ln B_{21}$ reaches a value close to 20 denoting very strong evidence in favour of the EH[1].

This result formally confirms the informal statement of Carriero, Favero and Kaminska (2005) that once uncertainty is added to the picture the EH cannot be rejected. Moreover, this result shows that the amount of uncertainty needed is considerably smaller than that implied by their simulation experiment performed with an unrestricted VAR.

Finally, this evidence explains the anomaly found by Campbell and Shiller (1987), that the EH is statistically rejected but the theoretical spread based on its validity is closely correlated with the actual spread. As stressed above, the theory is strongly supported by the data, which explains the high correlation, but also the EH is perturbed by some noise which leads the Wald test to reject the exact restrictions.

[1]The fact that the Bayes factor is maximized by the same value of σ maximizing the marginal likelihood of the RVAR is obvious, as long as the competing prior (the UVAR) does not depend on that parameter.

4.3 Robustness

To evaluate the robustness of our results, we next extend the testing framework in two dimensions, providing statistical results recursively through time, and adding macroeconomic information to the picture. Before doing so, we study the effect of alternative parametrizations of δ on our results. Finally we also look at the likelihood ratios which is unusual in the adopted Bayesian framework, but still useful to confirm the results derived so far.

4.3.1 Alternative parametrizations

As mentioned in section 3, a value of $\delta = 10$ is sufficiently large to ensure that the prior does not add any information to that of the likelihood, and the posterior mean of α is numerically identical to the OLS estimate. Still, we may be interested in what happens if we increase the value δ above 10 (i.e. increasing the looseness of the loose prior).

Table 2 and Figure 5 display Bayes factors computed at different values of σ and δ. An increase in δ rescales our results, but does not change their qualitative pattern.

Table 2: Bayes factors ($2 \ln B_{21}$) for different σ, δ.

	$\delta = 10$	$\delta = 100$	$\delta = 200$	$\delta = 500$
$\sigma \to 0$	-29.031	-15.408	-11.268	-5.794
$\sigma = 0.13$	19.619	33.339	37.493	42.988
$\sigma = \delta$	-0.002	$-3.19e-004$	$-1.60e-004$	$-6.45e-006$

As δ increases, the magnitude of the Bayes factor at the peak (i.e. when $\sigma = 0.13$) increases, while its value at the extremes remains the same: as $\sigma \to 0$ the value of $2 \ln B_{21}$ goes to $-\infty$ (i.e. B_{21}converges to 0), while as $\sigma = \delta$ it goes to 0 (i.e. B_{21}converges to 1). Thus results are independent from the particular choice of δ we make.

4.3.2 Macroeconomic information

Whereas the EH implies that monetary policy affects long-term rates by influencing expectations about future short-term rates, central banks also look at bond markets to extract information about inflation expectations. Therefore, policy is likely to respond to bond market conditions which may introduce an obvious misspecification to our framework: the omission of macroeconomic variables to which the monetary authority reacts.

Fuhrer (1996) uses a simple Taylor-rule type reaction function, the EH and reduced-form equations for output and inflation, to solve for the reaction function coefficients that deliver long-term rates consistent with the EH. He finds that modest and smoothly evolving time-variation in parameters of the reaction function is sufficient to reconcile the expectations model with the long-bond data. Favero (2005) extends Fuhrer's framework to derive standard errors for long-term rates consistent with the EH. Roush (2001) argues that previous work on the EH has failed to sufficiently account for interactions between monetary policy and bond markets in the determination of long- and short-term interest rates and using a VAR model with macro and financial variables finds strong evidence supporting a term structure channel for policy that is consistent with the EH. Ang and Piazzesi (2003) find that macro factors explain a significant amount of the variation in bond yields: in particular they explain most of the forecast variance of short-term rates at long forecast horizons, and of long-term rates at short forecast horizons.

Thus we extend our framework to include macroeconomic information. In particular, we add the CPI inflation rate and the unemployment rate to both the restricted and the unrestricted VAR. Both series are provided by the Federal Reserve of St.Louis and are not subject to revision, so they can be used to produce forecasts in real time, as we do in Section 5. Of course, the inclusion of new variables requires some additional restrictions to hold under the EH. Appendix D shows how we derived the additional restrictions implied by the EH on this augmented VAR.

Performing the same analysis in this extended framework does not change the results (see Figure 6). The estimated tightness decreases to a value of $\sigma = 0.11$, and at this point the value of $2 \ln B_{21}$ is above 20, even higher than in the previous case.

4.3.3 Time

It is worth checking whether our results are stable through time. Figure 7 panel A plots Bayes factors computed recursively from 1984:1 until 2004:8. Figure 7 Panel B does the same with the RVAR augmented with macroeconomic information. As is clear, $2 \ln B_{21}$ is always above 10, providing strong evidence in favour of the EH.

4.3.4 Likelihood ratios

Figure 8 plots twice the natural logarithm of the likelihood ratio of the RVAR versus the UVAR as a function of the EH prior tightness. Results confirm those obtained with the Bayes factor: low values of σ imply a significant loss in the likelihood, and this explains the common result of rejection in the literature. On the other hand, for higher values of σ the loss in the likelihood becomes lower and eventually zero, providing evidence in favor of the EH.

5 Explaining Long Term Rate Dynamics

The previous section provided clear statistical evidence in favour of the EH. This section complements this result with economic evidence. In particular, we will show that the behaviour of the U.S. 10-year interest rate from the 1970s onwards has been entirely consistent with the statement of the EH. To demonstrate this we first construct a theoretical, EH-consistent long-term rate and then contrast it with the actual, realized long-term rate.

5.1 The EH-consistent long term rate

Recall that under the EH the long term rate $R^*_{t,T}$ is given by

$$R^*_{t,T} = \frac{1-\gamma}{1-\gamma^T} \sum_{i=0}^{T-1} \gamma^i E_t r_{t+i} + TP_T, \tag{23}$$

where the star denotes we are under the null of the EH. In our framework the expectational term $E_t r_{t+i}$ can be obtained by the linear projection of the estimated VAR. This avoids both problems related to the common strategy when testing the EH: we do not use ex-post data to proxy for expectations, and we avoid the simultaneity problems inherent to the single equation approach. This approach has been developed by Campbell and Shiller (1987).

However, Campbell and Shiller (1987) use information from the whole sample to simulate investors' expectations while investors can use only historically available information to generate predictions of short-term rates. Indeed, as long as expectations are formed given the information at time t, the estimation window should not contain data beyond that date.

Therefore, we compute $R^*_{t,T}$ using a recursive estimation / projection scheme, such that at each point in time only the available information is used to first estimate the VAR, and then to project it forward. In particular, our procedure works as follows. i) The first estimation is performed over the sample 1966:1 1970:12. All the subsequent estimations are performed over the sample 1966:1 1970:12+i where i is the number of iterations already executed. ii) Using the posterior of the coefficients obtained at point i) the VAR is projected forward and posterior of the variables $E_t r_{t+i}$ and $R^*_{t,T}$ are obtained. iii) Then we move forward one period, adding one data point to the estimation window, and go back to point i). This recursive estimation / projection scheme provides time series of the posterior distributions of the variables $E_t r_{t+i}$ and $R^*_{t,T}$. As long as we are under the null of the EH, these variables would also yield the posterior distribution of the term premium as $TP_T = R^*_{t,T} - (1-\gamma)\sum \gamma^i E_t r_{t+i}$.

5.2 An economic test of the EH

Our recursive estimation/projection scheme provides a simple but effective test of the Expectations Hypothesis. Indeed, a test for the pure EH (i.e. with $TP_T = 0$) is immediately performed simply by checking whether the actual long-term rate could be a plausible draw from the posterior distribution of the EH-consistent long-term rate, i.e. if $R_{t,T}$ lies within some credible bounds of the posterior distribution of $R^*_{t,T}$. This procedure is very similar to that of Carriero, Favero and Kaminska (2005), but with the subtle difference that here the uncertainty does not arise from estimation, but is modeled within the theory.

The implied distribution of $R^*_{t,T}$ is plotted in Figure 9. Interestingly, at the beginning of the sample there are several instances in which $R^*_{t,T}$ has an asymmetric distribution. This comes from the fact that in those periods the draws of the coefficients are such that the VAR is nearly unstable. The forecasts explode and as a result the mean goes very far from the median. Notwithstanding this initial instability, the actual long-term rate $(R_{t,T})$ almost[2] always lies within the 2.5% and the 97.5% percentiles of the posterior distribution of the EH-consistent long-term rate $\left(R^*_{t,T}\right)$. The EH-consistent and the actual long-term rate are highly correlated, but of course they do not perfectly coincide. Again, this explains both the common result of rejection and the Campbell and Shiller (1987) anomaly. The traditional framework to test the EH proxies ex-ante expected rates with the ex-post realized rates and therefore it understates the amount of uncertainty that individuals face when forecasting future short-term rates up to 120-month ahead, which is huge, as measured by the width of the bounds around the EH consistent rate.

Therefore, once we take into account the uncertainty involved in predicting short-term rates, the EH cannot be rejected. This result would also hold if the actual and the EH-consistent long-term rates were less clearly correlated: Figure 9 shows that the gap between the 2.5% and the 97.5% percentiles is considerably wider than the difference between the actual and the EH-consistent long-term rates. Even if the actual long-term rate had behaved much more differently from EH-consistent one, it still could be consistent with the theory.

Thus the dynamics of the 10-year rate in the last 35 years is entirely consistent with the EH, and this adds an economic validation to our statistical evidence.

[2] Except for some instances all occurring during the reserves targeting era between the end of the 1970s and the beginning of the 1980s

6 Does the Expectations Hypothesis Help in Forecasting Short Term Interest Rates?

Once the EH has been validated on statistical and economic grounds, it is natural to ask whether it can be used for policy. An obvious use of the theory is forecasting: as the EH relates future short-term and actual long-term interest rates, it should provide additional useful information about future short-term rates by extracting it from the actual long-term rate. In this section we study whether the EH prior produces significant improvements in forecast accuracy. As the Bayes factor can be interpreted as a summary of forecasting performance, we expect that imposing the EH restrictions would yield advantages in terms of forecasting.

6.1 Forecast gains of the EH prior

We compare the forecasting performance of the RVAR versus the UVAR, when both include macroeconomic information. There are many criteria available to compare predictive accuracy, for simplicity we choose absolute forecast error (AFE) and squared forecast error (SFE). A first assessment about predictive accuracy can be done by inspecting the following regression equations:

$$AFE_{t,h}^{UVAR} - AFE_{t,h}^{EH} = \alpha + u_{t,h}, \tag{24}$$
$$SFE_{t,h}^{UVAR} - SFE_{t,h}^{EH} = \beta + v_{t,h}, \tag{25}$$

where h indexes the forecast horizon. Figures 10 and 11 plot the rolling estimates of the coefficients α and β together with their HAC standard error bounds. The performance of the UVAR is significantly worse than that of the EH-restricted VAR up to 6 months-ahead for the AFE loss function and from 3 up to 6 months-ahead for the SFE loss function. In order to interpret this as a valid test of equal predictive accuracy the error terms should be normally distributed, but this is not the case in the data.[3] This issue cannot be solved using asymptotic tests of equal predictive accuracy[4] as the two models are nested and so forecast errors would coincide asymptotically under the null.

This problem is solved by implementing the test recently developed by Giacomini and White (2004). This new test (GW) is based on conditional expectations of forecasts rather than on unconditional expectations and can handle the comparison of both nested and non-nested models and forecasts obtained by nonparametric and Bayesian

[3] The null of normality is rejected in almost all instances.

[4] For example, that proposed by Diebold and Mariano (1995).

estimation. Results of rolling GW test for both SFE and AFE are plotted in Figure 12. The performance of the UVAR is significantly different to that of the EH-restricted VAR at all horizons from 1 up to 6 months-ahead for both the loss functions. The sign of the differences in forecast accuracy can be deduced by Figures 10 and 11 and is positive, implying a better forecasting performance of the EH-restricted VAR.

6.2 Adding the EH to a Taylor rule: how to extract additional information from the term structure

The UVAR augmented with macroeconomic variables can be interpreted as the reduced form of a model featuring an interest rate rule, for example a Taylor rule.

A large, growing body of empirical literature has established interest rate rules as a convenient way to model and interpret monetary policy (Taylor, 1993, Clarida, Gali and Gertler, 1998, 1999, 2000). When allowing for persistence in the short rate interest rate rules responding to inflation and the output gap tend to track the data well (Rudebush 2002, Söderlind, Söderström and Vredin, 2005). They are also capable of explaining the high inflation in the seventies in terms of an accommodating behaviour towards inflation in the pre-Volcker era. Thus an interest rate-rule model is the natural benchmark to evaluate the accuracy of short-term rate forecasts.

Our results provide clear evidence that the EH-restricted VAR outperforms the UVAR in forecasting short-term rates. This result can be interpreted as evidence that, if we add to an interest rate rule based VAR also the EH restrictions, then additional information about future short-term rates contained into the long-term rate can be extracted and exploited to increase significantly the accuracy of the forecasts.

Notice that the pattern of results in Figures 10-12 shows a significant increase in the accuracy of the EH prior after the end of the reserves-targeting period. This provides evidence that while in the Volcker era interest rates were too volatile for the EH to be useful in forecasting, in the Greenspan era the stabilization of interest rates makes the EH useful in extracting information from long-term rates.

To conclude, as the EH relates expected future short-term and actual long-term rates, it provides additional useful information about future short-term rates by extracting it from the actual long-term rate. Of course, this additional information can be extracted only if the EH does hold, which is what we found in the data.

7 Conclusions

By definition any economic theory is a simplification of reality, and as such it cannot hold exactly even if the theory is “true”. Indeed, in the real world there is always some kind of noise which blurs any equilibrium.

This study develops an extended version of the EH in which transitory deviations from the theory may occur. To model these deviations we derive a set of restrictions on a VAR and then we let them hold with some degree of uncertainty. This amounts to deriving from the EH a prior for the coefficients of the VAR. Then, we can contrast this prior with an unrestricted, loose prior representing a world in which the EH does not hold. To make the comparison, we use the Bayes factor, which is a summary of the evidence provided by the data in favour of one theory as opposed to another.

To ensure the robustness of our results, we extend the testing framework towards two dimensions, providing statistical results recursively through time, and including macroeconomic information into the picture. Indeed, the EH implies that monetary policy affects long-term rates by influencing expectations on future short-term rates, but also central banks look at bond markets to get informed about inflation expectations and so incidental policy reactions to bond market conditions are likely to occur.

Beyond statistical evidence, we also provide economic evidence. In particular, we perform an economic test by checking the consistency of the observed long-term rate with the EH, and we study whether the EH may lead to some improvements in forecasting short-term rates.

Our results show that the EH is strongly supported by the data and is entirely consistent with the dynamics of the US 10-year rate from the seventies onwards. Moreover, the proposed framework is able to explain the very common result of rejection of the EH, to solve the anomaly found out by Campbell and Shiller (1987) and to significantly improve forecast accuracy of a Taylor-rule based model in predicting future short-term rates.

8 Appendices

A. Derivation of the EH restrictions

Stack the VAR as:

$$\begin{bmatrix} \Delta r_t \\ \Delta r_{t-1} \\ \Delta r_{t-2}. \\ S_t \\ S_{t-1} \\ S_{t-2} \end{bmatrix} = \begin{bmatrix} k_1 \\ 0 \\ 0 \\ k_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 & a_3 & b_1 & b_2 & b_3 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ c_1 & c_2 & c_3 & d_1 & d_2 & d_3 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta r_{t-1} \\ \Delta r_{t-2} \\ \Delta r_{t-3}. \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ 0 \\ 0 \\ u_{2t} \\ 0 \\ 0 \end{bmatrix},$$

or, more succinctly:

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

The trace statistic for the null of no cointegration (with the intercept both in cointegrating equation and in the VAR) is well above the critical value (207.811 while the 1% critical value is 20.04).

Recalling (3), the EH would put on the VAR the following set of nonlinear restrictions:

$$\begin{aligned} g'z_t &= \sum_{i=1}^{\infty} \gamma^i h' \left(\sum_{n=0}^{i-1} A^n C + A^i z_t \right) + TP_\infty, \\ &= h' \left[\sum_{i=1}^{\infty} \gamma^i \sum_{n=0}^{i-1} A^n C \right] + \sum_{i=1}^{\infty} \gamma^i h' A^i z_t + TP_\infty, \\ &= h' \left[\sum_{i=1}^{\infty} \gamma^i (I - A^i) \right] (I - A)^{-1} C + \sum_{i=1}^{\infty} \gamma^i h' A^i z_t + TP_\infty, \end{aligned}$$

where g' and h' are selector vectors with 6 elements, all of which are zero except for the 4th element of g' and the first element of h' which are unity. Since the above expression has to hold in general, it holds also for $TP_\infty = -h' \left[\sum_{i=1}^{\infty} \gamma^i (I - A^i) \right] (I - A)^{-1} C$ and[5] for any z_t :

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

[5] Notice that by adding the restriction $TP_t = 0$ we could test the PURE EH.

As this VAR is stable we can exploit the properties of geometric series to write:

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

Postmultiplying by $(I - \gamma A)$ provides the following set of linear restrictions:

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

i.e.:

$$g' \begin{bmatrix} 1-\gamma a_1 & -\gamma a_2 & -\gamma a_3 & -\gamma b_1 & -\gamma b_2 & -\gamma b_3 \\ -\gamma & 1 & 0 & 0 & 0 & 0 \\ 0 & -\gamma & 1 & 0 & 0 & 0 \\ -\gamma c_1 & -\gamma c_2 & -\gamma c_3 & 1-\gamma d_1 & -\gamma d_2 & -\gamma d_3 \\ 0 & 0 & 0 & -\gamma & 1 & 0 \\ 0 & 0 & 0 & 0 & -\gamma & 1 \end{bmatrix} = h' \begin{bmatrix} \gamma a_1 & \gamma a_2 & \gamma a_3 & \gamma b_1 & \gamma b_2 & \gamma b_3 \\ \gamma & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 & 0 & 0 \\ \gamma c_1 & \gamma c_2 & \gamma c_3 & \gamma d_1 & \gamma d_2 & \gamma d_3 \\ 0 & 0 & 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma & 0 \end{bmatrix}.$$

As g' and h' select respectively the 4th and 1st row we obtain:

$$\begin{bmatrix} -\gamma c_1 & -\gamma c_2 & -\gamma c_3 & 1-\gamma d_1 & -\gamma d_2 & -\gamma d_3 \end{bmatrix} = \begin{bmatrix} \gamma a_1 & \gamma a_2 & \gamma a_3 & \gamma b_1 & \gamma b_2 & \gamma b_3 \end{bmatrix}.$$

Thus the EH imposes the following constraints on the individual coefficients of the VAR:

$$\begin{bmatrix} a_1 + c_1 \\ b_1 + d_1 \\ a_2 + c_2 \\ b_2 + d_2 \\ a_3 + c_3 \\ b_3 + d_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/\gamma \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

B. Representation of the EH prior in terms of the vector of coefficients rather than in terms of the restrictions

Write the VAR as:

$$\begin{aligned}
y &= \Xi\alpha + \varepsilon, \\
\overbrace{\underset{MT\times 1}{\begin{bmatrix} \Delta r_t \\ S_t \end{bmatrix}}}^{y} &= \overbrace{\underset{MT\times M(pM+1)}{\begin{bmatrix} I_M \otimes \underset{T\times(pM+1)}{X} \end{bmatrix}}}^{\Xi} * \underset{M(pM+1)\times 1}{\alpha} + \overbrace{\underset{MT\times 1}{\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}}}^{\varepsilon}, \\
X &= \begin{bmatrix} \Delta r_{t-1} & S_{t-1} & \Delta r_{t-2} & S_{t-2} & \Delta r_{t-3} & S_{t-3} & 1 \end{bmatrix}, \\
\alpha &= \begin{bmatrix} a_1 & b_1 & a_2 & b_2 & a_3 & b_3 & k_1 & c_1 & d_1 & c_2 & d_2 & c_3 & d_3 & k_2 \end{bmatrix}', \\
\varepsilon &\sim N\left(0, \Omega = \underset{MT\times MT}{\Sigma_u \otimes I_T}\right),
\end{aligned}$$

where $M = 2$ is the number of equations, $p = 3$ is the number of lags included, and T is the sample size.

The generic form of a normal prior with fixed variance for the vector of coefficient α would be:

$$\alpha \sim N(\alpha_0, \Sigma_{\alpha 0}),$$

The unrestricted VAR corresponds the following loose prior:

$$\alpha \sim N(\alpha_0 = \underset{14\times 1}{\mathbf{0}}, \Sigma_{\alpha 0} = \delta I_{14}),$$

and for $\delta = 10$ the posterior mean of α is identical to the OLS estimator.

Now consider the set of restrictions implied on the unrestricted VAR by the EH:

$$\begin{bmatrix} a_1 + c_1 \\ b_1 + d_1 \\ a_2 + c_2 \\ b_2 + d_2 \\ a_3 + c_3 \\ b_3 + d_3 \end{bmatrix} \sim N\left(\mu_{EH_0} = \begin{bmatrix} 0 \\ 1/\gamma \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma_{EH_0} = \begin{bmatrix} \sigma & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma \end{bmatrix} \right).$$

Denoting with α_{EH} the vector of coefficients of the VAR when it satisfies the EH-restrictions we can write:

$$H\alpha_{EH} \sim N\left(\mu_{EH_0}, \Sigma_{EH_0}\right), \tag{26}$$

where

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

The RVAR consists of the VAR plus the EH-restrictions:

$$\begin{cases} y = \Xi\alpha_{EH} + \varepsilon \\ H\alpha_{EH} \sim N\left(\mu_{EH_0}, \Sigma_{EH_0}\right) \end{cases}.$$

There is an alternative way to write the EH restrictions. The generic form of a normal prior satisfying the EH restrictions would be:

$$\alpha_{EH} \sim N(\alpha_{EH0}, \Sigma_{\alpha_{EH0}}),$$

which implies:

$$H\alpha_{EH} \sim (H\alpha_{EH0}, H\Sigma_{\alpha_{EH0}}H'). \tag{27}$$

Under the EH both (26) and (27) must hold, so there is the following relation between the prior moments of the vector of restrictions and those of the vector of coefficients:

$$\begin{aligned} \Sigma_{EH_0} &= H\Sigma_{\alpha_{EH0}}H', \\ \mu_{EH_0} &= H\alpha_{EH0}. \end{aligned}$$

The above system has no unique solution since there are 14 coefficients and 6 restrictions, 8 coefficients are not restricted and H is not invertible. To solve this problem simply set a loose normal prior with mean 0 and variance δ (the same of the UVAR coefficients) on the unrestricted coefficients. This provides an invertible H without affecting the EH restrictions and the analysis.

The restrictions become:

$$H_2\alpha_{EH} \sim N\left(\mu_{2EH_0}, \Sigma_{2EH_0}\right),$$

with:

$$\mu_{2EH_0} = \left[\begin{array}{cccccccccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\gamma & 0 & 0 & 0 & 0 & 0 \end{array}\right]',$$

$$\Sigma_{2EH_0} = \left[\begin{array}{cccccccccccccc}
\delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta
\end{array}\right],$$

$$H_2 = \left[\begin{array}{cccccccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right].$$

With δ sufficiently big (which is the case in the paper) the form of the EH restrictions is not affected and this specification is equivalent to the preceding one.

Now it is possible to invert the restriction matrix and to get an explicit prior for α_{EH}:

$$\alpha_{EH} \sim N\left(\alpha_{_{EH0}}, \Sigma_{\alpha_{EH0}}\right),$$

where

$$\alpha_{_{EH0}} = H_2^{-1}\mu_{2EH_0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\gamma & 0 & 0 & 0 & 0 & 0 \end{bmatrix}',$$

$$\begin{aligned} \Sigma_{\alpha_{EH0}} &= H_2^{-1}\Sigma_{2EH_0}H_2'^{-1} \\ &= \begin{bmatrix} \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta \end{bmatrix}. \end{aligned}$$

So the RVAR can be written as:

$$\begin{cases} y = \Xi\alpha_{EH} + \varepsilon \\ \alpha_{EH} \sim N\left(\alpha_{_{EH0}}, \Sigma_{\alpha_{EH0}}\right) \end{cases}.$$

The prior correlation matrix of the coefficients is:

$$\left[\begin{array}{cccccccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right].$$

As $\sigma \to \infty$ this matrix converges to the correlation matrix of the UVAR (i.e. the identity matrix), while as $\sigma \to 0$ it converges to the correlation matrix of the VAR under the exact EH restrictions.

To estimate the VAR define:

$$\begin{aligned}
-v_{EHt} &= \alpha_{EH} - \alpha_{EH_0} \sim N\left(0, \Sigma_{\alpha_{EH0}}\right) \Longrightarrow \\
\alpha_{EH_0} &= \alpha_{EH} + v_{EHt},\ v_{EHt} \sim N\left(0, \Sigma_{\alpha_{EH0}}\right)
\end{aligned}$$

then plug in:

$$\begin{aligned}
y_{EH} &= \Xi_{EH}\alpha_{EH} + \varepsilon_{EH}, \\
\overbrace{\underset{(MT+M(pM+1))\times 1}{\begin{bmatrix} \Delta r_t \\ S_t \\ \alpha_{EH_0} \end{bmatrix}}}^{y_{EH}} &= \overbrace{\underset{(MT+M(pM+1))\times M(pM+1)}{\begin{bmatrix} \underset{MT\times M(pM+1)}{\left[I_M \otimes \underset{T\times(pM+1)}{X}\right]} \\ \left[I_{M(pM+1)}\right] \end{bmatrix}}}^{\Xi_{EH}} * \underset{M(pM+1)\times 1}{\alpha_{EH}} + \overbrace{\underset{(MT+M(pM+1))\times 1}{\begin{bmatrix} u_{1t} \\ u_{2t} \\ v_{EHt} \end{bmatrix}}}^{\varepsilon_{EH}}, \\
\varepsilon_{EH} &\sim N\left(0, \Omega_{EH} = \underset{(MT+M(pM+1))\times(MT+M(pM+1))}{\begin{bmatrix} \Omega = \underset{MT\times MT}{\left[\Sigma_u \otimes I_T\right]} & 0 \\ 0 & \Sigma_{\alpha_{EH0}} \end{bmatrix}}\right).
\end{aligned}$$

The same procedure is applied to the competing model. The UVAR is:

$$\begin{cases} y = \Xi\alpha + \varepsilon \\ \alpha \sim N(\alpha_0 = \underset{14\times 1}{\mathbf{0}}, \Sigma_0 = \delta I_{14}) \end{cases}.$$

Define:

$$\begin{aligned} -v_t &= \alpha - \alpha_0 \sim N\left(0, \Sigma_{\alpha_0}\right) \Longrightarrow \\ \alpha_0 &= \alpha + v,\ v \sim N\left(0, \Sigma_{\alpha_0}\right), \end{aligned}$$

then plug in:

$$\begin{aligned} y_{UVAR} &= \Xi_{UVAR}\alpha + \varepsilon_{UVAR}, \\ \overbrace{\underset{(MT+M(pM+1))\times 1}{\begin{bmatrix} \Delta r_t \\ S_t \\ \alpha_0 \end{bmatrix}}}^{y_{UVAR}} &= \overbrace{\underset{(MT+M(pM+1))\times M(pM+1)}{\begin{bmatrix} \underset{MT\times M(pM+1)}{\left[I_M \otimes \underset{T\times(pM+1)}{X} \right]} \\ \left[I_{M(pM+1)} \right] \end{bmatrix}}}^{\Xi_{UVAR}} * \underset{M(pM+1)\times 1}{\alpha} + \overbrace{\underset{(MT+M(pM+1))\times 1}{\begin{bmatrix} u_{1t} \\ u_{2t} \\ v \end{bmatrix}}}^{\varepsilon_{UVAR}}, \\ \varepsilon_{UVAR} &\sim N\left(0, \Omega_{UVAR} = \underset{(MT+M(pM+1))\times(MT+M(pM+1))}{\begin{bmatrix} \Omega = \underset{MT\times MT}{\left[\Sigma_u \otimes I_T \right]} & 0 \\ 0 & \Sigma_{\alpha_0} \end{bmatrix}} \right). \end{aligned}$$

C. Posterior densities, marginal likelihoods, Bayes factor

Here we compute the posterior and the marginal likelihood of the vector of coefficients α. Results apply to both the models at hand ($RVAR$ and $UVAR$).

$$
\begin{aligned}
\overbrace{\underset{MT\times 1}{\begin{bmatrix} \Delta r_t \\ S_t \end{bmatrix}}}^{y} &= \overbrace{\underset{MT\times Mk}{\begin{bmatrix} I_M \otimes \underset{T\times k}{X} \end{bmatrix}}}^{\Xi} * \underset{Mk\times 1}{\alpha} + \overbrace{\underset{MT\times 1}{\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}}}^{\varepsilon}, \\
\alpha &\sim N(\underset{k\times 1}{\alpha_0}, \underset{k\times k}{\Sigma_{\alpha_0}}), \\
e &\sim N(0, \Omega),\ \Omega = \underset{MT\times MT}{[\Sigma_u \otimes I_T]}.
\end{aligned}
$$

where $M = 2$ is the number of equations, $p = 3$ is the number of lags included, $k = pM + 1$ is the number of regressors and T is the sample size. Here α_0 and Σ_{α_0} can be both the UVAR and the RVAR prior moments. In compact notation:

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

The prior density is:

$$p(\alpha) = (2\pi)^{-Mk/2} \left|\Sigma_{\alpha_0}\right|^{-1/2} \exp\left\{-1/2(\alpha - \alpha_0)'\Sigma_{\alpha_0}^{-1}(\alpha - \alpha_0)\right\},$$

the likelihood is[6]:

$$p(y|\alpha) = (2\pi)^{-MT/2} \left|\Omega\right|^{-1/2} \exp\left\{-1/2(y - \Xi\alpha)'\Omega^{-1}(y - \Xi\alpha)\right\},$$

a posterior density kernel is:

$$
\begin{aligned}
p(y|\alpha)p(\alpha) &= (2\pi)^{-M(T+k)/2} \left|\Omega\right|^{-1/2} \left|\Sigma_{\alpha_0}\right|^{-1/2} \\
&\quad \exp\left\{-1/2 \begin{bmatrix} (y - \Xi\alpha)'\Omega^{-1}(y - \Xi\alpha) \\ +(\alpha - \alpha_0)'\Sigma_{\alpha_0}^{-1}(\alpha - \alpha_0) \end{bmatrix}\right\}.
\end{aligned}
$$

[6] notice that: $\left|\Omega\right|^{-1/2} = \left|\Sigma_u \otimes I_T\right|^{-1/2} = (\left|\Sigma_u\right|^T \left|I_T\right|^M)^{-1/2} = \left|\Sigma_u\right|^{-T/2}$.

Now define[7]:

$$\Sigma_{\bar{\alpha}} = \left[\Sigma_{\alpha_0}^{-1} + \Xi'\Omega^{-1}\Xi\right]^{-1},$$
$$\bar{\alpha} = \Sigma_{\bar{\alpha}} * \left[\Sigma_{\alpha_0}^{-1}\alpha_0 + \Xi'\Omega^{-1}y\right].$$

Using the above definitions and completing the square yields:

$$(y - \Xi\alpha)'\Omega^{-1}(y - \Xi\alpha) + (\alpha - \alpha_0)'\Sigma_{\alpha_0}^{-1}(\alpha - \alpha_0)$$

$$\begin{aligned} = \;& y'\Omega^{-1}y - y'\Omega^{-1}\Xi\alpha - \alpha'\Xi'\Omega^{-1}y + \alpha'\Xi'\Omega^{-1}\Xi\alpha \\ & +\alpha'\Sigma_{\alpha_0}^{-1}a - \alpha'\Sigma_{\alpha_0}^{-1}a_0 - \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 \end{aligned}$$

$$\begin{aligned} = \;& y'\Omega^{-1}y - [y'\Omega^{-1}\Xi + \alpha_0{}'\Sigma_{\alpha_0}^{-1}]\alpha - \alpha'[\Xi'\Omega^{-1}y + \Sigma_{\alpha_0}^{-1}a_0] \\ & +\alpha'[\Xi'\Omega^{-1}\Xi + \Sigma_{\alpha_0}^{-1}]\alpha + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 \end{aligned}$$

$$= y'\Omega^{-1}y - [\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha}]'\alpha - \alpha'[\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha}] + \alpha'[\Sigma_{\bar{\alpha}}^{-1}]\alpha + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0$$
$$= y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\alpha - \alpha'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + \alpha'\Sigma_{\bar{\alpha}}^{-1}\alpha + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0.$$

This can be rewritten as[8]:

$$\begin{aligned} & (y - \Xi\alpha)'\Omega^{-1}(y - \Xi\alpha) + (\alpha - \alpha_0)'\Sigma_{\alpha_0}^{-1}(\alpha - \alpha_0) \\ = \;& y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + (\bar{\alpha} - \alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha} - \alpha) + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0, \end{aligned}$$

[7] Notice that:

$$\begin{aligned} \Sigma_{\bar{\alpha}} &= \left[\Sigma_{\alpha_0}^{-1} + \Xi'\Omega^{-1}\Xi\right]^{-1} = \left[\Sigma_{\alpha_0}^{-1} + (I_M \otimes X)'(\Sigma_u \otimes I_T)(I_M \otimes X)\right]^{-1} \\ &= \left[\Sigma_{\alpha_0}^{-1} + I_M'\Sigma_u I_M \otimes X'I_T X\right]^{-1} = \left[\Sigma_{\alpha_0}^{-1} + \Sigma_u \otimes X'X\right]^{-1}, \\ \bar{\alpha} &= \Sigma_{\bar{\alpha}} * \left[\Sigma_{\alpha_0}^{-1}\alpha_0 + \Xi'\Omega^{-1}y\right] = \Sigma_{\bar{\alpha}} * \left[\Sigma_{\alpha_0}^{-1}\alpha_0 + (I_M \otimes X)'(\Sigma_u \otimes I_T)y\right] \\ &= \Sigma_{\bar{\alpha}} * \left[\Sigma_{\alpha_0}^{-1}\alpha_0 + (I_M'\Sigma_u \otimes X'I_T)y\right] = \Sigma_{\bar{\alpha}} * \left[\Sigma_{\alpha_0}^{-1}\alpha_0 + (\Sigma_u \otimes X')y\right]. \end{aligned}$$

[8] Since:

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

so a posterior density kernel can be also written as follows:

$$\begin{aligned}
p(y|\alpha)p(\alpha) &= (2\pi)^{-M(T+k)/2} \left|\Omega\right|^{-1/2} \left|\Sigma_{\alpha_0}\right|^{-1/2} \\
&\quad \exp\left\{ -1/2 \left[\begin{array}{c} (\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha) \\ y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 \end{array} \right] \right\} \\
&= (2\pi)^{-M(T+k)/2} \left|\Omega\right|^{-1/2} \left|\Sigma_{\alpha_0}\right|^{-1/2} \\
&\quad \exp\left\{ -1/2[(\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha) + Q] \right\},
\end{aligned}$$

where:

$$Q = y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0.$$

Forgetting constants:

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

which shows that $\bar{\alpha}, \Sigma_{\bar{\alpha}}$ are the moments of the posterior. The posterior properly normalized density is:

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

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

$$ML = \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} p(y|\alpha)p(\alpha)\ d\alpha_1 ... d\alpha_{Mk} = \int_{\Re^{Mk}} p(y|\alpha)p(\alpha) d\alpha$$

---

we have that:

$$\begin{aligned}
& y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\alpha - \alpha'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + [\alpha'\Sigma_{\bar{\alpha}}^{-1}\alpha] + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 \\
= \ & y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\alpha - \alpha'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + \left[\begin{array}{c} (\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha) + \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + \\ -(\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha) \end{array} \right] + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 \\
= \ & y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\alpha - \alpha'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + (\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha) \\
& + \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} - (\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha) + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 \\
= \ & y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}[\alpha + (\bar{\alpha}-\alpha)] - [(\bar{\alpha}-\alpha)' + \alpha']\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} \\
& + (\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha) + \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 \\
= \ & y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} - \bar{\alpha}\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + (\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha) + \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 \\
= \ & y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + (\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha) + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0.
\end{aligned}$$

$$= \int_{\Re^{Mk}} (2\pi)^{-M(T+k)/2} |\Omega|^{-1/2} |\Sigma_{\alpha_0}|^{-1/2} \exp\left\{ -1/2 \left[\begin{array}{c} (y-\Xi\alpha)'\Omega^{-1}(y-\Xi\alpha)+ \\ (\alpha-\alpha_0)'\Sigma_{\alpha_0}^{-1}(\alpha-\alpha_0)] \end{array} \right] \right\} d\alpha$$

$$\begin{aligned} &= \int_{\Re^{Mk}} (2\pi)^{-M(T+k)/2} |\Omega|^{-1/2} |\Sigma_{\alpha_0}|^{-1/2} \\ &\quad \exp\left\{-1/2[(\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha)+Q]\right\} d\alpha \\ &= (2\pi)^{-M(T+k)/2} |\Omega|^{-1/2} |\Sigma_{\alpha_0}|^{-1/2} \exp\{-Q/2\} \\ &\quad \exp \int_{\Re^{Mk}} \left\{-1/2[(\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha)]\right\} d\alpha. \end{aligned}$$

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 recognize a posterior kernel in the above expression and exploit the fact that the posterior properly normalized density integrates to one:

$$\int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} p(\alpha|y)\ d\alpha_1 ... d\alpha_{Mk} = \int_{\Re^{Mk}} p(\alpha|y) d\alpha = 1 \Longrightarrow$$

$$\begin{aligned} 1 &= \int_{\Re^{Mk}} (2\pi)^{-Mk/2} |\Sigma_{\bar{\alpha}}|^{-1/2} \exp\left\{-1/2 \left(\bar{\alpha}-\alpha\right)' \Sigma_{\bar{\alpha}}^{-1} \left(\bar{\alpha}-\alpha\right)\right\} d\alpha \\ \Longrightarrow & \frac{1}{(2\pi)^{-Mk/2} |\Sigma_{\bar{\alpha}}|^{-1/2}} = \int \exp\left\{-1/2 \left(\bar{\alpha}-\alpha\right)' \Sigma_{\bar{\alpha}}^{-1} \left(\bar{\alpha}-\alpha\right)\right\}. \end{aligned}$$

The marginal likelihood is thus:

$$\int_{\Re^{Mk}} p(y|\alpha)p(\alpha)d\alpha = (2\pi)^{-MT/2} |\Omega|^{-1/2} \frac{|\Sigma_{\alpha_0}|^{-1/2}}{|\Sigma_{\bar{\alpha}}|^{-1/2}} \exp\{-Q/2\},$$

where:

$$Q = y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0.$$

From this it is immediate to derive the Bayes factor of the RVAR against the UVAR:

$$BF = \left[\frac{\frac{\left|\Sigma_{\alpha}^{priorRVAR}\right|}{\left|\Sigma_{\alpha}^{postRVAR}\right|}}{\frac{\left|\Sigma_{\alpha}^{priorUVAR}\right|}{\left|\Sigma_{\alpha}^{postUVAR}\right|}} \right]^{-1/2} \exp\left\{ \frac{Q^{UVAR} - Q^{RVAR}}{2} \right\}.$$

D. VAR Augmentation

Our VAR becomes:

$$\Delta r_t = k_1 + a(L)\Delta r_{t-1} + b(L)S_{t-1} + \psi(L)\pi_{t-1} + \varphi(L)y_{t-1} + u_{1t},$$
$$S_t = k_2 + c(L)\Delta r_{t-1} + d(L)S_{t-1} + \rho(L)\pi_{t-1} + \chi(L)y_{t-1} + u_{2t}.$$

Stack the VAR as:

$$\begin{bmatrix} \Delta r_t \\ \Delta r_{t-1} \\ \Delta r_{t-2}. \\ S_t \\ S_{t-1} \\ S_{t-2} \\ \pi_t \\ \pi_{t-1} \\ \pi_{t-2} \\ y_t \\ y_{t-1} \\ y_{t-2} \end{bmatrix} = \begin{bmatrix} k_1 \\ 0 \\ 0 \\ k_2 \\ 0 \\ 0 \\ k_3 \\ 0 \\ 0 \\ k_4 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 & a_3 & b_1 & b_2 & b_3 & \psi_1 & \psi_2 & \psi_3 & \varphi_1 & \varphi_2 & \varphi_3 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ c_1 & c_2 & c_3 & d_1 & d_2 & d_3 & \rho_1 & \rho_2 & \rho_3 & \chi_1 & \chi_2 & \chi_3 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ e_1 & e_2 & e_3 & f_1 & f_2 & f_3 & g_1 & g_2 & g_3 & n_1 & n_2 & n_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ h_1 & h_2 & h_3 & i_1 & i_2 & i_3 & l_1 & l_2 & l_3 & m_1 & m_2 & m_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & o & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta r_{t-1} \\ \Delta r_{t-2} \\ \Delta r_{t-3}. \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \\ \pi_{t-1} \\ \pi_{t-2} \\ \pi_{t-3} \\ y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ 0 \\ 0 \\ u_{2t} \\ 0 \\ 0 \\ u_{3t} \\ 0 \\ 0 \\ u_{4t} \\ 0 \\ 0 \end{bmatrix}.$$

More succinctly:

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

Repeating the same steps of appendix A we get:

$$g' = \sum_{i=1}^{T} \gamma^i h' A^i.$$

For T going to infinity, since also this augmented VAR is cointegrated, this converges to:

$$g' \underset{T\to\infty}{\longrightarrow} h'\gamma A(I - \gamma A)^{-1}.$$

Postmultiplying provides a set of linear restrictions:

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

i.e:

$$
g' \begin{bmatrix}
1/\gamma - a_1 & -a_2 & -a_3 & -b_1 & -b_2 & -b_3 & -\psi_1 & -\psi_2 & -\psi_3 & -\varphi_1 & -\varphi_2 & -\varphi_3 \\
-1 & 1/\gamma & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 1/\gamma & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-c_1 & -c_2 & -c_3 & 1/\gamma - d_1 & -d_2 & -d_3 & -\rho_1 & -\rho_2 & -\rho_3 & -\chi_1 & -\chi_2 & -\chi_3 \\
0 & 0 & 0 & -1 & 1/\gamma & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 1/\gamma & 0 & 0 & 0 & 0 & 0 & 0 \\
-e_1 & -e_2 & -e_3 & -f_1 & -f_2 & -f_3 & 1/\gamma - g_1 & -g_2 & -g_3 & -n_1 & -n_2 & -n_3 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 1/\gamma & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1/\gamma & 0 & 0 & 0 \\
-h_1 & -h_2 & -h_3 & -i_1 & -i_2 & -i_3 & -l_1 & -l_2 & -l_3 & 1/\gamma - m_1 & -m_2 & -m_3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -o & -1 & 1/\gamma & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1/\gamma
\end{bmatrix} =
$$

$$
= h' \begin{bmatrix}
a_1 & a_2 & a_3 & b_1 & b_2 & b_3 & \psi_1 & \psi_2 & \psi_3 & \varphi_1 & \varphi_2 & \varphi_3 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
c_1 & c_2 & c_3 & d_1 & d_2 & d_3 & \rho_1 & \rho_2 & \rho_3 & \chi_1 & \chi_2 & \chi_3 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
e_1 & e_2 & e_3 & f_1 & f_2 & f_3 & g_1 & g_2 & g_3 & n_1 & n_2 & n_3 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
h_1 & h_2 & h_3 & i_1 & i_2 & i_3 & l_1 & l_2 & l_3 & m_1 & m_2 & m_3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & o & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}.
$$

So the EH imposes the following constraints on the individual coefficients of the VAR:

$$
\begin{aligned}
& \begin{bmatrix} -c_1 & -c_2 & -c_3 & 1/\gamma - d_1 & -d_2 & -d_3 & -\rho_1 & -\rho_2 & -\rho_3 & -\chi_1 & -\chi_2 & -\chi_3 \end{bmatrix} \\
= & \begin{bmatrix} a_1 & a_2 & a_3 & b_1 & b_2 & b_3 & \psi_1 & \psi_2 & \psi_3 & \varphi_1 & \varphi_2 & \varphi_3 \end{bmatrix}.
\end{aligned}
$$

References

[1] Amihud Y., Mendelson H., 1991, "Liquidity, Maturity, and the Yields on U.S. Treasury Securities", *Journal of Finance* **46**: 1411-1425.

[2] Ang A., Piazzesi M., 2003, "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables", *Journal of Monetary Economics* **50**: 745-787.

[3] Boudoukh J., Whitelaw R.F., 1991, "The Benchmark Effect in the Japanese Government Bond Market", *Journal of Fixed Income* September: 52-59.

[4] Campbell J., 1995, "Some Lessons from the Yield Curve", *Journal of Economic Perspectives.* **9** **3**: 129-152.

[5] Campbell J., Shiller R., 1987, "Cointegration and Tests of Present Value Models", *Journal of Political Economy.* **95:** 1062-1088.

[6] Campbell J., Shiller R., 1991, "Yield Spreads and Interest Rate Movements: A Bird's Eye View", *Review of Economic Studies.* **58** **3**: 495-514.

[7] Carriero A., Favero C.A., Kaminska I., 2005, "Financial factors, Macroeconomic Information and the Expectations Theory of the Term Structure of interest Rates", *Journal of Econometrics*, forthcoming.

[8] Clarida R., Gali J., Gertler M., 1998, "Monetary Policy Rules in Practice: Some International Evidence", *European Economic Review.* **42**.

[9] Clarida R., Gali J., Gertler M., 1999, "The Science of Monetary Policy: A new-Keynesian Perspective", *Journal of Economic Literature.* **XXXVII** **4**: 1661-1707.

[10] Clarida R., Gali J., Gertler M., 2000, "Monetary Policy Rules and Macroeconomic Stability: evidence and some Theory", *The Quarterly Journal of Economics.* **115** **1**: 147-180.

[11] Cochrane J., 2001, *Asset Pricing* (Princeton: Princeton University Press).

[12] Cochrane J., Piazzesi M., 2005, "Bond Risk Premia", *American Economic Review* **95**, **1**: 138-160.

[13] Cornell B., Shapiro A., 1990, "The Mispricing of U.S. Treasury Bonds: a Case Study", *Review of Financial Studies.* **2**: 297-310.

[14] Cox J., Ingersoll J., Ross S., 1981, "A Reexamination of the Traditional Hypotheses About the Term Structure of Interest Rates", *Journal of Finance* **36**: 321-346.

[15] Daves P., Ehrhardt M.C., 1993, "Liquidity, Reconstruction. and the Value of U.S. Treasury Strips", *Journal of Finance.* **48**: 315-329.

[16] Doan T., Litterman R., Sims C., 1984, "Forecasting and Conditional Projection Using Realistic Prior Distributions", *Econometric Reviews.* **3**: 1-100.

[17] Diebold F.X., Mariano R.S., 1995, "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics* **13 3**: 253-63.

[18] Diebold F.X., Rudebusch G.D., Aruoba S.B., 2003, "The Macroeconomy and the Yield Curve: a Dynamic Latent Factor Approach", *NBER Working Paper.* No. 10616.

[19] Duffee G., 1996, "Idiosyncratic Variation of Treasury Bill Yields", *Journal of Finance.* **51**: 527-551.

[20] Elton E.J., 1999, "Expected return, Realized Return and Asset Pricing Tests", *Journal of Finance.*

[21] Elton E.J., Green C., 1998, "Tax and Liquidity Effects in Pricing Government Bonds", *Journal of Finance* **53**: 1533-1562.

[22] Fama E., Bliss R.R., 1987, "The Information in Long-Maturity Forward Rates", *American Economic Review.* **77**: 680-692.

[23] Favero C.A., 2005, "Taylor rules and the Term Structure", Forthcoming in *Journal of Monetary Economics*

[24] Fisher I., 1930, *The Theory of Interest* (New York: Macmillan).

[25] Fisher M., Gilles C., 1998, "Around and Around: the Expectations Hypothesis", *Journal of Finance* **53**: 365-383.

[26] Fuhrer J.C., 1996, "Monetary Policy Shifts and Long-Term Interest Rates", *Quarterly Journal of Economics.* **111 4**: 1183-1209.

[27] Geweke J., 1998, "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication", *FED of Minneapolis Research Department Staff Report 249*

[28] Giacomini R., White H., 2004, "Tests of Conditional Forecast Accuracy", *mimeo UCSD*

[29] Hicks J., 1953, "The Long-Term Dollar Problem", *Oxford EP*

[30] Jeffreys H., 1935, "Some Tests of Significance, Treated by the Theory of Probability", *Proceedings of the Cambridge Philosophy Society.* **31**: 203-222.

[31] Jeffreys H., 1961, *Theory of Probability.* (Oxford: Oxford University Press)

[32] Johansen S., 1995, *Likelihood-based Inference in Cointegrated Vector Auto-regressive Models.* Oxford University Press, Oxford UK.

[33] Kass R.E., Raftery A., 1995, "Bayes Factors", *Journal of the American Statistical Association*, **90 430**: 773-795.

[34] Keynes J.M., 1930, *Treatise on Money.* (New York: Harcourt, Brace and Howe Inc.).

[35] Ingram B., Whitemann C., 1994, "Supplanting the Minnesota prior. Forecasting Macroeconomic Time Series Using Real Business Cycle Priors", *Journal of Monetary Economics.* **34**: 497-510.

[36] Longstaff F.A., 1990, "Time Varying Term Premia and Traditional Hypotheses about the Term Structure", *The Journal of Finance* **45**, 1307-1314.

[37] Longstaff F.A., 1992, "Are Negative Option Prices Possible? The Callable U.S. Treasury Bond Puzzle", *Journal of Business.* **65**: 571-592.

[38] Longstaff F.A., 2000, "Arbitrage and the Expectations Hypothesis", *Journal of Finance.* **55 2:** 989-994.

[39] Mankiw N.G., Miron J., 1986, "The Changing Behaviour of the Term Structure of Interest Rates", *The Quarterly Journal of Economics.* **101**: 211-221.

[40] McCallum B., 1994, "Monetary Policy and the Term Structure of Interest Rates", *NBER Working Paper.* No. 4938.

[41] McCulloch J.H., 1993, "A Reexamination of Traditional Hypoteheses About the Term Structure: a Comment", *Journal of Finance* **49**: 1861-82.

[42] Roush J., 2001, "Evidence Uncovered: Long-Term Interest Rates, Monetary Policy and the Expectations Theory", *mimeo Board of Governors of the Federal Reserve System.*

[43] Rudebusch G.D., 1995, "Federal Reserve Interest Rate Targeting, rational expectations, and the term structure", *Journal of Monetary Economics.* **35**: 245-274.

[44] Rudebusch G.D., 2002, "Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia", *Journal of Monetary Economics.* **49**: 1161-1187.

[45] Shiller R., 1979, "The Volatility of Long Term Interest Rates and Expectations Models of the Term Stucture", *Journal of Political Economy.* **87**: 1190-1219.

[46] Shiller R., 1981, "Alternative Tests of Rationals Expectations Models: the Case of the Term Structure", *Journal of Econometrics* **16**: 71-87.

[47] Sims C., Zha T., 1998, "Bayesian Methods for Dynamic Multivariate Models", *International Economic Review.* **39**: 949-968.

[48] Söderlind P., Söderström U., Vredin A., 2005, "Dynamic Taylor Rules and the Predictability of Interest Rates", *Macroeconomic Dynamics*, forthcoming.

[49] Taylor J.B., 1993, "Discretion Versus Policy Rules in Practice", *Carnegie-Rochester Conference Series on Public Policy.* **39**. 195-214.

[50] Theil H., 1971, *Principles of Econometrics.* (New York; Wiley).

[51] Thornton D.L., 2005, "Tests of the Expectations Hypothesis: Resolving the Campbell-Shiller Paradox", *Journal of Money, Credit, and Banking.* Forthcoming.

[52] Zellner A., 1988, "Bayesian Analysis in Econometrics", *Journal of Econometrics.* **37**: 27-50.

.

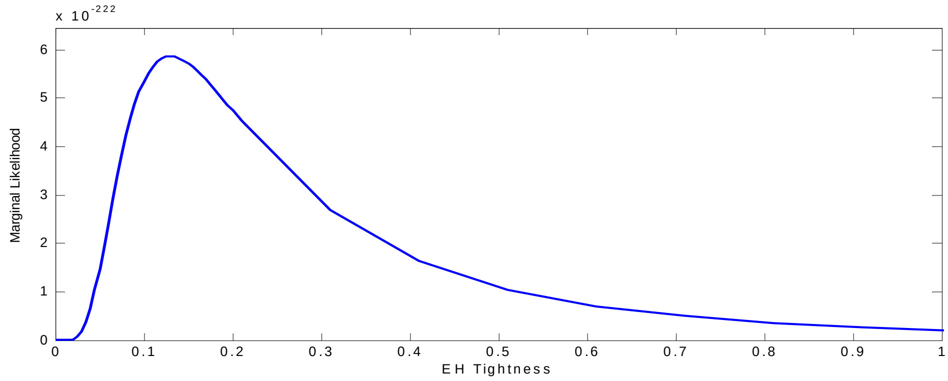

Figure 1: Marginal Likelihood of the Restricted VAR as a function of the EH prior tightness σ

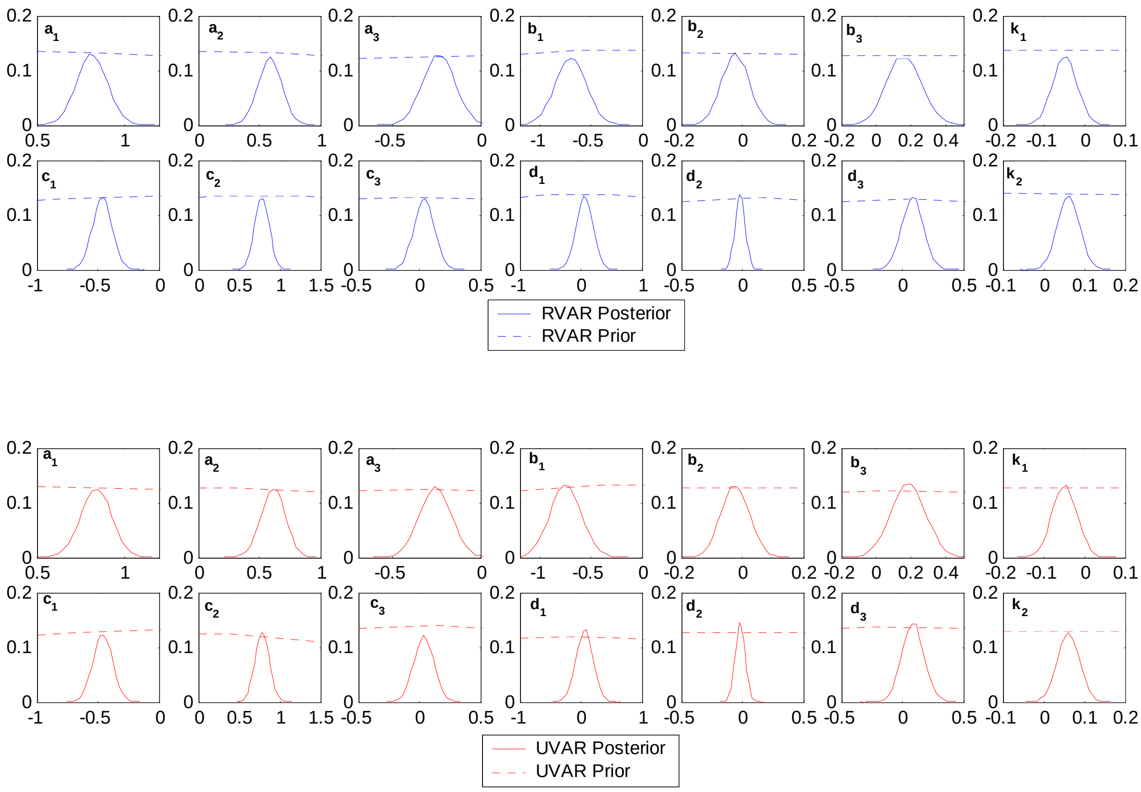

Figure 2: Prior and posterior distributions for the VAR coefficients under the Expectations Hypothesis prior (Restricted VAR) and under the loose prior (Unrestricted VAR)

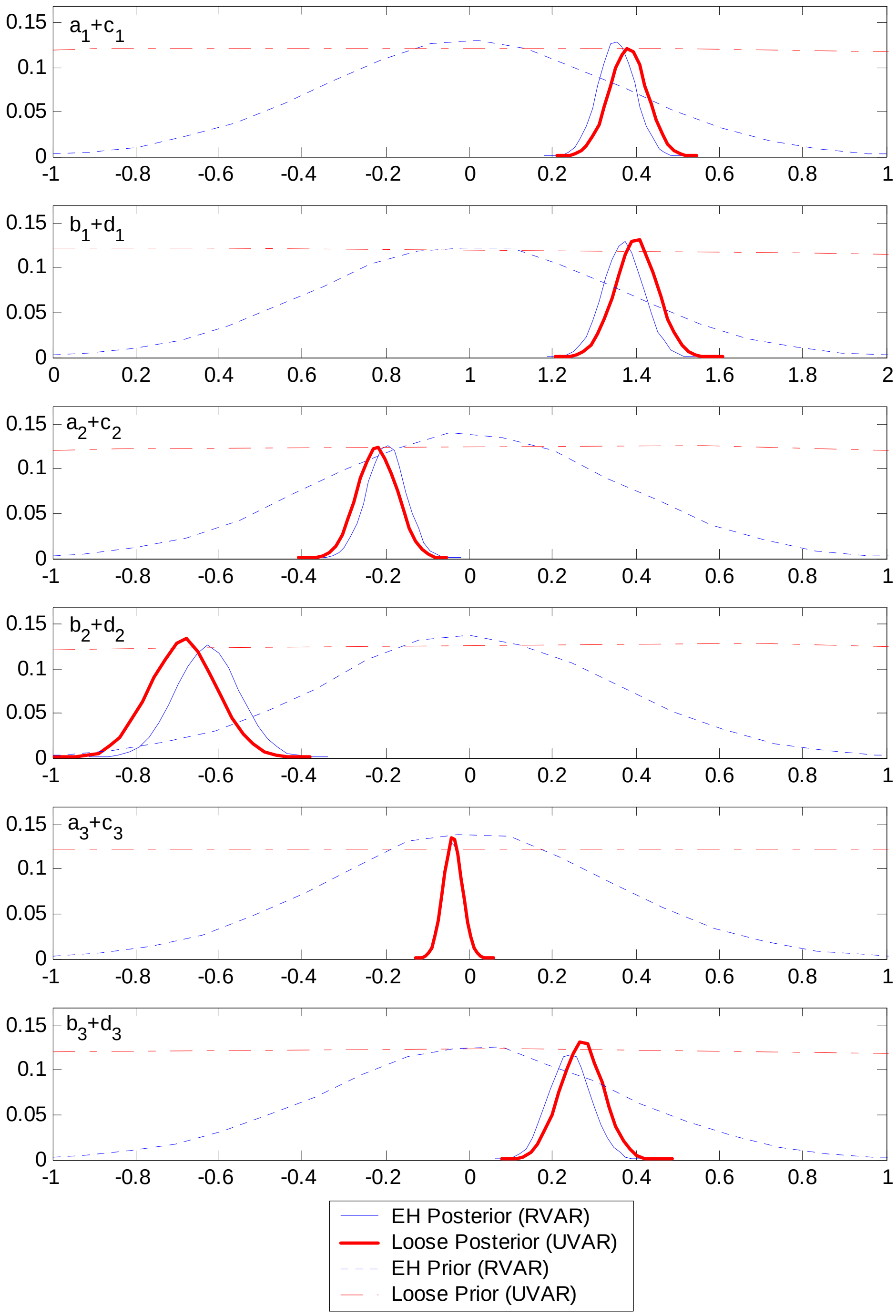

Figure 3: Prior and posterior distributions of the Expectation Hypothesis restrictions under the Expectation Hypothesis prior (Restricted VAR) and the loose prior (Unrestricted VAR)

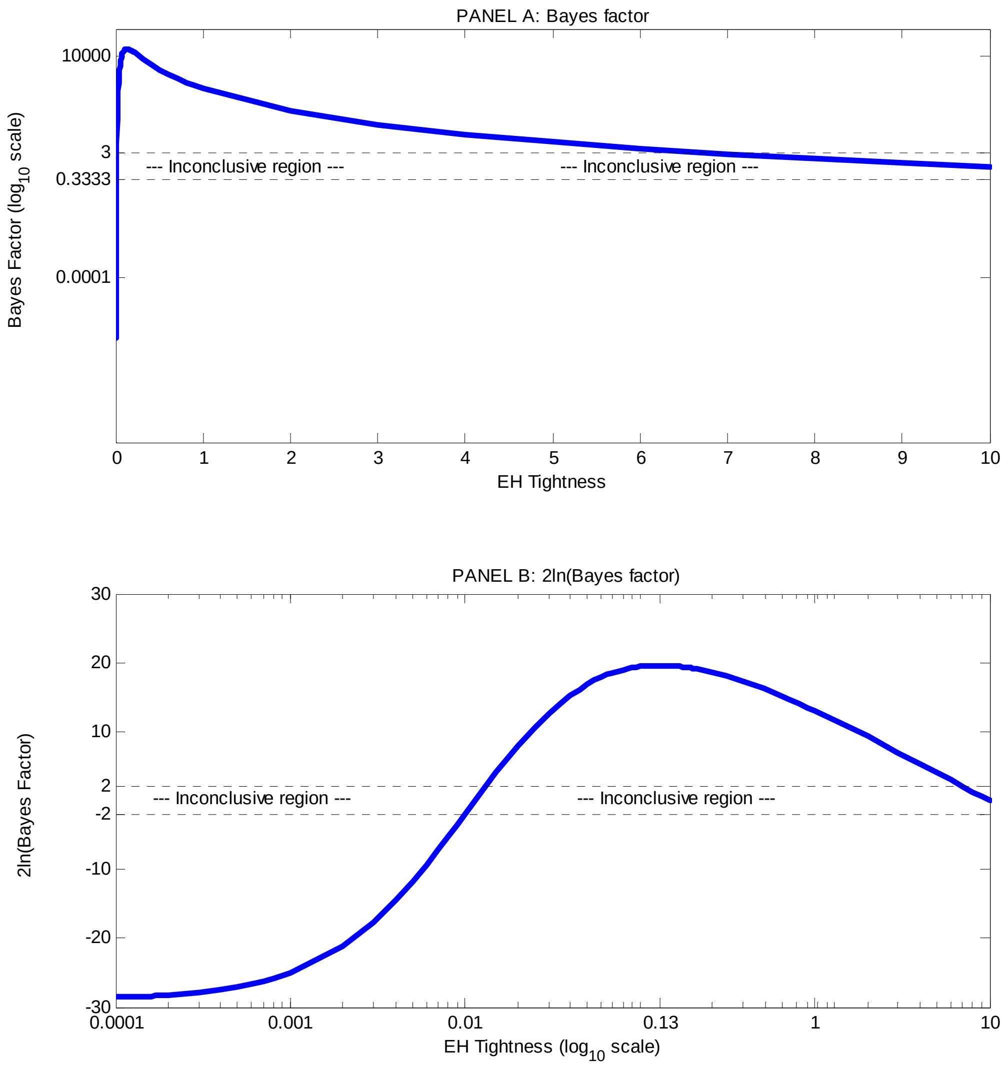

Figure 4: Bayes factor and twice its natural logarithm for the EH-restricted versus the unrestricted VAR, as a function of the EH prior tightness σ.The y axis in panel A and the x axis is in panel B are in base 10 logarithmic scale.

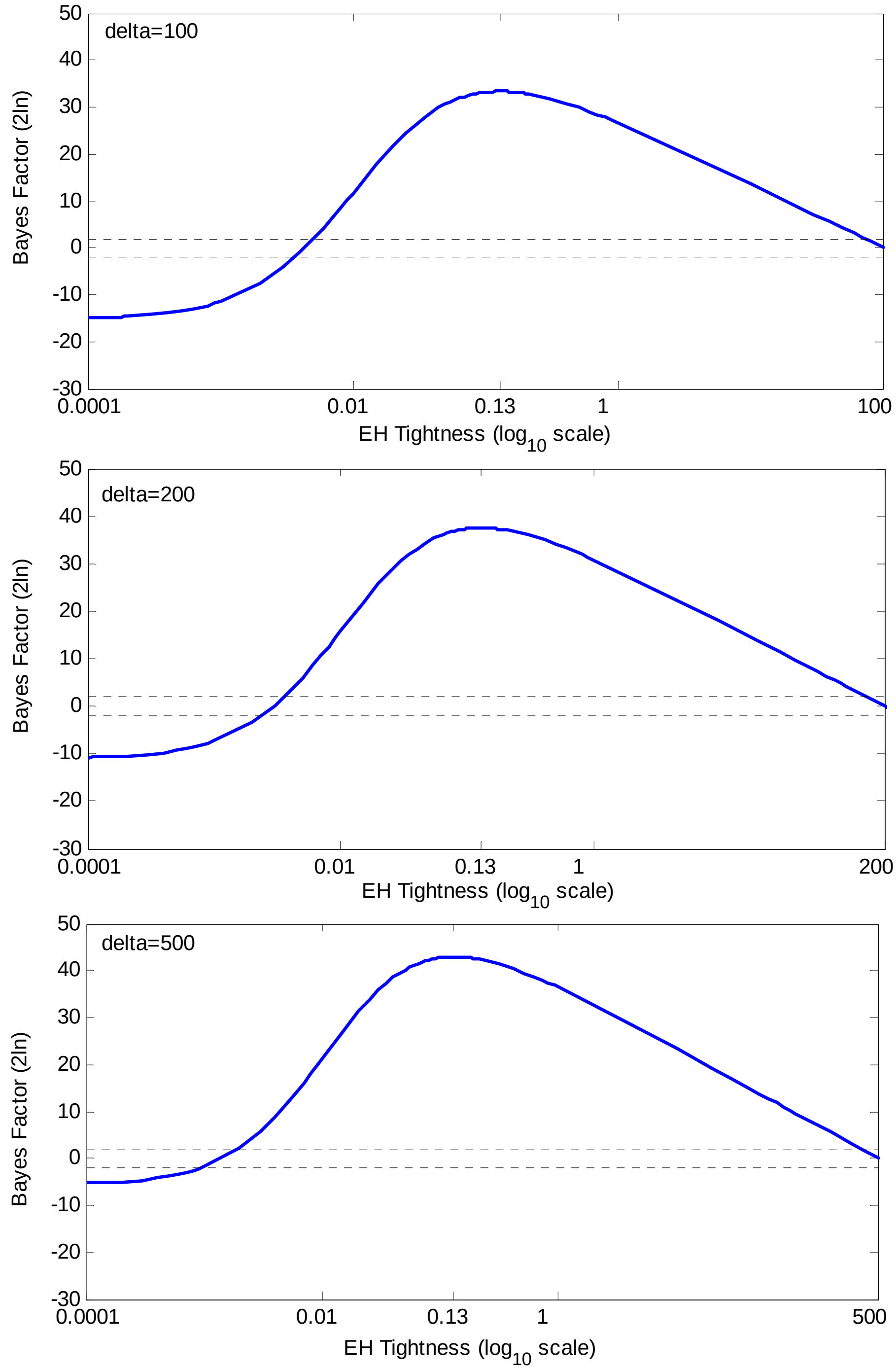

Figure 5: Robustness to alternative values of the variance of unrestricted coefficients δ. Bayes factor (twice its natural logarithm) for the EH-restricted versus the unrestricted VAR, as a function of the EH prior tightness σ. The x axis is in base 10 logarithmic scale.

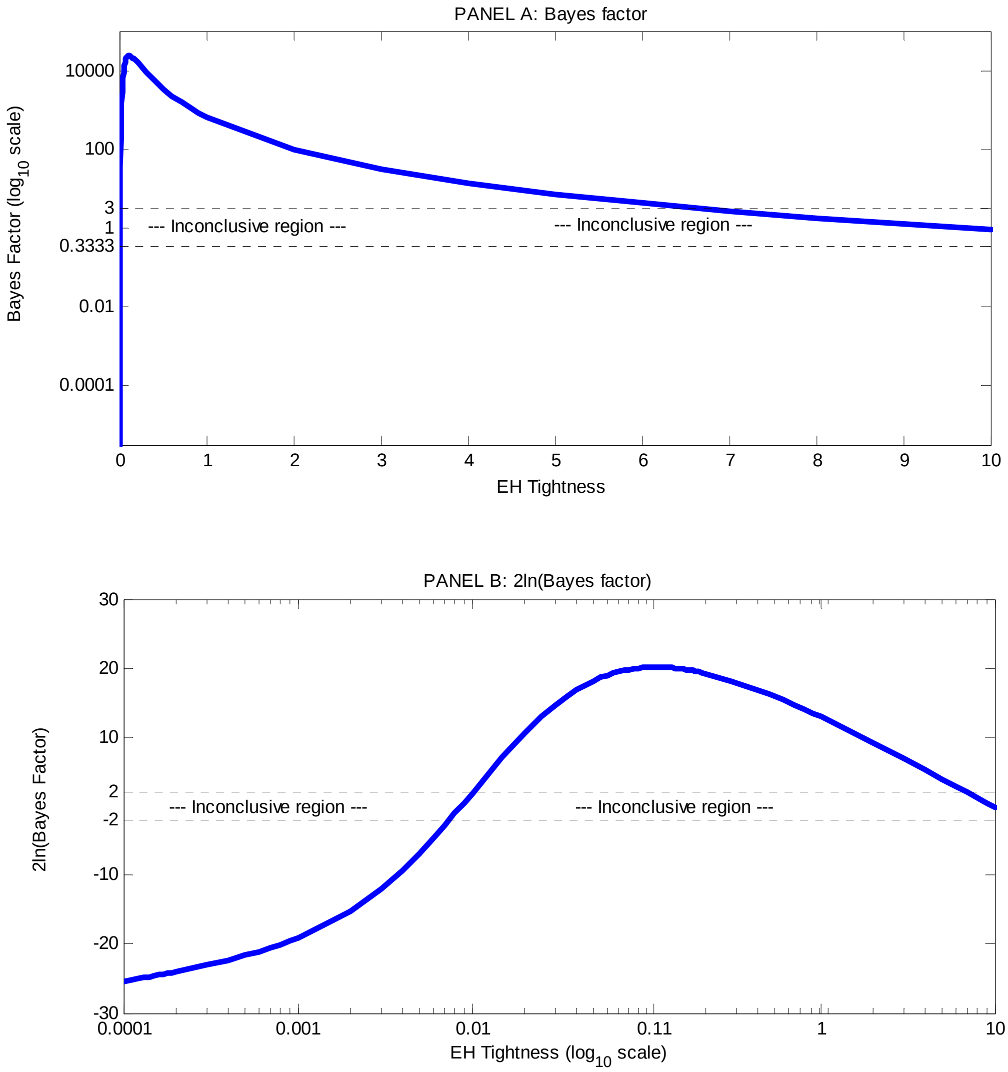

Figure 6: Bayes factor and twice its natural logarithm for the EH-restricted versus the unrestricted VAR, as a function of the EH prior tightness σ.The y axis in panel A and the x axis is in panel B are in base 10 logarithmic scale.

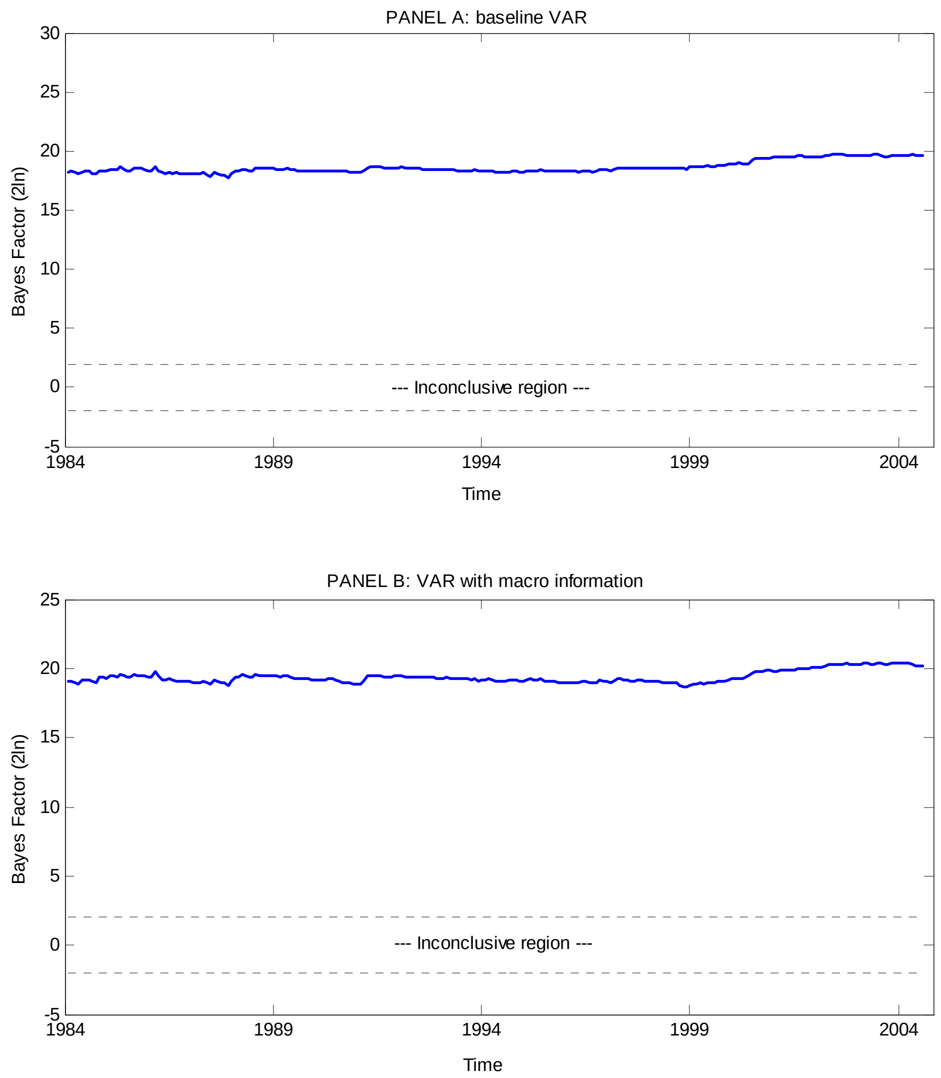

Figure 7: Robustness through time. Bayes factor (twice its natural logarithm) for the EH-restricted versus the unrestricted VAR as a function of time. The EH prior tightness is fixed at its estimated value $\sigma = 0.13$

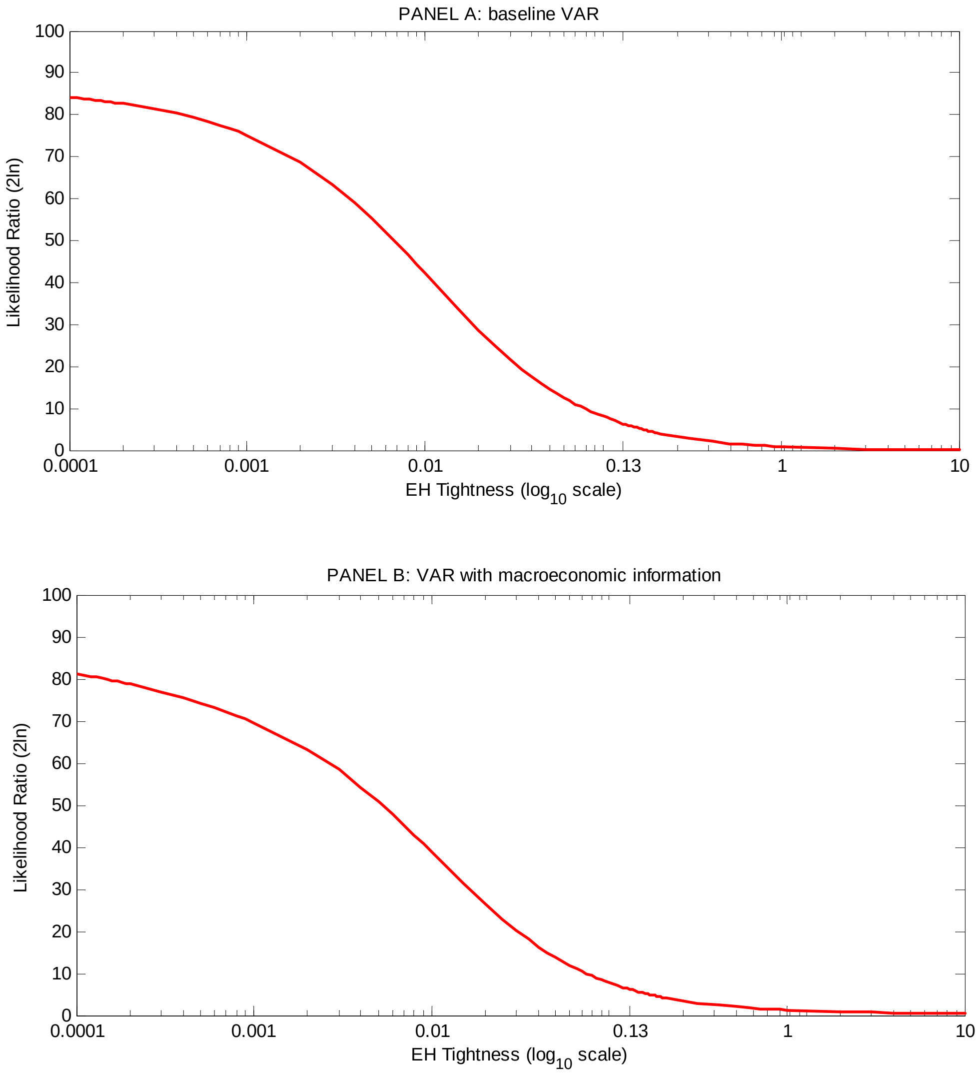

Figure 8: Twice the natural logarithm of likelihood ratios for the EH-restricted versus the unrestricted VAR, as a function of the EH prior tightness σ. The x axis is in base 10 logarithmic scale.

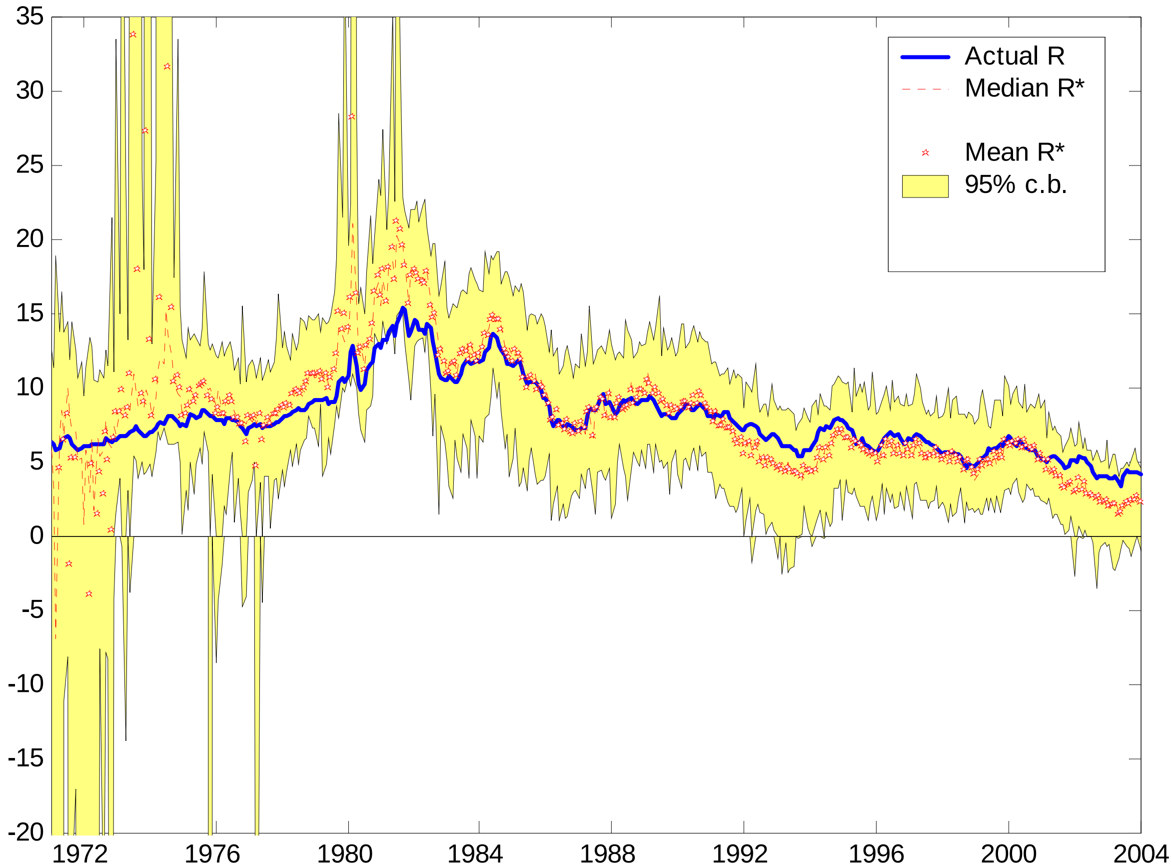

Figure 9: Economic test of the Expectations Hypothesis. The distribution of the theoretical, EH-consistent long term rate $R^*_{t,T}$ is obtained by a recursive estimation/projection scheme, such that at each point in time only the available information is used to estimate the VAR and then to project it forward.The procedure works as follows: i) The first estimation is performed over the sample 1966:1 1970:12. All the subsequent estimations are performed over the sample 1966:1 1970:12+i where i is the number of iterations already executed. ii) Using the posterior of the coefficients obtained at point i) the VAR is projected forward and posterior of the variables $E_t r_{t+i}$ and $R^*_{t,T}$ are obtained . iii) Then we move forward one period, adding one data point to the estimation window, and go back to point i).

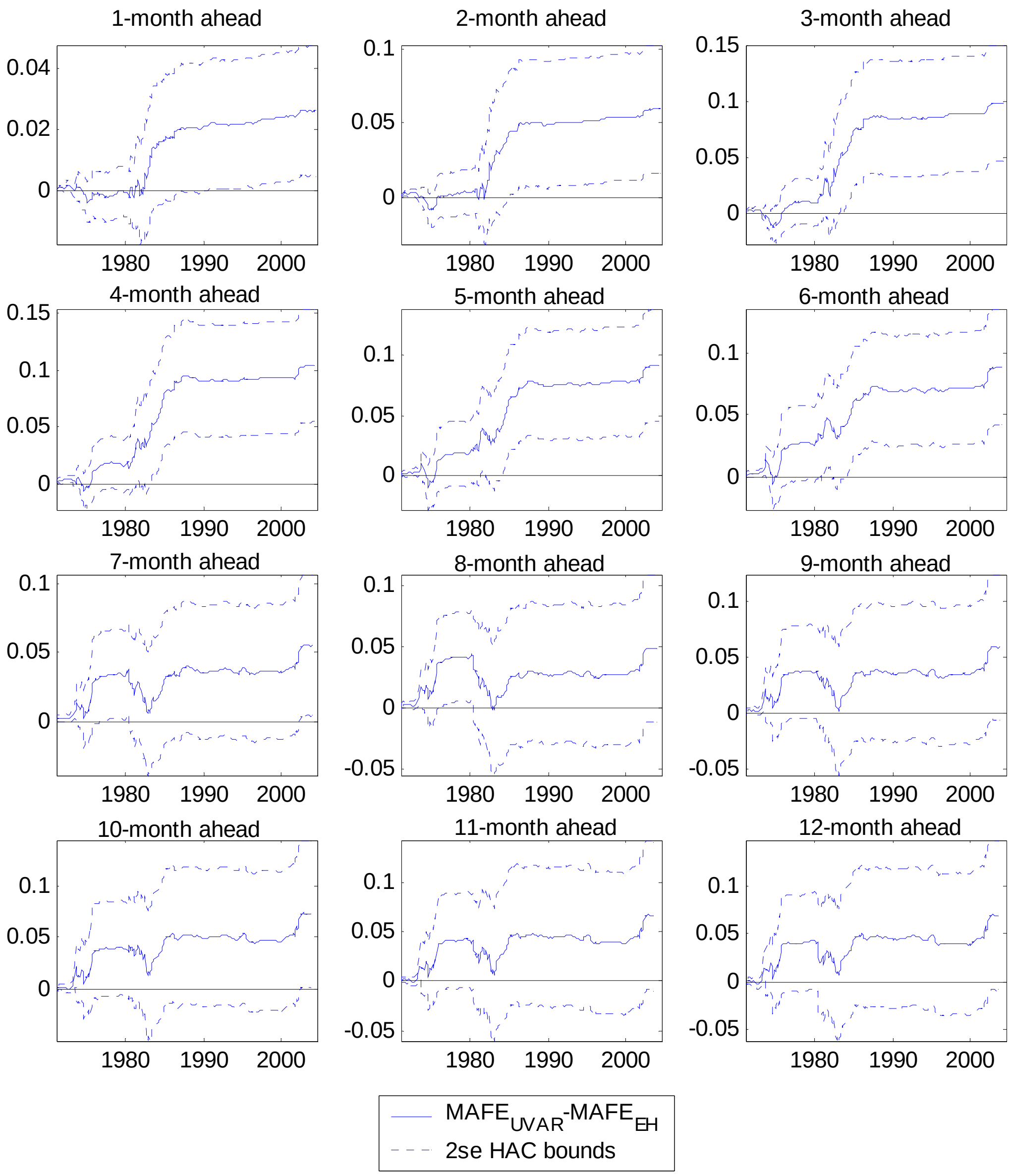

Figure 10: Forecast accuracy evaluation. Excess of Absolute Forecast Errors of the UVAR with respect to the EH. Rolling (5 year window) estimates with 2 HAC standard errors bounds of the parameter α in the following regression equation: $AFE_{h,UVAR} - AFE_{h,EH} = \alpha + u_t$, where h is the number of step-ahead.

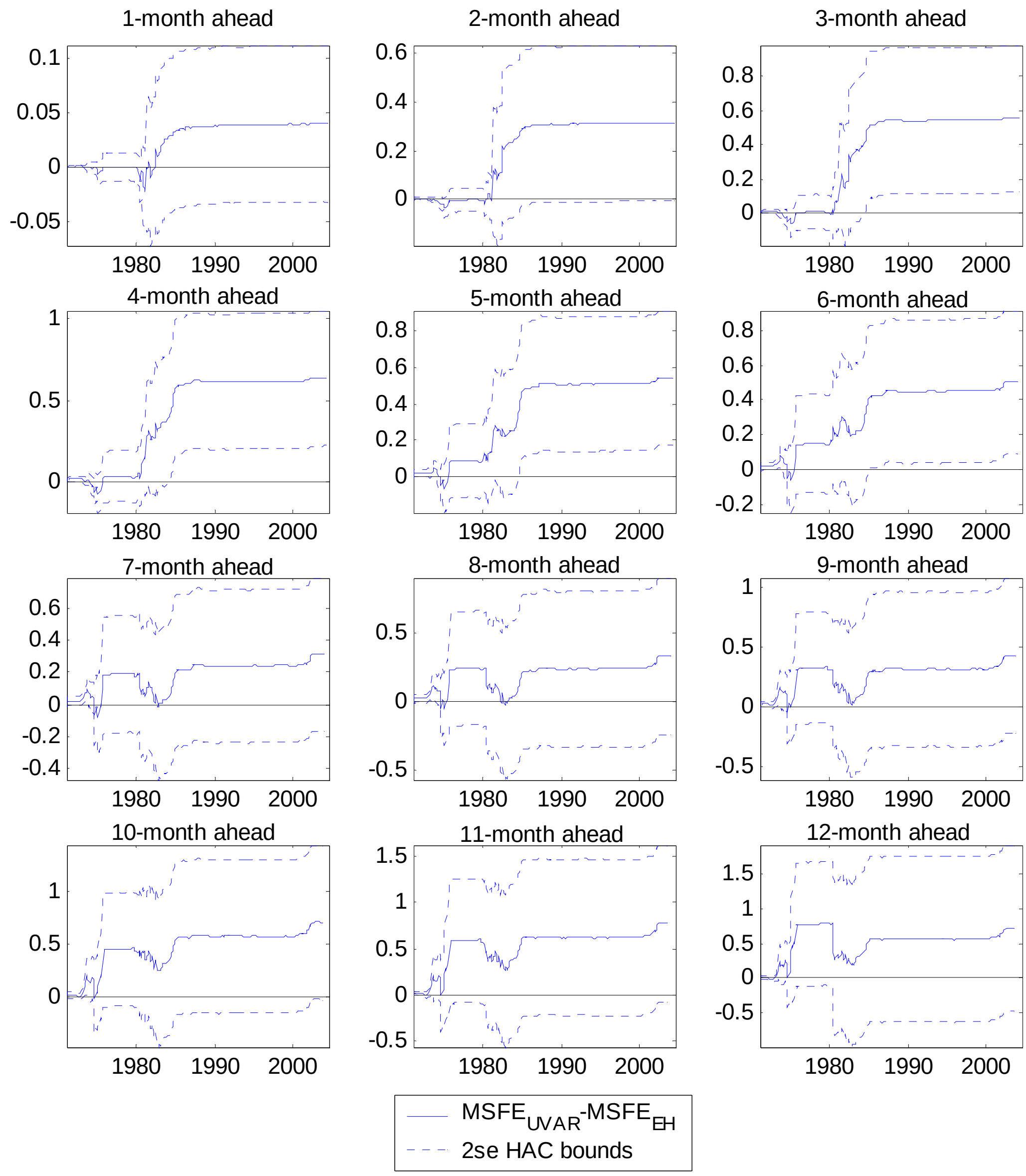

Figure 11: Forecast accuracy evaluation. Excess of Squared Forecast Errors of the UVAR with respect to the EH. Rolling (5 year window) estimates with 2 HAC standard errors bounds of the parameter β in the following regression equation: $SFE_{h,UVAR} - SFE_{h,EH} = \beta + v_t$, where h is the number of step-ahead.

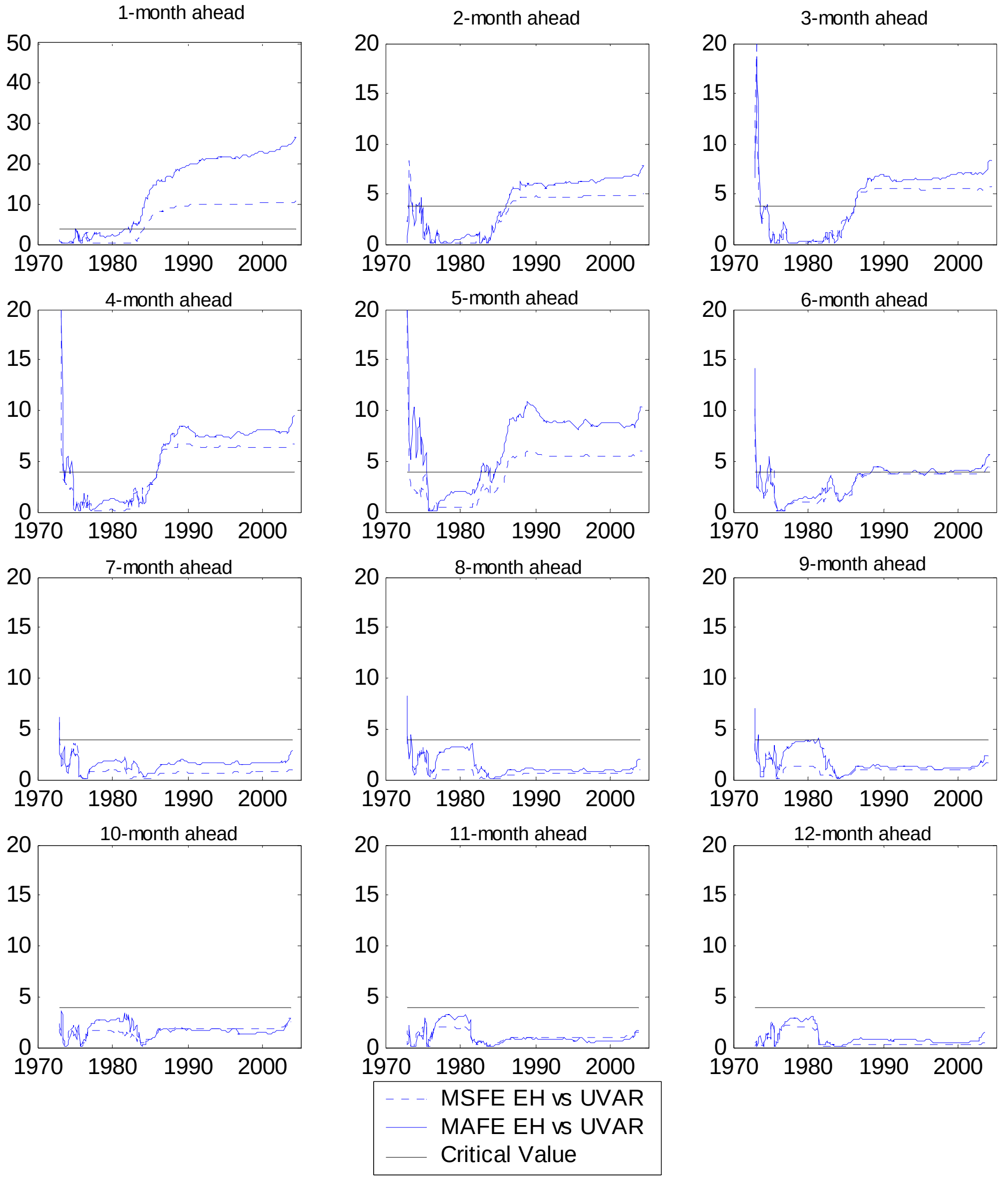

Figure 12: Forecast accuracy evaluation. Rolling (5-year window) Giacomini and White (2004) test statistic for the null of equal conditional predictive accuracy of the UVAR and the EH. The statistic is computed both for the MAFE (solid) and the MSFE (dotted) loss function. If the statistic is above the critical value, the null of equal accuracy can be rejected.