Università Commerciale Luigi Bocconi - Milano

Ciclo XIX

Dottorato di ricerca in ECONOMIA

PARAMETERS UNCERTAINTY AND TESTS OF PRESENT VALUE MODELS

Tesi di Dottorato di :

Andrea Carriero

Matr. n°: 935304DT

Anno Accademico 2005-2006

Parameters Uncertainty and Tests of Present Value Models

Andrea Carriero

20 January 2007

Acknowledgements

I would like to thank Carlo Favero, my main advisor, for his suggestions and constant support during my research. Special thanks go also to Fabio Canova, Massimiliano Marcellino, and Ulf Söderström. I should also mention my graduate fellows at Bocconi University, Marco Aiolfi, Agostino Consolo, and Iryna Kaminska.

Milan, 20 January 2007

Andrea Carriero

Abstract

This thesis is composed of 3 essays.

The first paper, "Financial Factors, Macroeconomic Information and the Expectations Theory of the Term Structure of Interest Rates", provides a new test of the Expectation Theory (ET) of the term structure. In particular we use financial factors and macroeconomic information to construct a test of the theory based on simulating investors' effort to use the model in 'real-time' to forecast future monetary policy rates. We build on this framework to measure the term premium and the implicit monetary policy maker's reaction function. The application of our approach to a monthly sample of US data from the eighties onward leads us to conclude that the deviation from the ET are very rarely significant and that fluctuations of risk premia are not large.

In the second paper, "A Bayesian Framework for the Expectations Hypothesis. How to Extract Additional Information From the Term Structure of Interest Rates", I propose a way to use the Expectations Hypothesis (EH) of the term structure of interest rates without imposing it dogmatically. I do 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 arising from 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.

In the third paper, "Explaining US-UK Interest Rate Differentials: a Reassessment of the Uncovered Interest Rate Parity in a Bayesian Framework", I test the Uncovered Interest Rate Parity (UIRP) allowing for transitory deviations from it. These deviations may arise from variations in risk premia, errors in expectations, and linearization errors.

Abstract

Importantly, this approach comprises the traditional one as a special case, which is derived simply setting the deviations to zero. When the noise is set to zero the UIRP is rejected, but if we allow for some degree of noise the UIRP is strongly supported by the data. Thus the UIRP relation does not hold exactly, but holds on average, with a stationary risk premium as opposed to a constant one. This result implies that analyzing the effects of policy experiments under the null of the UIRP may be both safe and useful.

Contents

Chapter 1

Financial Factors, Macroeconomic Information and the Expectations Theory of the Term Structure of Interest Rates

1.1 Introduction

The[1] objective of this paper is to provide new evidence on the expectations theory (ET) of the term structure of interest rates.

How is this possible?

Our starting point is the widely cited work by Campbell and Shiller(1987)(CS), where they implement a bi-variate vector autoregressive (VAR), which is different from the bulk of the available literature which rejects the ET within a single-equation, limited information approach (see, for example, Campbell,1995, Fama and Bliss,1987, and Cochrane,2001). CS implement a test which still rejects the ET but their analysis of the data leads them to conclude that there is an important element of truth to the expectations theory of the term structure.

We develop on the CS framework along three dimensions: the use of a testing method based on a real-time procedure in which the econometrician is given the same information available to market participants when they make their decisions on portfolio allocation,

[1] This paper is coauthored with Carlo Favero and Iryna Kaminska, and has been published in the Journal of Econometrics, Vol. 131, pp. 339-358. It benefited of very helpful comments and constructive criticisms from Giampiero Gallo and three anonymous referees.

the specification of the implicit monetary policy maker's reaction function, the measurement of the risk premium in case of rejection of the null of the ET.

First, CS test the restrictions imposed by the ET on a VAR model in the spread between long and short term interest rates and the change of short-term interest rates and by using only in-sample information. Such procedure cannot simulate the investors' effort to use the model in 'real time' to forecast future monetary policy rates: the information from the whole sample is used to estimate parameters while investors can use only historically available information to generate (up to n-period ahead) predictions of policy rates. Moreover, the within sample test understates the uncertainty of agents who forecast policy rates by out-of-sample projections. In this paper we use the present value framework to generate real time forecast for future policy rates. At each point in time we estimate, using the historically available information, a model and then we use it to project out-of-sample policy rates up to the nth-period ahead. Given the path of simulated future policy rates, we can construct yield to maturities consistent with the Expectations Theory. Using the historically available information on uncertainty we perform dynamic stochastic simulations and construct confidence bounds around the ET-consistent long-term rates. These bounds reflect explicitly the uncertainty associated with out-of-sample projections. It then becomes natural to test the ET by checking if the observed long-term rates fluctuate within the bounds.

Second, by having an explicit model for the short-rate in their testing framework CS circumvent one of the main assumptions of the single-equation approach to the ET, namely the use of ex-post realized returns as a proxy for ex-ante expected returns. In a recent paper, Elton (1999) clearly asserts that there is ample evidence against the belief that information surprises tend to cancel out over time and hence realized returns cannot be considered as an appropriate proxy for expected returns. Interestingly, Campbell (2001) finds strong effects of expectations errors on the single-equation tests, which are confirmed by a number of papers which concentrates on expectations errors by relating them to "peso problems" or to the very low predictability of short term interest rates. In a famous study Mankiw and Miron, 1986, using data on a three and six month maturity, found evidence in favor of the expectation theory prior to the founding of the Federal Reserve System in 1915. They show that the shift in regime occurred with the founding of the Fed led to a remarkable decrease in the predictability of short-term interest rates. Rudebusch, 1995, and Balduzzi et al., 1997, expand on this evidence by looking at more recent data. As a consequence of the use of ex-post realized returns as a proxy for ex-ante expected returns the single-equation approach cannot identify if the empirical failure of the model is due to systematic expectations errors, or to shifts in the risk premia. CS have an implicit model to construct expectations, they find much milder evidence against the ET but they do not exploit their model to construct a measure of risk premium.

By implementing our simulation based procedure we can explicitly measure deviations from the ET and, under the null that our proposed model delivers expected future policy rates not different from those expected by the market, interpret them as a measure of risk premium.

Third, on a different, but clearly related, ground McCallum(1994) is the first to argue that the limited information approach might cause bias in the estimates due to simultaneity. He shows that the anomalous empirical findings based on a single equation evidence can be rationalized with the expectations theory by a recognition of an exogenous term premium plus the assumption that monetary policy involves smoothing of the policy rates together with the responses to the prevailing level of the spread. Interestingly, the bi-variate framework considered by CS matches exactly the scenario used by McCallum to illustrate the simultaneity bias in the single-equation approach. However, McCallum himself notes that a reaction function according to which the Fed reacts to the spread only represents a simplification relative to the actual behaviour of the Fed, which almost certainly responds to recent inflation and output or employment movements, as well as to the spread. In fact, both the financial literature and the macroeconomic literature point to potential mis-specification of the simple reaction function used by CS.

There is ample empirical evidence that a three-factor model is needed to accurately describe the term structure and that the use of term structure related factors is of considerable help in modelling monetary policy rates (see, for example, Ang and Piazzesi(2003)), it is easy to see that in the CS approach only two factors are considered. The success of Taylor rules (Taylor,1993, Clarida, Gali and Gertler, 1998, 1999, 2000) points out an obvious potential misspecification of the original Campbell-Shiller framework: the omission of macroeconomic variables to which the monetary policy maker reacts. We shall assess potential mis-specification effects by using an extended VAR which includes three factors for the term structure and macroeconomic variables used in Taylor rules.

The paper is organized as follows. Section 1 illustrates the testing framework by contrasting the Present Value approach with our simulation based alternative. Section 2 illustrates our testing framework and our extension of the information set. Section 3 presents the empirical evidence. Section 4 contains an assessment of the robustness of our results to the use of a different sample and of a different method for updating parameter estimates upon accrual of new information. Section 5 concludes.

1.2 Testing framework

We introduce our testing framework by comparative evaluation of the traditional present value approach and of proposed simulation based approach.

1.2.1 The Present Value approach

We describe the Present Value approach by adopting the linearized expectations model of Shiller (1979) in the bi-variate framework proposed by CS.

We start by imposing a no-arbitrage condition, according to which the expected one-period holding returns from long-term bonds must be equal the risk-free short term interest rate plus a term premium. For long term bonds bearing a coupon C, $H_{t,T}$ is a non-linear function of the yield to maturity $R_{t,T}$. Shiller (1979) proposes a linearization which takes the approximation in the neighborhood $R_{t,T} = R_{t+1,T} = \bar{R} = C$, in which case we have:

$$E[H_{t,T} \mid I_t] = E\left[\frac{R_{t,T} - \gamma_T R_{t+1,T}}{1-\gamma_T} \mid I_t\right] = r_t + \phi_{t,T} \tag{1.1}$$

where $H_{t,T}$ is the one-period holding return of a bond with maturity date T, I_t is the information set available to agents at time t, r_t is the short term interest rate,γ_T is a constant of linearization which depends on the maturity of the bond and $\phi_{t,T}$ is a term premium defined over a one-period horizon for holding the bond with residual maturity $T-t$. Consider the above expression for a very long term bond , by recursive substitution, under the terminal condition that at maturity the price equals the principal,we obtain:

$$R_{t,T} = R^*_{t,T} + E[\Phi_{t,T} \mid I_t] = \frac{1-\gamma}{1-\gamma^{T-t}} \sum_{j=0}^{T-t-1} \gamma^j E[r_{t+j} \mid I_t] + E[\Phi_{t,T} \mid I_t] \tag{1.2}$$

where $\lim_{T\longrightarrow\infty} \gamma_T = \gamma = 1/(1+\bar{R})$ and $\Phi_{t,T}$ is the term premium over the whole life of the bond:

$$\Phi_{t,T} = \frac{1-\gamma}{1-\gamma^{T-t}} \sum_{j=0}^{T-t-1} \gamma^j \phi_{t+j,T}$$

CS tests the ET[2] by using equation (1.2) in considering the case of the risk free rate and a very long term bond. In such case, the null of the ET is imposed in strong form

[2] In fact CS use de-meaned-variables, that is equivalent to test a weak form of the Expectations Theory, in the sense that de-meaning eliminates a constant risk premium.

by imposing that $E[\Phi_{t,T} \mid I_t]$ is zero and in weak form by imposing that $E[\Phi_{t,T} \mid I_t]$ is captured by a constant. CS consider de-meaned variables, and hence test a weak form of the ET by considering the following restriction:

$$R_{t,T} = R^*_{t,T} \approx (1-\gamma) \sum_{j=0}^{T-t-1} \gamma^j E[r_{t+j} \mid I_t] \tag{1.3}$$

which could be re-written in terms of spread between long and short-term rates, $S_{t,T} = R_{t,T} - r_t$:

$$S_{t,T} = S^*_{t,T} = \sum_{j=1}^{T-t-1} \gamma^j E[\Delta r_{t+j} \mid I_t] \tag{1.4}$$

(1.4) shows that a necessary condition for the ET to hold puts constraints on the long-run dynamics of the spread. In fact, the spread should be stationary being a weighted sum of stationary variables. Obviously, stationarity of the spread implies that, if yields are non-stationary, they should be cointegrated with a cointegrating vector (1,-1). However, the necessary and sufficient conditions for the validity of the ET impose restrictions both on the long-run and the short run dynamics.

Assuming[3] that $R_{t,T}$ and r_t are cointegrated with a cointegrating vector (1,-1), CS construct a bivariate stationary VAR in the first difference of the short-term rate and the spread :

$$\begin{aligned} \Delta r_t &= a(L)\Delta r_{t-1} + b(L)S_{t-1} + u_{1t} \\ S_t &= c(L)\Delta r_{t-1} + d(L)S_{t-1} + u_{2t} \end{aligned} \tag{1.5}$$

Stack the VAR as:

$$\begin{bmatrix} \Delta r_t \\ . \\ . \\ \Delta r_{t-p+1} \\ S_t \\ . \\ . \\ S_{t-p+1} \end{bmatrix} = \begin{bmatrix} a_1 & . & . & a_p & b_1 & . & . & b_p \\ 1 & . & . & 0 & 0 & . & . & 0 \\ 0 & . & . & 0 & 0 & . & . & 0 \\ 0 & . & 1 & 0 & 0 & . & . & 0 \\ c_1 & . & . & c_p & d_1 & . & . & d_p \\ 0 & . & . & 0 & 1 & . & . & 0 \\ 0 & . & . & 0 & 0 & . & . & 0 \\ 0 & . & . & 0 & 0 & . & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta r_{t-1} \\ . \\ . \\ \Delta r_{t-p} \\ S_{t-1} \\ . \\ . \\ S_{t-p} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ . \\ . \\ 0 \\ u_{2t} \\ . \\ . \\ 0 \end{bmatrix} \tag{1.6}$$

[3]In fact, the evidence for the restricted cointegrating vector which constitutes a necessary condition for the ET to hold is not found to be particularly strong in the original CS work.

This can be written more succinctly as:

$$z_t = Az_{t-1} + v_t \tag{1.7}$$

The ET null puts a set of restrictions which can be written as :

$$g'z_t = \sum_{j=1}^{T-1} \gamma^j h' A^{j'} z_t \tag{1.8}$$

where $g\prime$ and $h\prime$ are selector vectors for S and Δr correspondingly (i.e. row vectors with 2p elements, all of which are zero except for the p+1st element of $g\prime$ and the first element of $h\prime$ which are unity). Since the above expression has to hold for general z_t, and, for large T, the sum converges under the null of the validity of the ET, it must be the case that:

$$g' = h'\gamma A(I - \gamma A)^{-1} \tag{1.9}$$

which implies:

$$g'(I - \gamma A) = h'\gamma A \tag{1.10}$$

and we have the following constraints on the individual coefficients of VAR(1.5):

$$\{c_i = -a_i, \forall i\}, \{d_1 = -b_1 + 1/\gamma\}, \{d_i = -b_i, \forall i \neq 1\} \tag{1.11}$$

The above restrictions are testable with a Wald test. By doing so using US data between the fifties and the eighties Campbell and Shiller (1987) rejected the null of the ET. However, when CS construct a theoretical spread $S^*_{t,T}$, by imposing the (rejected) ET restrictions on the VAR they find that, despite the statistical rejection of the ET, $S^*_{t,T}$ and $S_{t,T}$ are strongly correlated.

1.2.2 A new testing framework with an extended information set

We extend the CS approach along two dimensions: the specification of the VAR and the construction of a test based on information available in real time.

Both the financial and the macroeconomic empirical literature suggest that the parsimonious model consisting of the spread and the change in the short-term rate could be in fact too parsimonious to fit the data. The financial literature has shown that the construction of a satisfactory model of the term structure requires at least three factors, usually labelled as level, slope and curvature. Rethinking the CS empirical work in this framework makes clear that they have included in their bivariate VAR some proxy for the level and the slope of the term structure, but they have omitted the curvature. Interest rate rules, which feature (very) persistent policy rates responding to central bank's perceptions of (expected) inflation and output gaps (Taylor,1993, Clarida, Gali and Gertler, 1998, 1999, 2000) not only track the data well but are also capable of explaining the high inflation in the seventies in terms of an accommodating behaviour towards inflation in the pre-Volcker era.

Interestingly, Fuhrer(1996) uses a simple Taylor-rule type reaction function, the expectations model and reduced-form equations for output and inflation to solve for the reaction function coefficients that delivers long-term rates consistent with the Expectations Theory. He finds that modest and smoothly evolving time-variation in the reaction functions parameters is sufficient to reconcile the expectations model with the long-bond data. Favero(2002) extends Fuhrer framework to derive standard errors for long-term rates consistent with the ET. Our approach of extending the VAR framework is also related to recent work by Roush(2003). Roush considers a VAR model with macro and financial variables to show that the expectations theory of the term structure holds conditional on an exogenous change in monetary policy. The paper adds to the picture the important issue of identification but it does not provide evidence on the impact of the extension of the original CS information set on the outcome of the test for the unconditional validity of cross equation restrictions; moreover, the attention is limited to the within-sample evidence.

The bivariate CS approach has an implicit reaction function according to which the only determinant of policy rates are long-term rates, therefore we have a potential misspecification due to the omission of macroeconomic factors.

However, we think that our main contribution is not the augmentation of the original dimension of the VAR but the proposal of a new approach to test the ET based on information available in real time. To show our point, consider a cointegrated VAR framework, in which the original set of variables used by CS is extended by including a vector of variables $\mathbf{X}$. Such vector includes financial factors and macroeconomic variables. At each point in time we estimate, using the historically available information, the following model:

$$\begin{aligned}
\Delta r_t &= a_0 + a_1(L)\Delta r_{t-1} + a_2(L)S_{t-1} + a_3(L)\mathbf{X}_{t-1} + u_{1t} \\
S_t &= b_0 + b_1(L)\Delta r_{t-1} + b_2(L)S_{t-1} + b_3(L)\mathbf{X}_{t-1} + u_{2t} \\
\mathbf{X}_t &= c_0 + c_1(L)\Delta r_{t-1} + c_2(L)S_{t-1} + c_3(L)\mathbf{X}_{t-1} + u_{3t}
\end{aligned}$$

$$\begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} \backsim N\left[0, \Sigma\right]$$

We then simulate the estimated model forward, to obtain projection for all the relevant policy rates and to construct ET-consistent spreads as follows:

$$\hat{S}^{*}_{t,T} = \sum_{j=1}^{T-t-1} \gamma^{j} E[\Delta r_{t+j} \mid \Omega_t] \tag{1.12}$$

where, $E[\Delta r_{t+j} \mid \Omega_t]$ are the VAR-based projections for the future changes in policy rates, hence Ω_t is the information set used by the econometrician to predict on the basis of the estimated VAR model . Given this simulation based version of the ET consistent spread we can also construct a confidence interval around it. Confidence intervals around simulated series are usually constructed by adopting stochastic simulation techniques. In a standard stochastic simulation the model is simulated forward repeatedly for N draws of its stochastic components. In general, there are two sources of uncertainty: residuals and coefficient uncertainty. Residuals are drawn from a multivariate normal distribution $N\left(0, \hat{\Sigma}\right)$ where $\hat{\Sigma}$ is the estimated variance-covariance matrix of residuals of our VAR. Similarly, VAR coefficients are drawn from a multivariate normal distribution with the vector mean given by the point estimates of coefficient and the variance-covariance matrix given by the parameters' variance-covariance matrix. However, the confidence interval constructed by allowing for residuals and coefficient uncertainty will be a confidence interval for the evolution of $\sum_{j=1}^{T-t-1} \gamma^{j}[\Delta r_{t+j} \mid \Omega_t]$ which is very different, and certainly larger, than a confidence interval for $\hat{S}^{*}_{t,T} = \sum_{j=1}^{T-t-1} \gamma^{j} E[\Delta r_{t+j} \mid \Omega_t]$[4] However, it is immediate to construct bounds for $\hat{S}^{*}_{t,T}$ by performing the stochastic simulation allowing only for coefficients uncertainty. While future realized policy rates are affected both by parameters uncertainty and shocks, future expected policy rates are not affected by shocks, hence the only source of uncertainty for the ET consistent spread is parameters' uncertainty. ET consistent yields are calculated applying equation (3.10) to each

[4] We thank a referee for pointing this out. In fact, bounds constructed by allowing both for residuals and coefficients uncertainty could be thought of as a simulation equivalent of the volatility bounds proposed by Shiller(1979).

of the N simulated paths of future expected short-term rates: among these, the 0.5th, 0.05th, and 0.95th quantiles represent respectively the median ET-consistent yield and its 90% confidence bounds. The estimation window is then enlarged by one observation and simulation horizon is shifted one period ahead and the same steps are repeated.

Importantly, in implementing our procedure the econometrician uses the same information available to market participants in real-time. Future policy rates at time t are constructed using information available in real time for parameters estimation and forward projection of the model. Point forecasts and their confidence bounds define a region inside which the actual long term rates should lie if the ET holds.

Interestingly, by simulating the VAR coefficients from a multivariate normal distribution we treat parameters as random and the data fixed. This is obviously different from the classical bootstrap or Monte Carlo methods that simulate data for fixed parameters. This might give rise to a Bayesian interpretation of our paper, according to which our relevant distribution is the posterior for the relevant parameters when the prior is non-informative and the sample error covariance matrix is assumed to coincide with the true distribution. Importantly, in this intepretation the flat prior imposes stationarity of the VAR[5]. This is the reason why, as in Campbell-Shiller, we always adopt a stationary representation of our system, in which the variables are transformed either by taking first differences or by taking stationary linear combinations of non-stationary variables (cointegrating vectors)[6]. As a result of our specification choice very few (less than one per cent) of the simulated VAR long-run matrices contain one eigenvalue that lie on or outside the unit circle, we discard these simulations before constructing the relevant distribution[7]. Using a full Bayesian framework with appropriate specification of prior distributions is on our agenda for research.

By combining (1.4) and (3.10) , we have:

[5] We thank (again) an anonymous referee for pointing out this interpretation to us.

[6] For example, our VAR does never contain the level of the term structure , which clearly a very persistent variable, but rather the spread between the level and the intercept which we interpret as cointegrating vector with parameters (-1,1). The intercept is then considered in first differences.

[7] In our simulations we also impose the constraint that nominal yields are non negative.

$$S_{t,T} = \sum_{j=1}^{T-t-1} \gamma^j E_t[\Delta r_{t+j} \mid I_t] + E[\Phi_{t,T} \mid I_t] \quad (1.13)$$

$$\widehat{S^*_{t,T}} = \sum_{j=1}^{T-t-1} \gamma^j E_t[\Delta r_{t+j} \mid \Omega_t] \quad (1.14)$$

$$S_{t,T} - \widehat{S^*_{t,T}} = \left(\sum_{j=1}^{T-t-1} \gamma^j E_t[\Delta r_{t+j} \mid I_t] - \sum_{j=1}^{T-t-1} \gamma^j E_t[\Delta r_{t+j} \mid \Omega_t] \right) + E[\Phi_{t,T} \mid I_t]$$

$$S_{t,T} - \widehat{S^*_{t,T}} = \xi_t + E[\Phi_{t,T} \mid I_t] \quad (1.15)$$

Equation (1.15) makes clear that deviation of $S_{t,T}$ from $\widehat{S^*_{t,T}}$ can be explained by the effect of the risk premia or by differences between model based forecasts, which are derived by using the information set used by the econometrician Ω_t, and agents' expectations, which are formed given the information set I_t , unknown to the econometrician. Under the assumption that the first term is negligible, (statistically) significant deviations of $S_{t,T}$ from $\widehat{S^*_{t,T}}$ do offer a measurable counterpart of the risk premium.

1.3 The Empirical Evidence

We shall present our empirical evidence in three sub-sections. The first sections discusses our data-set, and our choice of sample for estimation and simulation, the second section presents the replica of the CS procedure on our data-set and an application of our simulation based procedure on the CS specification, while the third section illustrates the extension of the original specification to include financial factors and macroeconomic variables.

1.3.1 The data-set

Our basic data set consists of a set of zero-coupon equivalent US yields, provided by Brousseau, V. and B. Sahel (1999). In particular, we consider data on zero-coupon equivalent yields for US data measured at the following maturities[8]:

1-month, 2-month, 3-month, 6-month, 9-month, 1-year, 2-year, 3-year, 5-year, 7-year, 10-year.

[8] The data were kindly made available by the ECB, and they are posted on Favero's website at the following address: http:/www.igier.uni-bocconi.it/personal/favero
in the section working papers

From this data set we construct financial factors by estimating at each point of our time series t, by non-linear least squares, on the cross-section of eleven yields, the following Nelson-Siegel model:

$$y_{t,t+k} = L_t + SL_t \frac{1 - \exp\left(-\frac{k}{\tau_1}\right)}{\frac{k}{\tau_1}} + C_t \left(\frac{1 - \exp\left(-\frac{k}{\tau_1}\right)}{\frac{k}{\tau_1}} - \exp\left(-\frac{k}{\tau_1}\right) \right) \tag{1.16}$$

which is implicit in the instantaneous forward curve:

$$f_{tk} = L_t + SL_t \exp\left(-\frac{k}{\tau_1}\right) + C_t \frac{k}{\tau_1} \exp\left(-\frac{k}{\tau_1}\right) \tag{1.17}$$

The parameter τ_1 is kept constant over time[9], as this restriction decreases the volatility of the β parameters, making them more predictable in time. As discussed in Diebold and Li (2002) the above interpolant is very flexible and capable of accommodating several stylized facts on the term structure and its dynamics. In particular, L_t, SL_t, C_t, which are estimated as parameters in a cross-section of yields, can be interpreted as latent factors. L_t has a loading that does not decay to zero in the limit, while the loading on all the other parameters do so, therefore this parameter can be interpreted as the long-term factor, the level of the term-structure. The loading on SL_t is a function that starts at 1 and decays monotonically towards zero; it may be viewed a short-term factor, the slope of the term structure. In fact, $r_t^{rf} = L_t + SL_t$ is the limit when k goes to zero of the spot and the forward interpolant. We naturally interpret r_t^{rf} as the risk-free rate. Obviously SL_t, the slope of the yield curve, is nothing else than the minus the spread in Campbell-Shiller. C_t is a medium term factor, in the sense that their loading start at zero, increase and then decay to zero (at different speed). Such factor captures the curvature of the yield curve. In fact, Diebold and Li show that it tracks very well the difference between the sum of the shortest and the longest yield and twice the yield at a mid range (2-year maturity). The repeated estimation of loadings using a cross-section of yields at different maturities allows to construct a time-series for our factors. We report in Figure 1.1 the three factors, while Figure 1.2 shows the goodness of fit of the Nelson and Siegel interpolation for all yields considered in our sample. The extreme good performance of the Nelson-Siegel interpolant for our observed data shows that the fact that we have fitted the Nelson-Siegel model to zero coupon equivalent yields rather

[9] We restrict τ_1 at the value of 0.87, which is the median, over the time series, of the estimated value of τ_1 in a four parameter version of the Nelson-Siegel interpolant.

than to individual yields should not be a cause of concern for the problem at hand.

Note that the fact that we use zero-coupon equivalent yields has a relevant implication for the CS linearization, which should be applied taking the limit of the relevant formuale when γ approaches 1. In particular, we have:

$$\begin{aligned} R_{t,T} &= R^*_{t,T} + E[\Phi_{t,T} \mid I_t] = \lim_{\gamma \to 1} \frac{1-\gamma}{1-\gamma^{T-t}} \sum_{j=0}^{T-t-1} \gamma^j E[r_{t+j} \mid I_t] + E[\Phi_{t,T} \mid I_t] \\ &= \frac{1}{T-t} \sum_{j=0}^{T-t-1} E[r_{t+j} \mid I_t] + E[\Phi_{t,T} \mid I_t] \end{aligned} \qquad (1.18)$$

and, given that $R^*_{t,T} = \frac{1}{T-t} \sum_{j=0}^{T-t-1} E[r_{t+j} \mid I_t]$, we then have

$$S^*_{t,T} = R^*_{t,T} - r_t = \sum_{j=1}^{T-t} \left(1 - \frac{j}{T-t}\right) E\Delta[r_{t+j} \mid I_t]$$

Our empirical analysis will be based on a simulation sample starting at the beginning of the eighties. One of the main points of our paper is to construct expected future policy rates by considering explicitly the central bank reaction functions, so we have chosen the initial date of the sample for simulation to concentrate on an era of homogenous monetary policy, i.e. the Volcker-Greenspan era. In fact, there is ample empirical evidence that, from the beginning of the eighties onward, the Fed engaged in interest rate targeting and that the behaviour of policy rates can be successfully described by a Taylor rule. The traditional argument of a Taylor rule are expected inflation and some measure of the output gap. Our framework for simulating policy rates is geared to mimic the decisions of agents in real-time. Orphanides (2001) has shown that data revisions could generate misleading inference. For this reason, as suggested by Evans(2003), we consider as macroeconomic factors variables which are not subject to revision: the CPI inflation and unemployment rate.

We presents our empirical evidence in three parts: a replica on our data-set of the original Campbell-Shiller results, an application of our simulation based procedure on the CS specification, the extension of the original specification to include financial factors and macroeconomic variables.

1.3.2 Testing the ET with a bivariate VAR

The discussion of the measurement of financial factors makes clear that the closest model to CS original specification in our framework is the following:

$$\begin{bmatrix} \Delta r_t^{rf} \\ S_t \end{bmatrix} = A(L) \begin{bmatrix} \Delta r_{t-1}^{rf} \\ S_{t-1} \end{bmatrix} + u_t \tag{1.19}$$

where $r_t^{rf} = L_t + SL_t$,and $S_t = -SL_t$.Our specification differs from CS in that they take the one-month rate as the short term rate and the yield to maturity on 10-year bonds as the long term rate. Interestingly the level factor,L_t, is the asymptote of the term structure, hence cross equation restrictions on the VAR hold exactly for the spread constructed by using this factor while they are just approximate for the spread constructed using a 10-year yield [10] We also estimate our model recursively, allowing for a smooth evolving path in the estimated coefficients. This procedure might capture historical shifts in market perceptions of the policy target for inflation, which have been shown (Kozicki and Tinsley, 2001) to be important to achieve a satisfactory specification of agents' expectations. We report the results of the application of the CS testing methodology, based on a four-lag VAR, in Figure 1.3. Figure 1.3 reports the results of the test for the ET cross-equation restrictions, which is conducted recursively after using the sample 1974:4 1991:12 for initialization. The ET restrictions are consistently rejected, however,as in the original work of CS the actual spread has a correlation of with the spread obtained by imposing the invalid restrictions of .96. this is the evidence that leads Campbell and Shiller to conclude that "... deviations from the present value model for bonds are transitory...", however no measurement of the risk premium is explicitly provided by the two authors.

We report in Figure 1.4 the results of our simulation based test of the ET. We use our model to simulate ET consistent 10-year yields to maturity and their associated confidence intervals. Figure 1.4 ET consistent yields to maturity along with their associated confidence interval and the actual yields. Under the null of the ET the observed yields should fall within the bounds. In fact, the actual yields lie consistently above the simulated ones, but they are outside the 90 per cent confidence intervals, constructed under the null of the ET, only in a short subsample covering the period 1991-1994. Interestingly, a positive risk difference between actual and simulated yields is what we should observe in the presence of risk premium, when the impact of the difference between the

[10] As a matter of fact we have tested that for simulation based on our VAR specification ten year is sufficiently far in the future to give a good approximation of infinite.

information sets used by the agents and the econometrician is negligible. Overall, we attribute the difference between the results of our simulation based methodology and the traditional CS to the fact that the tests for the cross-equations restrictions understates the uncertainty faced by the agents in real time and therefore uses a too tight statistical criterion. Our evidence of non-rejection of the EH is consistent with the evidence proposed by CS of the very high correlation between the actual spread and the spread obtained by simulating imposing the restrictions (rejected by the Wald test). Our results confirm the much less strong evidence against the EH generated by models in which expectations are explicitly derived rather than taking the ex-post realized returns as a proxy for ex-ante expected returns. Interestingly, Bekaert and Hodrick (2000) find the same results from a different perspective: the use of small sample distribution of the relevant tests in VAR models leads to much less strong evidence against the ET.

We believe that it is important to assess this first set of results against those obtained by enlarging the information set of the VAR following the available empirical evidence form studies on the term structure and on the empirical analysis of monetary policy. In particular, the difference between actual and simulated rates is sizeable when significant and we think that it would be interesting to see how this distance is affected by the enlargement of the information set which we shall implement in the next section.

1.3.3 Testing the ET in a model with financial factors and macroeconomic variables

Our VAR with financial factor and macroeconomic variables takes the following specification:

$$\begin{bmatrix} \Delta r_t^{rf} \\ -S_t \\ C_t \\ \pi_t \\ UN_t \end{bmatrix} = A(L) \begin{bmatrix} \Delta r_{t-1}^{rf} \\ -S_{t-1} \\ C_{t-1} \\ \pi_{t-1} \\ UN_{t-1} \end{bmatrix} + u_t \tag{1.20}$$

We consider the three factors obtained via the application of the Nelson-Siegel interpolant together with CPI inflation, π_t, and the unemployment rate,

UN_t, which are our proxies for the variables normally entered as arguments of Taylor rules. Importantly, our macroeconomic variables are not subject to revision, consistently with our intention of using the model to replicate the decision process of agents in real time. As in the VAR with financial factors our representation is stationary and it allows for the cointegrating relationship which constitute a necessary condition for the ET to hold, being also consistent with the presence of a stationary risk premium[11]. Estimation is conducted on the same sample with the two variables VAR and, on the basis of the traditional lag selection criteria, we adopt a VAR of length two.[12] The results of the recursive within sample test and of the simulation based out-of-sample procedure are reported respectively in Figure 1.5 and Figure 1.6. The results of the Wald tests are very similar to those obtained in the basic model. However, the enlargement of the information set generates some notable modification in the simulation based procedure. In fact, the difference simulated yields get much closer to actual yields and there no evidence of violation of the ET. The peak in the differences between observed yields and simulated yields under the null of ET is in 1994, a period which has been widely cited in the literature as featuring an episode of "inflation scare" (see, for example, Rudebusch,1998). We interpret these results as evidence for the importance of the VAR enlargement to achieve a better identification of the expectations for the future path of the financial and macroeconomic variables relevant to determine monetary policy.

1.4 Robustness

The results on the size and the significance of risk premium delivered by our simulation based approach call for some robustness analysis. In particular, we want to make sure that our sample initialization is not inappropriate in that our initial VAR estimates are not contaminated by large residuals. In fact, after the Volcker disinflation, the volatility of macroeconomic variables has decreased remarkably in the eighties. We conduct our robustness check by concentrating on our five variable VAR specification, by considering as a benchmark the recursive estimation approach with initial sample 1974:6-1991:12 discussed in the previous section and by considering as an alternative estimation strategy a rolling estimation with initialization 1974:6 1991:12 and a fixed window of 210 observations. The alternative estimation method is chosen to evaluate the impact of our choice of initialization for the recursive estimation. In fact, the last sample for our rolling estimation approach is 1984:6-2001:12 and covers a very different period from the initial one in terms of (unconditional) volatility of all variables included in the

[11] The trace statistics for the null of at most four cointegrating vectors yieded an observed values of 6.35, for the estimation on the full sample and of 5.2 for the estimation on the shortest sample used in the recursive approach, while the five per cent critical value is 3.76 (We allowed for a constant restricted to belong to the cointegrating vector)

[12] The lag length criteria do not uniformly favour two lags for all possible sample splits. So we have analyzed the robustness of our results to the adoption of a four-lags VAR. The evidence, available upon request, shows that moving from a lag length of two to a lag length of four leaves our results unaltered.

VAR. Moreover, our rolling estimation could also provide evidence against the potential objection that some estimates (see, for example, Bernanke-Mihov(1998)) suggest that the starting period of the Volcker Greenspan era should be located at the beginning of the 1984, and simulation and tests based on post 1984 data could be different from those based on pre 1984 data.

We find the results of the application of the Wald tests and of the simulation based procedure, reported in Figures 1.7 and 1.8, interesting.

The uniform rejection of the theory obtained by the recursive approach based on the initialization on the large sample is not confirmed by the rolling approach, which does not lead to rejection of the theory for an estimation sample of 210 observations ending after the end of 1999. Very differently, the results of the simulation based approach in the five variables VAR are very robust to the two different estimation strategies. We report in Figure 1.8 the difference between actual 10-year yields and 10-year yields simulated under the null of the ET, obtained by projections based on rolling and recursive estimation for the five variables VAR and the two variables VAR. The results derived using the five factor models are very robust to the choice of the rolling and recursive estimation techniques, delivering differences that reach their peaks during the inflation scare of 1994. The results from the two variables VAR are instead sensitive to the estimation technique. In this case the rolling method delivers series which fluctuate at a level consistently lower than the recursive technique and closer to the series obtained from the five variables VAR. This evidence can be naturally interpreted as indicating mis-specification caused by omitted variables in the more parsimonious model. Interestingly, the results from the five variables VAR are consistent with the evidence, originally reported in CS, that the correlation between the actual spread and the spread obtained under the null of EH is very high even when the null is rejected. Our interpretation of these facts is that the uncertainty faced by the agents in simulating the model to obtain path for the relevant variables to forecast monetary policy is rather stable in a sufficiently parameterized model, even if the coefficients in the estimated VAR do vary over time.

1.5 What have we learned? A discussion of our results and their relation to the literature

In this paper we have simulated the real time decision of agents who forecast policy rates by projecting forward a model including financial factors and macro variables to generate long-term rates consistent with the expectations theory along with a confidence interval reflecting the uncertainty associated to out-of-sample forecasting. Our evidence shows that, for different specifications of the information set, the observed long-term yields are, with very few exceptions, contained in the confidence interval generated by our model. Our procedure delivers an observable counterpart of the deviation of the long-term rates from those consistent with the ET. Upon significance of such deviations we can interpret this variable as a proxy for risk premium under the null hypothesis that model based forecasts are not different from agents' expectations. Our empirical results show that a better specification of the VAR used to forecast future monetary policy delivers more credible estimates of the risk premium.

The standard response in finance to the empirical rejection of the Expectations Theory has been modelling the term structure based on the assumption that there are no riskless arbitrage opportunities among bonds of various maturities. The standard model is based on three components: a transition equation for the state vector relevant for pricing bonds, made traditionally of latent factors, an equation which defines the process for the risk-free one-period rate and a relation which associates the risk premium with shocks to the state vector, defined as a linear function of the state of the economy. In such structure, the price of a j-period nominal bond is a linear function of the factors. Unobservable factors and coefficients in the bond pricing functions are jointly estimated by maximum likelihood methods (see, for example, Chen and Scott(1993)). This type of models usually provides a very good within sample fit of different yields but do not perform well in forecasting. Duffee(2002) shows that the forecasts produce by no-arbitrage models with latent factors do not outperform the random walk model.

Recently the no-arbitrage approach has been extended to include some observable macroeconomic factors in the state vector and to explicit allow for a Taylor-rule type of specification for the risk-free one period rate. Ang and Piazzesi(2002) and Ang, Piazzesi and Wei(2003) show that the forecasting performance of a VAR improves when no-arbitrage restrictions are imposed and that augmenting non-observable factors models with observable macroeconomic factors clearly improves the forecasting performance. Hordahl et al.(2003) and Rudebusch and Wu(2003) use a small scale macro model to interpret and parameterize the state vector; forecasting performance is improved and models have also some success in accounting for the empirical failure of the Expectations Theory.

No-arbitrage models with observable factors feature a complicated parameterization and cannot accommodate time variation in the parameters of the state vector relevant for pricing bonds. Within this approach, the failure of ET is entirely abscribed to the presence of a time-varying risk premium, which is modelled as a linear function of the state of the economy. There is a lot in common between the latest developments of the no-arbitrage approach and the approach to the term structure proposed in our paper. We share the view on the importance of augmenting the information set with macroeconomic and financial factors to model the yield curve but we concentrate directly on a VAR model for all the relevant factors and we derive risk premium as a residual. The main cost of our approach is that our derived proxy for the risk premium is valid only under the assumption that the difference between the agents information set and the econometrician' information set does not lead to different future projected short-term rates. The main advantage is a much more parsimonious (and linear) parameterization, which easily accommodates time-variation in the parameters describing the state vector relevant for pricing bonds. Our findings suggests that the importance of fluctuations of risk premia in explaining the deviation from the ET might be reduced when some forecasting model for short-term rates is adopted and a proper evaluation of uncertainty associated to policy rates forecast is considered. We believe that improving the forecasting model for policy rates within a no-arbitrage approach is an important step to assess the relative weight of forecasting errors and risk premia in explaining deviations from the Expectations Theory. This is on our agenda for future research.

1.6 References

Ang A. and M.Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, Journal of Monetary Economics 50, 745-787.

Ang A., Piazzesi M. and M.Wei, 2003, What does the yield curve tell us about GDP growth?, mimeo, Columbia University

Balduzzi P., Bertola G., and S.Foresi, 1997, A model of target changes and the term structure of interest rates, Journal of Monetary Economics 24, 371-399.

Bernanke B. S. and I. Mihov,1998, Measuring monetary policy, Quarterly Journal of Economics 113 3, 869-902.

Campbell J.,1995, Some Lessons from the Yield Curve, Journal of Economic Perspectives 93, 129-152.

Campbell J., Lo A. and C. MacKinlay, 1997, The Econometrics of Financial Markets, Princeton University Press.

Campbell,J., and R. Shiller, Cointegration and Tests of Present Value Models, Journal of Political Economy 95, 1987, 1062-1088.

Chen R.R. and L. Scott, 1993, Maximum Likelihood estimation for a multi-facor equilibrium model of the term structure of interest rates, Journal of Fixed Income 3, 14-31.

Clarida R., J. Gali and M. Gertler,1998, Monetary policy rules in practice: some international evidence, European Economic Review 42, 1033-1067.

Clarida R., J. Gali and M. Gertler,1999, The science of monetary policy: A new-Keynesian perspective, Journal of Economic Literature XXXVII 4, 1661-1707.

Clarida R., J. Gali and M. Gertler, 2000, Monetary policy rules and macroeconomic stability: evidence and some theory, The Quarterly Journal of Economics 115 1, 147-180.

Cochrane J., 2001, Asset Pricing , Princeton University Press.

Diebold F.X. and C. Li, 2002, Forecasting the Term Structure of Government Bond Yields , mimeo, University of Pennsylvania.

Diebold F.X., G.D. Rudebusch and S.B.Aruoba, 2003, The Macroeconomy and the yield curve, mimeo, University of Pennsylvania.

Duffee G.R.,2002, Term Premia and Interest Rate Forecast in Affine Models, Journal of Finance 57, 1, 405-443.

Elton E.J., 1999, Expected return, realized return and asset pricing tests, Journal of Finance 54, 1199-1220.

Evans C. , 2003, Real-time Taylor rules and the federal funds futures market, Economic Perspectives Federal Reserve Bank of Chicago, 44-55.

Fama E., and R.R. Bliss ,1987, The Information in Long-Maturity Forward Rates, American Economic Review 77, 680-692.

Favero C.A.,2002, Taylor rules and the Term Structure, IGIER Working Paper 195.

Fuhrer J.C.,1996, Monetary Policy Shifts and Long-Term Interest Rates, Quarterly Journal of Economics 111 4,1183-1209.

Hordahl P., O.Tristani and D.Vestin,2003, A joint econometric model of macroeconomic and term structure dynamics, mimeo ECB.

Kozicki S. and P.A. Tinsley, 2001, Shifting endpoints in the term structure of interest rates, Journal of Monetary Economics 47, 613-652.

Johansen S.,1995, Likelihood-based inference in cointegrated vector auto-regressive models, Oxford University Press, Oxford.

Mankiw N.G. and J.Miron,1986, The changing behaviour of the term structure of interest rates, The Quarterly Journal of Economics, 101, 211-221.

McCallum,1994, Monetary Policy and the Term Structure of Interest Rates, NBER Working Paper 4938.

Nelson C.R. and A.F. Siegel,1987, Parsimonious modelling of yield curves, Journal of Business 60, 473-489.

Orphanides A.,2000, The quest for prosperity without inflation, European Central Bank Working Paper 15.

Roberds W. and C.H. Whiteman,1996, Endogenous Term Premia and anomalies in the Term Structure of Interest Rates: Explaining the Predictability Smile , Federal Reserve Bank of Atlanta, Working Paper 96-11.

Roush J.,2003, Evidence uncovered: Long-term interest rates, Monetary Policy and the Expectations Theory , mimeo, Board of Governors of the Federal Reserve System.

Rudebusch G.,1998, Monetary Policy and the Term Structure of Interest Rates: An Overview of Some Recent Research. In Monetary Policy and the Term Structure of Interest Rates, eds. Riccardo Rovelli and Ignazio Angeloni, New York: St. Martin's Press, 263-271.

Rudebusch G.D G.D.,1995, Federal reserve interest rate targeting, rational expectations, and the term structure, Journal of Monetary Economics 35, 245-274.

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

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

Shiller,R.,Campbell,J., and Schoenholtz, 1983, K. Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates, Broking Papers in Economic Activity 1, 173-217.

Taylor J.B. ,1993, Discretion versus policy rules in practice, Carnegie-Rochester Conference Series on Public Policy 39, 195-214.

1.7 Figures

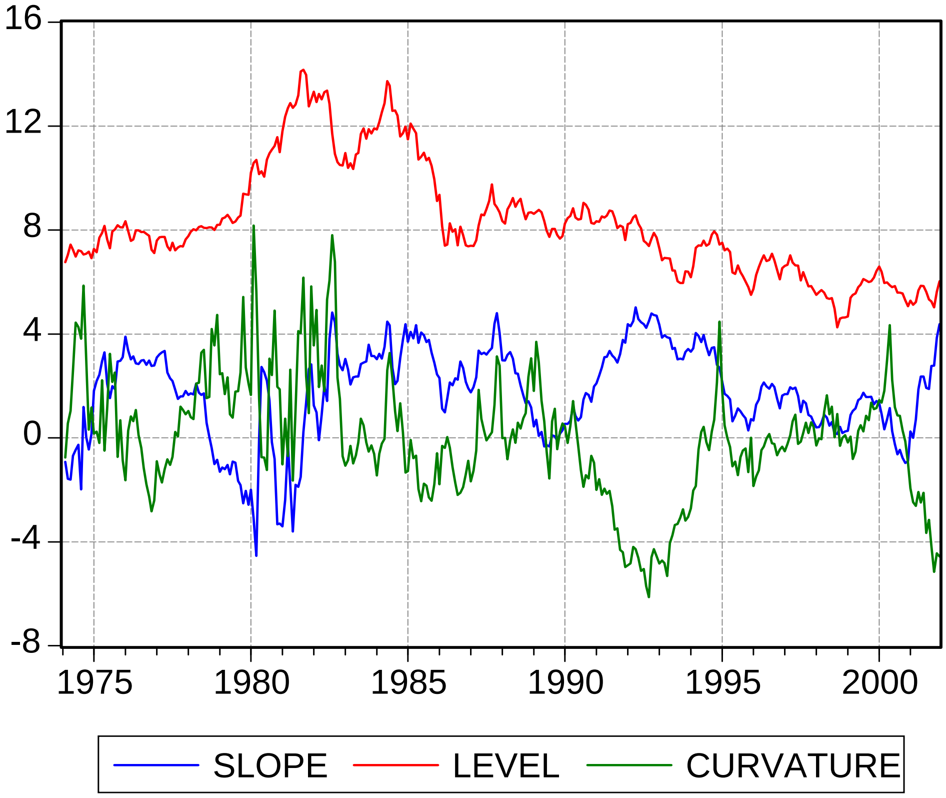

Figure 1.1: The time series of the three Nelson-Siegel factors for the US yield curve

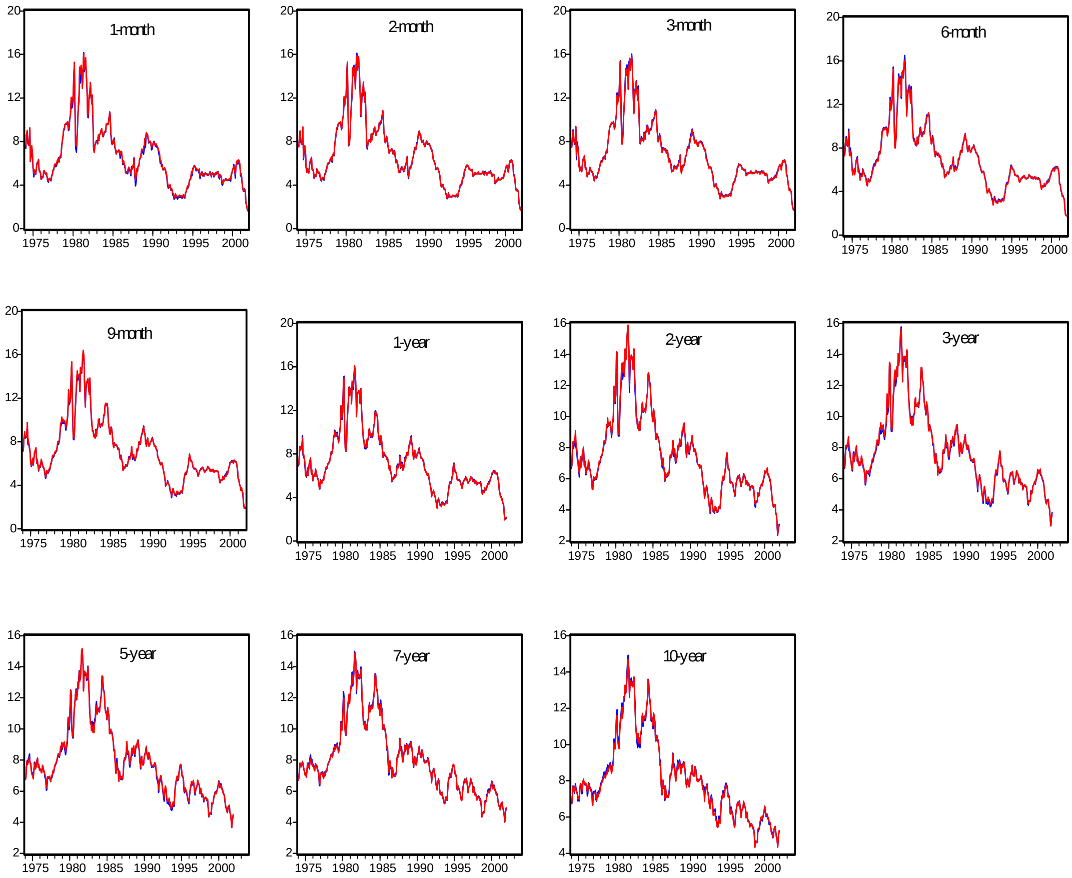

Figure 1.2: The time series of yields at different maturities and the Nelson-Siegel interpolants

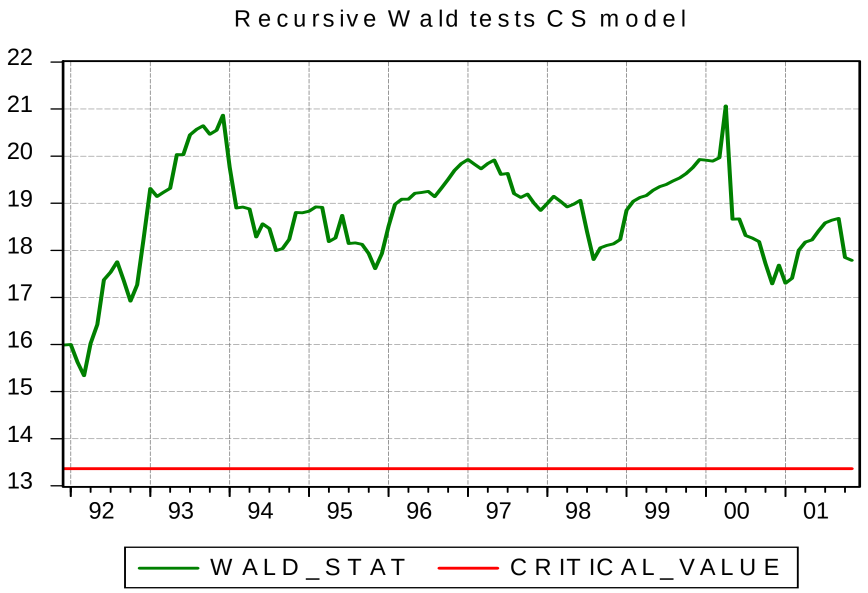

Figure 1.3: Recursive tests (and five per cent critical value) for the validity of the cross-equation restrictions implied by the Expectations Theory in a four-lags VAR with two financial factors (change in policy rates and slope of the yield curve, as in CS).

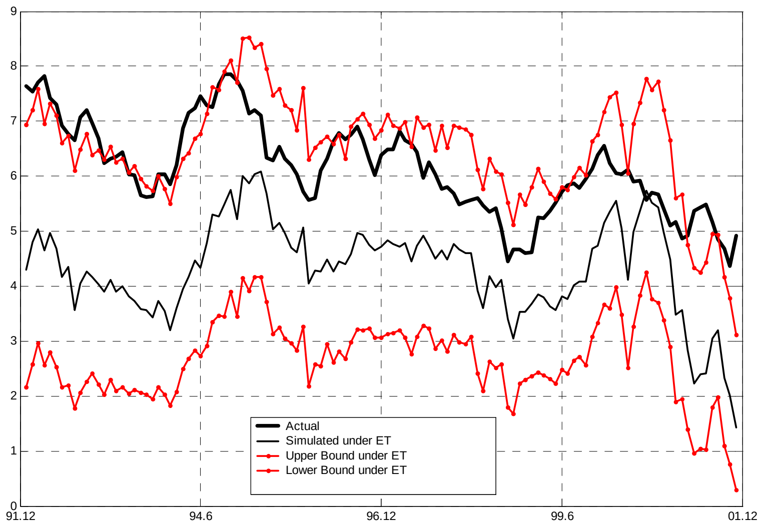

Figure 1.4: Simulated ET-consistent 10-year yields to maturity based on the CS model, with lower and upper bond of its 90% Confidence Interval

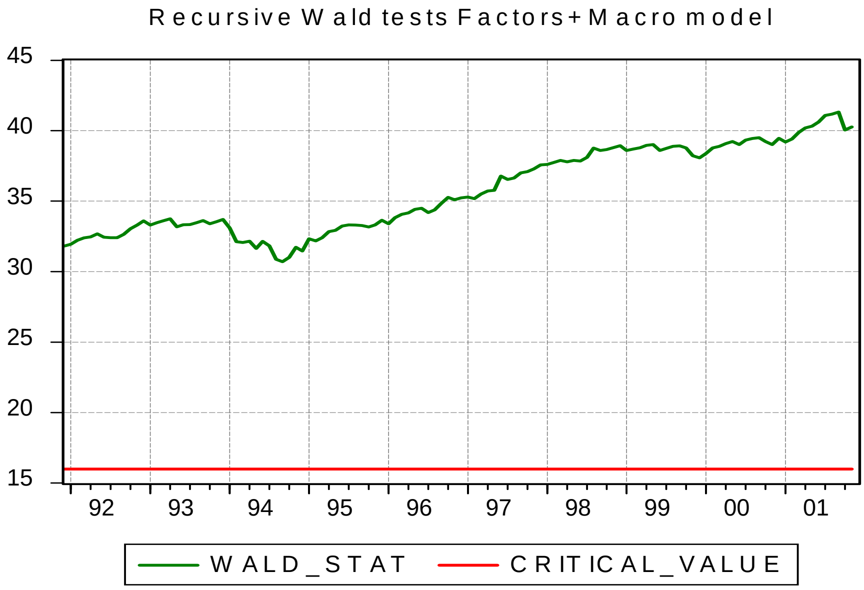

Figure 1.5: Recursive tests (and five per cent critical value) for the validity of the cross-equation restrictions implied by the Expectations Theory in a VAR with three financial factors and two macroeconomic variables.

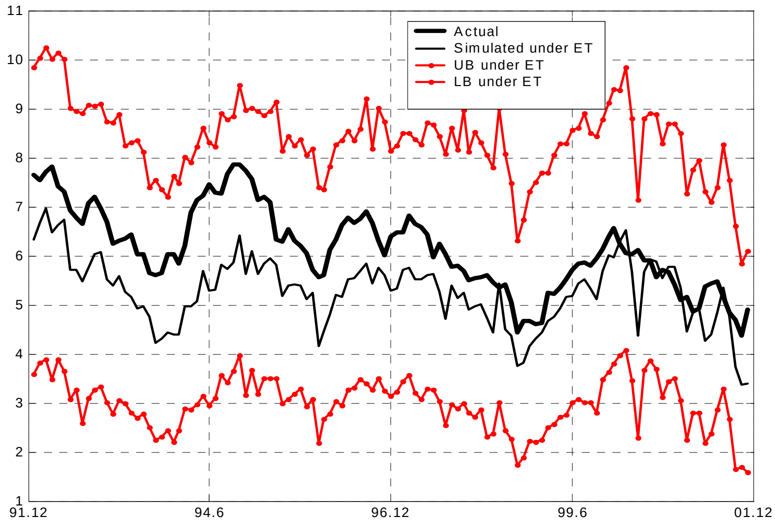

Figure 1.6: Simulated ET-consistent 10-year yields to maturity based on the model with financial factors and macroeconomic variables, with lower and upper bond of its 90% Confidence Interval

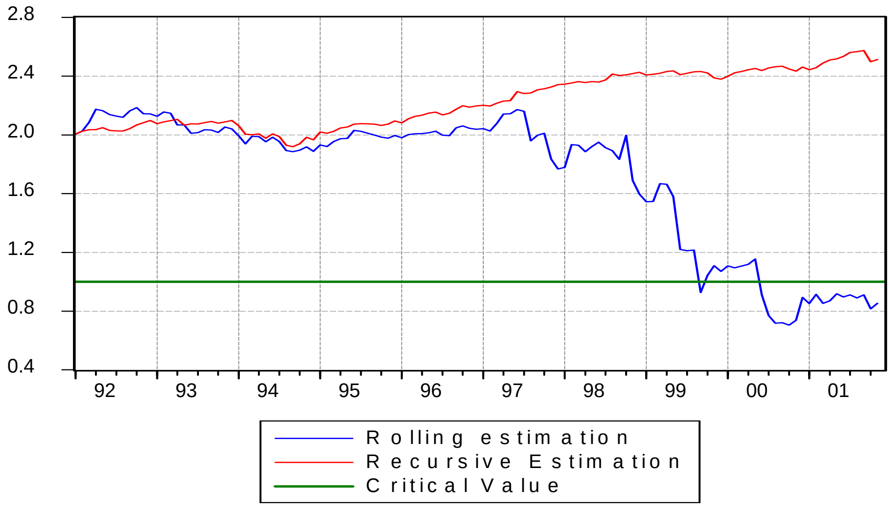

Figure 1.7: Wald tests for the EH restrictions on the VAR with financial factors and macroeconomic variables. The reported tests, scaled by their 95 per cent critical value, are recursively computed for all end sample points from 1992:1 to 2001:12. Initial sample points are chosen by twodifferent methods: Recursive estimation is based on anchoring the first observation to 1974:6 , Rollling estimation is results based on a rolling estimation with initialization 1974:6 1991:12 and a fixed window of 210 observations.

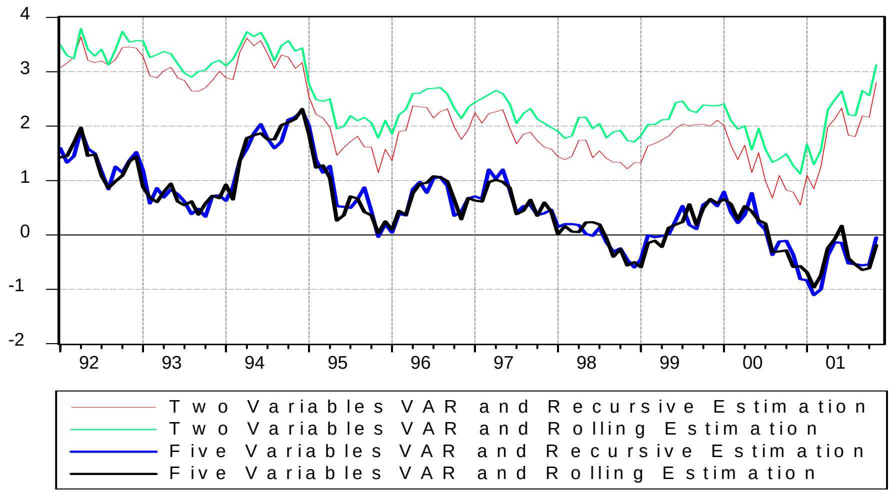

Figure 1.8: Time-series of the difference between actual and simulated yields under ET . Yields are simulated based respectively on recursive and rolling estimation of a five-variables VAR, and of a two-variables VAR

Chapter 2

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

2.1 Introduction

The[1] 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. 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) imposes the EH to disentangle expected from unexpected movements in interest rates.

[1] This is my job-market paper. I thank Carlo Favero for his guidance, and Fabio Canova, Massimiliano Marcellino, Nicola Scalzo, Ulf Söderström for their comments and suggestions.

However, the empirical evidence casts serious doubts on the appropriatness of using the EH for these forecasting and simulation excercises. Indeed, the EH has been widely tested, and almost invariably rejected. Examples are Fama (1984), Fama and Bliss (1987), Stambaugh (1988), Campbell and Shiller (1991), Campbell (1995), Backus et al. (2001), Beakert 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 those results still are 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. Once cast in this framework, the EH holds on average and can be safely and effectively used for forecasting and simulation.

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.

From a statistical point of view, modeling the EH with uncertainty may improve over the traditional way of imposing it exactly. As stressed out 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 from a time varying but stationary term premium and expectations errors. These deviations may be thought of as a stationary disturbance around the EH relation. When restrictions are derived from the EH, the stochasticity of the disturbance transfers in the restrictions which then become inherently fuzzy.

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 uncertainty. The Bayes factor provides evidence that the EH holds on average, i.e. after integrating out the effect of the deviations which may affect it in the short run. This finding suggests that the EH prior can be safely imposed on the data to perform simulation excercises. This result 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).

As data indicate clearly 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.

The use of the EH prior in forecasting 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.

2.2 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.

2.2.1 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 $r_t, r_{t+1}, ...$:

$$R_t = \frac{1-\gamma}{1-\gamma^T} \sum_{i=0}^{T-1} \gamma^i E_t\left(r_{t+i}\right) + TP_T. \tag{2.1}$$

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 (3.1) relates R_t to the present value of future short-term interest rates discounted by $\bar{R}$.

Rearranging (3.1) gives an expression involving the spread $S_t = R_t - r_t$ and the change in short-term rate $\Delta r_t = r_t - r_{t-1}$:

$$S_t = \sum_{i=1}^{T-1} \gamma^i E_t(\Delta r_{t+i}) + \gamma^T (R_t - TP_T) + TP_T. \tag{2.2}$$

As $T \to \infty$ this simplifies to

$$S_t = \sum_{i=1}^{\infty} \gamma^i E_t(\Delta r_{t+i}) + TP_\infty. \tag{2.3}$$

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

2.2.2 Expectations Hypothesis restrictions

Now, following Campbell and Shiller (1987), consider a VAR representation for Δr_t and S_t:

$$\begin{bmatrix} \Delta r_t \\ S_t \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} + \begin{bmatrix} a(L) & b(L) \\ c(L) & d(L) \end{bmatrix} \begin{bmatrix} \Delta r_{t-1} \\ S_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \tag{2.4}$$

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 (3.3) it is possible to derive the following set of p restrictions implied on the VAR in equation (2.4):

$$\begin{cases} a_j + c_j = 0, \ \forall\ j = 1...p \\ b_1 + d_1 = 1/\gamma \\ b_j + d_j = 0, \ \forall\ j = 2...p \end{cases}. \tag{2.5}$$

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

$$\alpha = \begin{bmatrix} a_1 & b_1 & ... & ... & a_p & b_p & k_1 & c_1 & d_1 & ... & ... & c_p & d_p & k_2 \end{bmatrix}', \tag{2.6}$$

the $2p$ restrictions in (2.5) can be compactly written as:

$$H\alpha = \mu_{EH}, \tag{2.7}$$

where

$$H = \left[\begin{array}{cccc} I_{2p} & \underset{2p\times 1}{0} & I_{2p} & \underset{2p\times 1}{0} \end{array} \right], \tag{2.8}$$

and

$$\boldsymbol{\mu}_{EH} = \left[\begin{array}{ccc} 0 & \frac{1}{\gamma} & \underset{1\times(2p-2)}{0} \end{array} \right]'. \tag{2.9}$$

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 (2.7) via a Wald test, and find strong rejection. As documented below in Section 4, the rejection result is confirmed also using more recent data. However, Campbell and Shiller (1987) show that imposing the rejected restrictions on the VAR in equation (2.4) does not yield 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.

2.3 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 Subection 3.1 and 3.2 we introduce the noise around the EH restrictions and discuss about its economic interpretation. Subection

3.3 derives the prior from the EH, Subection 3.4 derives posteriors and the marginal likelihood. Subection 3.5 extends the framework to priors on second order parameters. Subection 3.6 describes Bayesian inference.

2.3.1 Adding uncertainty

By definition any economic theory is a simplification of reality, and as such it cannot 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 (2.7). Formally, let the uncertainty introduced by this noise be measured by the parameter σ. The resulting set of stochastic constraints is:

$$H\alpha \sim N\left(\mu_{EH}, \sigma I_p\right). \tag{2.10}$$

The hyperparameter σ is the tightness of the restrictions. By controlling this hyperparameter the researcher 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 (2.4) restricted according equation (2.7). When $\sigma \to \infty$ the uncertainty about the restrictions is so high that they are not binding, and this provides a model equivalent to the VAR in equation (2.4) 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.

2.3.2 Interpreting noise

In this subsection we discuss why the set of restrictions in equation (2.7) may hold only up to some noise, as shown in equation (2.10). 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 bonds returns at all maturities. Finally, the EH relation in equation (3.1) 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 (3.1) 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 (3.1) 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 transfers in 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 (3.1) 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 wich is not necessarily the same used by the agents. Again, these sources of noise transfer in the restrictions making them inherently fuzzy.

2.3.3 A prior from the EH

The set of restriction (2.10) can be thought of in a Bayesian perspective as a prior on the coefficients of the VAR in equation (2.4). Define $y = vec([\Delta r \;\; S])$, $\Xi = I_2 \otimes [\Delta r_{-1} \; S_{-1} \; ... \; \Delta r_{-p} \; S_{-p} \; 1]$, and $\varepsilon = vec([u_1 \; u_2])$. 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:

$$y = \Xi\alpha + \varepsilon. \tag{2.11}$$

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 (2.4), 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 (2.11) and the restrictions in (2.10) as the $EH(\sigma)$ model:

$$\begin{cases} y = \Xi\alpha + \varepsilon \\ H\alpha \sim N\left(\mu_{EH}, \sigma I_p\right) \end{cases} . \tag{2.12}$$

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:

$$\begin{cases} y = \Xi\alpha + \varepsilon \\ \alpha \sim N\left(\alpha_{EH}, \Sigma_{EH}\right) \end{cases} , \tag{2.13}$$

where

$$\alpha_{EH} = \left[\underset{1\times(2p+1)}{0} \quad 0 \quad 1/\gamma \quad \underset{1\times(2p-1)}{0} \right]', \tag{2.14}$$

and

$$\Sigma_{EH} = \begin{bmatrix} \delta I_{2p} & \underset{2p\times1}{0} & -\delta I_{2p} & \underset{2p\times1}{0} \\ \underset{1\times2p}{0} & \delta & \underset{1\times2p}{0} & 0 \\ -\delta I_{2p} & \underset{2p\times1}{0} & (\sigma+\delta)I_{2p} & \underset{2p\times1}{0} \\ \underset{1\times2p}{0} & 0 & \underset{1\times2p}{0} & \delta \end{bmatrix}. \tag{2.15}$$

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 uninformativeness[2].

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 (2.15):

$$Corr(\alpha) = \begin{bmatrix} I_{2p} & \underset{2p\times1}{0} & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} I_{2p} & \underset{2p\times1}{0} \\ \underset{1\times p}{0} & 1 & \underset{1\times2p}{0} & 0 \\ \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} I_{2p} & \underset{2p\times1}{0} & I_{2p} & \underset{2p\times1}{0} \\ \underset{1\times2p}{0} & 0 & \underset{1\times2p}{0} & 1 \end{bmatrix}. \tag{2.16}$$

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

Notice that depending on the value of the tightness parameter σ we move from the exact restrictions case ($\sigma = 0$) to the unrestricted VAR ($\sigma \to \infty$). 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 $\sigma \to \infty$, the correlation across the relevant couples of coefficients goes to zero and the correlation matrix approaches that of a VAR without cross equation restrictions.

2.3.4 Posteriors and marginal likelihood

The $EH(\sigma)$ model described in equation (2.13) features exact closed form solutions for the coefficients posterior densities and the marginal likelihood. The posterior density is normal with variance:

$$\Sigma_{\bar{\alpha}} = \left(\Sigma_{EH}^{-1} + \Xi'\Omega^{-1}\Xi\right)^{-1}, \tag{2.17}$$

and mean

$$\bar{\alpha} = \Sigma_{\bar{\alpha}} \left(\Sigma_{EH}^{-1}\alpha_{EH} + \Xi'\Omega^{-1}y\right). \tag{2.18}$$

The marginal likelihood is:

$$p(y) = (2\pi)^{-T} \left|\Omega\right|^{-1/2} \left|\Sigma_{\bar{\alpha}}\right|^{1/2} \left|\Sigma_{EH}\right|^{-1/2} \exp\left(-Q/2\right), \tag{2.19}$$

where

$$Q = y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + \alpha_{EH}{}'\Sigma_{EH}^{-1}\alpha_{EH}. \tag{2.20}$$

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

2.3.5 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 (2.17)-(2.20). More generally, we could specify a prior also on the variance matrix of the disturbances. A natural choice would be a Wishart prior:

$$\Sigma_u^{-1} \sim Wi(v_0, \Sigma_{u0}^{-1}). \tag{2.21}$$

Uninformativeness is achieved by setting $v_0 = 0$ and $\Sigma_{u0} = 0_{2\times 2}$. If we assume independency between α and Σ_u equations (2.21) and (2.13) constitute the so called independent Normal-Wishart prior, which requires simulation methods to compute posteriors and marginal likelihoods. For details see Appendix C, Geweke (2005), Koop (2003).

2.3.6 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 $\frac{pr(H_2|\mathbf{D})}{pr(H_1|\mathbf{D})} = \frac{pr(\mathbf{D}|H_2)}{pr(\mathbf{D}|H_1)}\frac{pr(H_2)}{pr(H_1)}$, so that the transformation is simply multiplication by

$$B_{21} = \frac{pr(\mathbf{D}|H_2)}{pr(\mathbf{D}|H_1)}. \tag{2.22}$$

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.

2.4 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 out 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 (2.4) under the stochastic constraints given by equation (2.10) for several different of 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. These deviations may be thought of as a stationary disturbance around the EH relation.

We apply our framework to US data. First, we gauge the appropriateness of the VAR in equation (2.4) 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.

2.4.1 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 (2.4) 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 Shwartz criterion (with a maximum lag length of 13). Residuals feature some outliers which can be 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. OLS estimates of the VAR in (2.4) are reported in the 3rd column of Table 2. The Wald test strongly rejects the set of restrictions in (2.7), consistently with Campbell and Shiller (1987). Notwithstanding the rejection of the restrictions, for future reference we also perform estimation of the VAR in equation (2.4) restricted according to the set of restrictions in equation (2.7). $FIML$ estimates of the restricted VAR are reported in the 1st column of Table 2.

2.4.2 Bayesian estimation

Now we turn to the Bayesian estimation of the $EH(\sigma)$ described by equation (2.13).

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 $EH(\infty)$[3]. As shown in Table 2, the $EH(0)$ is equivalent to the VAR in equation (2.4) restricted according to equation (3.3). 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 (2.4) 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 2.1. As is clear from the graph, the marginal likelihood is hump-shaped and features a neat 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 higest 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 them by simulation[4]. For details see appendix C. The marginal likelihood for this case is also reported in Figure 2.1. 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 keep using the prior with simple fixed covariance matrix of the errors.

---

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

[4] To 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

2.4.3 Bayes factors

The neat 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 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. Indeed, 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[5]. 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 by OLS the VAR, and then we use these OLS estimates to elicitate a prior for the competing model[6]. 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.

Results of this analysis are reported in Figure 2.2, which plots the Bayes factor (B_{21}) as a function of the EH prior tightness σ, together with the inconclusive region[7]. If we allow for very little noise, letting $\sigma \to 0$, the Bayes factor supports the $UVAR$. This is the common result of rejection. Indeed, 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

[5] Importantly, the Lindlay 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 Lindlay paradox effect this would work against and not in favour of it.

[6] The 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.

[7] Notice 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 2.1 and 2.2 as the Bayes factor is graphed in logs.

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.

2.4.4 Discussion

In this section we estimated a continuum of models ranging from 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 mean 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 excercises. 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 data indicate clearly 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.

2.5 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 rates. This means that using the EH as a prior allows to extract additional information from the term structure of interest rates.

2.5.1 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 2.3 plots the difference between the $in-det$ statistic proposed by Doan et al. (1984) obtained by using the EH prior and that obtained by using the simple unrestricted VAR described in equation (2.4). The difference in the $in-det$ statistic provides an overall assessment of the percentage gain in using the former model rather than the latter[8]. Positive values of the percentage difference imply a positive gain in 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 form 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 (2.4).

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 1-month horizon is about 2%, and at 12-month horizon is about 4%, for intermediate horizons it goes up to 7%.

[8] Specifically, 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 multipying by 100.

2.5.2 Forecasts comparisons

The analysis performed so far is intended mostly as an initial inspection at 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 open the question 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. Then, 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, i.e. 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 about the prior mean on the first lag of the variables which 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 (2.4), 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[9].

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 lenghts of the estimation window used for the forecasting excercise. 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 wider estimation windows improves the precision in estimating the model used to forecast but yields less 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 form 1.45 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 worse performance at the very long horizons when a 8 years estimation window is used.

Third, the Minnesota prior is on average the best model for predicting the spread. When using a 8 years 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 Minnesota produces significant and high gains, which may be up to 13.38 percent. Also in this case, there are nosignificant differences at the very short horizons.

[9] Reagarding 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.

2.5.3 Discussion

In this section we investigated whether using the EH as prior information may yield significant imporovements 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 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.

When the EH prior is compared to a more competitive benchmark, i.e. 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 Cambpell 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 random walk assumption has only a statistical justification, while the EH prior has been derived from economic theory.

2.6 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 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 appropriatness of using the EH for these forecasting and simulation excercises.

This paper 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.

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 uncertainty. This result explains both the common result of rejection, and the anomalous 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 excercises.

Using the EH as a prior allows to extract additional information from the term structure of interest rates.

Indeed, our forecasting excericise 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 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.

2.7 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 (2.4) as:

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

where the ˜ indicates that the variables are taken in deviations form their mean. Define A the coefficient matrix, v_t the vector of disturbances, and z_t as the vector containing $\Delta r_t,\ S_t$ 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:

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

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:

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

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.:

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

B. Derivation of the EH prior

Define $y = vec\left([\Delta r \quad S]\right)$, $\Xi = I_2 \otimes [\Delta r_{-1}\ S_{-1}\ ...\ \Delta r_{-p}\ S_{-p}\ 1]$, and $\varepsilon = vec\left([u_1\ u_2]\right)$. 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.:

$$\overbrace{\underset{2T\times 1}{\begin{bmatrix} \Delta r \\ S \end{bmatrix}}}^{y} = \overbrace{\underset{2T\times 2(2p+1)}{\begin{bmatrix} I_2 \otimes \underset{T\times(2p+1)}{X} \end{bmatrix}}}^{\Xi} * \underset{2(2p+1)\times 1}{\alpha} + \overbrace{\underset{2T\times 1}{\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}}}^{\varepsilon},$$

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

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:

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

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

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

where

$$H = \left[\begin{array}{cccc} I_{2p} & \underset{2p\times 1}{0} & I_{2p} & \underset{2p\times 1}{0} \end{array}\right],$$

and

$$\boldsymbol{\mu}_{EH} = \left[\begin{array}{ccc} 0 & \frac{1}{\gamma} & \underset{1\times(2p-2)}{0} \end{array}\right]'.$$

Adding the noise to each of the restrictions yields:

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

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:

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

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 arbitrarly high variance δ on the unrestricted coefficients.

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

with:

$$H_2 = \begin{bmatrix} I_{2p} & \underset{2p\times1}{0} & \underset{2p\times2p}{0} & \underset{2p\times1}{0} \\ \underset{1\times2p}{0} & 1 & \underset{1\times2p}{0} & 0 \\ I_{2p} & \underset{2p\times1}{0} & I_{2p} & \underset{2p\times1}{0} \\ \underset{1\times2p}{0} & 0 & \underset{1\times2p}{0} & 1 \end{bmatrix}, \ \mu_{2EH} = \begin{bmatrix} \underset{2p\times1}{0} \\ 0 \\ \underset{2p\times1}{\mu_{EH}} \\ 0 \end{bmatrix}, \ \Sigma_2 = \begin{bmatrix} \delta I_{2p} & \underset{2p\times1}{0} & \underset{2p\times2p}{0} & \underset{2p\times1}{0} \\ \underset{1\times2p}{0} & \delta & \underset{1\times2p}{0} & 0 \\ \underset{2p\times2p}{0} & \underset{2p\times1}{0} & \sigma I_{2p} & \underset{2p\times1}{0} \\ \underset{1\times2p}{0} & 0 & \underset{1\times2p}{0} & \delta \end{bmatrix}.$$

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

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

where

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

and

$$\Sigma_{EH} = H_2^{-1}\Sigma_2 H_2'^{-1} = \begin{bmatrix} \delta I_{2p} & \underset{2p\times1}{0} & -\delta I_{2p} & \underset{2p\times1}{0} \\ \underset{1\times2p}{0} & \delta & \underset{1\times2p}{0} & 0 \\ -\delta I_{2p} & \underset{2p\times1}{0} & (\sigma+\delta)I_{2p} & \underset{2p\times1}{0} \\ \underset{1\times2p}{0} & 0 & \underset{1\times2p}{0} & \delta \end{bmatrix}.$$

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

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

The correlation marix is:

$$Corr(\alpha) = \begin{bmatrix} I_{2p} & \underset{2p\times1}{0} & \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} I_{2p} & \underset{2p\times1}{0} \\ \underset{1\times p}{0} & 1 & \underset{1\times2p}{0} & 0 \\ \frac{-\sqrt{\delta}}{\sqrt{\sigma+\delta}} I_{2p} & \underset{2p\times1}{0} & I_{2p} & \underset{2p\times1}{0} \\ \underset{1\times2p}{0} & 0 & \underset{1\times2p}{0} & 1 \end{bmatrix}$$

C. Derivation of Posterior Densities and Marginal Likelihood.

Consider a generic VAR with a normal prior on the coefficients:

$$\begin{aligned} \underset{MT\times 1}{y} &= \underset{MT\times Mk}{\Xi}\ \underset{Mk\times 1}{\alpha} + \underset{MT\times 1}{\varepsilon}, \\ \alpha &\sim N(\alpha_0, \Sigma_{\alpha_0}), \\ e &\sim N(0, \Omega = \underset{MT\times MT}{\Sigma_u \otimes I_T}.), \end{aligned}$$

Here M is the number of equations, p is the number of lags included, $k = Mp + 1$ is the number of regressors and T is the sample size, while α_0 and Σ_{α_0} are the coefficients prior moments. The prior density is:

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

the likelihood is[10]:

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

a posterior density kernel is:

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

Now define[11]:

$$\Sigma_{\bar{\alpha}} = \left(\Sigma_{\alpha_0}^{-1} + \Xi'\Omega^{-1}\Xi\right)^{-1},$$

$$\bar{\alpha} = \Sigma_{\bar{\alpha}}\left(\Sigma_{\alpha_0}^{-1}\alpha_0 + \Xi'\Omega^{-1}y\right).$$

Using the above definitions and completing the square yields:

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

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

[11] Notice that:

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

$$= y'\Omega^{-1}y - y'\Omega^{-1}\Xi\alpha - \alpha'\Xi'\Omega^{-1}y + \alpha'\Xi'\Omega^{-1}\Xi\alpha + \alpha'\Sigma_{\alpha_0}^{-1}a - \alpha'\Sigma_{\alpha_0}^{-1}a_0 - \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 =$$

$$= y'\Omega^{-1}y - \left(y'\Omega^{-1}\Xi + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\right)\alpha - \alpha'\left(\Xi'\Omega^{-1}y + \Sigma_{\alpha_0}^{-1}a_0\right) + \alpha'\left(\Xi'\Omega^{-1}\Xi + \Sigma_{\alpha_0}^{-1}\right)\alpha + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 =$$

$$= y'\Omega^{-1}y - \left(\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha}\right)'\alpha - \alpha'\left(\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha}\right) + \alpha'\left(\Sigma_{\bar{\alpha}}^{-1}\right)\alpha + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 =$$

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

This can be rewritten as[12]:

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

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

Therefore, the posterior density kernel can be also written as follows:

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

where:

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

---

[12] Since:

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

we have that:

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

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:

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

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:

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

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

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

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

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:

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

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

The marginal likelihood is thus:

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

In our case $M = 2$, $\Sigma_{\alpha_0} = \Sigma_{EH}$, $\alpha_0 = \alpha_{EH}$, so the formula computed above simplify to:

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

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

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

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

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:

$$\Sigma_u^{-1} \sim Wi(\upsilon_0, \Sigma_{u0}^{-1}).$$

Uninformativeness for this prior is achieved by setting $\upsilon_0 = 0$ 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:

$$\begin{aligned} p(\alpha, \Sigma_u) &= p(\alpha)p(\Sigma_u), \\ \alpha &\sim N(\alpha_{EH}, \Sigma_{EH}), \\ \Sigma_u^{-1} &\sim Wi(\upsilon_0, \Sigma_{u0}^{-1}). \end{aligned}$$

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

$$\begin{aligned} \alpha|y, \Sigma_u^{-1} &\sim N(\bar{\alpha}, \Sigma_{\bar{\alpha}}), \\ \Sigma_u^{-1}|y, \alpha &\sim Wi(\upsilon_0 + T, (\Sigma_{u0} + S)^{-1}), \end{aligned}$$

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.

2.8 References

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.

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.

Campbell J., 1995, "Some Lessons from the Yield Curve", *Journal of Economic Perspectives*, Vol. 9, pp. 129-152.

Campbell J., Shiller R., 1987, "Cointegration and Tests of Present Value Models", *Journal of Political Economy*, Vol. 95, pp.1062-1088.

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.

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.

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.

Clews R., 2002. “Asset Prices and Inflation” *Quarterly Bulletin Bank of England*, Summer 2002, 178-85.

Cochrane J., Piazzesi M., 2005, "Bond Risk Premia", *American Economic Review,* Vol. 95, pp. 138-160.

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.

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

ECB Monthly Bulletin. http://www.ecb.int/pub/mb/html/index.en.html

Fama E.F,1984, "The Information in the Term Structure", *Journal of Financial Economics*, Vol.13, pp. 509-528.

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

Fisher I., 1930, *The Theory of Interest*, Macmillan, New York.

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.

Geweke, J. (2005) *Contemporary Bayesian Econometrics and Statistics.* Wiley.

Giacomini R., White H., 2006, "Tests of Conditional Forecast Accuracy", *Econometrica*, Vol. 74, pp. 1545-1578.

Hicks J., 1953, *Value and Capital*, Oxford University Press, Oxford.

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

Kass R. E., Raftery A., 1995, "Bayes Factors", *Journal of the American Statistical Association*, Vol. 90, pp. 773-795.

Keynes J.M., 1930, *Treatise on Money*, Harcourt, Brace and Howe Inc., New York.

Koo, G., 2003, *Bayesian Econometrics*, Wiley, Chicester.

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.

Lewis K.K., 1989, "Changing beliefs and systematic rational forecast errors with evidence from foreign exchange", *American Economic Review*, Vol. 79, pp. 269-276.

Litterman R.B., 1986, "Forecasting with Bayesian Vector Autoregressions-Five Years of Experience", *Journal of Business & Economic Statistics*, Vol. 4, pp. 25-38.

Mishkin F.S., 1988, "The Information in the Term Structure: Some Further Results", *Journal of Applied Econometrics*, Vol. 3, pp. 307-14.

Rudebusch G.D., 1998, "Do Measures of Monetary Policy in a VAR Make Sense?", *International Economic Review*, Vol. 39, pp. 907-31.

Scholtes C., 2002, “On Market-Based Measures of Inflation Expectations”, *Quarterly Bulletin of Bank of England*, Spring 2002, pp. 67-77.

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

Sims C.A., 2003, "Probability Models for Monetary Policy Decisions", manuscript, Princeton University.

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.

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,

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.

Theil H., Goldberger A.S., 1961, "On Pure and Mixed Estimation in Economics", *International Economic Review*, Vol. 2, pp. 65–78.

Todd R.M., 1984, "Improving Economic Forecasting with Bayesian Vector Autoregression", *Federal Reserve Bank of Minneapolis Quarterly Review* , pp.18–29.

White H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity", *Econometrica*, Vol. 48, pp. 817-838.

2.9 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

	Restricted VAR		Unrestricted VAR		VAR with EH prior		
	$FIML$	$EH(0)$	OLS	$EH(\infty)$	$EH(\sigma^*)$	$\%d_{EH(0)}$	$\%d_{EH(\infty)}$
a_1	0.2901	0.2901	0.4528	0.4528	0.4293	48.0%	-5.18%
b_1	0.1058	0.1057	0.2591	0.2591	0.2378	124%	-8.23%
a_2	0.0914	0.0914	-0.0036	-0.0036	0.0146	-84.0%	-509%
b_2	-0.0610	-0.0610	-0.2781	-0.2781	-0.2372	288%	-14.6%
a_3	0.2054	0.2054	0.2016	0.2016	0.1987	-3.28%	-1.46%
b_3	-0.0232	-0.0232	0.0409	0.0409	0.0212	-191%	-48.2%
k_1	-0.0234	-0.0234	-0.0219	-0.0219	-0.0218	-6.73%	-0.45%
c_1	-0.2901	-0.2901	-0.0298	-0.0298	-0.0674	-76.7%	125%
d_1	0.8884	0.8885	1.1338	1.1338	1.0997	23.7%	-3.00%
c_2	-0.0914	-0.0914	-0.2433	-0.2433	-0.2142	134%	-11.9%
d_2	0.0610	0.0610	-0.2861	-0.2861	-0.2208	-461%	-22.8%
c_3	-0.2054	-0.2054	-0.2115	-0.2115	-0.2162	5.25%	2.22%
d_3	0.0232	0.0232	0.1256	0.1256	0.0940	305%	-25.1%
k_2	0.0093	0.0094	0.0117	0.0117	0.0118	27.0%	1.37%

The first 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 second group of columns provides estimates of the VAR without restrictions.The first column reports the OLS estimates, the second column 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 restricted and the unrestricted VAR.

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

	96-obs		120-obs		144-obs	
h	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 untrestricted VAR. The column h reports the forecast horizon. For the single variables, the 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

	96-obs		120-obs		144-obs	
h	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. For the single variables, the 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).

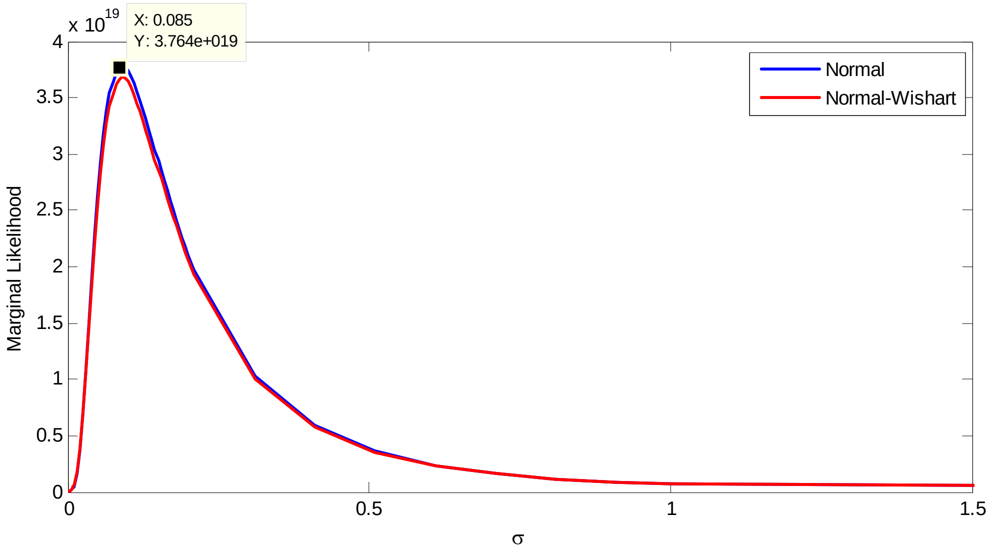

Figure 2.1: Marginal likelihood of the $EH(\sigma)$ model, as a function of σ. The blue line is based on the Normal prior with fixed variance matrix for the disturbances, while the red line is based on the independent Normal-Wishart prior.

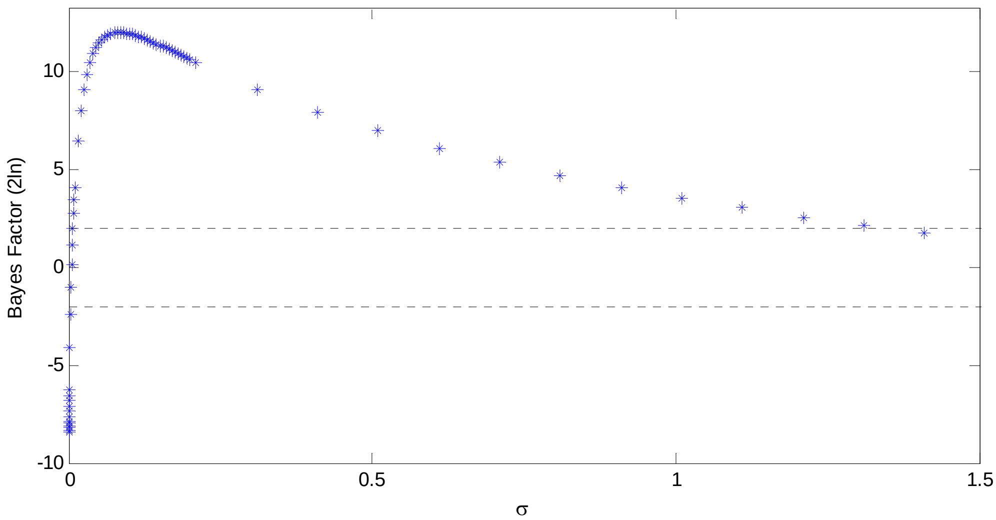

Figure 2.2: Twice the log of the Bayes factor of the $EH(\sigma)$ model agains the $UVAR$, as a function of σ. The two dotted lines represent the so-called inconclusive region. Below the dotted lines the data provide evidence in favour of the $UVAR$. Above the dotted lines the data provide evidence in favour of the $EH(\sigma)$ model. Within the dotted lines the data do not provide evidence in favor of any of the two models.

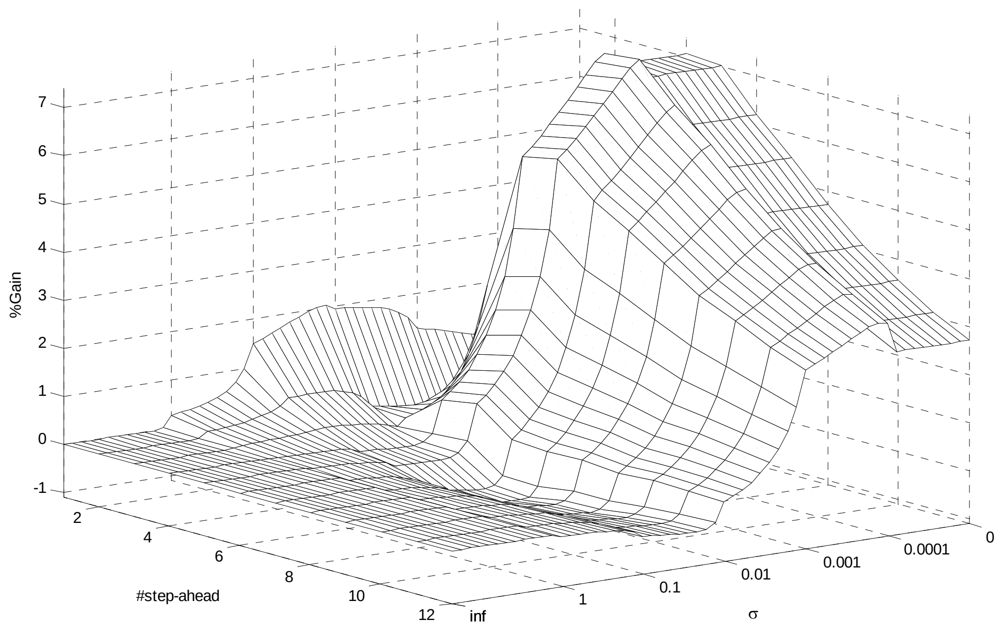

Figure 2.3: Percentage gain in forecasting using the VAR with the EH prior rather than the unrestricted VAR. The x axis reports the EH prior tightness, the y axis the forecast horizon. The percentage gain is the difference in the multivariate in-det statistic. The in-det statistic is given by the log determinant of the forecast error covariance matrix divided by the number of forecasted variables..

Chapter 3

Explaining US-UK Interest Rate Differentials: a Reassessment of the Uncovered Interest Rate Parity in a Bayesian Framework

3.1 Introduction

The[1] Uncovered Interest Rate Parity (UIRP) states that interest rate differentials among two currencies reflect expectations on future exchange rate movements plus a constant risk premium. From a theoretical point of view, the UIRP can be considered a cornerstone of international finance, and it is a key assumption of most important exchange rate determination theories.

Despite its importance in macroeconomics and finance the UIRP has received little empirical support (see Hodrik, 1987, Engel, 1996, Froot and Thaler, 1990). More recently Baillie and Bollerslev (2000) and Bekaert and Hodrick (2001) have argued that doubtful statistical inference may have contributed to the strong rejections of the UIRP, while Bekaert, Wei and Xing (2002) show that the evidence against the UIRP is mixed, that deviations from the UIRP are less pronounced than previously documented.

In the first part of the paper we test the UIRP at the 10-year horizon by deriving from it linear restrictions on a VAR in the interest rate differential and the variation of the spot exchange rate. When tested with a Wald test on US-UK data, the theory is

[1]This paper has been published in the Oxford Bulletin of Economics and Statistics, vol. 68, pp. 879-899. I thank Anindya Banerjee, Carlo Favero, Ulf Söderström, and two anonymous referees.

strongly rejected. This is in line with previous studies testing the UIRP for the US-UK at the 10-year horizon (Alexius, 2001, Chinn and Meredith 2004).

Several authors have analyzed the possible causes of failure of the UIRP. In a recent paper Cochrane and Piazzesi (2005) provide neat evidence of significant time variation in risk premiums implying predictability of yields. Bekaert, Wei and Xing (2002) show that deviations from the UIRP are consistent with a time varying but stationary risk premium. A number of researchers have identified possible explanations for deviations from rationality, 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). Clarida and Taylor (1997), develop a framework featuring the testable implication that deviations from the UIRP, whether due to risk aversion or to nonrational expectations, are realizations of a stationary stochastic process. Interestingly, they conclude that the restrictions implied by their framework are not rejected. However, they do not test the UIRP restrictions explicitly. We shall see how our framework allows to do so, and also enables us to measure the relevance of these deviations and to assess how far we are from a world in which in any period risk premiums are constant and all agents are fully rational. Flood and Taylor (1994) study the UIRP within a single equation approach and show that some evidence in favor of the UIRP arises as data are averaged through time. Indeed, taking five-, ten-, twenty-year averages they found a strong proportionality between average exchange rate depreciation and average movements in the fundamentals and conclude that this is indicative of the fact that fundamental relations apply as time goes by, while they feature noise in the short run.

These considerations open the ground to the second part of the paper, in which we test the UIRP allowing for transitory deviations from it. Such deviations may arise from time varying but stationary risk premia, errors in expectations, and linearization errors, and are modelled as a zero-mean noise around the UIRP restrictions. Importantly, this approach comprises the traditional one as a special case, which is derived simply setting the noise to zero. As a result, we show that when the deviations are set to zero the UIRP is rejected, but if we allow for some degree of noise the UIRP is strongly supported by the data. We also provide an informal economic test showing that the behaviour of the U.S. / U.K. 10-year interest rate differential from the eighties onwards has been entirely consistent with the UIRP. This provides evidence that the UIRP relation does not hold exactly, on a period-by-period basis, but holds on average, i.e. after integrating out the effect of the transitory deviations which may affect it. As the amount of noise needed to accept the UIRP is small, this result implies that analyzing the effects of policy experiments under the null of the UIRP may be both safe and useful.

The paper is organized as follows: Section 2 introduces the basic framework, Section 3 builds on it to derive 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 interest rate differential. Section 6 concludes. In the Appendices we derive some results used throughout the paper.

3.2 The Uncovered Interest Rate Parity

Let $i_{t,T}$ be the T-month interest rate in the home country and $i^*_{t,T}$ that of the foreign country. The UIRP states that the interest rate differential between the two countries D_t depends on the expected percentage variations in the exchange rate $E_t\Delta e_{t+i}$ plus a constant (i.e. dependent on maturity only) risk premium RP_T:

$$D_t = i_{t,T} - i^*_{t,T} = \frac{1}{T}\sum_{i=1}^{T} E_t\Delta e_{t+i} + RP_T. \tag{3.1}$$

Our data set consists of the spread between the 10-year U.S. and U.K. government bonds yields, at a constant maturity rate, and the U.K. / U.S. foreign exchange spot rate. The two series are average monthly figures going from 1979:1 to 2005:6, provided by the Federal Reserve of St.Louis and by the Bank of England.

Consider the following VAR for D_t and Δe_t:

$$\begin{aligned} \Delta e_t &= k_1 + a_1\Delta e_{t-1} + a_2\Delta e_{t-2} + a_3\Delta e_{t-3} + b_1 D_{t-1} + b_2 D_{t-2} + b_3 D_{t-3} + u_{1t}, \\ D_t &= k_2 + c_1\Delta e_{t-1} + c_2\Delta e_{t-2} + c_3\Delta e_{t-3} + d_1 D_{t-1} + d_2 D_{t-2} + d_3 D_{t-3} + u_{2t}, \end{aligned} \tag{3.2}$$

where the lag length has been chosen via the Bayesian information criterion performed over the whole sample with a maximum lag length of 13. Both Wald and Likelihood Ratio sequential lag exclusion tests confirm this choice by clearly rejecting specifications with richer dynamics.

To gauge the appropriateness of the VAR in (3.2) we checked for parameter stability and conducted a battery of tests on the residuals. Recursive residuals and parameter estimates are stable in the second equation, while there is some suspect for instability for some parameters of the first equation at the end of 1992. Still, Chow breakpoint tests performed recursively in each month do not signal presence of structural breaks at the 5% confidence level. Diagnostic tests on the residuals provide evidence in favor of nonautocorrelation and homoscedasticity of the disturbances. In particular the LM test statistic (reported in Johansen, 1995 p.22) for the null of no autocorrelation up to order 4 is well below the critical value (p-value 0.601). The evidence against heteroscedasticity is less clear cut, as the p-value of the White (1980) test is 0.066. There are some concerns about normality. Indeed when testing for joint normality the null is strongly rejected due to excess kurtosis in the residuals. The minimum amount of lags needed for the excess kurtosis to disappear is 9. As normality is not necessary for the asymptotic validity of Wald tests we stay with our more parsimonious 3-lag specification and we keep the 9-lag specification for a robustness check.

From equation (3.1) it is possible to derive a set of restrictions implied by the UIRP on the VAR (3.2). In appendix A we show that these restrictions are given by:

$$\begin{bmatrix} \frac{1}{T}a_1 + c_1 \\ \frac{1}{T}b_1 + d_1 \\ \frac{1}{T}a_2 + c_2 \\ \frac{1}{T}b_2 + d_2 \\ \frac{1}{T}a_3 + c_3 \\ \frac{1}{T}b_3 + d_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{3.3}$$

which can be compactly written as:

$$H\alpha = \mu_{UIRP_0}, \tag{3.4}$$

where:

$$\begin{aligned} \alpha &= \begin{bmatrix} a_1 & b_1 & a_2 & b_2 & a_3 & b_3 & k_1 & c_1 & d_1 & c_2 & d_2 & c_3 & d_3 & k_2 \end{bmatrix}', \\ H &= \begin{bmatrix} \frac{1}{T}I_6 & \underset{6\times1}{\mathbf{0}} & I_6 & \underset{6\times1}{\mathbf{0}} \end{bmatrix}, \ \mu_{UIRP_0} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}'. \end{aligned}$$

We do not have restrictions on the constants of the VAR as we allowed for a constant risk premium. Notice that the UIRP implies that the coefficients attached to the same variable in the two equations must be perfectly negatively correlated.

When tested with a Wald test, the set of restrictions (3.4) is strongly rejected, as the Chi-square (6) statistic is 83.29, corresponding to a p-value of 0. The rejection of the joint restrictions is due to the combinations of the coefficients on the lags of D_t which are all neatly significant, while all the combinations regarding the coefficients on the lags of Δe_t would be accepted one by one. This latter result is explained by the fact that the lags of Δe_t are not significant in the equation for D_t, while they are significant but small in the equation for Δe_t. As a robustness check we implement the Wald test also on the VAR with 9-lags, and also for this case restrictions are neatly rejected.

These results are in line with previous findings in the literature testing the UIRP at the 10-year horizon. Alexius (2001) provides tests of the UIRP at the 10-year horizon for 13 OECD countries vis-a-vis US, showing that there is favorable evidence in more than half of the cases, but still for the US-UK system the UIRP is clearly rejected. Also Chinn and Meredith (2004) find favorable evidence for the UIRP at the 10-year horizon for some G-7 countries but reject the UIRP for the pound. Bekaert, Wei and Xing (2002) also use a VAR framework to test the UIRP jointly with the Expectations Theory of the Term Structure, and find only marginal evidence in favor of the UIRP at the 5-year horizon for the US-UK system.

3.3 The Uncovered Interest Rate Parity as a Set of Uncertain Restrictions

In this section we develop a version of the UIRP featuring some noise in the short run. This version of the UIRP is more general and comprises the traditional one as a special case. Our approach is the same of Carriero (2005) and is similar in spirit to that of Del Negro and Schorfheide (2005).

3.3.1 Adding uncertainty

Suppose the UIRP does hold, but only on average, i.e. some noise causes temporary departures from the UIRP restrictions in (3.4). Formally, let the uncertainty introduced by this noise be measured by the parameter σ. The resulting set of stochastic constraints is:

$$\begin{bmatrix} \frac{1}{T}a_1 + c_1 \\ \frac{1}{T}b_1 + d_1 \\ \frac{1}{T}a_2 + c_2 \\ \frac{1}{T}b_2 + d_2 \\ \frac{1}{T}a_3 + c_3 \\ \frac{1}{T}b_3 + d_3 \end{bmatrix} \sim N \left(\mu_{UIRP_0} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma_{UIRP_0} = \begin{bmatrix} \sigma & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma \end{bmatrix} \right), \tag{3.5}$$

which can be compactly written as:

$$H\alpha \sim N\left(\mu_{UIRP_0}, \Sigma_{UIRP_0}\right). \tag{3.6}$$

The parameter σ is the tightness of the restrictions imposed by the UIRP: a large value of σ implies that the UIRP restrictions hold with a lot of uncertainty, while as σ decreases the UIRP restrictions become more binding and eventually become certain. Of course, if we allow for very little variation (or no variation) this essentially implies ruling out any noise and imposing the UIRP restrictions to hold exactly, while allowing for very large uncertainty around the restrictions would lead to an insignificant version of the UIRP. 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 3.4 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.

3.3.2 Interpreting noise

In this subsection we discuss why the set of restrictions in equation (3.4) may hold only up to some noise, as shown in equation (3.6). There are several potential sources of noise. Some noise directly affects the UIRP relation, while some additional noise arises when URIP restrictions are derived within the VAR framework.

First, the UIRP relation in equation (3.1) cannot be considered an exact relation, as it comes from a linearization which ignores the Jensen Inequality term.

Second, we could have 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).

Third, it may well be the case that the risk premium is not constant but features some movement around its mean. Cochrane and Piazzesi (2005) provide neat evidence of significant time variation in risk premiums implying predictability of yields. Bekaert, Wei and Xing (2002) show that deviations from the UIRP are hard to reconcile with a short-term market frictions or market inefficiency story, while they are consistent with a time varying but stationary risk premium. For these reasons we think this is the main and economically most relevant source of noise which may affect the UIRP.

These sources of noise may be thought of as a stationary error term appended to equation (3.1) as in Clarida and Taylor (1997), who develop a framework featuring the testable implication that deviations from the UIRP, whether due to risk aversion or to nonrational expectations, are realizations of a stationary stochastic process. Interestingly, they conclude that the restrictions implied by their framework are not rejected. 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 risk premiums are constant and all agents are fully rational.

Clearly, with a stationary disturbance, equation (3.1) 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 transfers in 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 sum in equation (3.1) is computed. Moreover, agents' expectations are not observable. Even in the case of full rationality there could be a discrepancy between the unobserved expectations and the linear projection used to proxy for them. Again, these sources of noise transfer in the restrictions.

To conclude, the UIRP restrictions in (3.4) may be affected from a time varying but stationary risk premium, expectations errors, and linearization errors. These deviations may imply the UIRP restrictions hold only up to some noise, yielding the null hypothesis in equation (3.6).

3.3.3 The UIRP in a Bayesian framework

The set of restriction (3.6) can be thought of in a Bayesian perspective as a prior on the coefficients of the VAR in equation (3.2). Therefore, we can test the UIRP by using the approach developed by Jeffreys (1935). 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 (3.2), while the second theory imposes the restrictions derived from the UIRP. Rewrite the VAR in equation (3.2) in the following way:

$$y = \Xi\alpha + \varepsilon, \tag{3.7}$$

with:

$$\begin{aligned} y &= vec\left(\left[\begin{array}{cc} \Delta e_t & D_t \end{array}\right]\right), \\ \Xi &= I_2 \otimes \left[\begin{array}{ccccccc} \Delta e_{t-1} & D_{t-1} & \Delta e_{t-2} & D_{t-2} & \Delta e_{t-3} & D_{t-3} & 1 \end{array}\right], \\ \varepsilon &= vec\left(\left[\begin{array}{cc} u_{1t} & u_{2t} \end{array}\right]\right) \sim N\left(0,\ \Omega\right),\ \Omega = \Sigma_u \ \otimes I_N, \end{aligned}$$

where N is the sample size and where y, ε and α are $2N \times 1$, $2N \times 1$, and 14×1 vectors, and Ξ and Σ_u are $2N \times 14$, 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:

$$\alpha \sim N(\alpha_0 = \underset{14\times1}{\mathbf{0}}, \Sigma_{\alpha_0} = \delta I_{14}). \tag{3.8}$$

We will refer to the VAR (3.7) with the loose prior (3.8) 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[2]. With our data, a value of $\delta = 100$ is large enough to ensure that this is the case[3].

The second theory imposes the restriction scheme (3.6) implied by the UIRP. We will refer to equation (3.6) as the UIRP prior and to the system consisting of the VAR (3.7) and equation (3.6) as the UIRP-restricted VAR (RVAR):

$$\begin{gathered} y = \Xi\alpha + \varepsilon, \\ H\alpha \sim N\left(\mu_{UIRP_0}, \Sigma_{UIRP_0}\right). \end{gathered} \tag{3.9}$$

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

$$\alpha \sim N\left(\alpha_{UIRP_0}, \Sigma_{\alpha_{UIRP_0}}\right), \tag{3.10}$$

with

$$\alpha_{UIRP_0} = \left[\begin{array}{ccc} \underset{1\times8}{\mathbf{0}} & 1 & \underset{1\times5}{\mathbf{0}} \end{array}\right]', \tag{3.11}$$

[2] To have exact coincidence between the OLS estimates and the Bayesian estimates we should impose $\delta = \infty$, which is the calibration implicit in the classical tradition. However, it is impossible to impose such a calibration in practice so we specify a finite value for δ large enough to ensure that differences between the OLS estimates and the Bayesian estimates are numerically insignificant.

[3] More precisely, with $\delta = 100$ the difference between OLS and Bayesian estimates is at most of an order of 0.59%.

and

$$\Sigma_{\alpha_{UIRP_0}} = \begin{bmatrix} \delta I_6 & \mathbf{0}' & -\frac{\delta}{T} I_6 & \mathbf{0}' \\ \mathbf{0} & \delta & \mathbf{0} & 0 \\ -\frac{\delta}{T} I_6 & \mathbf{0}' & \left(\sigma + \frac{\delta}{T^2}\right) I_6 & \mathbf{0}' \\ \mathbf{0} & 0 & \mathbf{0} & \delta \end{bmatrix}, \tag{3.12}$$

where $\mathbf{0}$ is a 1×6 vector of zeros. Equations (3.7) and (3.10) lead to the following alternative representation of the RVAR in equation (3.9):

$$\begin{aligned} y &= \Xi\alpha + \varepsilon, \\ \alpha &\sim N\left(\alpha_{UIRP_0}, \Sigma_{\alpha_{UIRP_0}}\right). \end{aligned} \tag{3.13}$$

To clarify the role played by the tightness parameter σ it is worth to look at the prior correlation matrix of the coefficients under the UIRP-prior, which is easily derived from (3.12):

$$Corr(\alpha_{UIRP_0}) = \begin{bmatrix} I_6 & \mathbf{0}' & \frac{-\sqrt{\frac{1}{T^2}\delta}}{\sqrt{\sigma+\frac{1}{T^2}\delta}} I_6 & \mathbf{0}' \\ \mathbf{0} & 1 & \mathbf{0} & 0 \\ \frac{-\sqrt{\frac{1}{T^2}\delta}}{\sqrt{\sigma+\frac{1}{T^2}\delta}} I_6 & \mathbf{0}' & I_6 & \mathbf{0}' \\ \mathbf{0} & 0 & \mathbf{0} & 1 \end{bmatrix}. \tag{3.14}$$

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 UIRP 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.

Both the UVAR and the RVAR are linear regression models subject to a set of stochastic linear restrictions on the regression coefficients, with a fixed variance matrix of the errors Σ_u. 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 0.09 by maximizing the marginal likelihood of the model. Turning to the correlation matrix of the coefficients, the parametrization $\sigma = 0.09$, $\delta = 100$ 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.

3.4 Statistical Evidence

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

3.4.1 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 $\frac{pr(H_2|\mathbf{D})}{pr(H_1|\mathbf{D})} = \frac{pr(\mathbf{D}|H_2)}{pr(\mathbf{D}|H_1)} \frac{pr(H_2)}{pr(H_1)}$, so that the transformation is simply multiplication by:

$$B_{21} = \frac{pr(\mathbf{D}|H_2)}{pr(\mathbf{D}|H_1)}, \tag{3.15}$$

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 (3.15) 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:

$$pr(\mathbf{D}|H_i) = (2\pi)^{-MN/2}\,|\Omega|^{-1/2}\,\frac{|\Sigma_{\alpha_{i0}}|^{-1/2}}{|\Sigma_{\bar{\alpha}_i}|^{-1/2}}\exp\{-Q_i/2\}, \tag{3.16}$$

for $i = 1, 2$, where $Q_i = y'\Omega^{-1}y - \bar{\alpha}_i'\Sigma_{\bar{\alpha}i}^{-1}\bar{\alpha}_i + \alpha_{0i}'\Sigma_{\alpha_{0i}}^{-1}\alpha_{0i}$,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 $\alpha_{20} = \alpha_{UIRP_0}$, $\bar{\alpha}_2 = \bar{\alpha}_{UIRP}$, $\Sigma_{\alpha_{20}} = \Sigma_{\alpha_{UIRP_0}}$, $\Sigma_{\bar{\alpha}_2} = \Sigma_{\bar{\alpha}_{UIRP}}$. Thus the Bayes factor for the RVAR versus the UVAR is:

$$B_{21} = \left[\frac{\frac{|\Sigma_{\alpha_{UIRP_0}}|}{|\Sigma_{\bar{\alpha}_{UIRP}}|}}{\frac{|\Sigma_{\alpha_0}|}{|\Sigma_{\bar{\alpha}}|}}\right]^{-1/2} \exp\left\{\frac{Q_{UVAR} - Q_{RVAR}}{2}\right\}. \tag{3.17}$$

3.4.2 Results

Figure 3.1 plots (twice) the natural logarithm of the Bayes factor ($2\ln B_{21}$) as a function of the UIRP prior tightness σ together with the inconclusive region. The inconclusive region for $2\ln B_{21}$ ranges from -2 to 2: in this region neither $2\ln B_{21} > 2$ nor $2\ln B_{12} = -2\ln B_{21} > 2 \Rightarrow 2\ln B_{21} < -2$. Thus, in this region 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 ($B_{21} = 0.0608$ and $2\ln B_{21} = -5.6015$). This confirms the results of the Wald test described in Section 2. Indeed, letting the tightness go to zero amounts to imposing the UIRP without noise. Therefore our general framework nests the traditional one as a special case, and is consistent with the empirical findings rejecting the UIRP.

On the other hand, allowing for very large departures from the UIRP 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 UIRP, 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 UIRP prior tightness $\sigma = 0.09$ the value of $2\ln B_{21}$ reaches a value above 35 denoting very strong evidence in favour of the UIRP[4]. This evidence explains the anomaly that the Wald test rejects the UIRP but the theoretical interest rate differential based on its validity is closely correlated with the actual one, as we shall see in the next section. As stressed above, the theory is strongly supported by the data, which explains the high correlation, but also the UIRP is perturbed by some noise which leads the Wald test to reject the exact restrictions. As discussed in Subsection 3.2, such noise might be due to time varying but stationary term premia, errors in expectations, and linearization errors.

These results are in line with those of Clarida and Taylor (1997), who develop a framework featuring the testable implication that deviations from the UIRP, whether due to risk aversion or to nonrational expectations, are realizations of a stationary stochastic process and conclude that the restrictions implied by their framework are not rejected. However, they do not test the UIRP restrictions explicitly. Our framework allows to assess the relevance of these deviations. Indeed, the world with fully rational expectations and constant risk premiums is represented by the model with $\sigma = 0$, while the unrestricted world, in which the UIRP does not hold, is represented by the unrestricted model, i.e. $\sigma = \delta$. The estimated σ indicates we are closer to the former.

These results also confirm the intuition of Flood and Taylor (1994) who study the UIRP within a single equation approach and show that some evidence in favor of the UIRP arises as data are averaged trough time and conclude that this is indicative of the fact that fundamental relations apply as time goes by, while they feature noise in the short run.

There is an appealing interpretation also in terms of forecasting. Indeed, the Bayes factor can also be interpreted as a measure of performance in forecasting, and in this case it signals that the optimal forecast is that of the model with $\sigma = 0.09$. Now suppose we want to forecast using classical VARs (i.e. without imposing priors). Obviously,

[4]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.

neither the exactly restricted VAR nor the unrestricted VAR would provide an optimal forecast. The former is implicitly assuming $\sigma = 0$ while the latter is implicitly assuming $\sigma = \delta$. The optimal forecast would be given by pooling the two models in order to implicitly obtain the model with $\sigma = 0.09$. In this case the optimal pooling weights should be $(\delta - 0.09)/\delta = 99.91\%$ for the exactly restricted VAR and $0.09/\delta = 0.09\%$ for the unrestricted VAR[5], so a very high weight is given to the restricted model.

To conclude, our evidence shows that the UIRP relation does not hold exactly, on a period-by-period basis, but holds on average, i.e. after integrating out the effect of the deviations which may affect it. Therefore, analyzing the effects of policy experiments under the null of the UIRP may be both safe and useful.

3.4.3 Robustness

In this subsection we check for the robustness of our results. First, we look at likelihood ratios between the competing models and we assess robustness trough time. Then we redo our exercise specifying a prior also on second order parameters and we address the concerns related with the so called Lindley paradox.

Likelihood ratios and stability

In order to check for robustness and interpret our results within a classical framework, we look at the likelihood ratios between the two models. Figure 3.2 plots twice the natural logarithm of the likelihood ratio of the RVAR versus the UVAR as a function of the UIRP 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 result of rejection. 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 UIRP. We assess the stability of our results trough time by computing recursively the Bayes factor over the whole sample, starting with a 5-year window. As is clear in Figure 3.3, the evidence in favor of the UIRP is stable over time.

Priors on second order parameters

The priors used so far feature a fixed covariance matrix of the errors ($\Omega = \Sigma_u \otimes I_N$), and imply the existence of closed form solutions for posteriors and marginal likelihoods of the two models at hand (see appendix C). It is worth to check the robustness of our

[5]Indeed, the resulting average σ would be $\bar{\sigma} = \sigma(\text{exactly restricted } VAR) * 99.91\% + \sigma(\text{unrestricted } VAR) * 0.09\% = 0 * 99.91\% + 100 * 0.09\% = 0.09$.

results if we specify a prior also on the matrix of second order parameters. To do so we specify an uninformative Wishart prior on the variance matrix of the errors: $\Sigma_u^{-1} \sim Wi(\upsilon_0, \Sigma_{u0}^{-1})$. Uninformativeness for this prior is achieved by setting $\upsilon_0 = 0$ and $\Sigma_{u0} = 0_{2\times 2}$. This provides us with an independent Normal-Wishart prior on our models. Such prior is standard in the Bayesian literature, and requires simulation methods to compute posteriors and marginal likelihoods. For details see Appendix C. We plot results in Figure 3.4. As seen in the picture, our results are robust to this extension. The main difference which arises is an increase in the Bayes factor for the smallest values of σ, while the shape of the Bayes factor is virtually identical in the neighborhoods of the peak value.

Lindley paradox

The most important concern about our results is related to the so-called Lindley paradox[6] (i.e. a prior with a very high variance is likely to be significantly disadvantaged respect to a tighter one). In our exercise, the UVAR should feature a very loose prior by definition. Indeed, in order to interpret the Bayes factor as a test of uncertain restrictions, the UVAR prior must be loose enough to be uninformative. Any UVAR prior imposing some restrictions on the data would not be appropriate as alternative hypothesis because it would not comprise the RVAR as a special case. In this case the Bayes factor would end up in contrasting two alternative models, not nested one into the other, and the whole exercise could not be interpreted as a test of restrictions.

Therefore, it is key to assess the relevance of the Lindley paradox effect on our results. To do so, we specify an alternative prior for the competing model, call it UVAR* prior, which features much less variance than the original one. We use pre-sample observations to estimate the VAR and then we use second moments of these estimates to calibrate the UVAR* prior. In particular, using data form 1979:1 to 1983:12 we estimate by OLS the individual variance of each coefficient of the unrestricted VAR and then we assign this value to the corresponding diagonal element δ of the UVAR* coefficients prior variance matrix. By doing so, the values assigned to the prior variance of the UVAR* coefficients now range from 0.001 to 0.53, which are much smaller than the value of 100 used in the previous case. On the other side, the RVAR prior is left completely unaltered, featuring an unrestricted variance of the coefficients of 100. This creates a scenario in which the looser prior becomes the RVAR and so allows to assess the relevance and the magnitude of the Lindley paradox effect.

Results of this analysis are reported in Figure 3.5. As we are using a different sample

[6] We thank an anonymous referee for pointing this out.

(the first five years are left out and used to calibrate the UVAR* prior) the peak value for the tightness σ changes, decreasing to 0.05. The value of $2\ln B_{21}$ at the peak decreases from 35 to a value of 7.5. Therefore, there is a significant effect due to the reduction of the looseness in the competing model. However, a value of $2\ln B_{21}$=7.5 (corresponding to $B_{21} = 42.5$) is very high and still provides strong evidence in favor of the RVAR according to Table 1. Therefore the Lindley paradox has a relevant effect but our results are robust after controlling for it[7].

3.5 Explaining Interest Rate Differential Dynamics

In this section we provide evidence that the 10-year US-UK interest rate differential from the 1980s onwards has been entirely consistent with the UIRP. To do so we first construct a theoretical, UIRP-consistent interest rate differential and then contrast it with the actual, realized one.

Recall that under the UIRP the interest rate differential D_t^* is given by

$$D_t^* = \frac{1}{T}\sum_{i=1}^{T} E_t \Delta e_{t+i} + RP_T, \tag{3.18}$$

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

Therefore, we compute D_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 1979:1 1981:12. All the subsequent estimations are performed over the sample 1979:1 1981:12+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 $E_t\Delta e_{t+i}$ and D_t^* are obtained. iii) Then we move forward one period, adding one data point to the estimation window,

[7] The fact that the Bayes factor goes below 0 as σ approaches δ is not surprising. Indeed this result simply shows that the new, tighter UVAR* prior has an higher marginal likelihood compared to the old, loose UVAR prior (recall that for $\sigma \longrightarrow \delta$ the RVAR prior has still the same marginal likelihood of the UVAR prior). In the original exercise, as $\sigma \to \delta$, $2\ln B_{21}$ converges to 0, but now this cannot happen, as the two models are no more nested. This clarifies why the exercise performed in this subsection can be interpreted as a robustness check but not as a test of the uncertain restrictions.

and go back to point i). This recursive estimation / projection scheme provides time series of the posterior distributions of the variables $E_t\Delta e_{t+i}$ and D_t^* . As long as we are under the null of the UIRP, these variables would also provide the posterior distribution of the risk premium as $RP_T = D_t^* - \frac{1}{T}\sum E_t\Delta e_{t+i}$.

Our recursive estimation/projection scheme provides a simple but effective test of the UIRP. Indeed, a test for the UIRP is immediately performed simply by checking whether the actual interest rate differential could be a plausible draw from the posterior distribution of the UIRP-consistent interest rate differential, i.e. if D_t lies within some credible bounds of the posterior distribution of D_t^*. This procedure is very similar to that used by Carriero, Favero and Kaminska (2006), but with the subtle difference that here the uncertainty does not arise from estimation, but is modelled within the theory.

The implied distribution of D_t^* is plotted in Figure 3.6 together with the actual differential D_t. The actual differential (blue solid line) almost always lies within the 2.5% and the 97.5% percentiles of the posterior distribution of the UIRP-consistent interest rate differential (the yellow area). The UIRP-consistent and the actual series are highly correlated, but of course they do not perfectly coincide. The traditional framework to test the UIRP proxies ex-ante expected exchange rates with the ex-post realized exchange rates and therefore it understates the amount of uncertainty that individuals face when forecasting future exchange rates up to 120-month ahead, which is huge, as measured by the width of the bounds around the UIRP-consistent interest rate differential D_t^*.

Once we take into account the uncertainty involved in predicting exchange rates, the UIRP cannot be rejected. This result would also hold if the actual and the UIRP-consistent differentials were less clearly correlated: Figure 3.6 shows that the gap between the 2.5% and the 97.5% percentiles is considerably wider than the difference between the actual and the UIRP-consistent differentials. Even if the actual differential had behaved much more differently from the UIRP-consistent one, it still could be consistent with the theory. Thus, the dynamics of the 10-year interest rate differential was entirely consistent with the UIRP, and this adds an economic validation to our statistical evidence.

3.6 Conclusions

In this paper we tested the Uncovered Interest Rate Parity (UIRP) allowing for transitory deviations from it. In the first part of the paper we tested the UIRP at the 10-year horizon by deriving from it linear restrictions on a VAR in the interest rate differential and the variation of the spot exchange rate. When tested with a Wald test on US-UK data, the theory is strongly rejected. This result is in line with previous findings in the literature testing the UIRP for the US-UK system at the 10-year horizon (Alexius 2001, Chinn and Meredith 2004). In the second part of the paper we tested the UIRP allowing for transitory deviations from it. These deviations may arise from time varying but stationary risk premia, errors in expectations, and linearization errors, and are modelled as a zero-mean noise around the UIRP restrictions. Importantly, this approach comprises the traditional one as a special case, which is derived simply setting the noise to zero.

When the deviations are set to zero the UIRP is rejected, but if we allow for some degree of noise the UIRP is strongly supported by the data. This means that we are not far from a world in which agents are fully rational and risk premiums are constant. If we would want to optimally forecast changes in the spot exchange rate and in the interest rate differentials we should use a pooled forecast giving a very high weight on a model wich imposes the UIRP and a very low weight to an unrestricted model. Beyond statistical evidence, the paper provides economic evidence showing that the dynamics of the 10-year US-UK interest rate differential was entirely consistent with the UIRP.

There are several potential deviations from the UIRP. Previous results in the literature suggest that the most important deviation could be variation in risk premia. Indeed, Cochrane and Piazzesi (2005) provide neat evidence of significant time variation in risk premiums implying predictability of yields. Bekaert, Wei and Xing (2002) show that deviations from the UIRP are hard to reconcile with a short-term market frictions or market inefficiency story, while they are consistent with a stationary risk premium. Thus, we think that a time varying risk premium is the main and economically most relevant source of noise which may affect the UIRP in the short run. In this light, our results can be interpreted as evidence that the UIRP does not hold with a constant risk premium, but it may well hold with a stationary risk premium.

To conclude, this paper provided evidence that the UIRP does not hold exactly, on a period-by-period basis, but holds on average, i.e. after integrating out the effect of the deviations which may affect it. This result implies that analyzing the effects of policy experiments under the null of the UIRP may be both safe and useful.

3.7 Appendices

A: UIRP restrictions

Stack the VAR as:

$$\begin{bmatrix} \Delta e_t \\ \Delta e_{t-1} \\ \Delta e_{t-2} \\ D_t \\ D_{t-1} \\ D_{t-2} \end{bmatrix} = \begin{bmatrix} k_1 \\ 0 \\ 0 \\ k_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 & a_3 & b_1 & b_2 & b_3 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ c_1 & c_2 & c_3 & d_1 & d_2 & d_3 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta e_{t-1} \\ \Delta e_{t-2} \\ \Delta e_{t-3} \\ D_{t-1} \\ D_{t-2} \\ D_{t-3} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ 0 \\ 0 \\ u_{2t} \\ 0 \\ 0 \end{bmatrix},$$

or, more succinctly:

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

The trace statistic for the null of no cointegration (with the intercept in the cointegrating equation) is well above the critical value (63.42 while the 5% critical value is 20.26).

The UIRP states that:

$$D_t = i_{t,T} - i^*_{t,T} = \frac{1}{T}\sum_{i=1}^{T} E_t \Delta e_{t+i} + RP_T.$$

Thus, the validity of the UIRP would put on the VAR the following set of nonlinear restrictions:

$$\begin{aligned} g' z_t &= \frac{1}{T}\sum_{i=1}^{T} h' \left(\sum_{n=0}^{i-1} A^n C + A^i z_t \right) + RP_T \\ &= h' \frac{1}{T} \left(\sum_{i=1}^{T}\sum_{n=0}^{i-1} A^n C + \sum_{i=1}^{T} A^i z_t \right) + RP_T \\ &= h' \frac{1}{T} \left(T \sum_{n=0}^{i-1} A^n C + \sum_{i=1}^{T} A^i z_t \right) + RP_T \end{aligned}$$

where $g = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}'$ and $h = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}'$.

Since the above expression has to hold in general, it holds also removing the mean from all the variables, and for any z_t:

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

As the eigenvalues of A all lie within the unit circle, this converges to:

$$g' = h'\frac{1}{T}A(I - A^{T+1})(I - A)^{-1},$$

thus

$$g'(I - A) = h'\frac{1}{T}A(I - A^{T+1}).$$

As this VAR is stable, for $T = 120$ the term $I - A^{T+1}$ can be safely approximated by $I - A^{T+1} \approx I$.

$$g'(I - A) = h'\frac{1}{T}A,$$

i.e.

$$g'\begin{bmatrix} 1-a_1 & -a_2 & -a_3 & -b_1 & -b_2 & -b_3 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ -c_1 & -c_2 & -c_3 & 1-d_1 & -d_2 & -d_3 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} = h'\frac{1}{T}\begin{bmatrix} a_1 & a_2 & a_3 & b_1 & b_2 & b_3 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ c_1 & c_2 & c_3 & d_1 & d_2 & d_3 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

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

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

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

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

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

Write the VAR as:

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

where $M = 2$ is the number of equations, $p = 3$ is the number of lags included, and N 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 $\delta = 100$ the posterior mean of α is identical to the OLS estimator.

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

$$
\begin{bmatrix} \frac{1}{T}a_1 + c_1 \\ \frac{1}{T}b_1 + d_1 \\ \frac{1}{T}a_2 + c_2 \\ \frac{1}{T}b_2 + d_2 \\ \frac{1}{T}a_3 + c_3 \\ \frac{1}{T}b_3 + d_3 \end{bmatrix} \sim N\left(\mu_{UIRP_0} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma_{UIRP_0} = \begin{bmatrix} \sigma & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma \end{bmatrix} \right).
$$

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

$$H\alpha_{UIRP} \sim N\left(\mu_{UIRP_0}, \Sigma_{UIRP_0}\right), \tag{3.19}$$

where

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

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

$$\begin{cases} y = \Xi\alpha_{UIRP} + \varepsilon \\ H\alpha_{UIRP} \sim N\left(\mu_{UIRP_0}, \Sigma_{UIRP_0}\right) \end{cases}.$$

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

$$\alpha_{UIRP} \sim N(\alpha_{UIRP0}, \Sigma_{\alpha_{UIRP0}}),$$

which implies:

$$H\alpha_{UIRP} \sim (H\alpha_{UIRP0}, H\Sigma_{\alpha_{UIRP0}}H'). \tag{3.20}$$

Under the UIRP both (3.19) and (3.20) must hold, so there is the following relation between the prior moments of the vector of restrictions and those of the vector of coefficients:

$$\begin{aligned} \Sigma_{UIRP_0} &= H\Sigma_{\alpha_{UIRP0}}H', \\ \mu_{UIRP_0} &= H\alpha_{UIRP0}. \end{aligned}$$

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 UIRP restrictions and the analysis.

The restrictions become:

$$H_2\alpha_{UIRP} \sim N\left(\mu_{2UIRP_0}, \Sigma_{2UIRP_0}\right),$$

with:

$$\mu_{2UIRP_0} = \left[\begin{array}{cccccccccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]',$$

$$\Sigma_{2UIRP_0} = \left[\begin{array}{cccccccccccccc}
\delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\delta & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\delta & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta
\end{array}\right],$$

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

With δ sufficiently big (which is the case in the paper) the form of the UIRP 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_{UIRP}:

$$\alpha_{UIRP} \sim N\left(\alpha_{UIRP0}, \Sigma_{\alpha_{UIRP0}}\right),$$

where

$$\alpha_{UIRP0} = H_2^{-1}\mu_{2UIRP_0} = \left[\begin{array}{cccccccccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right]',$$

$$\Sigma_{\alpha_{UIRP0}} = H_2^{-1}\Sigma_{2UIRP_0}H_2'^{-1}$$

$$= \left[\begin{array}{cccccccccccccc}
\delta & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{\delta}{T} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{\delta}{T} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{\delta}{T} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{\delta}{T} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{\delta}{T} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{\delta}{T} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-\frac{\delta}{T} & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\frac{\delta}{T^2} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\frac{\delta}{T} & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\frac{\delta}{T^2} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -\frac{\delta}{T} & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\frac{\delta}{T^2} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\frac{\delta}{T} & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\frac{\delta}{T^2} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -\frac{\delta}{T} & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\frac{\delta}{T^2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -\frac{\delta}{T} & 0 & 0 & 0 & 0 & 0 & 0 & \sigma+\frac{\delta}{T^2} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta
\end{array}\right].$$

So the RVAR can be written as:

$$\begin{cases} y = \Xi\alpha_{UIRP} + \varepsilon \\ \alpha_{UIRP} \sim N\left(\alpha_{UIRP0}, \Sigma_{\alpha_{UIRP0}}\right) \end{cases}$$

To estimate the VAR define:

$$\begin{aligned} -v_{UIRPt} &= \alpha_{UIRP} - \alpha_{UIRP_0} \sim N\left(0, \Sigma_{\alpha_{UIRP0}}\right) \Longrightarrow \\ \alpha_{UIRP_0} &= \alpha_{UIRP} + v_{UIRPt},\ v_{UIRPt} \sim N\left(0, \Sigma_{\alpha_{UIRP0}}\right) \end{aligned}$$

then plug in:

$$
\begin{aligned}
y_{UIRP} &= \Xi_{UIRP}\alpha_{UIRP} + \varepsilon_{UIRP}, \\
\overbrace{\underset{(MN+M(pM+1))\times 1}{\begin{bmatrix} \Delta e_t \\ D_t \\ \alpha_{UIRP_0} \end{bmatrix}}}^{y_{UIRP}} &= \overbrace{\underset{(MN+M(pM+1))\times M(pM+1)}{\begin{bmatrix} \underset{MN\times M(pM+1)}{\left[I_M \otimes \underset{N\times(pM+1)}{X} \right]} \\ \left[I_{M(pM+1)} \right] \end{bmatrix}}}^{\Xi_{UIRP}} * \underset{M(pM+1)\times 1}{\alpha_{UIRP}} + \overbrace{\underset{(MN+M(pM+1))\times 1}{\begin{bmatrix} u_{1t} \\ u_{2t} \\ v_{UIRPt} \end{bmatrix}}}^{\varepsilon_{UIRP}}, \\
\varepsilon_{UIRP} &\sim N\left(0, \Omega_{UIRP} = \underset{(MN+M(pM+1))\times(MN+M(pM+1))}{\begin{bmatrix} \Omega = \underset{MN\times MN}{[\Sigma_u \otimes I_N]} & 0 \\ 0 & \Sigma_{\alpha_{UIRP0}} \end{bmatrix}} \right).
\end{aligned}
$$

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:

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

then plug in:

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

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 $UVAR$).

$$\overbrace{\underset{MN\times 1}{\begin{bmatrix}\Delta e_t \\ D_t\end{bmatrix}}}^{y} = \overbrace{\underset{MN\times Mk}{\begin{bmatrix} I_M \otimes \underset{N\times k}{X}\end{bmatrix}}}^{\Xi} * \underset{Mk\times 1}{\alpha} + \overbrace{\underset{MN\times 1}{\begin{bmatrix}u_{1t} \\ u_{2t}\end{bmatrix}}}^{\varepsilon},$$

$$\begin{aligned} \alpha &\sim N(\underset{k\times 1}{\alpha_0}, \underset{k\times k}{\Sigma_{\alpha_0}}), \\ e &\sim N(0,\Omega),\ \Omega = \underset{MN\times MN}{[\Sigma_u \otimes I_N]}. \end{aligned}$$

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 N 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:

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

the likelihood is[8]:

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

a posterior density kernel is:

$$\begin{aligned} p(y|\alpha)p(\alpha) &= (2\pi)^{-M(N+k)/2} |\Omega|^{-1/2} |\Sigma_{\alpha_0}|^{-1/2} \\ &\quad \exp\left\{-1/2\begin{bmatrix} (y-\Xi\alpha)'\Omega^{-1}(y-\Xi\alpha) \\ +(\alpha-\alpha_0)'\Sigma_{\alpha_0}^{-1}(\alpha-\alpha_0)\end{bmatrix}\right\}. \end{aligned}$$

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

Now define[9]:

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

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

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[10]:

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

[9] Notice that:

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

[10] 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:

$$\begin{aligned} p(y|\alpha)p(\alpha) &= (2\pi)^{-M(N+k)/2} |\Omega|^{-1/2} |\Sigma_{\alpha_0}|^{-1/2} \\ & \exp\left\{ -1/2 \left[\begin{array}{c} (\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha) \\ y'\Omega^{-1}y - \bar{\alpha}'\Sigma_{\bar{\alpha}}^{-1}\bar{\alpha} + \alpha_0{}'\Sigma_{\alpha_0}^{-1}\alpha_0 \end{array} \right] \right\} \\ &= (2\pi)^{-M(N+k)/2} |\Omega|^{-1/2} |\Sigma_{\alpha_0}|^{-1/2} \\ & \exp\left\{ -1/2[(\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha) + Q] \right\}, \end{aligned}$$

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:

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

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:

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

$$= \int_{\Re^{Mk}} (2\pi)^{-M(N+k)/2} |\Omega|^{-1/2} |\Sigma_{\alpha_0}|^{-1/2} \exp\left\{-1/2\left[\begin{array}{c}(y-\Xi\alpha)'\Omega^{-1}(y-\Xi\alpha)+\\(\alpha-\alpha_0)'\Sigma_{\alpha_0}^{-1}(\alpha-\alpha_0)]\end{array}\right]\right\} d\alpha$$

$$\begin{aligned} &= \int_{\Re^{Mk}} (2\pi)^{-M(N+k)/2} |\Omega|^{-1/2} |\Sigma_{\alpha_0}|^{-1/2} \\ &\quad \exp\left\{-1/2[(\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha)+Q]\right\} d\alpha \\ &= (2\pi)^{-M(N+k)/2} |\Omega|^{-1/2} |\Sigma_{\alpha_0}|^{-1/2} \exp\{-Q/2\} \\ &\quad \exp\int_{\Re^{Mk}} \left\{-1/2[(\bar{\alpha}-\alpha)'\Sigma_{\bar{\alpha}}^{-1}(\bar{\alpha}-\alpha)]\right\} d\alpha. \end{aligned}$$

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:

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

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

The marginal likelihood is thus:

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

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:

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

The above results apply to the case with fixed variance matrix of the errors ($\Omega = \Sigma_u \otimes I_N$) as in Theil (1971) and Litterman (1986). Alternatively, we could specify a prior also on this matrix. In particular, the prior used in Subsection 4.3 is an independent Normal-Wishart:

$$\begin{aligned} p(\alpha, \Sigma_u) &= p(\alpha)p(\Sigma_u), \\ \alpha &\sim N(\alpha_0, \Sigma_{\alpha 0}), \\ \Sigma_u^{-1} &\sim Wi(v_0, \Sigma_{u0}^{-1}). \end{aligned}$$

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

$$\begin{aligned} \alpha | y, \Sigma_u^{-1} &\sim N(\bar{\alpha}, \Sigma_{\bar{\alpha}}), \\ \Sigma_u^{-1} | y, \alpha &\sim Wi(v_0 + N, (\Sigma_{u0} + S)^{-1}), \end{aligned}$$

where the generic element of the matrix S is $s_{ij} = (y_i - \Xi_i \alpha)(y_j - \Xi_j \alpha)$, and $i, j = 1, .., M$ signal the subvector or submatrix composed by the N rows associated with the i-th and j-th equation (so for example in our case $y_1 = \Delta e_t$ and $y_2 = D_t$). Uninformativeness for this prior is achieved by setting $v_0 = 0$ and $\Sigma_{u0} = 0_{M \times M}$.

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.

3.8 References

Alexius, A. (2001). 'Uncovered interest parity revisited', *Review of International Economics,* Vol. 9, pp. 505-517.

Baillie, R. T. and Bollerslev, T. (2000). 'The forward premium anomaly is not as bad as you think', *Journal of International Money and Finance,* Vol. 19, pp. 471-488.

Bekaert, G. and Hodrick, R. J. (1993). 'On biases in the measurement of foreign exchange risk premiums', *Journal of International Money and Finance,* Vol. 12, pp. 115-138.

Bekaert, G., Wei, M. and Xing, Y. (2002). *Uncovered interest rate parity and the term structure*, NBER Working Paper No. 8795.

Carriero, A. (2005). *Resurrecting the expectations hypothesis: how to extract additional information from the term structure of interest rates*, IEP Bocconi Working Paper No. 0502.

Carriero, A., Favero, C. and 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.

Chinn, M. D. and Meredith, G. (2004). 'Monetary policy and long horizon uncovered in terest rate parity', *IMF Staff Papers,* Vol. 51, pp. 409-430.

Clarida, R. and 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.

Cochrane, J. H. and Piazzesi, M. (2005). 'Bond Risk Premia', *American Economic Review*, Vol. 95, pp. 138-160.

De Long, J. B., Shleifer, A., Summers L.H. and Waldmann, R. (1990). 'Noise Trader Risk in Financial Markets', *Journal of Political Economy*, Vol. 98, pp. 703-738.

Del Negro, M. and Schorfheide, F. (2005). *Monetary policy analysis with potentially misspecified models*, ECB Working Paper Series No. 475.

Engel, C. (1996). 'The forward discount anomaly and the risk premium: a survey of recent evidence', *Journal of Empirical Finance,* Vol. 3, pp. 123-192.

Flood, R. P. and Taylor, M. P. (1994). 'Exchange Rate Economics: What's Wrong with the Conventional Macro Approach?' in Gali, J, Giovannini, A. and Frankel J. (eds), *Foreign Exchange Market Microstructure*, Chicago University Press for NBER, Chicago, pp. 262-301.

Frankel, J. A. and 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*,Oxfprd University Press, Oxford, pp. 73-128.

Froot, K. A. and Frankel, J. A. (1989). 'Forward discount bias: is it an exchange risk premium?', *Quarterly Journal of Economics,* Vol. 104, pp. 139-161.

Froot, K. A. and Thaler, R. H (1990). 'Foreign exchange', *Journal of Economic Perspectives,* Vol. 4, pp. 179-192.

Hodrick, R. J. (1987). *The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange Markets*, Harwood Academic Publishers, Chur.

Jeffreys H. (1935). 'Some tests of significance, treated by the theory of probability', *Proceedings of the Cambridge Philosophy Society*, Vol. 31, pp. 203-222.

Johansen, S. 1995. *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

Kass R. E. and Raftery A. (1995). 'Bayes Factors', *Journal of the American Statistical Association*, Vol. 90, pp. 773-795.

Lewis, K. K. (1989). 'Changing beliefs and systematic rational forecast errors with evidence from foreign exchange', *American Economic Review,* Vol. 79, pp. 269-276.

Taylor, M. P. and Allen, H.L. (1992). 'The use of technical analysis in the foreign exchange market', *Journal of International Money and Finance,* Vol. 11, 304-314.

Theil, H. (1971). *Principles of Econometrics*, Wiley, New York.

White, H. (1980). 'A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity', *Econometrica*, Vol. 48, pp. 817-838.

3.9 Figures

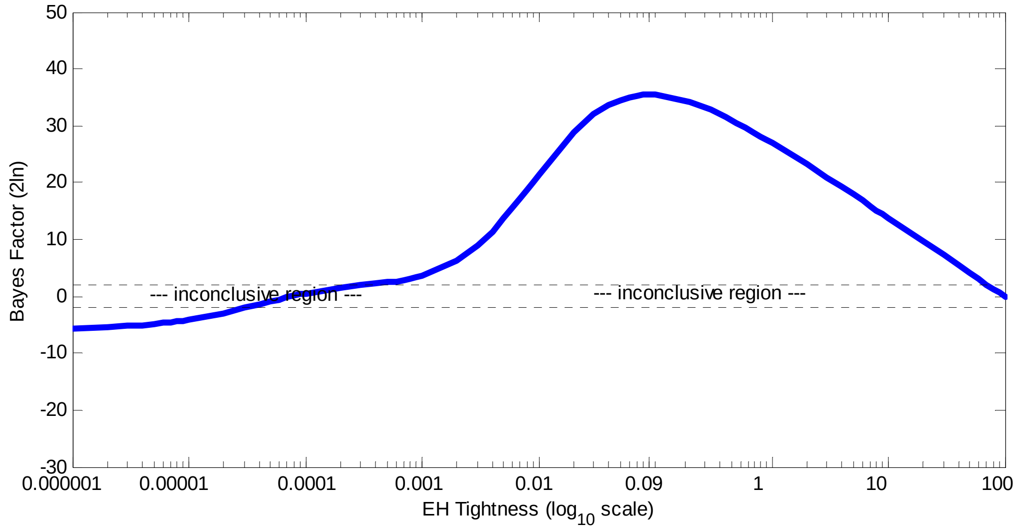

Figure 3.1: Bayes factor (twice its natural logarithm) for the UIRP-restricted versus the unrestricted VAR, as a function of the UIRP prior tightness σ. The x axis is in base 10 logarithmic scale.

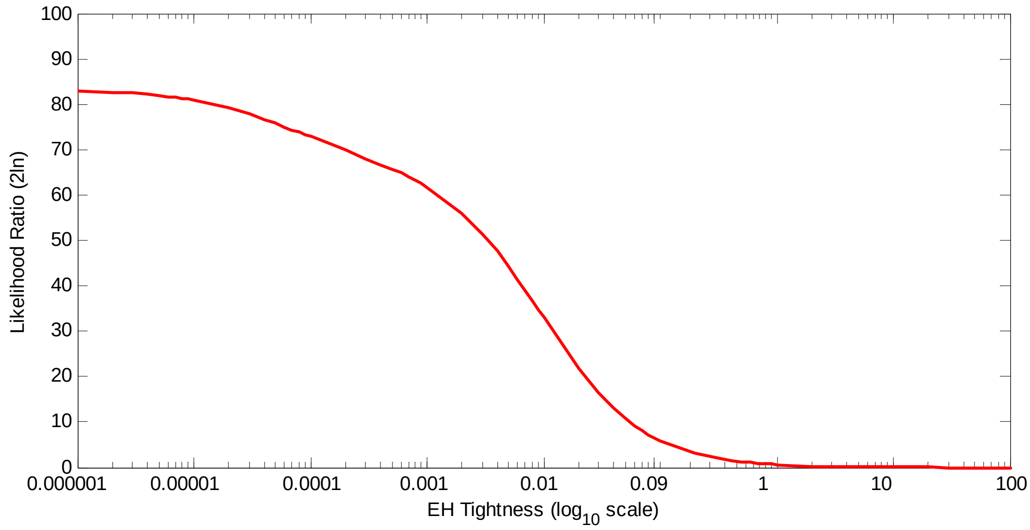

Figure 3.2: Twice the natural logarithm of likelihood ratios for the UIRP-restricted versus the unrestricted VAR, as a function of the UIRP prior tightness σ. The x axis is in base 10 logarithmic scale.

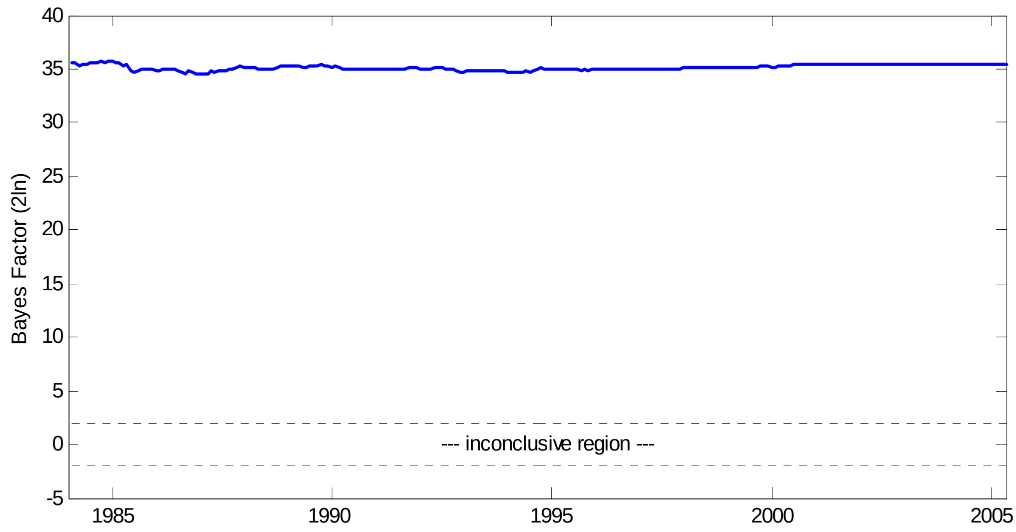

Figure 3.3: Robustness through time. Bayes factor (twice its natural logarithm) for the UIRP-restricted versus the unrestricted VAR as a function of time. The UIRP prior tightness is fixed at its estimated value $\sigma = 0.09$.

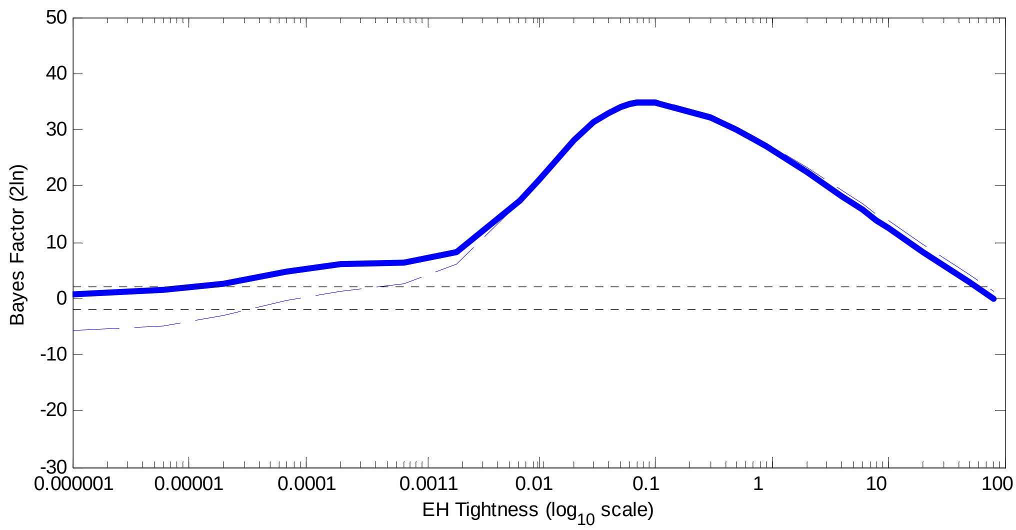

Figure 3.4: Independent Normal-Wishart prior. The solid line is the Bayes factor (twice its natural logarithm) when a Wishart prior is assigned to the variance matrix of the errors ($\Sigma_u^{-1} \sim Wi(v, \Sigma_{u0}^{-1})$ with $v_0 = 0$ and $\Sigma_{u0} = 0_{2\times 2}$). The dotted line reports the Bayes factor depicted in Figure 3.1, i.e. with Σ_u fixed. The x axis is in base 10 logarithmic scale.

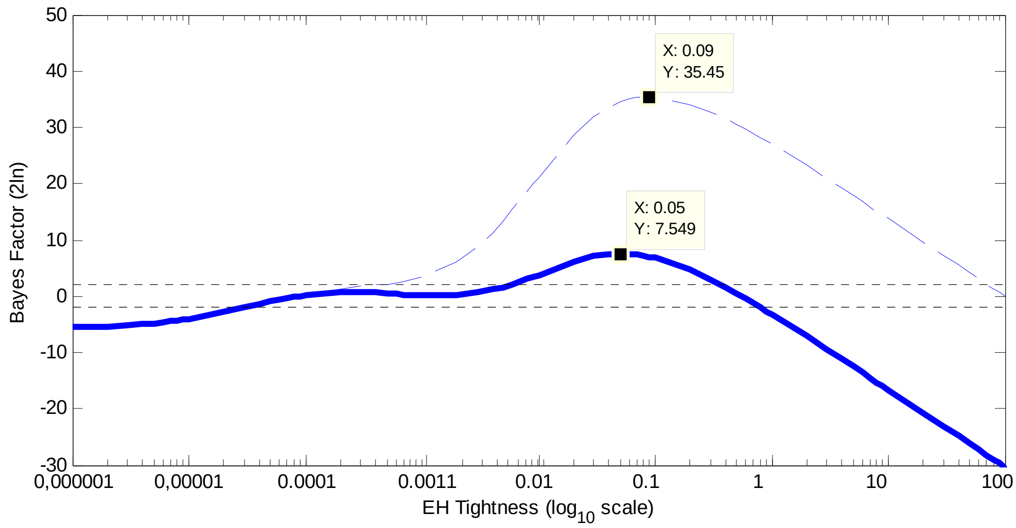

Figure 3.5: Robustness to Lindley paradox. The solid line is the Bayes factor (twice its natural logarithm) of the UIRP-restricted VAR versus the alternative competing model UVAR*. The dotted line reports the Bayes factor depicted in Figure 3.1, i.e. the Bayes factor of the UIRP-restricted VAR versus the baseline competing model UVAR.

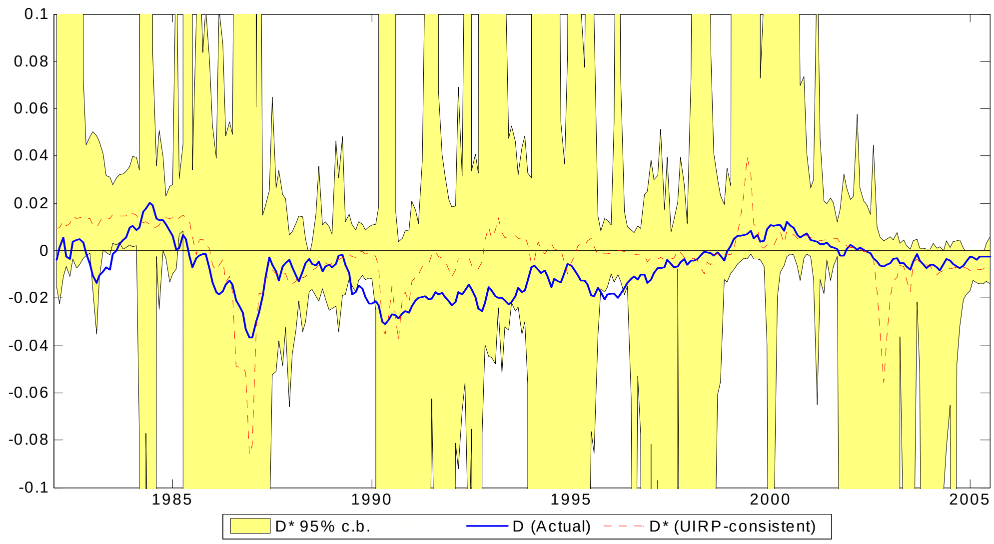

Figure 3.6: Economic test of the UIRP. The distribution of the theoretical, UIRP-consistent interest rate differential D_t^* is obtained by a recursive estimation/projection scheme, such that at each point in time only the available information is used to estimate the VAR and then to project it forward.