NO-ARBITRAGE NEAR-COINTEGRATED VAR(p)
TERM STRUCTURE MODELS,
TERM PREMIA AND GDP GROWTH

Caroline JARDET [1]
Banque de France

Alain MONFORT [2]
CNAM, CREST and Banque de France

Fulvio PEGORARO[3]
Banque de France and CREST

First version : September, 2008. This version: June 2009.

Abstract

**No-arbitrage Near-Cointegrated VAR(p) Term Structure Models,
Term Premia and GDP Growth**

   Macroeconomic questions involving interest rates generally require a reliable joint dynamics of a large set of variables. More precisely, such a dynamic modelling must satisfy two important conditions. First, it must be able to propose reliable predictions of some key variables. Second, it must be able to propose a joint dynamics of some macroeconomic variables, of the whole curve of interest rates, of the whole set of term premia and, possibly, of various decompositions of the term premia. The first condition is required if we want to disentangle the respective impacts of, for instance, the expectation part of the term premium of a given long-term interest rate on some macroeconomic variable. The second condition is necessary if we want to analyze the interactions between macro-variables with some global features of the yield curve (short part, long part, level, slope and curvature) or with, for instance, term premia of various maturities.

   In the present paper we propose to satisfy both requirements by using a Near-Cointegrated modelling based on averaging estimators, in order to meet the first condition, and the no-arbitrage theory, in order to meet the second one. Moreover, the dynamic interactions of this large set of variables is based on the statistical notion of New Information Response Function, recently introduced by Jardet, Monfort and Pegoraro (2009b). This technical toolkit is then used to propose a new approach to two important issues: the "conundrum" episode and the puzzle of the relationship between the term premia on long-term yields and future economic activity.

Keywords: Averaging estimators, Near-Cointegrated VAR(p) model, Term structure of interest rates; Term premia; GDP growth; No-arbitrage affine term structure model; New Information Response Function.

JEL classification: C51, E43, E44, E47, G12.

[1]Banque de France, Financial Economics Research Service [DGEI-DEMFI-RECFIN; E-mail: Caroline.JARDET@banque-france.fr].

[2]CNAM, CREST, Laboratoire de Finance-Assurance [E-mail: monfort@ensae.fr], and Banque de France, Financial Economics Research Service [DGEI-DEMFI-RECFIN.

[3]Banque de France, Financial Economics Research Service [DGEI-DEMFI-RECFIN; E-mail: Fulvio.PEGORARO@banque-france.fr] and CREST, Laboratoire de Finance-Assurance [E-mail: pegoraro@ensae.fr].

1 Introduction

Macroeconomic questions involving interest rates generally require a reliable joint dynamics of a large set of variables. More precisely, such a dynamic modelling must satisfy two important conditions. First, it must be able to propose reliable predictions of some key variables. Second, it must be able to propose a joint dynamics of some macroeconomic variables, of the whole curve of interest rates, of the whole set of term premia and, possibly, of various decompositions of the term premia. The first condition is required if we want to disentangle the respective impacts of, for instance, the expectation part of the term premium of a given long-term interest rate on some macroeconomic variable. The second condition is necessary if we want to analyze the interactions between macro-variables with some global features of the yield curve (short part, long part, level, slope and curvature) or with, for instance, term premia of various maturities.

In the present paper we propose to satisfy both requirements by using a Near-Cointegrated modelling based on averaging estimators [see B. Hansen (2007, 2008, 2009) and Jardet, Monfort and Pegoraro (2009a)], in order to meet the first condition, and the no-arbitrage theory, in order to meet the second one. Moreover, the dynamic interactions of this large set of variables is based on the statistical notion of New Information Response Function, recently introduced by Jardet, Monfort and Pegoraro (2009b). This technical toolkit is then used to propose a new approach to two important issues: the "conundrum" episode and the puzzle of the relationship between the term premia on long-term yields and future economic activity. Let us now describe more precisely these technical points as well as these empirical issues.

The foundation of our approach is based on a careful VAR modelling of the dynamics of some basic observable variables, namely $Y_t = (r_t, R_t, G_t)'$, where r_t is a short rate, R_t a long rate and G_t is the Log of the real gross domestic product (GDP). These variables are also considered in the pioneering paper of Ang, Piazzesi and Wei (2006) [APW (2006), hereafter] whose model constitutes a benchmark of our study. The observability of the basic variables allows for a crucial step of econometric diagnostic tests on stationarity, cointegration, and number of lags. In our application, based on quarterly observations of US yields and real GDP, this analysis shows a unique cointegration relationship, namely the spread $S_t = R_t - r_t$, and then leads to a cointegrated VAR(3), denoted by CVAR(3), for the variables $X_t = (r_t, S_t, g_t = \Delta G_t)'$, which are exactly the ones appearing in the VAR(1) model considered by APW (2006).

Nevertheless, the presence of unit roots is wrongly induced by the high persistence of (stationary) interest rates and this phenomenon rises the huge "discontinuity" problem, already stressed by Cochrane and Piazzesi (2008), that is, the dramatic difference between long-run predictions based on the CVAR(3) and an unconstrained VAR(3) model (for X_t). An additional problem is the well known bias that stands out when estimating without constraints dynamic models which are "nearly non-stationary". Among the possible ways of tackling these problems, we choose the one based on the local-to-unity asymptotic properties of predictions based on averaging estimators [see B. Hansen (2007, 2008, 2009)]. More precisely, we consider the averaging among the estimators of the cointegrated and unconstrained VAR(3) models (for X_t). In order to motivate and to evaluate the prediction performances of these averaging estimators we propose a Monte Carlo study

comparing these performances with those of natural competitors, namely bias-corrected estimators like Indirect Inference estimator, Bootstrap estimator, Kendall's estimator and Median-unbiased estimator. We thus are lead to a Near-Cointegrated VAR(3) model, or NCVAR(3), in which the averaging parameter is obtained by optimizing the out-of-sample predictions of a variable of interest[4]. Given the application we have in mind, this variable is chosen to be the expectation part of a long-term yield. Our empirical application fully confirms the theoretical results in Hansen (2009), since the NCVAR(3) model allows for a large reduction of the out-of-sample root mean square-forecast errors (RMSFE) in interest rates.

The second technical component of our paper is the derivation of an affine term structure of interest rates based on an exponential-affine stochastic discount factor (SDF), including stochastic market prices of risk depending on present and lagged values of X_t. The parameters of the SDF are estimated by a least square fitting of the whole yield curve. In the empirical application the fitting is very good and, moreover, the out-of-sample RMSFE in the prediction of yields of various maturities at various horizons is much better than the one of competing models [VAR(1), VAR(3), CVAR(3), AR(1)], the reduction reaching 45% for the 10-year horizon. Moreover, this affine modelling allows for a simple recursive computation of term premia and of their decompositions in terms of forward term premia and in terms of expected risk premia.

The first application is devoted to the "conundrum" episode. The rise of federal funds rates (f.f.r.) of 425 basis points and the low and relative stable level of the 10-year interest rate, observed between June 2004 and June 2006 on the U.S. market is described as a "conundrum" by the Federal Reserve Chairman Alan Greenspan in February 2005, given that, during three previous episodes of restrictive monetary policy (in 1986, 1994 and 1999), the 10-year yield on US zero-coupon bonds strongly increased along with the fed funds target.

Among several finance and macro-finance models [see, for instance, Hamilton and Kim (2002), Bernanke, Reinhart and Sack (2004), Favero, Kaminska and Sodestrom (2005), Kim and Wright (2005), Ang, Piazzesi and Wei (2006), Bikbov and Chernov (2006), Dewachter and Lyrio (2006), Dewachter, Lyrio and Maes (2006), Rudebusch, Swanson and Wu (2006), Rosenberg and Maurer (2007), Rudebusch and Wu (2007, 2008), Chernov and Mueller (2008), Cochrane and Piazzesi (2008), and the survey proposed by Rudebusch, Sack and Swanson (2007)], some have indicated that the reason behind the coexistence of increasing f.f.r. and stable long-rates is found in a reduction of the term premium, that offsets the upward revision to expected future short rates induced by a restrictive monetary policy. Our analysis is based on a reliable measure of the term premia (on the long-term bond) and also on a decomposition of that measure in terms of forward term premia at different horizons, and in terms of expected risk premia attached to the one-period holding of bonds at different maturities. This analysis, which also considers a comparison with other recent periods showing a rise of the short rate (1994 and 1999), leads us to results in line with Cochrane and Piazzesi (2008).

The second application deals with the relationship between term premia (on the long-term yield) and future economic activity. Some works [Hamilton and Kim (2002), and Favero, Kaminska and Sodestrom (2005)] find a positive relation between term premium and economic activity. In contrast, Ang, Piazzesi and Wei (2006), Rudebusch, Sack and Swanson (2007), and Rosenberg and Maurer (2007) find that the term premium has no predictive power for future GDP growth.

---

[4]It is important to point out that the averaging estimator strategy does not imply any parameter or model uncertainty of the investor [like, for instance, in L.P. Hansen and Sargent (2007, 2008)]. It is a procedure adopted by the econometrician to improve the out-of-sample forecast performances of the model.

Practitioner and private sector macroeconomic forecaster views agree on the decline of the term premium behind the conundrum but suggest a relation of negative sign between term premium and economic activity [see Rudebusch, Sack and Swanson (2007), and the references there in, for more details]. This negative relationship is usually explained by the fact that a decline of the term premium, maintaining relatively low and stable long rates, may stimulate aggregate demand and economic activity, and this explanation implies a more restrictive monetary policy to keep stable prices and the desired rate of growth. Therefore, policy makers seems to have no precise indication about the stimulating or shrinking effect of term premia on gross domestic product (GDP) growth.

In the present paper we provide a dynamic analysis of the relationship between the spread and future economic activity. In addition, we are interested in disentangling the effects of a rise of the spread due to an increase of its expectation part, and a rise of the spread caused by an increase of the term premium. For that purpose, we propose a new approach based on a generalization of the Impulse Response Function, called New Information Response Function (NIRF) [see Jardet, Monfort and Pegoraro (2009b)]. This approach allows us to measure the dynamic effects of a new (unexpected) information at date $t = 0$, regarding any state variables, any yield to maturity or any linear filter of that variables, on any variables. Like in most studies found in the literature, we find that an increase of the spread implies a rise of the economic activity. We find similar results when the rise of the spread is generated by an increase of its expectation part. In contrast, an increase of the spread caused by a rise of the term premium induces two effects on future output growth: the impact is negative for short horizons (less than one year), whereas it is positive for longer horizons. Therefore, our results suggest that the ambiguity found in the literature regarding the sign of the relationship between the term premium and future activity, could come from the fact that the sign of this relationship is changing over the period that follows the shock. In addition, we propose an economic interpretation of this fact.

The paper is organized as follows. Section 2 proposes a motivation for the use of averaging estimators based on their prediction performances. Section 3 describes the data and Section 4 introduces the Near-cointegration methodology, that is, describes the cointegration analysis of $Y_t = (r_t, R_t, G_t)'$, specifies the CVAR(3) model for $X_t = (r_t, S_t, g_t)'$, stresses the discontinuity problem and presents its solution based on a averaging estimators. Moreover, we present the empirical performances of the NCVAR(3) model in terms of out-of-sample forecast of the short rate and long rate, and we compare them to some competing model in the literature. Section 5.1 shows how the Near-cointegrated model can be completed by a no-arbitrage affine term structure model, and Section 5.2 presents the risk sensitivity parameter estimates, along with the risk sensitivity parameter estimates of the other competing models considered in the empirical analysis. We study in Section 5.3 the empirical performances of our model in terms of in-sample fit of the yield curve, yield curve out-of-sample forecasts and Campbell-Shiller regressions. In Section 6, we present decompositions of the term premia in terms of forward and expected risk term premia, and we show how these measures can be used to accurately analyse the recent "conundrum" episode. Section 7 presents the impulse response analysis based on the notion of New Information Response Function. Section 8 concludes, Appendix 1 gives further details about unit root analysis and Appendix 2 derives the yield-to-maturity formula. In Appendix 3 we gather additional tables and graphs.

2 Persistence, bias, prediction and averaging estimators

2.1 The "discontinuity" and the "bias" problems

Many macroeconomic time series are persistent, that is highly serially autocorrelated. It is the case, in particular, for interest rates which will be central variables of this study. Because of the persistence property of these time series, we will necessarily have to face two important problems, namely the "discontinuity" problem and the "bias" problem. The discontinuity problem is the huge difference between predictions, in particular long-run predictions, based on models taking into account unit root and cointegration constraints and on unconstrained models. In the former class of models the predictions stay close to actual values whereas in the latter class the forecasts move toward the marginal mean [see Cochrane and Piazzesi (2008) and Section 3 of the present paper]. The bias problem is a very old one. Kendall (1954) stressed that the OLS estimator of the correlation ρ in the AR(1) model is severely downward biased in finite sample, especially when this correlation is close to one, and he proposed an approximation of this bias. Since this pioneering paper, many other studies have considered this problem [see e.g. Marriott and Pope (1954), Evans and Savin (1981) and Shaman and Stine (1988)]. In the following sections we will present some simulation studies, in order to evaluate these problems, and we will propose some possible solution [see Jardet, Monfort and Pegoraro (2009a) for details].

2.2 Finite sample distribution and bias

There exists a large literature which considers the behavior of the OLS estimator $\hat{\rho}_T$ of the autoregressive coefficient ρ when ρ is close to one, by introducing various "near to unit root asymptotics". Although this literature provides interesting theoretical results, it fails to give a clear message to practitioners because these results heavily depend on elements which are difficult to take into account in practice, like the behavior of the initial values (of the AR process) or the rate of convergence [see Elliott (1999), Elliott and Stock (2001), Muller and Elliott (2003), Giraitis and Phillips (2006), Magdalinos and Phillips (2007) and Andrews and Guggenberger (2007)]. So, to get a precise idea of the finite sample behavior of $\hat{\rho}_T$, we use simulation techniques in the simple AR(1) model $y_t = \mu(1-\rho) + \rho y_{t-1} + \varepsilon_t$, $t \in \{1, \ldots, T\}$, where the ε_t's are independently distributed as $N(0, \sigma^2)$, $y_0 = \mu$ and $T = 160$, which is a typical sample size in empirical studies based on quarterly data (a bivariate model will be also considered in Section 2.3). The empirical pdf of the OLS estimator $\hat{\rho}_T$ of ρ (which does not depend on μ and σ^2) are given in figure 1 for $\rho \in \{0.91, 0.95, 0.99\}$.

These distributions are clearly left skewed and far from the asymptotic distribution $N(\rho, (1 - \rho^2)/T)$. If we focus on the bias $b_T(\rho) = E_\rho(\hat{\rho}_T)$, we can first evaluate it by the Kendall's approximation $-(1 + 3\rho)/T$, but the exact bias presented in figure 2 is even much worse when ρ is close to one. For instance, for $\rho = 0.99$ the Kendall's approximation of the bias is 0.025 while the true value is 0.034. It is clear that this bias may have dramatic consequences for predictions since the behavior of ρ^q and $(\rho - 0.034)^q$ are very different for ρ close to one and q large (for instance, when $q = 20$ quarters). Figure 2 also gives a quadratic spline approximation of $b_T(\rho)$ which will be useful in Section 2.3.
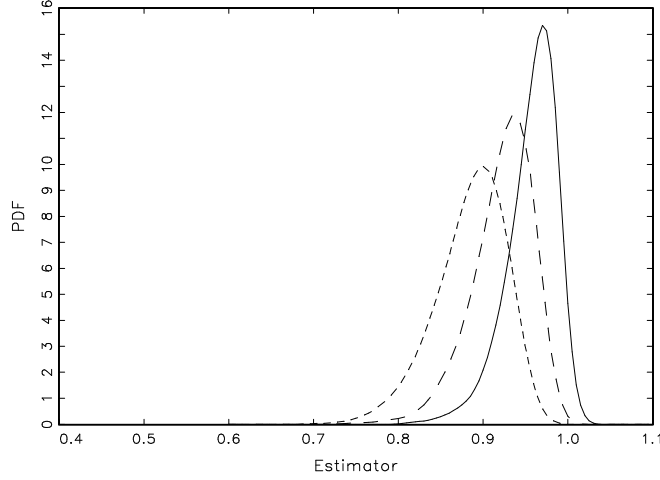
4

Figure 1:
PDF of the OLS estimators of ρ with sample size $T = 160$.
$\rho = 0.91$ (short dashes), 0.95 (dashes), 0.99 (solid line).

2.3 Prediction performances of bias-corrected and averaging estimators

It seems natural, given the severe finite sample bias of the OLS estimator $\hat{\rho}_T$ and the focus of this paper on interest rates forecasts, to consider different bias-corrected estimators and to evaluate their prediction performances. The bias-corrected estimators considered here are:

i) the indirect inference estimator [see Gourieroux, Monfort and Renault (1993), Gourieroux and Monfort (1996) and Gourieroux, Touzi and Renault (2000)] defined by:

$$\hat{\rho}_T^I = e_T^{-1}(\hat{\rho}_T), \tag{1}$$

where $e_T(\rho) = E_\rho(\hat{\rho}_T) = \rho + b_T(\rho)$.

ii) the bootstrap bias-corrected estimator [see Hall (1997), chap. 1] defined by:

$$\hat{\rho}_T^B = 2\hat{\rho}_T - e_T(\hat{\rho}_T). \tag{2}$$

iii) the Kendall's estimator:

$$\hat{\rho}_T^K = \hat{\rho}_T - \left(-\frac{1 + 3\hat{\rho}_T}{T} \right) = \left(1 + \frac{3}{T} \right) \hat{\rho}_T + \frac{1}{T}. \tag{3}$$

iv) the median-unbiased estimator [see Andrews (1993)]:

$$\hat{\rho}_T^M = m_T^{-1}(\hat{\rho}_T), \tag{4}$$

where $m_T(\rho)$ is the median of $\hat{\rho}_T$ when ρ is the true value.
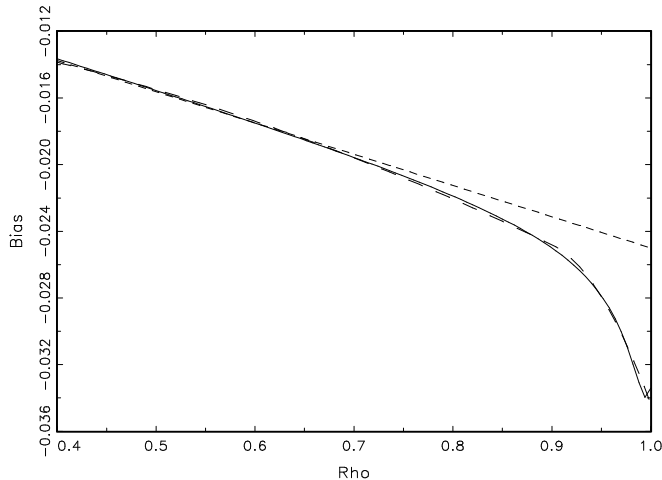
5

Figure 2:
Bias of the OLS estimator $\hat{\rho}_T$ with sample size $T = 160$.
Exact (solid line), Spline approximation (dashes), Kendall's approximation (short dashes).

The estimators $\hat{\rho}_T^I$, $\hat{\rho}_T^B$, and $\hat{\rho}_T^K$ can be shown to be mean-unbiased at order T^{-1}, and $\hat{\rho}_T^M$ is exactly median-unbiased. Figure 3 shows these estimators as functions of $\hat{\rho}_T$ and we see that the more important correction is provided by $\hat{\rho}_T^I$. In practice, all these estimators will be truncated at one. We will also consider another kind of estimators, namely the class of "averaging estimators" proposed by B. Hansen (2009) and defined, in our context, by:

$$\hat{\rho}_T^A(\lambda) = (1 - \lambda) + \lambda\hat{\rho}_T, \ \ 0 \leq \lambda \leq 1 \,. \tag{5}$$

Figure 4 shows the bias of these various estimators when $\rho = 0.99$ (based on 5×10^5 simulations) and we see that, among the bias-corrected estimators, the best correction is provided by the indirect inference one, which is in line with the results given in Gourieroux, Phillips and Yu (2007) and Phillips and Yu (2009). Figure 5 provides the root mean square error (RMSE) of these estimators and it is clear that the optimal averaging estimator (obtained for $\lambda \approx 0.15$) is much better than all other ones: $\hat{\rho}_T^A(0.15)$ provides a RMSE at least five times smaller than the one obtained by the bias-corrected estimators and nine times smaller than the OLS estimator.

Nevertheless, since we are mainly interested in forecast performances, we also compare these estimators in terms of root mean squared forecast errors (RMSFE) ratio, computed again from 5×10^5 estimations and out-of-sample predictions. Each ratio is calculated by dividing the RMSFE of each estimator by the one obtained from the true model. Figures 6 and 7 provide these RMSFE ratios for prediction horizons $q = 1$ and $q = 20$, respectively. The performance of the optimal averaging estimator (obtained with $\lambda \approx 0.25$ both for $q = 1$ and $q = 20$) is by far the best one: the percentage of increase of the RMSFE (compared to the one obtained from the true model) is
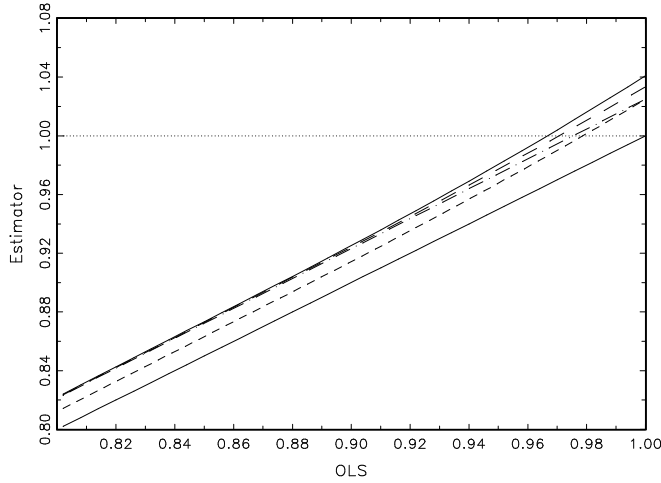
6

Figure 3:
Bias-corrected estimators with sample size $T = 160$.
Indirect Inference (upper solid line), Bootstrap (dashes), Median (short dashes),
Kendall (dots and dashes), OLS (lower solid line).

about four times smaller than for the bias-corrected estimators and six times smaller for the OLS estimator.

Similar conclusions are obtained for different values of T and ρ [see Jardet, Monfort and Pegoraro (2009a)] and these results confirm those obtained by B. Hansen (2009) short-run predictions using a near to unit root approach in an univariate AR model.

Let us also consider a "near-cointegrated" bivariate model, defined by:

$$
\begin{aligned}
y_{1t} &= (1 - \rho) + \rho y_{1,t-1} + \varepsilon_{1t} & (6) \\
y_{2t} &= 2 y_{1t} + \varepsilon_{2t} & (7)
\end{aligned}
$$

with $\rho \in \{0.97, 0.98, 0.99\}$, where ε_{1t} and ε_{2t} are independent standard Gaussian white noises. In order to define an averaging estimator we consider, on the one hand, an unconstrained VAR(1) model

$$
y_t = \nu + A y_{t-1} + \eta_t \,, \tag{8}
$$

where the unconstrained OLS estimators of ν and A are denoted by $\hat{\nu}_T^{(u)}$ and $\hat{A}_T^{(u)}$ and, on the other hand, we consider an error correction model imposing one cointegration relationship:

$$
\Delta y_t = \mu + \alpha(y_{1,t-1} - \beta y_{2,t-1}) + \xi_t \,, \tag{9}
$$

and the associated constrained estimators of ν and A:

$$
\hat{\nu}_T^{(c)} = \hat{\mu}_T \text{ and } \hat{A}_T^{(c)} = I + \hat{\alpha}_T(1, -\hat{\beta}_T) \,,
$$

7

Figure 4:
Bias with $\rho = 0.99$ and $T = 160$.
$\hat{\rho}_T^A(\lambda)$ (solid curve), $\hat{\rho}_T^I$ (solid line),
$\hat{\rho}_T^B$ (dashes), $\hat{\rho}_T^M$ (short dashes).
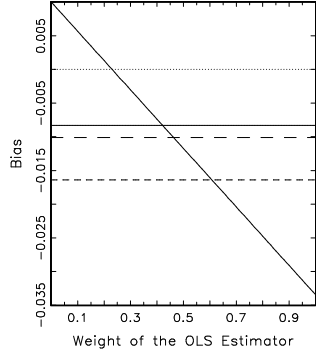

Figure 5:
RMSE with $\rho = 0.99$ and $T = 160$.
$\hat{\rho}_T^A(\lambda)$ (solid curve), $\hat{\rho}_T^I$ (solid line),
$\hat{\rho}_T^B$ (dashes), $\hat{\rho}_T^M$ (short dashes).

Figure 6:
RMSFE ratio with $\rho = 0.99$, $T = 160$, $q = 1$.
$\hat{\rho}_T^A(\lambda)$ (solid curve), $\hat{\rho}_T^I$ (solid line),
$\hat{\rho}_T^B$ (dashes), $\hat{\rho}_T^M$ (short dashes).

Figure 7:
RMSFE ratio with $\rho = 0.99$, $T = 160$, $q = 20$.
$\hat{\rho}_T^A(\lambda)$ (solid curve), $\hat{\rho}_T^I$ (solid line),
$\hat{\rho}_T^B$ (dashes), $\hat{\rho}_T^M$ (short dashes).

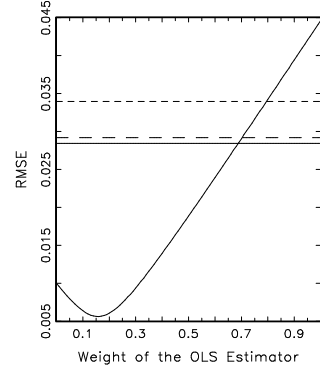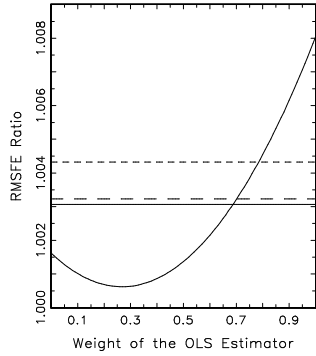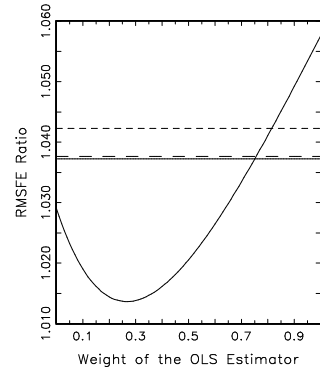where $\hat{\beta}_T$ is obtained by regressing y_{1t} on y_{2t}, while $\hat{\mu}_T$ and $\hat{\alpha}_T$ are obtained by regressing Δy_t on $(1, y_{1,t-1} - \hat{\beta}_T y_{2,t-1})$.

The class of averaging estimators is:

$$\hat{\nu}_T(\lambda) = (1-\lambda)\hat{\nu}_T^{(c)} + \lambda\hat{\nu}_T^{(u)},$$

$$\hat{A}_T(\lambda) = (1-\lambda)\hat{A}_T^{(c)} + \lambda A_T^{(u)}, \ \ 0 \le \lambda \le 1, \tag{10}$$

and the predictions of y_{T+h} at T are :

$$\hat{y}_{T,h}(\lambda) = [I - \hat{A}_T(\lambda)]^{-1}[I - \hat{A}_T^h(\lambda)]\hat{\nu}_T(\lambda) + \hat{A}_T^h(\lambda)y_T . \tag{11}$$

Figures 8 to 11 provide the RMSFE ratios for y_1 and y_2 when $q = 1$ and $q = 20$, each figure providing these ratios as functions of λ and $\rho \in \{0.97, 0.98, 0.99\}$. We see that we still have a clear minimum for a λ between 0 and 1, that this minimum depends on ρ and that for $q = 20$ the minimum values of λ are similar for y_1 and y_2. However, for $q = 1$ these values are different and, therefore, the choice of the variable of interest have some impact.

Finally, all these examples suggest the following strategy that will be adopted in this paper. If we denote by $\underline{y_T} = (y_1, \ldots, y_T)$ the observations and by $g(\underline{y_t})$, $t \in \{1, \ldots, T\}$ a variable of interest that we want to predict accurately at horizon h, the strategy we suggest is as follows :

- define a sequence of increasing windows $\{1, \ldots, t\}$, with $t \in \{t_0, \ldots, T-q\}$;

- for each t compute the unconstrained estimator $\hat{\theta}_t^{(u)}$ and the constrained estimator $\hat{\theta}_t^{(c)}$ of the parameter θ;

- for each t compute the class of averaging estimators $\hat{\theta}_t(\lambda) = (1-\lambda)\hat{\theta}_t^{(c)} + \lambda\hat{\theta}_t^{(u)}$, the corresponding predictions $\hat{g}_{t,q}(\lambda)$ of $g(\underline{y_{t+q}})$ and the prediction error $g(\underline{y_{t+q}}) - \hat{g}_{t,q}(\lambda)$;

- compute $Q_T(\lambda, q) = \sum_{t=t_0}^{T-q} [g(\underline{y_{t+q}}) - \hat{g}_{t,q}(\lambda)]^2$;

- calculate $\lambda^*(q) = argmin_{\lambda \in [0,1]} Q_T(\lambda, q)$;

- compute $\hat{\theta}_T(\lambda^*(q))$.

Also note that it would be possible to minimize a criterion taking into account several variables of interest and/or several prediction horizons.

3   Description of the Data

The data set that we consider in the empirical analysis contains 174 quarterly observations of U. S. zero-coupon bond yields, for maturities 1, 2, 3, 4, 8, 12, 16, 20, 24, 28, 32, 36 and 40 quarters, and U. S. real GDP, covering the period from 1964:Q1 to 2007:Q2. The yield data are obtained from Gurkaynak, Sack, and Wright (2007) [GSW (2007), hereafter] data base and from their estimated Svensson (1994) yield curve formula. In particular, given that GSW (2007) provide interest rate values at daily frequency, each observation in our sample is given by the daily value observed at the end of each quarter. The same data base is used by Rudebusch, Sack, and Swanson (2007) [RSS (2007), hereafter] in their study on the implications of changes in bond term premiums on economic

Figure 8:
RMSFE ratio for y_1 when $T = 160$, $q = 1$.
$\rho = 0.97$ (solid), $\rho = 0.98$ (dashes),
$\rho = 0.99$ (short dashes).

Figure 9:
RMSFE ratio for y_2 when $T = 160$, $q = 1$.
$\rho = 0.97$ (solid), $\rho = 0.98$ (dashes),
$\rho = 0.99$ (short dashes).

Figure 10:
RMSFE ratio for y_1 when $T = 160$, $q = 20$.
$\rho = 0.97$ (solid), $\rho = 0.98$ (dashes),
$\rho = 0.99$ (short dashes).

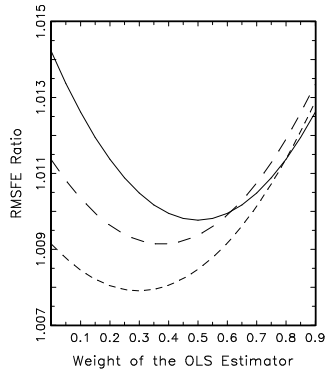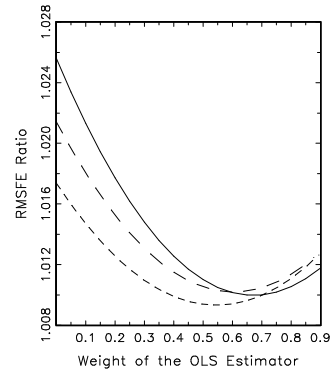Figure 11:
RMSFE ratio for y_2 when $T = 160$, $q = 20$.
$\rho = 0.97$ (solid), $\rho = 0.98$ (dashes),
$\rho = 0.99$ (short dashes).
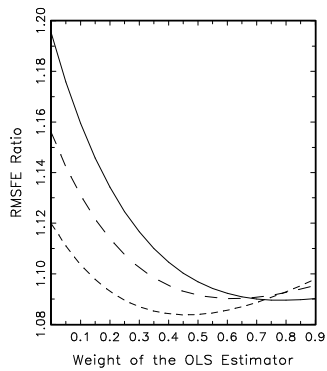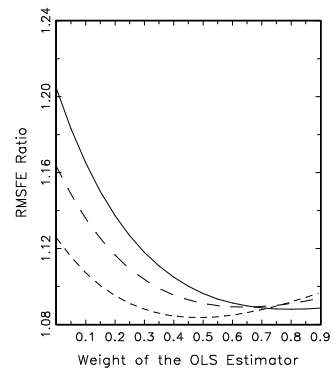
activity. Observations about real GDP are seasonally adjusted, in billions of chained 2000 dollars, and taken from the FRED database (GDPC1).

In the data base they provide, GSW (2007) do not propose (over the entire sample period,

Yields	1-Q	4-Q	8-Q	12-Q	16-Q	20-Q	40-Q	G_t	g_t
Mean	0.015	0.016	0.016	0.017	0.017	0.017	0.018	8.709	0.008
Std. Dev.	0.007	0.007	0.007	0.006	0.006	0.006	0.006	0.385	0.008
Skewness	1.029	0.846	0.842	0.863	0.886	0.909	0.982	-0.025	-0.085
Kurtosis	4.691	4.136	4.008	3.880	3.782	3.720	3.643	1.830	4.487
Minimum	0.002	0.003	0.003	0.004	0.005	0.006	0.009	7.990	-0.020
Maximum	0.040	0.039	0.039	0.039	0.038	0.038	0.037	9.352	0.039
ACF(1)	0.910	0.932	0.940	0.946	0.951	0.955	0.959	0.981	0.268
ACF(4)	0.760	0.788	0.805	0.817	0.826	0.831	0.842	0.925	0.093
ACF(8)	0.513	0.581	0.627	0.658	0.679	0.693	0.717	0.853	-0.167
ACF(12)	0.335	0.426	0.494	0.538	0.566	0.585	0.616	0.785	-0.170
ACF(16)	0.240	0.307	0.365	0.404	0.430	0.448	0.482	0.718	0.004
ACF(20)	0.224	0.252	0.283	0.308	0.325	0.336	0.356	0.655	0.127

Table 1: Summary Statistics on U.S. Quarterly Yields, log-GDP (G_t) and one-quarter GDP growth rate (g_t) observed from 1964:Q1 to 2007:Q2 [Gurkaynak, Sack and Wright (2007) data base]. ACF(k) indicates the empirical autocorrelation with lag k expressed in quarters.

ranging from 1961 to 2007), yields with maturities shorter than one year. Moreover, they calculate yields with 8, 9 and 10 years to maturity only after (mid-)August, 1971. Our construction of the interest rate time series with 3, 6 and 9 months to maturity, based on the Svensson (1994) formula estimated by GSW (2007), is justified by the fact that they estimate this formula using Treasury notes and bonds with at least three months to maturity. The construction of the three long-term interest rate time series before 1971 is justified [as indicated by RSS (2007, footnote 26), for the 10-years yield-to-maturity] by the fact that (even if there were few bond observations with these maturities), the reconstructed time series are highly correlated with other well known and widely used time series [like, for instance, the FRED interest rates data base (Treasury Constant Maturity interest rates), or the McCulloch and Kwon (1993) data base]. Moreover, in order to be coherent with the literature and, in particular, with the majority of the papers concerned with the predictive ability of the term spread for GDP [see, for instance, Fama and Bliss (1987), and Ang, Piazzesi and Wei (2006)], we have decided to start the sample period in 1964.

Summary statistics about the yields (expressed on a quarterly basis), the real log-GDP and its first difference are presented in Table 1. The average yield curve is upward sloping, and interest rates with larger standard deviation, skewness and kurtosis are those with shorter maturities. Furthermore, yields are highly autocorrelated with an autocorrelation which is, for any given lag, increasing with the maturity and, for any given maturity, decreasing with the lag. The high persistence in log-GDP strongly reduces when we move to its first difference (the one-quarter GDP growth rate).

The short rate (r_t) and the long rate (R_t) used in this paper are, respectively, the 1-quarter and 40-quarter yields, and the log-GDP at date t is denoted by G_t. These three variables, collected in the vector Y_t, constitute the information that investors use to price bonds.

4   Near-Cointegration Analysis

The purpose of this section is to present the first two steps of the modelling procedure we follow to specify and implement the Near-Cointegrated VAR(p) class of affine term structure models. In

particular, in Section 4.1, we apply a cointegration analysis to the autoregressive dynamics of the vector $Y_t = (r_t, R_t, G_t)'$, suggested by classical and efficient unit root tests presented in Section 4.1.1 (first step). This econometric procedure lead us to a vector error correction model (with two lags) for ΔY_t, that we can write as a Cointegrated VAR(3) for $X_t = (r_t, S_t, g_t)'$, the spread $S_t = R_t - r_t$ being the cointegrating relationship (Section 4.1.2).

This specification has the advantage to explain the persistence in interest rates better than the unconstrained counterpart given by a VAR(3) model for X_t, but has two important drawbacks. First, it assumes the non-stationarity of interest rates, while a wide literature on nonlinear models indicates that they are highly persistent but stationary [see, for instance, Gray (1996) and Ang and Bekaert (2002), and the references therein]. Second, as indicated by Cochrane and Piazzesi (2008), interest rate forecasts over long horizons, coming from alternative CVAR and VAR specifications, have very different behaviors because of the discontinuity problem induced by the presence or not of unit roots. As a consequence, important differences are found about the term premia extraction. The methodology we follow to solve this problem is presented in Section 4.2. More precisely, this discontinuity problem is discussed in Section 4.2.1 and the methodology we follow to solve it is presented in Section 4.2.2 (second step). The third step of the modelling procedure is presented in Section 5, where we introduce the parametric exponential-affine Stochastic Discount Factor (SDF), we obtain the yield-to-maturity formula and we estimate risk sensitivity parameters by Constrained NLLS (CNLLS).

4.1 A Vector Error Correction Model of the State Variables

4.1.1 Unit Root Tests

The first step of our modelling start studying the presence of unit roots in the short rate, long rate and real log-GDP time series. We apply not only classical unit root tests, like the Augmented Dickey-Fuller (ADF) tests (t test and F test), and the Phillips-Perron (PP) test, but also the (so-called) efficient unit root tests proposed in the paper of Elliott, Rothenberg and Stock (1996) [Dickey-Fuller test with GLS detrending (denoted Dickey-Fuller GLS), and Point-Optimal test], and in the work of Ng and Perron (2001) (denoted Ng-Perron). It is well known that ADF and PP tests have size distortion and low power against various alternatives, and against trend-stationary alternatives when conventional sample size are considered [see, for instance, De Jong, Nankervis, Savin and Whiteman (1992a, 1992b), and Schwert (1989)]. For these reasons, we verify the presence of unit roots using also these efficient unit root tests which have more power against persistent alternatives, like the time series we analyze [see Table 1].

The results are the following. With regard to the short rate and the long rates, Table A.1 shows that for both series, and for all tests, we accept (at 5% or 10% level) the hypothesis of unit root without drift. As far as the real log-GDP level is concerned, the hypothesis of unit root is accepted at 10 % level and for every test when a constant is included in the test regression (see left panel of Table A.2). When, also a linear time trend is included in the test regression (see Table A.2, right panel), the hypothesis of unit root in the time series G_t is rejected at 1 % level by the ADF test, and at the 5 % level by the PP test. Nevertheless, when we consider the efficient unit root tests, the hypothesis of unit root is always accepted at 10% level and for each test. Given the better power properties of efficient unit root tests, with respect to ADF and PP tests, we are lead to accept the hypothesis of unit root in G_t. We have also applied unit root tests to the components of ΔY_t, and we always reject the unit root hypothesis.

The results presented above suggest that short rate, long rate and log-GDP are I(1) time series, thus, Y_t is a I(1) process [in the Engle and Granger (1987) sense, that is, a vectorial process in which all scalar components are integrated of the same order]. The purpose of the next section is to search for long-run equilibrium relationships (common stochastic trends) among the components of Y_t, using cointegration techniques.

4.1.2 Cointegration Analysis and State Dynamics Specification

We study the presence of cointegrating relationships among the short rate, long rate and log-GDP time series using the (VAR-based) Johansen (1988, 1995) Trace and Maximum Eigenvalue tests. First, we assume that the I(1) vector $Y_t = (r_t, R_t, G_t)'$ can be described by a 3-dimensional Gaussian VAR(p) process of the following type:

$$Y_t = \nu + \sum_{j=1}^{p} \Phi_j Y_{t-j} + \varepsilon_t \,, \tag{12}$$

where ε_t is a 3-dimensional Gaussian white noise with $\mathcal{N}(0, \Omega)$ distribution [Ω denotes the (3×3) variance-covariance matrix]; Φ_j, for $j \in \{1, \ldots, p\}$, are (3×3) matrices, while ν is a 3-dimensional vector. On the basis of several lag order selection criteria (and starting from a maximum lag of $p = 4$, in order to make the following estimation of risk-neutral parameters feasible), the lag length is selected to be $p = 3$ (see Table A.3), and the OLS estimation of the (unrestricted) VAR(3) model is presented in Table A.4. Then, we write the Gaussian VAR(3) model in the (equivalent) vector error correction model (VECM) representation :

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{j=1}^{2} \Gamma_j \Delta Y_{t-j} + \nu + \varepsilon_t \,,$$
$$\text{with} \quad \Pi = - \left[I_{3 \times 3} - \sum_{j=1}^{3} \Phi_j \right] \text{ and } \Gamma_j = - \sum_{i=j+1}^{3} \Phi_i \,, \tag{13}$$

and we determine the rank $r \in \{0, 1, 2, 3\}$ of the matrix Π using the (likelihood ratio) trace and maximum eigenvalue tests. The rank(Π) gives the number of cointegrating relations (the so-called cointegrating rank, that is, the number of independent linear combinations of the variables that are stationary), and $(3 - r)$ the number of unit roots (or, equivalently, the number of common trends). The results, presented in the first part of Table A.5, indicate that both tests accept the presence of one cointegrating relation ($r = 1$) at 5 % level, and, thus, they decide for the presence of two unit roots in the vector Y_t. Consequently, we can write $\Pi = \alpha \beta'$, where α and β are (3×1) vectors (the second part of Table A.5 provides the maximum likelihood parameter estimates of these matrices), and $\beta' Y_t$ will be $I(0)$ [see Engle and Granger (1987) and Johansen (1995)].

Observe that, the cointegration analysis is based on the model specification (13), in which the unrestricted constant term ν induces a linear trend in Y_t. Given the decomposition $\nu = \alpha \mu + \gamma$ (with μ a scalar determined so that the error correction term has a sample mean of zero, and γ a (3×1) vector), we have tested the null hypothesis $H_0 : \nu = \alpha \mu$ (the intercept is restricted to lie in the α direction) using the $\chi^2(2)$-distributed (under H_0) likelihood ratio statistic $\tilde{lr}$ taking the value 13.9354 which is larger than the chi-square 1 % quantile (with two degrees of freedom)

$\chi^2_{0.01}(2) = 9.21$. Consequently, the null hypothesis is rejected, which implies a drift in the common trends[5].

Moreover, in order to achieve economic interpretability of the cointegrating relation, we have tested the null hypothesis $H_0 : \beta = (-1, 1, 0)'$ (the spread between the long and the short rate is the cointegrating relation) using the likelihood ratio statistic lr^* taking the value[6] 3.276, which is smaller than the chi-square 5 % quantile (with two degrees of freedom) $\chi^2_{0.05}(2) = 5.99$. Consequently, the null hypothesis is accepted, and, therefore, the spread provides the long-run equilibrium relationship[7]. Least squares parameter estimates of model (13), when $\Pi = \alpha\beta'$, with $\beta = (-1, 1, 0)'$, and $\nu = \alpha\mu + \gamma$, are presented in Table A.6. Observe that, the same kind of model specification (a VECM with two lags in differences, one cointegrating relation given by the spread and an unrestricted constant term) is obtained when the 5-years yield is considered as the long rate, when the analysis is applied to the same sample period (1964:Q1 - 2001:Q4), or the same data base[8], as in APW (2006) [the results are available upon request from the authors].

In order to propose a direct comparison between the performances of our model (under the historical and the risk-neutral probability) and the one proposed by APW (2006), we rewrite model (13) in terms of the 3-dimensional state process $X_t = (r_t, S_t, g_t)'$, with $S_t = R_t - r_t$ and $g_t = G_t - G_{t-1}$:

$$X_t = \tilde{\nu} + \sum_{j=1}^{3} \tilde{\Phi}_j X_{t-j} + \tilde{\eta}_t, \qquad (14)$$

$$\text{with} \quad \tilde{\nu} = A\nu, \ A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\tilde{\Phi}_1 = \tilde{\Gamma}_1 + \tilde{\alpha}(0, 1, 0) + B, \ \tilde{\Phi}_2 = \tilde{\Gamma}_2 - \tilde{\Gamma}_1 B, \ \tilde{\Phi}_3 = -\tilde{\Gamma}_2 B,$$

$$\tilde{\Gamma}_i = A\Gamma_i A^{-1} \ \text{for} \ i \in \{1, 2\}, \ B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ \tilde{\alpha} = A\alpha,$$

and where $\tilde{\eta}_t$ is a 3-dimensional Gaussian white noise with $\mathcal{N}(0, \tilde{\Omega})$ distribution and $\tilde{\Omega} = \Sigma\Sigma' = A\Omega A'$ [the parameter estimates are presented in Table A.7, while the estimates of the APW (2006) state dynamics are organized in Table A.8], where Σ is assumed to be lower triangular. Note that the third column of $\tilde{\Phi}_3$ is a vector of zeros. This Cointegrated VAR (3) model [CVAR(3), hereafter]

---

[5]The likelihood ratio statistic is $\tilde{lr} = -T \sum_{k=2}^{3} \log[(1-\tilde{\lambda}_k)/(1-\lambda_k)]$, where $(\tilde{\lambda}_2, \tilde{\lambda}_3)$ and (λ_2, λ_3) are, respectively, the two smallest eigenvalues associated to the maximum likelihood estimation of the restricted (under H_0) and unrestricted model (13). The estimation of the two models leads to $(\tilde{\lambda}_2, \tilde{\lambda}_3) = (0.0962431, 0.032958)$ and $(\lambda_2, \lambda_3) = (0.039789, 0.008368)$.

[6]The likelihood ratio statistic is $lr^* = -T \log[(1-\lambda_1^*)/(1-\lambda_1)]$ ($\chi^2(2)$-distributed under the null), where λ_1^* is the largest eigenvalue associated to the maximum likelihood estimation of model (13) under H_0.

[7]Many authors have found cointegration between short-term and long-term interest rates, and the existence of long-run equilibrium relationships given by the spread [see Campbell and Shiller (1987), Engle and Granger (1987), Hall, Anderson and Granger (1992)].

[8]We are very grateful to Andrew Ang, Monika Piazzesi and Min Wei for providing us the data set.

can equivalently be represented in the following 9-dimensional VAR(1) form:

$$\tilde{X}_t = \tilde{\Phi}\tilde{X}_{t-1} + e_1[\tilde{\nu} + \tilde{\eta}_t],$$

$$\text{where} \quad \tilde{\Phi} = \begin{bmatrix} \tilde{\Phi}_1 & \tilde{\Phi}_2 & \tilde{\Phi}_3 \\ I_{3\times 3} & 0_{3\times 3} & 0_{3\times 3} \\ 0_{3\times 3} & I_{3\times 3} & 0_{3\times 3} \end{bmatrix}, \quad \tilde{X}_t = (X'_t, X'_{t-1}, X'_{t-2})', \tag{15}$$

and where e_1 is a (9×3) matrix equal to $(I_3, 0_3, 0_3)'$.

4.2 Near-Cointegrated VAR(p) Dynamics

4.2.1 A Discontinuity Problem

It is well known that moving from a stationary environment to a non-stationary one, implies various types of discontinuity problems, in particular in term of asymptotic behavior of the estimation or testing procedure (see e.g. Chan and Wei (1987), Phillips (1987, 1988), Phillips and Magdalinos (2007)) or in term of prediction (see e.g. Stock (1996), Kemp (1999), Diebold and Kilian (2000), Elliott (2006)). In the context of macro-finance VAR modelling, Cochrane and Piazzesi (2008) also noted very different long term interest rates predictions depending whether unit roots constraints are imposed or not (see figures 12 and 13). In the VAR context this discontinuity simply comes from the fact that the long run behavior of predictions is driven by roots of the determinant of the autoregressive matrix polynomial and that this behavior becomes very different as soon as at least one unit root is present.

As an illustration, we consider the q-step ahead short rate forecasts obtained from the CVAR(3) and an unconstrained VAR(3) model for X_t (see Table A.9 for its parameter estimates). The forecasts are displayed in figures 12 and 13, respectively, for $q = 1$, 4, 8, 12, 16 and 20 quarters. We observe that the forecasts of the short rate differ sharply depending on the considered model. More specifically, with a VAR(3) model, forecasts tend to quickly converge to the unconditional mean of the short rate as far as the forecast horizon increases. In contrast, when a unit root constraint is imposed (like in the CVAR(3) model), forecasts obtained at all horizons are very similar, and very close to the present short rate. This sharp difference is due to the fact that model (14) imposes a unit root in the determinant of the autoregressive lag operator, whereas in the unconstrained VAR(3) specification the largest root is found to be equal to 0.93 (see Table A.9).

4.2.2 Handling the Discontinuity Problem

The discontinuity problems can be tackled in different ways. First, it would be possible to try to extend to macrofinance models the switching regime approach which has been used successfully in pure finance models [see Bansal and Zhou (2002), Bansal, Tauchen and Zhou (2004), Dai, Singleton and Yang (2007) and Monfort and Pegoraro (2007)] and thus checking how persistence properties are transformed within each regime [see Evans (2003) and Ang, Bekaert and Wei (2008)]. Second, a bayesian approach would also be interesting provided that a sensitivity of the results to the choice of the prior (informative prior, Jeffreys or flat prior) is taken into account, since the behavior of the prior near the unit root is an important issue [see Sims and Uhlig (1991), Uhlig (1994)]. Third, we could try to use fractionally integrated processes and the generalized notion of cointegration in this
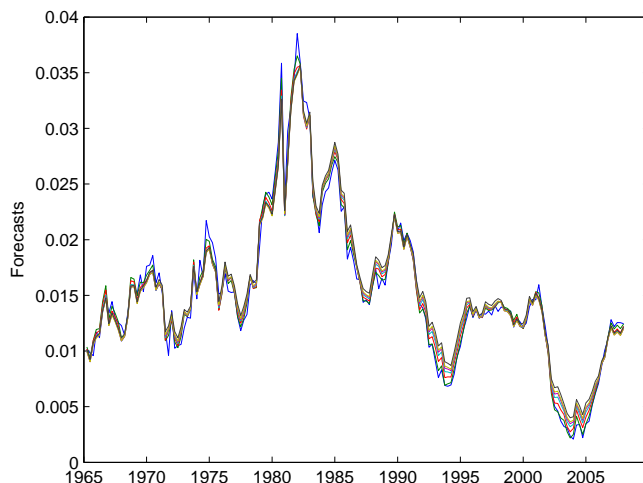
Figure 12:
q-step ahead short rate forecasts from the CVAR(3) model
$q = 1, 4, 8, 12, 16, 20, 40$ quarters.

framework and to cope with technical problems appearing in this kind of approach, in particular the possible slow rate of convergence of some estimators [see, among the others, Geweke and Porter-Hudak (1983), Sowell (1992), Agiakloglou, Newbold and Wohar (1993), Robinson (1995)].

In this paper we adopt a fourth approach resting on the averaging estimators considered in Section 2 and proposed by B. Hansen (2009). Hansen's results have been derived in a univariate and one-step-ahead framework and their generalization to a multivariate and multi-horizon setting raise difficult technical problems, in particular the multiplicity of the parameter paths leading to the constrained VAR at rates proportional to $1/T$. So we have decided to adopt a pragmatic approach and, extrapolating the Monte Carlo results of Section 2, we have checked empirically whether the out-of-sample mean square forecast errors, when forecasting some variable of interest at various horizons, are improved when using an average estimator based on the VAR(3) and CVAR(3) models. As explained below, our empirical findings thoroughly confirm Hansen (2009)'s theoretical results.

4.2.3   Averaging Estimations and Extraction of Short Rate Expectations

The Near-Cointegrated VAR(3) class of models for the state vector X_t is obtained in the following way: once we have estimated by OLS the vector of the unconstrained VAR(3) parameters θ_{var} (parameter estimates are presented in Table A.9) and the vector of parameters θ_{cvar} of the CVAR(3) model (see Table A.7), the of averaging estimators specifying the Near-Cointegrated VAR(3) class is given by:

$$\theta_{nc} = \theta_{nc}(\lambda) = \lambda\theta_{var} + (1 - \lambda)\theta_{cvar}, \tag{16}$$

with $\lambda \in [0, 1]$ a free parameter selected to minimize a criterion of interest. In particular, given our aim to provide a reliable measure of the term premia on long term bonds, we focus on minimizing
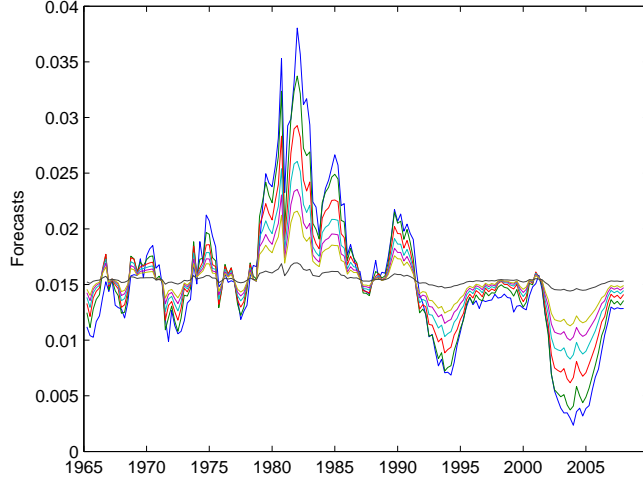
16

Figure 13:
q-step ahead short rate forecasts from the VAR(3) model
$q = 1, 4, 8, 12, 16, 20, 40$ quarters.

the prediction error of the associated expectation part. This choice could leave some uncertainty about the selected variable of interest and, thus, about the selected value of the averaging parameter specifying the near-cointegrated factor dynamics. We will see in Section 5.3.1 that, if we select λ, together with risk sensitivity parameters, by minimizing the yield curve fitting error, we find (in practice) for λ the same value as the one obtained with our preferred criterion.

Let us present now the criterion we consider to select the value of the averaging parameter. Given at date t a yield with residual maturity h, denoted by $R_t(h)$, we define its expectation term as $EX_t(h) = -\frac{1}{h} \log B_t^*(h)$ with $B_t^*(h) = E_t[\exp(-(r_t + r_{t+1} + ... + r_{t+h}))]$. The associated term premium is given by $TP_t(h) = R_t(h) - EX_t(h)$ (see Section 6 for a detailed presentation). For a given maturity h, the parameter $\lambda = \lambda(h)$ (say) is selected as the solution of the following problem:

$$\lambda^*(h) = arg \min_{\lambda \in [0,1]} \sum_t [\tilde{B}_t^*(h) - \hat{B}_t^*(h, \lambda)]^2 \qquad (17)$$

where, for each date t and residual maturity h, $\tilde{B}_t^*(h)$ is the observed realization of $\exp(-r_t - ... - r_{t+h-1})$ while $\hat{B}_t^*(h, \lambda)$ is the NCVAR(3) model implied $B_t^*(h)$, that is the model's forecast of $\exp(-r_t - ... - r_{t+h-1})$. The out-of-sample forecasts are performed during the period 1990:Q1 - 2007:Q2, using an expanding window for the estimation of models VAR(3) and CVAR(3). More precisely, we first estimate the parameters θ_{var} and θ_{cvar} over the period 1964:Q1 to 1989:Q4 and we calculate $\hat{B}_t^*(h, \lambda)$ with $t =$1989:Q4. Then, at each later point in time t, we re-estimate θ_{var} and θ_{cvar} taking into account the new observation and, in doing so, we replicate the typical behavior of an investor that incorporates new information over time (see also Favero, Kaminska and Sodestrom (2006)).

In table 2 we compare, for h ranging from 2 to 40 quarters, the RMFSE obtained from the NCVAR(3) model, with $\lambda^*(h)$ solution of (17), with those obtained by the CVAR(3), VAR(3), VAR(1) and AR(1) (based on the short rate) models. With regard to the NCVAR mechanism, when $\lambda^*(h) = 0$, the optimal forecasts of $B_t^*(h)$ are obtained from the CVAR(3) model, while, when

$\lambda^*(h) = 1$ the optimal forecasts come from the VAR(3) model. The case $0 < \lambda^*(h) < 1$ corresponds to predictions of $B_t^*(h)$ computed with the NCVAR(3) model, with a vector of parameters given by $\theta_{nc}^*(h) = \lambda^*(h)\theta_{var} + (1 - \lambda^*(h))\theta_{cvar}$. We observe that, for $h > 4$, the NCVAR(3) specification outperforms the VAR(3) and CVAR(3) models: there exist a $\lambda^*(h)$, strictly between 0 and 1, such that the average of the estimated parameters in the CVAR(3) and VAR(3) models improves the forecasts of $B_t^*(h)$ [see figure 14]. Even more, the NCVAR(3) model outperforms the (most competing) VAR(1) and AR(1) models (except for $h = 2$ for the AR(1) model); in particular, for long maturities, that is for short rate forecasts over long horizons, our model reduces their out-of-sample RMSFEs of 20-30%.

Since, in this work, one of the main objectives is to extract the term premium from the 40-quarters long term bond, we will assume that the NCVAR(3) state dynamics, driving term structure shapes over time and maturities, be specified by a $\lambda^*(40) = 0.2624$. This means that, the optimal extraction of the expectation part of the long term bond is obtained by a NCVAR model in which the weight of the CVAR(3) model is three times larger than the one of the VAR(3) model. This result could be interpreted not only as an indication of the high persistence in the short rate which asks for a larger weight for the model which mostly catch sources of serial dependence. But also, as a suggestion (given by the estimated value of λ) of a hierarchy among its dynamic properties. It seems that catching long term dynamics (persistence) in the short rate has the priority with respect to short term variability.

We will see in Section 5 that, even if we select λ by minimizing the fitting error of the yield curve, we will find $\lambda \approx 0.26$ again. This result seems to reinforce, at the same time, the reliability of our criterion, and also the above mentioned interpretation about the dominating role that persistence has in interest rates modelling. In order to deeply understand all the potentialities of the proposed NCVAR term structure model, we will also consider the case of a weighting parameter λ optimally selected on the basis of a criterion of interest like the forecast of state variables and yields over several horizons [see Sections 4.2.3 and 5.3].

	h	AR(1) (Vasicek)	$\lambda^*(h)$	NCVAR(3)	CVAR(3)	VAR(3)	VAR(1)
	2	0.0014	1.0000	0.0015	0.0015	0.0015	0.0016
	4	0.0071	1.0000	0.0064	0.0066	0.0064	0.0075
	8	0.0257	0.7911	0.0228	0.0237	0.0229	0.0261
	12	0.0476	0.5672	0.0423	0.0436	0.0429	0.0469
$\tilde{B}_t^*(h)$	16	0.0689	0.3471	0.0595	0.0607	0.0623	0.0667
	20	0.0882	0.1772	0.0736	0.0742	0.0806	0.0848
	28	0.1151	0.1782	0.0963	0.0979	0.1084	0.1114
	32	0.1229	0.2320	0.0991	0.1037	0.1135	0.1196
	36	0.1326	0.2598	0.1034	0.1121	0.1197	0.1304
	40	0.1411	0.2624	0.1012	0.1155	0.1224	0.1406

Table 2: Out-of-sample forecasts of $\tilde{B}_t^*(h) = \exp(-r_t - ... - r_{t+h-1})$. Table entries are associated RMSFEs. AR(1) (Vasicek) denotes forecasts of $\tilde{B}_t^*(h)$ using a Gaussian AR(1) process describing the dynamics of the (one-quarter) short rate. The time to maturities (h) are measured in quarters.
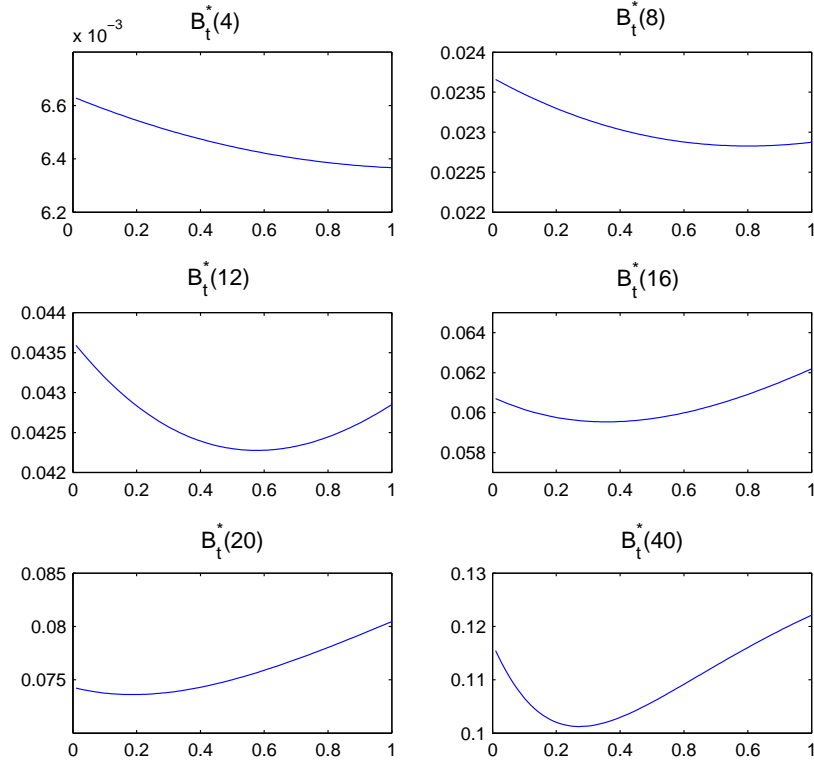
Figure 14:
X axis: λ; Y axis: RMSFE of $B_t^*(h)$ obtained from the NCVAR(3) model with vector of parameters $\theta_{nc} = \lambda\theta_{var} + (1 - \lambda)\theta_{cvar}$. $\lambda = 0$ corresponds to the CVAR(3) case, while $\lambda = 1$ corresponds to the VAR(3) case.

4.2.4 Short and Long Rate Out-of-Sample Forecasts with NCVAR(3) State Dynamics

In Section 4.2.2 we have seen that the specification of the expectation term of a zero-coupon bond $B_t(h)$, namely $B_t^*(h)$, is in general more precise when performed by our NCVAR(3) model. Moreover, besides the cases $h = 2$ and $h = 4$ quarters, $\lambda^*(h)$ is always inside the interval $[0, 1]$, indicating the advantage in using the averaging estimator to forecast $\tilde{B}_t(h)$, with respect to the extreme CVAR(3) and VAR(3) cases.

The purpose of the present section is to analyze the interest rates out-of-sample forecast performances that the NCVAR(3) state dynamics is able to produce. In particular, we study its ability to forecast the one-quarter short rate and the 10-years long rate in two main cases: a) when λ is selected to minimize, for each forecasting horizon q (say) and for each variable, the associated RMSFE; in this context λ is considered as a free parameter which gives a further degree of freedom in order to improve model's performances like, in this case, the forecast of a variable of interest

over a certain horizon; b) when the averaging parameter is fixed to $\lambda = 0.2624$, in order to establish the performances of the factor characterizing the yield-to-maturity formula of our selected term structure model [in Section 5.3 we will concentrate on the forecast of yields with maturities between 4 and 20 quarters]. As in Section 4.2.2, the out-of-sample forecast exercise is performed using an expanding window: we first estimate the parameters θ_{var} and θ_{cvar} over the period 1964:Q1 - 1989:Q4 and then, at each later point in time t, we re-estimate θ_{var} and θ_{cvar} taking into account the new observation.

The results, organized in Table 3, are presented for case a) and, then, for case b).

a) First, with regard to the optimal value of $\lambda = \lambda(q)$ (say) in the NCVAR(3) specification, we observe that, as far as q increases, $\lambda^*(q)$ decreases from $\lambda^*(q) = 1$ to $\lambda^*(q) = 0$. This result indicates that the minimization of the forecast error, when the forecasting horizon increases, gives an increasing weight to the CVAR(3) component and, thus, it indicates how important it is for obtaining reliable long-run forecasts. Second, our NCVAR(3) model outperforms, over both short and long forecasting horizons, the AR(1) and VAR(1) specifications. In particular, it is important to observe the remarkable performance about short rate long-horizon forecasts: the NCVAR(3) model reduces the RMSFE obtained by AR(1) and VAR(1) specifications of 45% when $q = 40$ quarters. This result, along with the forecast performance of the expectation term, highlights the ability of our approach to extract a reliable measure of term premia on long-term bonds.

	q	AR(1)	$\lambda^*(q)$	NCVAR(3)	NCVAR(3) $[\lambda = 0.2624]$	CVAR(3)	VAR(3)	VAR(1)
r_t	1	0.0015	1.0000	0.0015	0.0016	0.0016	0.0015	0.0016
	4	0.0044	0.7585	0.0039	0.0039	0.0040	0.0039	0.0045
	8	0.0066	0.5993	0.0061	0.0061	0.0063	0.0061	0.0065
	12	0.0075	0.4329	0.0069	0.0070	0.0071	0.0072	0.0072
	20	0.0081	0.0378	0.0067	0.0068	0.0067	0.0078	0.0078
	40	0.0106	0.0000	0.0058	0.0079	0.0058	0.0102	0.0107
R_t	1	0.0113	1.0000	0.0012	0.0012	0.0012	0.0012	0.0012
	4	0.0101	0.8894	0.0023	0.0024	0.0024	0.0023	0.0023
	8	0.0090	0.6469	0.0030	0.0031	0.0032	0.0031	0.0031
	12	0.0082	0.4707	0.0039	0.0040	0.0041	0.0041	0.0043
	20	0.0070	0.0000	0.0042	0.0045	0.0042	0.0056	0.0059
	40	0.0055	0.0000	0.0055	0.0064	0.0055	0.0081	0.0086

Table 3: Out-of-sample forecasts of the short and long rate. Table entries are RMSFEs. r_t denotes the (one-quarter) short rate and R_t is the 10-years interest rate. AR(1) denotes a Gaussian scalar autoregressive of order one process used to forecast, respectively, r_t and R_t. The forecasting horizons (q) are measured in quarters.

b) If we consider the forecast of the state variables obtained by the NCVAR(3) process with $\lambda = 0.2624$, the results we obtain are the following. With regard to the short rate, even if the averaging parameter is selected using the yield curve fitting criterion, the RMSFEs produced by our selected state process remain in general lower than those obtained by the AR(1) and VAR(1) models and, in particular, for long horizons. Indeed, for $q = 40$ quarters, our selected model reduces the RMSFE of the AR(1) and VAR(1) models of 25%. If we consider the long

rate, the forecast errors remain, in average, quite close to those obtained in case a), for short and middle forecasting horizons, while, for $q = 40$ quarters, they get slightly worse than the AR(1) forecasts, but still better than those obtained by the VAR(1) process.

5 Affine Term Structure Models

5.1 The Yield Curve Formula

In the previous sections we have specified (and estimated) the historical dynamics of the state variable X_t as a Near-Cointegrated Gaussian VAR(3) process with averaging parameter given by $\lambda^*(40) = 0.2624$. The following step is to derive the (arbitrage-free) yield-to-maturity formula by specifying a positive stochastic discount factor (SDF) $M_{t,t+1}$ for each period $(t, t+1)$. More precisely, we assume:

$$M_{t,t+1} = \exp\left[-\mu_0 - \mu_1'\tilde{X}_t + \Gamma_t'\eta_{t+1} - \tfrac{1}{2}\Gamma_t'\Gamma_t\right], \qquad (18)$$

where $\tilde{\eta}_{t+1} = \Sigma\eta_{t+1}$ and where $\Gamma_t = \gamma_0 + \gamma\tilde{X}_t = \gamma_0 + \gamma_1 X_t + \gamma_2 X_{t-1} + \gamma_3 X_{t-2}$ is the affine (multiple lags) stochastic risk sensitivity vector; the constant term γ_0 is a (3×1) vector and $\gamma = [\gamma_1 : \gamma_2 : \gamma_3]$ is a (3×9) matrix. $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ are called risk sensitivity coefficients or parameters. The absence of arbitrage opportunity (A.A.O.) restriction for the risk-free asset implies $r_t = \mu_0 + \mu_1'\tilde{X}_t$, where r_t is the one-period interest rate between t and $t+1$ (known at t). So, under the no-arbitrage restriction, we have $M_{t,t+1} = \exp\left[-r_t + \Gamma_t'\eta_{t+1} - \tfrac{1}{2}\Gamma_t'\Gamma_t\right]$. This specification is convenient computationally because $Var(\eta_{t+1}) = I$, however it depends on the arbitrary choice of Σ in the decomposition $\tilde{\Omega} = \Sigma\Sigma'$. A more intrinsic specification involves the innovation $\tilde{\eta}_{t+1}$ of X_{t+1}:

$$M_{t,t+1} = \exp\left[-r_t + \tilde{\Gamma}_t'\tilde{\eta}_{t+1} - \frac{1}{2}\tilde{\Gamma}_t'\tilde{\Omega}\tilde{\Gamma}_t\right]$$

where $\tilde{\Gamma}_t'\Sigma = \Gamma_t'$ or equivalently, $\tilde{\Gamma}_t'\tilde{\Omega} = \Gamma_t'\Sigma'$ or $\Sigma\Gamma_t = \tilde{\Omega}\tilde{\Gamma}_t$ (where $\tilde{\Gamma}_t = \tilde{\gamma}_0 + \tilde{\gamma}\tilde{X}_t = \tilde{\gamma}_0 + \tilde{\gamma}_1 X_t + \tilde{\gamma}_2 X_{t-1} + \tilde{\gamma}_3 X_{t-2}$). Now, given that under the A.A.O. the price $B_t(h)$ at date t of a zero-coupon bond (ZCB) maturing at $t + h$ can be written as $B(t, h) = E_t[M_{t,t+1} \ldots M_{t+h-1,t+h}]$, we have the following result.

Proposition 1: The price at date t of the zero-coupon bond with time to maturity h is:

$$B_t(h) = \exp(c_h'\tilde{X}_t + d_h), \qquad (19)$$

where c_h and d_h satisfies, for $h \geq 1$, the recursive equations:

$$\begin{cases} c_h &= -\tilde{e}_1 + \tilde{\Phi}'c_{h-1} + (\Sigma\gamma)'c_{1,h-1} \\ &= -\tilde{e}_1 + \tilde{\Phi}^{*'}c_{h-1}, \\ d_h &= c_{1,h-1}'(\tilde{\nu} + \Sigma\gamma_o) + \tfrac{1}{2}c_{1,h-1}'\Sigma\Sigma'c_{1,h-1} + d_{h-1}, \end{cases} \qquad (20)$$

and where :

$$\tilde{\Phi}^* = \begin{bmatrix} \tilde{\Phi}_1 + \Sigma\gamma_1 & \tilde{\Phi}_2 + \Sigma\gamma_2 & \tilde{\Phi}_3 + \Sigma\gamma_3 \\ I_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I_{3\times3} & 0_{3\times3} \end{bmatrix} \quad \text{is a } (9 \times 9) \text{ matrix},$$

with initial conditions $c_0 = 0, d_0 = 0$ (or $c_1 = -\tilde{e}_1, d_1 = 0$), where $\tilde{e}_1$ is the (9×1) vector with all entries equal to 0 except the first one equal to 1. $c_{1,h}$ indicates the vector of the first 3 components of the 9-dimensional vector c_h. If we adopt the parameterization $(\tilde{\Omega}, \tilde{\Gamma}_t)$ instead of (Σ, Γ_t) we just have to replace $\Sigma\Sigma'$ by $\tilde{\Omega}$ and $\Sigma\gamma_i$ by $\tilde{\Omega}\tilde{\gamma}_i$ $(i = 0, 1, 2, 3)$ [**Proof** : see Appendix 2].

Corollary 1: The yield with h periods to maturity at date t, denoted $R_t(h)$, is given by :

$$
\begin{aligned}
R_t(h) &= -\frac{1}{h} \log B_t(h) \\
\\
&= -\frac{c'_h}{h} \tilde{X}_t - \frac{d_h}{h}, \quad h \geq 1 .
\end{aligned}
\tag{21}
$$

So $R_t(h)$ is an affine function of the factor $\tilde{X}_t$, that is of the 3 most recent lagged values of the 3-dimensional factor X_{t+1}. The same procedure is clearly followed to derive the yield-to-maturity formulas of the competing term structure models based on unconstrained and cointegrated autoregressive factor dynamics.

5.2 Risk Sensitivity Parameter Estimates

The estimation of historical (VAR, CVAR and NCVAR) and risk sensitivity parameters follows a consistent two-step procedure, as adopted, among the others, by APW(2006), Monfort and Pegoraro (2007), and Garcia and Luger (2007). In Section 4 we have presented the (first step) least squares estimates of θ_{var} and θ_{cvar}, thanks to the observations of the 1-quarter and 40-quarters yields, and of the real GDP. Given these parameter estimates, and given the selected value of λ for the NCVAR(3) model, the (second step) estimation of the risk sensitivity parameters $\theta_\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ is obtained by constrained nonlinear least squares (CNLLS), using the observations on yields with maturities different from those used in the first step (that is, maturities ranging from 2-quarters to 36-quarters). A constraint is imposed in order to satisfy the arbitrage restriction on the 10-years yield (the long rate). In particular, the Constrained NLLS estimator is given by:

$$
\begin{cases}
\hat{\theta}_\gamma = Arg \min_{\theta_\gamma} S^2(\theta_\gamma) \\
\\
S^2(\theta_\gamma) = \sum_t \sum_h (\tilde{R}_t(h) - R_t(h))^2, \\
\\
\text{s. t. } \sum_t (\tilde{R}_t(40) - R_t(40))^2 = 0,
\end{cases}
\tag{22}
$$

where, for each date t and maturity h, $R_t(h)$ is the theoretical yield determined by formula (21), and $\tilde{R}_t(h)$ indicate the observed one. Risk sensitivity parameter estimates of the Near-Cointegrated VAR(3) factor-based term structure model are presented in Table A.10, while risk sensitivity parameter estimates (obtained by CNLLS) of the CVAR(3), VAR(3) and VAR(1) factor-based term structure models are presented in Table A.11.

5.3 Empirical Results

5.3.1 In-Sample Fit of the Yield Curve

The purpose of this section is to study the ability of our yield-to-maturity formula, driven by the NCVAR(3) factor (with $\lambda = 0.2624$), to explain the observed interest rates variability in terms of fitting performances over the maturities used in the estimation of the risk sensitivity parameters. In the following section (Section 5.3.2), we will study the performances of our model to forecast out-of-sample these interest rates. Moreover, in Section 5.3.3, with the purpose to further analyze the specification of our term structure model, we will test its ability to explain the observed violation of the Expectation Hypothesis theory (see, among the others, Campbell and Shiller (1991), and Dai and Singleton (2002,2003)). These results will be systematically compared with those obtained by the competing CVAR(3), VAR(3) and VAR(1) factor-based term structure models.

Let us start from fit performances: in the last four columns of Table 4, we compare the (annualized absolute) yield-to-maturity errors of our selected NCVAR(3) factor-based term structure model with the performances of the other competing term structure models. For each date t and for each estimated model, we compute, over the maturities used to estimate the risk sensitivity parameters $\theta_\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$, the pricing error in the following way:

$$PE_t = \frac{\sum_h |\tilde{R}_t(h) - R_t(h)|}{H} \, , \tag{23}$$

where $\tilde{R}_t(h)$ and $R_t(h)$ are, respectively, the (annualized) observed and model-implied yields, and where H denotes the number of maturities used to estimate θ_γ. Given the time series PE_t, we calculate (for each model) the associated mean, standard deviation, minimum and maximum value.

	NCVAR(3) [$\lambda^* = 0.2583$]	NCVAR(3) [$\lambda = 0.2624$]	CVAR(3)	VAR(3)	VAR(1)
Mean	16.19	16.26	16.91	16.86	18.76
Median	12.53	12.69	12.89	12.55	16.02
Std. Dev.	13.22	13.20	14.02	13.96	15.23
Min.	0.99	0.84	2.07	1.21	2.01
Max.	88.57	88.25	93.39	91.49	112.74

Table 4: Annualized Absolute Pricing Errors (Basis Points).

The indications that stand out from this (in-sample) term structure fit comparison are the following. First, if we compare the fit of the yield curve obtained by the CVAR(3) and VAR(3) term structure models, we observe that these two models perform equally well and they outperform the APW(2006) model. This similarity highlights the compatibility of the parameter's restriction characterizing the CVAR(3) model, with respect to the interest rates dynamics. Second, our preferred model, with an averaging parameter selected to minimize an error forecast criterion and not a fitting criterion like (22), has smaller mean, std. dev., minimum and maximum pricing errors than those obtained by the three competing models. This result indicates that, the parameter λ we add to solve the discontinuity problem and we select to improve the specification of the expectation term in the long-term bond leads, at the same time, to a better fitting of the yield curve (see also figures A.1 and A.2). The reason of this result is shown in the second column of Table 4, where we put pricing error statistics of the NCVAR(3) term structure model in which we have jointly

estimated the parameters θ_γ and λ using the CNLLS methodology of Section 5.2. We observe that the optimal value of the averaging parameter, for the criterion (22), is $\lambda^* = 0.2583$, which is very close to $\lambda^*(40) = 0.2624$.

5.3.2 Yields Out-of-Sample Forecasts

In this section we want to further corroborate the out-of-sample forecast ability of our term structure modelling, based on a NCVAR(3) factor dynamics. The exercise is based on the increasing-size window procedure used above, in which we re-estimate at each iteration the historical parameters θ_{cvar} and θ_{var} when a new observation is available, while, for ease of computation, risk sensitivity parameters are fixed to the values estimated over the entire sample period[9]. In particular, we aim at studying if our model is able to forecast well the yield curve, with respect to the competing models mentioned above, both in the case of a parameter λ selected to minimize the yield forecast error for a given maturity h and over a certain horizon q, and in the case of $\lambda = 0.2624$ [case i) and case ii), respectively, presented below]. The results, presented in Table 5, are the following.

i) With regard to the optimal value of the averaging parameter λ, considered as a function of the forecasting horizon q and of the time-to-maturity h [$\lambda^* = \lambda^*(h, q)$, say], we first observe that, for any $h \in \{4, 8, 12, 20\}$, $\lambda^*(h, .)$ decreases when q increases. In other words, for any considered yield-to-maturity, the weight of the CVAR(3) component in the minimization of the forecast error increases when the forecast horizon increases. This means that, as for the short and long rate forecasts analyzed in Table 3, the out-of-sample forecast of yields over increasing horizons, asks for a model (in the VAR setting) increasingly able to explain their serial dependence. Second, for $q \in \{1, 4, 8\}$ (short forecast horizons), as far as h increases, $\lambda^*(., q)$ decreases as well, indicating the increasing importance of the CVAR(3) model in the short run forecast of long-term yields. In particular, one may observes that, for $q = 1$, we move from $\lambda^*(4, 1) = 0.9171$ to $\lambda^*(20, 1) = 0.4892$. Third, for $q = 12$, the optimal value of λ is around 0.5 for any $h \in \{4, 8, 12, 20\}$, suggesting the equal importance of the CVAR(3) and VAR(3) components in the forecast over this particular horizon. Finally, for any $q > 12$ (medium and long forecast horizons), the CVAR(3) model has a dominating role in forecasting yields for any residual maturity $h \in \{4, 8, 12, 20\}$. In particular, for $q = 40$, $\lambda^*(h, q) = 0$ for any h.

   Let us now make a comparison of forecast performances between the NCVAR(3) term structure model with the AR(1) time series model and the VAR(1) term structure model. The conclusion standing out from Table 5 is that our NCVAR(3) affine model outperforms the most competing AR(1) model, as well as the VAR(1) term structure model, over any forecasting horizon q (except for $q = 1$) and any residual maturity h. Moreover, for long forecasting horizons ($q = 40$), when we move from the AR(1) time series model to the NCVAR(3) term structure model, the RMSFE reduces between 35% and 45% for residual maturities ranging from $h = 4$ and $h = 20$.

ii) We consider now the case of the NCVAR(3) term structure model with $\lambda(h, q) = 0.2624$ for any pair (h, q). What we interestingly observe, again from Table 5, is that our model still

---

[9]We have also performed forecast exercises estimating, at each iteration, historical and risk sensitivity parameters, and we have found that the ranking among the models was the same and the magnitude of associated RMSFEs was almost unchanged.

	q	AR(1)	$\lambda^*(q)$	NCVAR(3)	NCVAR(3) $[\lambda = 0.2624]$	CVAR(3)	VAR(3)	VAR(1)
$R_t(4)$	1	0.0015	0.9171	0.0017	0.0018	0.0019	0.0023	0.0021
	4	0.0041	0.7542	0.0039	0.0040	0.0041	0.0047	0.0046
	8	0.0062	0.6578	0.0060	0.0061	0.0062	0.0071	0.0064
	12	0.0071	0.5120	0.0069	0.0069	0.0071	0.0084	0.0073
	20	0.0080	0.0520	0.0066	0.0068	0.0066	0.0091	0.0083
	40	0.0108	0.0000	0.0058	0.0080	0.0057	0.0109	0.0114
$R_t(8)$	1	0.0015	0.6245	0.0018	0.0018	0.0020	0.0026	0.0022
	4	0.0038	0.6838	0.0036	0.0037	0.0038	0.0046	0.0042
	8	0.0056	0.6414	0.0054	0.0056	0.0058	0.0069	0.0058
	12	0.0066	0.5290	0.0064	0.0065	0.0067	0.0082	0.0067
	20	0.0077	0.0680	0.0063	0.0065	0.0062	0.0093	0.0079
	40	0.0106	0.0000	0.0057	0.0078	0.0056	0.0109	0.0112
$R_t(12)$	1	0.0015	0.5399	0.0018	0.0018	0.0020	0.0029	0.0023
	4	0.0035	0.6364	0.0033	0.0035	0.0036	0.0046	0.0039
	8	0.0050	0.6124	0.0050	0.0051	0.0053	0.0067	0.0054
	12	0.0061	0.5213	0.0060	0.0061	0.0063	0.0081	0.0062
	20	0.0074	0.0811	0.0060	0.0062	0.0059	0.0093	0.0075
	40	0.0104	0.0000	0.0058	0.0077	0.0056	0.0108	0.0109
$R_t(20)$	1	0.0015	0.4892	0.0018	0.0019	0.0022	0.0035	0.0027
	4	0.0031	0.5905	0.0030	0.0032	0.0034	0.0048	0.0038
	8	0.0043	0.5681	0.0043	0.0045	0.0046	0.0065	0.0048
	12	0.0053	0.4990	0.0053	0.0054	0.0055	0.0078	0.0056
	20	0.0067	0.1060	0.0056	0.0057	0.0053	0.0093	0.0070
	40	0.0098	0.0000	0.0061	0.0075	0.0056	0.0106	0.0105

Table 5: Out-of-sample forecasts of $R_t(h)$, with $h \in \{4, 8, 12, 20\}$ measured in quarters. Table entries are RMSFEs. AR(1) denotes a Gaussian scalar autoregressive of order one process used to forecast $R_t(h)$ for any $h \in \{4, 8, 12, 20\}$. Forecasting horizons (q) are measured in quarters.

outperforms the AR(1) time series model and the VAR(1) term structure model, for any $q > 1$ and any $h \in \{4, 8, 12, 20\}$, even with a fixed averaging parameter selected to optimally forecast the expectation part of the 40-quarters yield. This means that, our specification of the NCVAR yield curve model, focused on the extraction of the term premia from the long-term bond is, at the same time, able to forecast interest rates, for any forecasting horizon $q \in \{4, \ldots, 40\}$ and any residual maturity $h \in \{1, 4, 8, 12, 20, 40\}$, better than the most competing AR(1) time series model and VAR(1) yield curve model.

5.3.3 Campbell-Shiller Regressions

Let us now study the ability of our NCVAR(3) term structure model (with $\lambda = 0.2624$) to explain the empirically observed failure of the Expectation Hypothesis Theory (EHT, hereafter) by means of the well known Campbell and Shiller (1991) long-rate regressions. This violation is documented by the fact that, for any residual maturity h, regressing the yield variation $R_{t+1}(h-1) - R_t(h)$ onto the normalized spread $(R_t(h) - r_t)/(h-1)$ leads to a negative regression coefficient ϕ_h (say) while, if EHT was correct (under the assumption of constant risk premiums), this coefficient (in the population) should be equal to one for any h. Moreover, several empirical studies have documented

that ϕ_h becomes increasingly negative when h increases [see Campbell and Shiller (1991), Bansal and Zhou (2002), Dai and Singleton (2002), Monfort and Pegoraro (2007)]. We find confirmation of this stylized fact also in the GSW (2007) data base considered in our empirical analysis; indeed, the estimated slope coefficients $\phi_{h,T}$ (say) obtained from the above mentioned regression is always negative and moves from -0.494 to -2.567 when h increases from three to forty quarters (see the second column of Table 6).

Let us compare the ability of our term structure model to replicate these increasingly negative Campbell-Shiller regression coefficients, with the ability of the competing VAR(3) and VAR(1) term structure models. In order to understand how well the proposed term structure models replicate the violation of the EHT, we operate in the following way. First, we calculate, for each of them, the population slope coefficient ϕ_h given by the following relation:

$$\phi_h = \frac{Cov[R_{t+1}(h-1) - R_t(h), (R_t(h) - r_t)/(h-1)]}{Var[(R_t(h) - r_t)/(h-1)]}, \tag{24}$$

where we take the estimates of the model parameters as the true parameters of the data-generating process, and we verify if ϕ_h is increasingly negative. Second, in order to understand if small-sample bias affect the population slope coefficients generated by any of the models we consider, we conduct the following Monte-Carlo exercise: for any given residual maturity h, we simulate 500 samples of length 174 (the length of our sample of observations) from a given estimated model, we calculate the 5% quantiles (Confidence Bands, hereafter) of the small sample distribution of the (Monte-Carlo based) estimated slope coefficient, and then we verify if the sample slope coefficients lie well inside these Monte-Carlo confidence bands. If the estimated term structure model generates negative downward sloping population Campell-Shiller regression coefficients and if their empirical counterpart lie inside the small-sample Monte-Carlo confidence bands, then we consider this model as able to successfully match the violation of the EHT. From Table 6, we observe that our NCVAR(3) factor-based term structure model is the only one able, among the three models considered in the empirical analysis, to successfully replicate this stylized fact: the population slope coefficient is increasingly negative for any h (while the VAR(3) and VAR(1) specifications generate a positive ϕ_3 coefficient) and the sample coefficients lie inside the Confidence Bands (except for $h = 8$).

6 Unbiased Term Premia

6.1 Definition of Unbiased Term Premia

Let us consider $R_t(h)$ and r_t, that is, the yield of maturity h periods at time t, and the short rate ($R_t(1) = r_t$). The usual term yield premium corresponding to this maturity is defined as:

$$\widetilde{TP}_t(h) = R_t(h) - \frac{1}{h}E_t\sum_{j=0}^{h-1} r_{t+j}, \tag{25}$$

$\widetilde{EX}_t(h) = \frac{1}{h}E_t\sum_{j=0}^{h-1} r_{t+j}$ being the Expectation Hypothesis term. So, we have:

$$R_t(h) = \widetilde{EX}_t(h) + \widetilde{TP}_t(h),$$

h	sample $\phi_{h,T}$	NCVAR(3) ϕ_h	Confidence Bands	VAR(3) ϕ_h	Confidence Bands	VAR(1) ϕ_h	Confidence Bands
3	-0.49 [0.28]	-0.20	(-0.72, 0.88)	0.19	(-0.54, 1.06)	0.02	(-0.68, 1.18)
4	-0.74 [0.40]	-0.31	(-0.92, 0.76)	0.00	(-0.77, 0.93)	-0.05	(-0.78, 1.06)
8	-0.98 [0.68]	-0.58	(-0.90, 0.86)	-0.48	(-0.90, 0.89)	-0.38	(-0.57, 1.20)
12	-1.20 [0.82]	-0.88	(-1.46, 0.56)	-0.88	(-1.43, 0.61)	-0.71	(-1.16, 0.84)
20	-1.55 [0.93]	-1.46	(-2.36, 0.11)	-1.52	(-2.35, 0.06)	-1.35	(-2.24, 0.18)
40	-2.57 [1.19]	-2.74	(-4.33, -0.69)	-2.75	(-4.38, -0.88)	-2.70	(-4.44, -0.59)

Table 6: Campbell-Shiller long-rate regressions. The slope sample coefficients $\phi_{h,T}$ are estimated from the regression $R_{t+1}(h-1) - R_t(h) = \phi_{o,h} + \phi_{h,T}[R_t(h) - r_t]/(h-1) + u_{t+1,h}$, using the GSW (2007) data base of sample size $T = 174$ [Newey-West standard errors with 4 lags are in brackets; the residual maturity h is measured in quarters]. The slope population coefficients ϕ_h are obtained from the model taking the parameter estimates as the true parameters of the data-generating process. Confidence bands show the 5% quantiles of the estimated slope coefficients from 500 samples of length 174 quarters simulated from the model.

and a similar decomposition for the spread gives:

$$
\begin{aligned}
S_t(h) &= R_t(h) - r_t \\
&= \widetilde{EXS}_t(h) + \widetilde{TP}_t(h),
\end{aligned}
$$

where $\widetilde{EXS}_t(h) = \widetilde{EX}_t(h) - r_t$ is the Expectation Hypothesis Spread.

A drawback of this version of the term premium $\widetilde{TP}_t(h)$ is that it would not be zero under the hypothetic situation where the historical dynamics and the risk-neutral dynamics would be identical, i.e. in a hypothetic world without risk aversion. In a such a situation, the yield of maturity h would be:

$$
EX_t(h) = -\frac{1}{h} \log B_t^*(h) \tag{26}
$$

where:

$$
B_t^*(h) = E_t[\exp -(r_t + r_{t+1} + \ldots + r_{t+h-1})],
$$

and not $\widetilde{EX}_t(h)$. Therefore, a more natural definition of the yield term premium, called "unbiased" because it is exactly equal to zero when the risk neutral and historical worlds are identical, is:

$$
TP_t(h) = R_t(h) - EX_t(h). \tag{27}
$$

Note that $EX_t(h)$ is easily computed using the recursive equations (20), with $\gamma_0 = 0$ and $\gamma = 0$.

6.2 Yield term premia and forward term premia

The short-term forward rate of maturity h is:

$$f_t(h) = \log B_t(h-1) - \log B_t(h) \,,$$

where $B_t(h)$ is the price at time t of the zero-coupon bond of residual maturity h. Therefore, we have:

$$f_t(h) = \log \frac{E_t^{\mathbb{Q}}[\exp(-r_t - \ldots - r_{t+h-2})]}{E_t^{\mathbb{Q}}[\exp(-r_t - \ldots - r_{t+h-1})]} \,,$$

with the convention $r_t + \ldots + r_{t+h-2} = 0$ if $h = 1$. If the historical dynamics was identical to the risk-neutral one, this forward rate would be:

$$EX_t^f(h) = \log B_t^*(h-1) - \log B_t^*(h)$$

So, a natural term premium for the short-term forward rate $f_t(h)$, called the forward term premium, is:

$$TP_t^f(h) = f_t(h) - EX_t^f(h) \,. \tag{28}$$

Given that $R_t(h) = \frac{1}{h}\sum_{j=1}^{h} f_t(j)$, and since we obviously have:

$$EX_t(h) = \frac{1}{h}\sum_{j=1}^{h} EX_t^f(j) \,, \tag{29}$$

we get:

$$TP_t(h) = \frac{1}{h}\sum_{j=1}^{h} TP_t^f(j) \,. \tag{30}$$

So, we have the following decomposition of $R_t(h)$ or $S_t(h)$:

$$R_t(h) = \frac{1}{h}\sum_{j=1}^{h} EX_t^f(j) + \frac{1}{h}\sum_{j=1}^{h} TP_t^f(j) \,, \quad \text{and}$$

$$S_t(h) = \frac{1}{h}\sum_{j=1}^{h} EXS_t^f(j) + \frac{1}{h}\sum_{j=1}^{h} TP_t^f(j) \,, \tag{31}$$

$$\text{with } EXS_t^f(j) = EX_t^f(j) - r_t \,, \ \forall j \in \{1, \ldots, h\} \,.$$

In other words, the yield and spread term structures, $R_t(h)$ and $S_t(h)$, are obtained by summing the averages of the forward Expectations and of the Premium term structures.

The forward term premium $TP_t^f(h)$ is the premium of a FRA (forward rate agreement) in the short rate between $t+h-1$ and $t+h$, negotiated at time t at level $f_t(h)$. So this premium can be viewed as the price evaluated at t of the risk coming from the uncertainty of the short rate between

$t+h-1$ and $t+h$. Of course there is no uncertainty for $h = 1$, so $TP_t^f(1) = 0$, and the uncertainty is likely to increase with h. According to (20) we have:

$$
\begin{aligned}
f_t(h) &= \log B_t(h-1) - \log B_t(h) \\
&= (c_{h-1} - c_h)' \tilde{X}_t + d_{h-1} - d_h \\
&= (\tilde{e}_1 + (I - \tilde{\Phi}')c_{h-1} - (\Sigma\gamma)'c_{1,h-1})' \tilde{X}_t - c_{1,h-1}'(\tilde{\gamma} + \Sigma\gamma_0) - \frac{1}{2}c_{1,h-1}'\Sigma\Sigma'c_{1,h-1} \\
&= r_t + c_{h-1}'(I - \tilde{\Phi})\tilde{X}_t - c_{1,h-1}'\tilde{\nu} - \frac{1}{2}c_{1,h-1}'\Sigma\Sigma'c_{1,h-1} - c_{1,h-1}'\Sigma\Gamma_t
\end{aligned}
\tag{32}
$$

Similarly

$$
EX_t^f(h) = r_t + c_{h-1}^{*\prime}(I - \tilde{\Phi})\tilde{X}_t - c_{1,h-1}^{*\prime}\tilde{\nu} - \frac{1}{2}c_{1,h-1}^{*\prime}\Sigma\Sigma'c_{1,h-1}^*
$$

where c_{h-1}^* is obtained from (20) with $\gamma = 0$, and

$$
\begin{aligned}
TP_t^f(h) &= (c_{h-1} - c_{h-1}^*)'(I - \tilde{\Phi})\tilde{X}_t - (c_{1,h-1} - c_{1,h-1}^*)'\tilde{\nu} \\
&\quad -\frac{1}{2}c_{1,h-1}'\Sigma\Sigma'c_{1,h-1} + \frac{1}{2}c_{1,h-1}^{*\prime}\Sigma\Sigma'c_{1,h-1}^* - c_{1,h-1}'\Sigma\Gamma_t
\end{aligned}
\tag{33}
$$

or with the parameterization $(\tilde{\Omega}, \tilde{\Gamma}_t)$:

$$
\begin{aligned}
TP_t^f(h) &= (c_{h-1} - c_{h-1}^*)'(I - \tilde{\Phi})\tilde{X}_t - (c_{1,h-1} - c_{1,h-1}^*)'\tilde{\nu} \\
&\quad -\frac{1}{2}c_{1,h-1}'\tilde{\Omega}c_{1,h-1} + \frac{1}{2}c_{1,h-1}^{*\prime}\tilde{\Omega}c_{1,h-1}^* - c_{1,h-1}'\tilde{\Omega}\tilde{\Gamma}_t
\end{aligned}
\tag{34}
$$

In particular, if $\Gamma_t = 0$ $(\gamma = 0, \gamma_0 = 0)$, $TP_t^f(h) = 0$, and if Γ_t does not depend on t $(\gamma = 0)$ we have $c_{h-1} = c_{h-1}^*$, $TP_t^f(h) = -c_{1,h-1}'\Sigma\gamma_0$ $(= -c_{1,h-1}'\tilde{\Omega}\tilde{\gamma}_0)$, and

$$
\begin{aligned}
TP_t(h) &= \frac{1}{h}\sum_{j=1}^{h} TP_t^f(j) \\
&= -\frac{1}{h}\sum_{j=1}^{h} c_{1,j-1}'\Sigma\gamma_0
\end{aligned}
\tag{35}
$$

6.3 Yield term premia, risk premia and risk sensitivities

Formula (30) gives a decomposition of the yield term premium $TP_t(h)$ in terms of the forward term premia $TP_t^f(j)$. Anoher interesting decomposition of $TP_t(h)$ is based on the expected risk premia attached to future one-period holdings of the zero-coupon bond with residual maturity h at time t. Indeed we have:

$$
\begin{aligned}
R_t(h) &= -\frac{1}{h}\sum_{j=1}^{h} \log \frac{B_{t+j-1}(h-j+1)}{B_{t+j}(h-j)} \\
&= \frac{1}{h}\sum_{j=1}^{h} \rho_{t+j}(h-j+1)
\end{aligned}
\tag{36}
$$

where $\rho_{t+j}(h-j+1)$ is the geometric return between $t+j-1$ and $t+j$ of the zero-coupon bond of residual maturity $h-j+1$ at $t+j-1$ (or residual maturity h at t).

If the historical dynamics was identical to the risk-neutral one, the yield of residual maturity h would be:

$$
\begin{aligned}
EX_t(h) &= -\frac{1}{h}\sum_{j=1}^{h}\log\frac{B^*_{t+j-1}(h-j+1)}{B^*_{t+j}(h-j)} \\
&= \frac{1}{h}\sum_{j=1}^{h}\rho^*_{t+j}(h-j+1)\,.
\end{aligned}
\tag{37}
$$

So:

$$
\begin{aligned}
TP_t(h) &= R_t(h) - EX_t(h) \\
&= \frac{1}{h}\sum_{j=1}^{h}\left[\rho_{t+j}(h-j+1) - \rho^*_{t+j}(h-j+1)\right]\,.
\end{aligned}
\tag{38}
$$

Note that in equations (36), (37) and (38) the left hand sides are known at t whereas the terms of the sums in the right hand sides are random. So we get additional identities by taking the conditional expectation of both sides of the equations with respect to the historical distribution. In particular we get:

$$
\begin{aligned}
TP_t(h) &= \frac{1}{h}\sum_{j=1}^{h}E_t\left[\rho_{t+j}(h-j+1) - \rho^*_{t+j}(h-j+1)\right] \\
&= \frac{1}{h}\sum_{j=1}^{h}E_tE_{t+j-1}\left[\rho_{t+j}(h-j+1) - \rho^*_{t+j}(h-j+1)\right] \\
&= \frac{1}{h}\sum_{j=1}^{h}E_tRP_{t+j-1}(h-j+1)\,,
\end{aligned}
\tag{39}
$$

where $RP_{t+j-1}(h-j+1)$ is the risk premium at $t+j-1$ associated with the one-period holding between $t+j-1$ and $t+j$ of a zero-coupon bond of residual maturity $h-j+1$. The risk premia $RP_{t+j-1}(h-j+1)$ are equal to zero if the risk sensitivities Γ_t's (or $\tilde{\Gamma}_t$'s) are equal to zero. Using equations (20) we have:

$$
\begin{aligned}
\rho_{t+j}(h-j+1) &= d_{h-j} + c'_{h-j}\tilde{X}_{t+j} - d_{h-j+1} - c'_{h-j+1}\tilde{X}_{t+j-1} \\
&= c'_{h-j}\tilde{X}_{t+j} - (-\tilde{e}_1 - c'_{h-j}\tilde{\Phi} - c'_{1,h-j}\Sigma\gamma)\tilde{X}_{t+j-1} \\
&\quad - c'_{1,h-j}(\tilde{\nu} + \Sigma\gamma_0) - \frac{1}{2}c'_{1,h-j}\Sigma\Sigma'c_{1,h-j} \\
&= c'_{h-j}(\tilde{X}_{t+j} - \tilde{\Phi}\tilde{X}_{t+j-1} - e_1\tilde{\nu}) + \tilde{e}'_1\tilde{X}_{t+j-1} \\
&\quad - c'_{1,h-j}\Sigma\gamma\tilde{X}_{t+j-1} - c'_{1,h-j}\Sigma\gamma_0 - \frac{1}{2}c'_{1,h-j}\Sigma\Sigma'c_{1,h-j} \\
&= c'_{1,h-j}\tilde{\eta}_{t+j} + r_{t+j-1} - c'_{1,h-j}\Sigma\Gamma_{t+j-1} - \frac{1}{2}c'_{1,h-j}\Sigma\Sigma'c_{1,h-j}
\end{aligned}
\tag{40}
$$

30

Similarly:

$$\rho^*_{t+j}(h-j+1) = c^{*'}_{1,h-j}\tilde{\eta}_{t+j} + r_{t+j-1} - \frac{1}{2}c^{*'}_{1,h-j}\Sigma\Sigma'c^{*'}_{1,h-j} \tag{41}$$

where the $c^*_{1,h-j}$ are obtained from (20) with $\gamma = 0$. Finally

$$RP_{t+j-1}(h-j+1) = -c'_{1,h-j}\Sigma\Gamma_{t+j-1} - \frac{1}{2}c'_{1,h-j}\Sigma\Sigma'c_{1,h-j} + \frac{1}{2}c^{*'}_{1,h-j}\Sigma\Sigma'c^*_{1,h-j} \tag{42}$$

or with the alternative parameterization in $(\tilde{\Gamma}_t, \tilde{\Omega})$:

$$RP_{t+j-1}(h-j+1) = -c'_{1,h-j}\tilde{\Omega}\tilde{\Gamma}_{t+j-1} - \frac{1}{2}c'_{1,h-j}\tilde{\Omega}c_{1,h-j} + \frac{1}{2}c^{*'}_{1,h-j}\tilde{\Omega}c^*_{1,h-j} \tag{43}$$

Since $\Gamma_{t+j-1} = \gamma_0 + \gamma\tilde{X}_{t+j-1}$, and $c^*_{1,h-j}(\tilde{\Phi}, \tilde{\nu}, \Sigma)$, $c_{1,h-j}(\tilde{\Phi}, \tilde{\nu}, \Sigma, \gamma_0, \gamma)$ satisfy $c^*_{1,h-j}(\tilde{\Phi}, \tilde{\nu}, \Sigma) = c_{1,h-j}(\tilde{\Phi}, \tilde{\nu}, \Sigma, 0, 0)$, it is clear that the risk premia $RP_{t+j-1}(h-j+1)$ are equal to zero if $\gamma_0 = 0$ and $\gamma = 0$. The decomposition (39) becomes:

$$\begin{aligned} TP_t(h) &= \frac{1}{h}\sum_{j=1}^{h} E_t RP_{t+j-1}(h-j+1) \\ &= \frac{1}{h}\sum_{j=1}^{h} -c'_{1,h-j}\Sigma E_t\Gamma_{t+j-1} - \frac{1}{2}c'_{1,h-j}\Sigma\Sigma'c_{1,h-j} + \frac{1}{2}c^{*'}_{1,h-j}\Sigma\Sigma'c^*_{1,h-j}, \end{aligned} \tag{44}$$

and, thus, $TP_t(h)$ is decomposed in h terms, each term depending on t through the expectation $E_t\Gamma_{t+j-1}$ of the risk sensitivity vector. Roughly speaking a term of this decomposition will increase if the expected price of risk within the period is increasing.

Note that if the risk sensitivities do not depend on t ($\gamma = 0$), we have $c_{1,h} = c^*_{1,h}$, and therefore, $TP_t(h) = -\frac{1}{h}\sum_{j=1}^{h} c'_{1,h-j}\Sigma\gamma_0$ does not depend on t either. Comparing with (35) we get $-c'_{1,i-1}\Sigma\gamma_0 = TP_t^f(i) = E_t RP_{t+h-i}(i)$, $\forall t, i, h > i$, and the forward term premium of horizon i is equal to the expectation of the risk premium attached to any future one-period holding of a zero coupon bond with residual maturity i (at the beginning of the period). Moreover, in the general case, the contribution of the distant periods (large j) in the general expression of $TP_t(h)$ are likely to be almost constant in t since the $E_t\Gamma_{t+j-1}$ are likely to be close to their unconditional expectations. And, finally, the level of the contribution of the distant periods are likely to be small given that the uncertainty in the return of the bond is likely to decrease when its residual maturity decreases; in particular the last term ($j = h$) in (44) is equal to zero since $RP_{t+h-1}(1) = 0$.

6.4   Computation of the term premia

Figure A.3 shows the term premia computed from VAR(3), CVAR(3) and NCVAR(3) models. The three measures are very similar in their variations over the sample. The VAR(3) term premium is the more volatile (standard deviation equal to 2.63), with levels that range from -1 to 6, and its correlation with the 10-year spread is equal to 0.31. The CVAR(3) measure is much more stable (standard deviation equal to 1.19), but is highly correlated with the 10-year spread (correlation coefficient equal to 0.91). This result is related to the presence of a unit root in the short term

interest rate. When the short rate is considered as an I(1) non stationary process, the expectation part of the 10-year interest rate, $EX_t(40)$, is very close to the short rate. Therefore, the expectation part of the spread, $EXS_t(40)$, is close to zero, and the 10-year spread is nearly equal to the term premium. This result highlights one of the limits of the CVAR approach for computing the term premium. As argued earlier, our preferred measure of the term premium is the one obtained from the NCVAR(3) model. This measure is an average of the measures obtained from VAR(3) and CVAR(3) models. It is more stable than the VAR(3) term premium (standard deviation equal to 1.16), and it is less correlated with the spread than the CVAR(3) term premium (the correlation coefficient between the spread and the NCVAR(3) term premium is equal to 0.77).

In figures A.4 and A.5 we focus on the behavior of our preferred measure, the NCVAR(3) term premium. This measure shows similarities with other measures of the term premium found in the literature. Rudebusch, Sack and Swanson (2007) compare five term premia measures before focusing their attention on a measure based on the Kim-Wright approach (2005)[10] which appears to be representative of other measures. Figure A.4 shows that over the period 1990-2007, our NCVAR(3) measure displays similar features, including peaks and trough observed with the Kim-Wright measure. More particularly, both measures are very close between 1990 and 2002. The main difference between both term premia is observed during the period 2002-2004 during which, the NCVAR(3) term premium is substantially higher than the Kim-Wright's one[11]. In addition, we also note that the decrease in the term premium in 2004 is more pronounced with the NCVAR(3) model. For comparison purposes, we also present in figure A.4, the 10-year term premium obtained from a VAR(1) model similar to Ang, Piazzesi and Wei (2006). This measure tends to be lower than the two others. In addition, we also note that the three term premia have substantially decreased over the recent period, consistently with results found in the literature. Finally we also report in figure A.5 recessions (shaded bars) as dated by the National Bureau of Economic Research. We note that the term premium tends to increase in period of recession, and then appears to be contra-cyclical.

6.5 An Application to the *Conundrum*

Figure 15 presents the short term rate and the 10-year interest rate, along with its expectation part and term premium, during the three previous episodes of monetary policy tightening (shaded areas). We observe that in the three cases, the rise in the short term interest rate comes with an increase of $EX_t(40)$ and a decrease in $TP_t(40)$. The final effect on the 10-year interest rate depends on the extent of the changes in its two components $EX_t(40)$ and $TP_t(40)$. For the 1994 and 1999 episodes, the rise of $EX_t(40)$ exceeds the decline in $TP_t(40)$ in such a way that in both cases, the 10-year interest rate increases. In contrast, for the 2004 episode, the rise in $EX_t(40)$ seems to be offset by the decline in $TP_t(40)$, leading to stable 10-year interest rate. This inertia of the 10-year interest rate is described as a "conundrum" by Alan Greenspan given that during previous episodes of restrictive monetary policy, this rate increased along with the fed fund target.

The explanation of this phenomenon has to be found in the sharp decrease of the term premium between June 2004 and June 2006. In order to shed more light regarding this phenomenon, we

---

[10]This approach is based on a standard no-arbitrage continuous-time affine term structure model, in which the yield curve is driven by a three-dimensional latent factor.

[11]During this period, levels of the NCVAR(3) term premium are closer to those obtained with others methodologies, such as the Rudebusch and Wu (2008) measure (see Rudebusch, Sack and Swanson (2007) for further details).
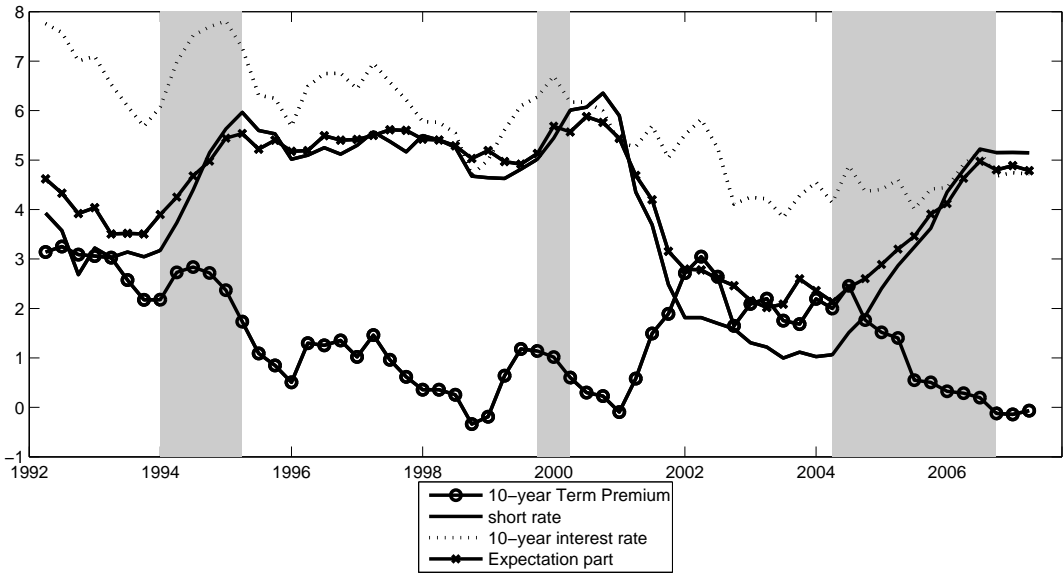
Figure 15:
Short rate, 10-year interest rates and its components over the three
previous monetary policy tightening episodes
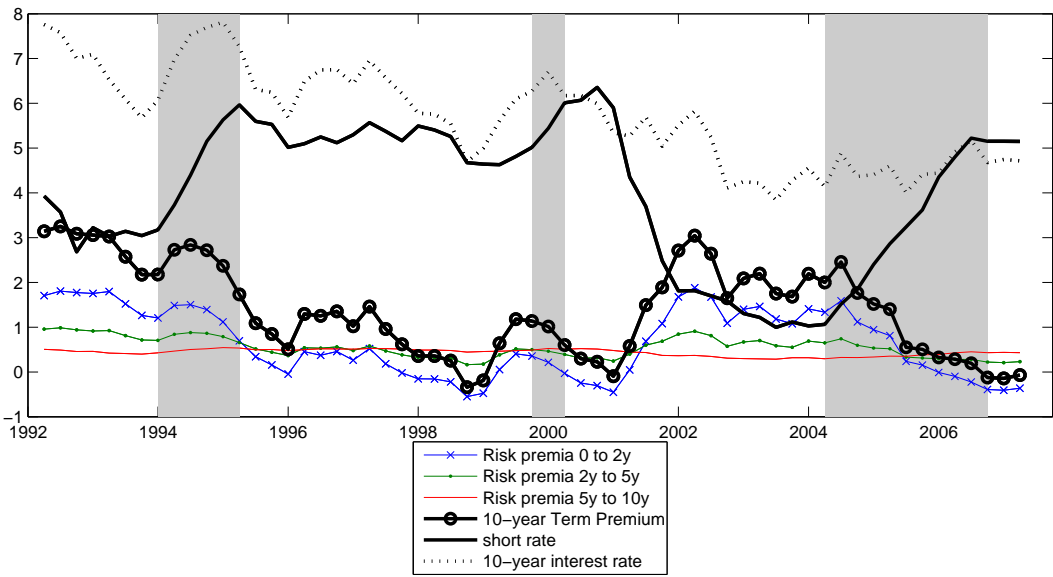Shaded areas: monetary policy tightening episodes.

Figure 16:
Expected risk premia and 10-year term premium
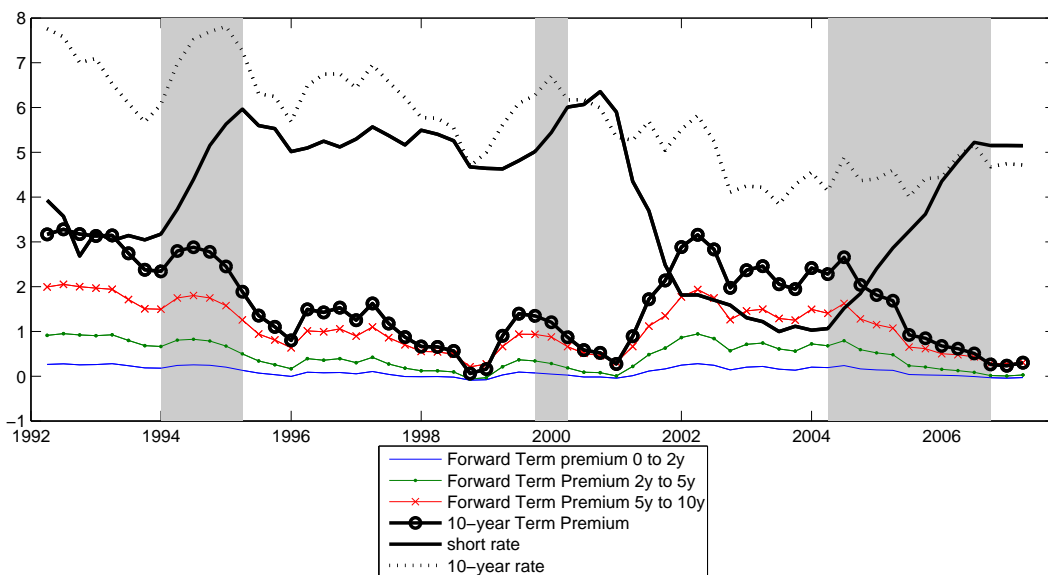Shaded areas: monetary policy tightening episodes.

Figure 17:
Forward term premia and 10-year term premium
Shaded areas: monetary policy tightening episodes.

present in figure 16 the term premium decomposition in risk premia as it is described in equation (44). For sake of readability, we aggregate the expected risk premia over the time intervals $(t, t+2y)$ (0 to 2y, in the graph), $(t + 2y, t + 5y)$ (2y to 5y), $(t + 5y, t + 10y)$ (5y to 10y), and we will call these aggregate measures as short-run, middle-run and long-run expected risk premia. The sum of these three components gives the 10-year term premium. Expected risk premia for a given period measures the expected risk of holding, over that period, a bond with residual maturity of 10 years at date t. As noted by Cochrane and Piazzesi (2008), this risk is closely linked to expected inflation over the period. Actually, a negative expected risk premium indicates that inflation is expected to be stable over the period. In other words, a negative risk premia, or at least a decreasing risk premia, can reflect the fact that the market believe in the (increased) credibility of the monetary authority in controlling inflation. Looking at figure 16 we see that the decreasing trend of the term premium observed during the rise of short term interest rate is mainly driven by the short-term expected risk premia. During the 1994 tightening, this short-term expected risk premia decreases but remains positive, except for one quarter 1996. During the 1999 episode, negative short-term expected risk premia are more frequent but do not exceed three quarters in 2000. In contrast, during the 2004 episode, period of negative short-term expected risk premia lasts at least six quarters. The observed increasing periods of negative short-term expected risk premia may reveal that the Fed became more and more credible for low and stable inflation over the three previous monetary policy tigthening episodes. Of course, other elements have probably intensified the decreasing trend of the risk premium, particularly in 2004 (foreign central banks intervention for instance), but the possibility an of increased credibility of the Fed cannot be rejected at a first glance. For that reason, we are tempted to adopt the views of Cochrane and Piazzesi (2008) questioning the puzzling feature

34

of the 2004 episode and argue that the 2004 episode is not different in nature but just in terms of the relative weights of the components of the long term interest rate. Finally, we also show in figure 17 the decomposition of the 10-year term premium, in terms of forward term premium (see equation (35)) aggregated over the same expected risk premia time intervals time intervals. We see that the 10-year term premium is mainly driven by the forward term premium spanning the period $(t + 5y, t + 10y)$. More particularly, this premium tends to decrease during restrictive monetary policy. Indeed, when short rate increases, inflation is expected to be lower in the future, reducing the volatility of future short rates. As a consequence, long term forward term premia decreases. We can expect that the credibility of the Fed in maintaining low and stable inflation is positively linked with the fall of the forward term premium. In figure 17, we observe that the decline in the 5-years to 10-years forward term premium is more pronounced during the 2004 tightening. Once again, this corroborate the idea according to which the credibility of the Fed has increased over the last decade.

7   New Information Response Functions

In what follows, the dynamics of the 3-dimensional state process $X_t = (r_t, S_t, g_t)'$ is given by the Near-Cointegrated VAR(3) model described in the previous sections. The optimal weight used to average the VAR(3) and CVAR(3) parameters is chosen to get the best prediction of $B_t^*(40)$ ($\lambda^*(40) = 0.2624$). Hence, our NCVAR(3) specification provides the best measure of the 10-year term premium.

In this section we are interested in measuring the differential impact on X_t, $t = 1, ..., T$ of a shock hitting a given variable. For that purpose, we follow in this section a new approach based on a generalization of the Impulse Response Function, called *New Information Response Function* and proposed by Jardet, Monfort and Pegoraro (2009b). The first two subsections rapidly summarize the methodology, while the last two present the responses to a shock on the spread, along with its expectation part and term premium component, and to a shock on the short term interest rate.

7.1   Definition of New Information Response Function

In this section, we generalize the standard notion of Impulse Response Function (IRF) to the notion of New Information Response Function ($NIRF$). Let us consider a n-dimensional VAR(p) process y_t, possibly non-stationary. We denote by η_t its innovation process. We want to measure the differential impact on y_t, $t = 1, ..., T$, of a new information I_0 at date $t = 0$ (by convention). Typically, this new information will be the value h_0 taken by some function $h(\eta_0)$ of the innovation of the process at $t = 0$. In order to measure this differential impact we use a definition introduced in the context of nonlinear models (see e.g. Gallant, Rossi and Tauchen (1993), Koop, Pesaran and Potter (1996), Gourieroux and Jasiak (1999)). More precisely, the NIRF is defined by:

$$NIRF(t) = E\left(y_t | I_0, y_{-p}\right) - E\left(y_t | y_{-p}\right), \ t \geq 0,$$

where $y_{-p} = (y'_{-1}, ..., y'_{-p})'$. Exploiting the linearity of the model we see that:

$$
\begin{aligned}
NIRF(t) &= E\left(y_t | h(\eta_0) = h_0, y_{-p} = 0\right) \\
&= E\left(y_t | \eta_0 = E(\eta_0 | h(\eta_0) = h_0), y_{-p} = 0\right)
\end{aligned}
$$

35

and:

$$NIRF(t) = D_t\delta \qquad (45)$$

with $\delta = E(\eta_0|h(\eta_0) = h_0)$, and D_t is the tth Markov matrix coefficient of the MA representation of y_t [see Jardet, Monfort and Pegoraro (2009b)].

This general definition of a $NIRF$ includes standard Impulse Response Functions. First, if the variance-covariance $V(\eta_0)$ matrix of η_0 is diagonal, it is usual to consider a shock of 1 on the jth component of η_0 and 0 on the others. In this case the new information is simply $\eta_0 = \delta = e_j$, (where e_j is the vector with components equal to zero except the jth equal to 1). Second, if $V(\eta_0)$ $=\Sigma$, it is usual to consider a shock of 1 on the jth component of a transformed vector ξ_0 defined by $\eta_0 = P\xi_0$, where $PP' = \Sigma$. In this case, the new information is $\eta_0 = \delta = P^{(j)}$, where $P^{(j)}$ is the jth column of P [$P^{(j)}$ can also be normalized in order to have its jth component equal to 1; see Jardet, Monfort and Pegoraro (2009b)]. Third, Pesaran and Shin (1998) also considered a "generalized" IRF, in which the new information is $\eta_{0j} = 1$ and therefore, in formula (45), $\delta = E(\eta_0|\eta_{0j} = 1) = Cov(\eta_0, \eta_{0j})/Var(\eta_{0j})$ (in the gaussian case).

But the New Information Response Function is useful in a much more general context [see Jardet, Monfort and Pegoraro (2009b) for further details], in particular when considering shocks on filtered variables.

7.2   Shocks on filtered variables

If we consider a m-dimensional process $\widetilde{y}_t$ obtained by applying a linear filter on y_t:

$$\widetilde{y}_t = F(L)y_t$$

where $F(L) = [F_1(L), ..., F_n(L)]$ is a $(m \times n)$ matrix of polynomials in the lag operator. The innovation of $\widetilde{y}_t$ at $t = 0$ is: $\widetilde{\eta}_0 = F(0)\eta_0$.

Therefore if the new information at $t = 0$ is $\widetilde{h}(\widetilde{\eta}_0) = \widetilde{h}_0$, the $NIRF$ is:

$$NIRF(t) = D_t\delta$$

with $\delta = E\left(\eta_0|\widetilde{h}(F(0)\eta_0) = \widetilde{h}_0\right)$. Obviously, the new information may also be made of an information on both η_0 and $\widetilde{\eta}_0$: $h(\eta_0) = h_0$, and $\widetilde{h}(\widetilde{\eta}_0) = \widetilde{h}_0$ or $h(\eta_0) = h_0$ and $\widetilde{h}(F(0)\eta_0) = \widetilde{h}_0$.

In the context of our model, the component of $\widetilde{y}_t$ may be, for instance, the expectation part of a spread of some maturity, or the term premium corresponding to some maturity. If the maturity if 40 quarters, the corresponding filter can be computed from the VAR coefficients only, otherwise it necessitates the affine term structure model.

7.3   Impulse responses to a shock on the 10-year spread

In this section we focus on the responses of the GDP, the yields of various maturities and their corresponding term premia and expectation components, to a unexpected increase in the spread equal to one at date $t = 0$.

For that purpose and following previous notations, we need to determine the value of the (3×1) vector δ such that $\delta = E(\eta_0|I_0)$, where η_0 is the innovation of the vector (r_t, S_t, g_t) and I_0 is the new information at date $t = 0$. Here, the new information I_0 includes, first of all, $\eta_{0,2} = 1$, where $\eta_{0,2}$ is the second component of η_0, that is, the innovation of the spread at date $t = 0$. In addition, we have

to remember that r_t and S_t are observed at the end of the period (end-of-quarter observations) and they drive an information covering a following period spanned by the residual maturity, whereas g_t is the growth rate of GDP between $t-1$ and t, observed at t, and driving an information spanning the two previous quarters. Therefore, a shock on the spread (or on any interest rate) occurring at date t (end of the quarter), should have no effect on the growth rate of real GDP between $t-1$ and t. Accordingly, we impose an additional restriction to ensure that the growth rate of real GDP does not respond instantaneously to a shock on the spread. More precisely, the information $\eta_{0,3} = 0$, where $\eta_{0,3}$ is the innovation of the one-quarter GDP growth at date $t = 0$, is included in I_0.

This means that, we have to find the value $\delta = E(\eta_0|\eta_{0,2} = 1, \eta_{0,3} = 0)$ or, in other words, the value of the first component of δ, that is the instantaneous expected response of the short rate when the spread increases by one unity whereas the growth rate of GDP remains at its past level: we have $\delta = (\beta, 1, 0)'$, where $\beta = E(\eta_{0,1}|\eta_{0,2} = 1, \eta_{0,3} = 0)$. In the gaussian case, β is the coefficient of $\eta_{0,2}$ in the theoretical regression of $\eta_{0,1}$ on $\eta_{0,2}$ and $\eta_{0,3}$. Figures 18 present the responses over 20 quarters of the real GDP, interest rates, term premia and expectation components of yield as defined by equation (26) and (27). Figure 18(a) indicates that an increase in the spread of 1 percentage point (that is 4 percentage points in annual basis) concurs with a decrease in the short term interest rate greater than 1 percentage point (4 percentage point in annual basis). The expectation component of the 10-year interest rate also decreases, but less than the short term interest rate. In contrast the differential impact on the 10-year term premium is initially positive, before becoming negative after about 10 quarters. Finally, the response of the 10 year interest rate is negative and ranges between 0 and -0.4 percentage point (that is a range between 0 and -1.6 percentage points in annual basis). As far as the yield curve is concerned [see figure 18(b)], we see that all the responses of the yields are negative with an amplitude that is growing as the maturity decreases. This suggests that the shock mainly affects the short end of the yield curve leading to a steepening of the curve.

In figure 18(c) we observe that after a slight decrease that does not exceed 1 quarter, the real GDP tends to increase until to reach its new steady state level. After 20 quarters, the real GPD has increased by 4%, corresponding to an average annual growth rate equal to 0.8%. This result confirms the well documented results in the literature that emphasizes the positive relationship between the slope of the yield curve and future activity.

There exists an extensive empirical literature relating the predictive power of the slope of the yield curve on subsequent real activity. Theoretically, one of the main explanation of this fact is related to countercyclicality of monetary policy. When the central bank lowers the short term interest rate two effects are expected. First, the long term interest rate tends to decrease, but less than the short term interest rate (because the central bank is expected to move to a contractionary policy in the future to respond to future increases in inflation). Second, with long term interest rates smaller, financing conditions improve and private investment increases, leading in turn to an increase of activity. According to this theory, the increase of the spread is mainly generated by the drop of the short term interest rate and the expectation part of the spread. More precisely, recall that the 10-year spread, $S_t(40) = S_t$, can be decomposed as:

$$S_t = TP_t(40) + EX_t(40) - r_t \qquad (46)$$

where $TP_t(40)$ is the 10-year term premium, and $EX_t(40)$ the expectation part of the 10-year interest rate defined by (27) and (26) respectively. $r_t = R_t(1)$ is the short term (one-quarter) interest rate. We denote by $EXS_t(40)$ the expectation part of the spread, defined by:

$$EXS_t(40) = EX_t(40) - r_t \qquad (47)$$

(a) 10-year spread and components
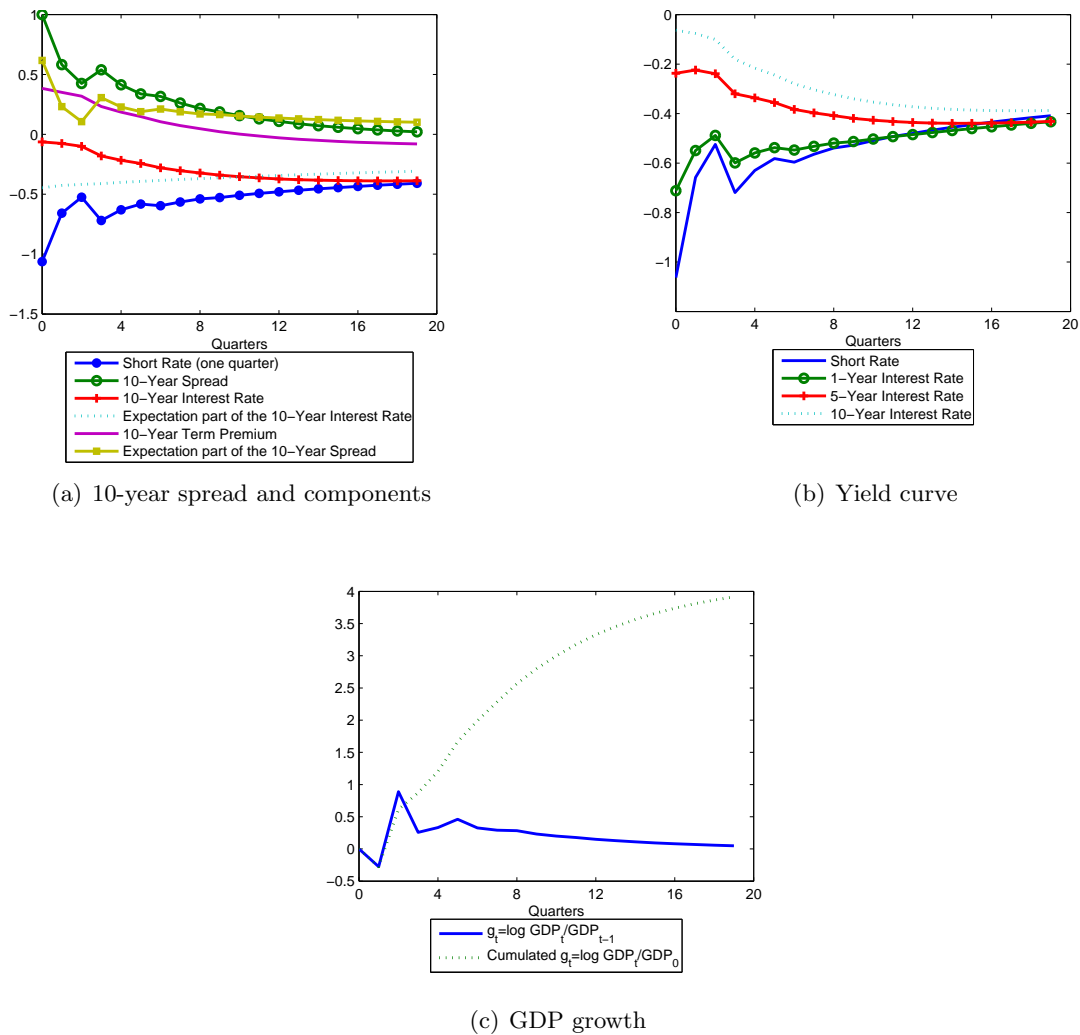
(b) Yield curve

(c) GDP growth

Figure 18: Responses to a shock on the 10-year spread

Therefore, we see that an increase in the spread can be generated by an increase in $EXS_t(40)$ or an increase in $TP_t(40)$ (or both). The "monetary policy explanation" of the predictive power of the spread is based on the fact that S_t increases in response to a decrease of r_t and an increase of $EXS_t(40)$. However, equation (46) indicates that an increase in the spread can also result from a rise in the term premium $TP_t(40)$, not necessarily related to monetary policy. For instance, any events that can affect the supply and demand for long term bonds are good candidate to explain a move on the term premium, and consequently the spread of interest rate. However, if the spread increases because of a rise in the term premium, the final effect on real activity is not clear. On one hand, an higher term premium, that is an higher long term interest rate, should deteriorate the financing conditions and then should reduce private investment and economic activity. In this case there is a negative relationship between the spread and future output growth. On the other hand, if the rise in the term premium and long term interest rate are due to an increase in the government

purchases, financed by the issue of long term bonds, one may expect that spending government policy should finally stimulate economic activity. In this case, the increase in the spread induces a rise in real output and the relationship between the spread and future activity is positive.

This ambiguity appears in the results of the literature. Actually, papers that try to determinate to what extent each component of the spread, that is $EXS_t(40)$ or $TP_t(40)$, helps to predict future activity, generally lead to different conclusions regarding the role of the term premium. Hamilton and Kim (2002), and Favero, Kaminska and Södeström (2005) tend to conclude to a positive and significant relationship between the term premium and future activity. In contrast, Ang, Piazzesi and Wei (2006), Rudebusch, Sack and Swanson (2007), and Rosenberg and Maurer (2007) do not find significant link between the level of the term premium and future output growth.

In what follows, we try to shed light on this debate by analyzing the dynamic effects of an increase in the spread on real activity, disentangling the effects of a rise in the spread due to an increase in its expectation part, and a rise in the spread caused by an increase in the term premium.

7.4 Impulse Responses to a shock on the term premium and the expectation part of the spread

Given the affine structure of our model, the expectation part of the spread $EXS_t(40)$ and the term premium $TP_t(40)$ are obtained by applying a linear filter on $y_t = (r_t, S_t, g_t)'$:

$$EXS_t(40) = F_{1,1}(L)r_t + F_{1,2}(L)S_t + F_{1,3}(L)g_t \tag{48}$$

$$TP_t(40) = F_{2,1}(L)r_t + F_{2,2}(L)S_t + F_{2,3}(L)g_t \tag{49}$$

Hence, the innovation at $t = 0$ of $EXS_t(40)$ and $TP_t(40)$, denoted by $\tilde{\eta}_{0,1}$ and $\tilde{\eta}_{0,2}$ respectively are:

$$\tilde{\eta}_{0,1} = F_{1,1}(0)\eta_{0,1} + F_{1,2}(0)\eta_{0,2} + F_{1,3}(0)\eta_{0,3} \tag{50}$$

$$\tilde{\eta}_{0,2} = F_{2,1}(0)\eta_{0,1} + F_{2,2}(0)\eta_{0,2} + F_{2,3}(0)\eta_{0,3} \tag{51}$$

where $\eta_{0,1}$, $\eta_{0,2}$ and $\eta_{0,3}$ are the innovation at $t = 0$ of r_t, S_t and g_t respectively. In addition, by construction, we have[12]:

$$\eta_{0,2} = \tilde{\eta}_{0,1} + \tilde{\eta}_{0,2}$$

7.4.1 Shock on the expectation part

We are interested in the dynamic effects of 1 percentage point increase in the spread that would be completely due to a 1 percentage point increase in the expectation part of the spread. More precisely, the new information I_0 includes $\eta_{0,2} = 1$, $\tilde{\eta}_{0,1} = 1$ and $\tilde{\eta}_{0,2} = 0$. We also assume that this increase has no instantaneous effect of the real GDP, that is I_0 also includes $\eta_{0,3} = 0$. Therefore, we have to determine the value of the vector $\delta = E(\eta_0|I_0) = E(\eta_0|\eta_{0,2} = 1, \tilde{\eta}_{0,1} = 1, \tilde{\eta}_{0,2} = 0, \eta_{0,3} = 0)$ where $\eta_0 = (\eta_{0,1}, \eta_{0,2}, \eta_{0,3})'$. From equation (50) we immediately obtain that $\eta_{0,1} = \frac{1 - F_{1,2}(0)}{F_{1,1}(0)}$. Then:

$$\delta = \left(\frac{1 - F_{1,2}(0)}{F_{1,1}(0)}, 1, 0 \right)'$$

Figures 19 show the impulse responses to a 1 percentage point shock on the expectation part of the spread (4 percentage point in annual basis). The response of the spread is mainly driven by

---

[12]This implies that the $F_{i,j}(0)$ verify $F_{1,1}(0) + F_{2,1}(0) = 0$, $F_{1,2}(0) + F_{2,2}(0) = 1$ and $F_{1,3}(0) + F_{2,3}(0) = 0$

its expectation part, the response of the 10-year term premium remaining very close to zero. In addition, we observe that the increase in the expectation part of the spread is mainly generated by a drop in the short term interest rate [see figure 19(a)]. More generally, figure 19(b) shows that this shock principally affects the short run of the yield curve (steepening of the yield curve). Figure 19(c) presents the responses of the real GDP (in log). We see that the real GDP tends to slightly decrease after one period before growing to its new long term steady state. Here the positive relationship between the spread and the subsequent values of GDP growth is confirmed. These results suggest that this shock can be interpreted as a monetary policy shock: the central bank decreases the short term interest rate, leading to a lower long term interest rate. Given that the decline in the long term interest rate is smaller (in absolute value) than the fall in the short term interest rate, the spread immediately increases. With lower long term interest rates, private investment tends to increase, as well as subsequent GDP.

We observe that responses to a spread shock, reported in previous section, seem very close to the ones obtained after a shock on the expectation part of the spread. This indicates that in our sample, rises and falls in the spread has been mainly generated by shocks on its expectation part.

7.4.2 Shock on the term premium

Now, we focus on dynamic effects of a 1 percentage point increase in the spread that is completely generated by a 1 percentage point increase in the term premium (4 percentage point in annual basis). Here the new information is $I_0 = \{\eta_{0,2} = 1, \tilde{\eta}_{0,1} = 0, \tilde{\eta}_{0,2} = 1, \eta_{0,3} = 0\}$.

From equation (51) we have $\eta_{0,1} = \frac{1 - F_{2,2}(0)}{F_{2,1}(0)}$.

Then:

$$\delta = \left(\frac{1 - F_{2,2}(0)}{F_{2,1}(0)}, 1, 0 \right)'$$

Figures 20 present the responses to the 10-year term premium shock. We observe that the impulse responses of the 10-year spread and the 10-year interest rate are mainly driven by the response of the term premium. The response of the short term interest rate is very flat and close to zero. More generally, the shock seems to affect principally the long end of the yield curve (see figure (20(b)). In addition, we observe that the shock have only slight effects on the expectation part of the spread and on the long term interest rate.

Regarding the response of real GDP (see figure (20(c)), we observe that in the first year that follows the shock, the real GDP tends to decrease. Then, real GDP increases until to reach a new long term steady state value that is higher than the previous one. Therefore, the relationship between the term premium part of the 10-year spread and future economic activity is negative for short horizon (smaller than one year), whereas it is positive for longer horizon.

Giving an economic interpretation to the term premium shock is not obvious because the only macroeconomic factor we take into account in our model is the GDP growth. More precisely, to be able to interpret more accurately the shock, we should incorporate more macroeconomic variables such as inflation, private investment or government spending. Notwithstanding, the shapes of impulses responses provide us some insight about the nature of the shock. Actually, the shock induces a higher long term interest rate that is followed by an increase in activity in the long run (with short term interest rates and expectations of future short term interest rates that remain relatively stable). We can conjecture that the term premium shock could be compared to a shock on government spending that would be financed by issue of long term bonds [see also Greenwood
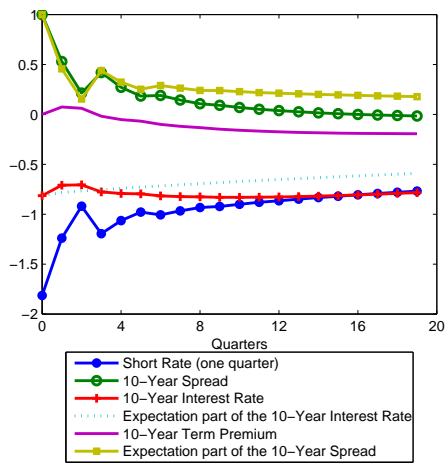
(a) 10-year spread and components
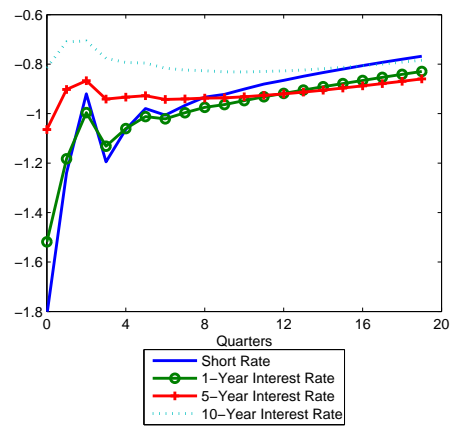
(b) Yield curve
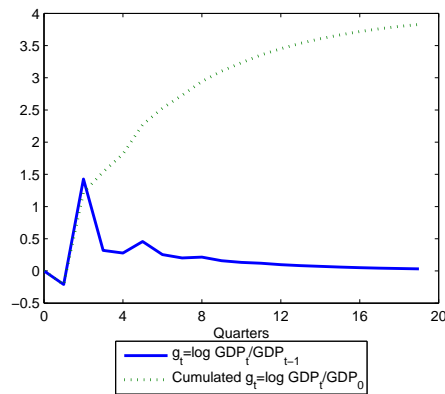
(c) GDP growth

Figure 19: Responses to a shock on the Expectation part of the 10-year spread

and Vayanos (2008)]. Such policy can generate two opposite effects on activity. First, higher long term interest rate tends to reduce private investment, and have negative effect on real GDP. Second, public investment tends to boost activity. Our results suggest that the first effect dominates in the short run, explaining the decreasing trend of real GDP during the first year, and is progressively offset by the second effect, leading the real GDP to increase in the long run. Of course, at this stage of our analysis we can only venture some interpretation that one has to verify with a more accurate macroeconomic (structural) model [see, for instance, Rudebusch and Swanson (2008a, 2008b)]. However, according to our result, the ambiguity found in the literature regarding the effect of the term premium component of the spread and future activity, could stem from the changing sign of this relationship over the period that follows the shock. Over short horizons, this relationship is negative, whereas it becomes positive for longer horizons.
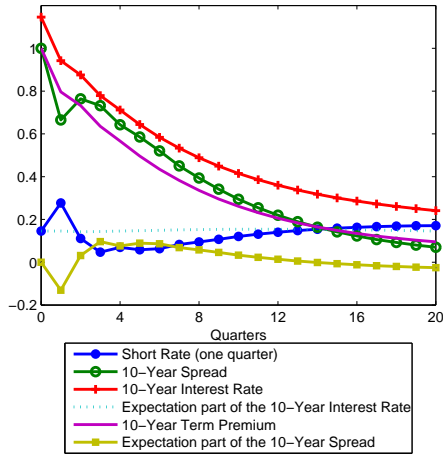
(a) 10-year spread and components
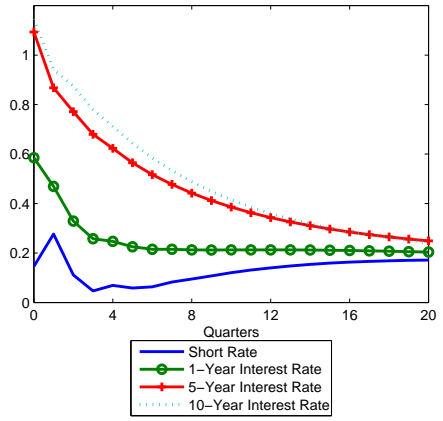
(b) Yield curve
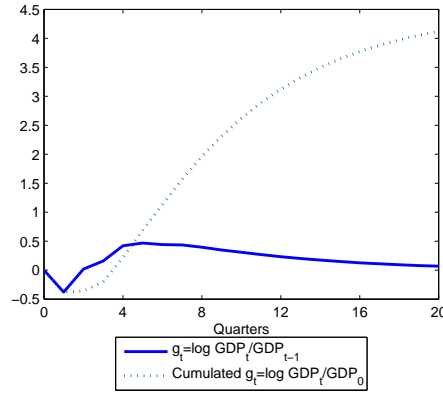
(c) GDP growth

Figure 20: Responses to a shock on the 10-year Term Premium

7.5 Shock on the short term interest rate

Finally, we focus on the dynamics effects of a decrease equal to one percentage point in the short term interest rate (4 percentage point in annual basis). The new information at date $t = 0$ is $I_0 = (\eta_{0,1} = -1, \eta_{0,3} = 0)$. Therefore, we have to determine the value of $\delta = E(\eta_0 | \eta_{0,1} = -1, \eta_{0,3} = 0)$. We have:

$$\delta = (-1, -\zeta, 0)'$$

where ζ is the coefficient of $\eta_{0,1}$ in the theoretical regression of $\eta_{0,2}$ on $\eta_{0,1}$ and $\eta_{0,3}$.

Figures 21 report the responses to the shock. Roughly speaking, an unexpected move on the short term interest rate can be interpreted as a monetary policy shock. We see that the responses to this shock are close to the one obtained with a shock on the expectation part of the spread (EXS shock hereafter). In particular, we observe that in both cases, the response of the spread seems to
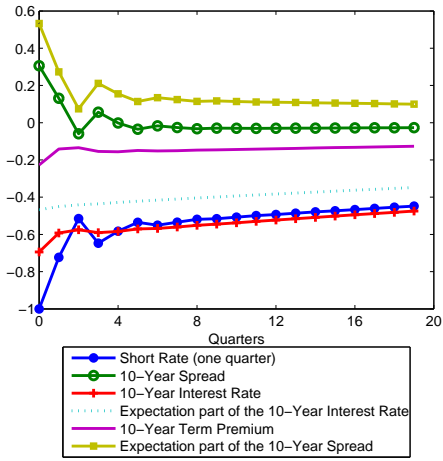
(a) 10-year spread and components
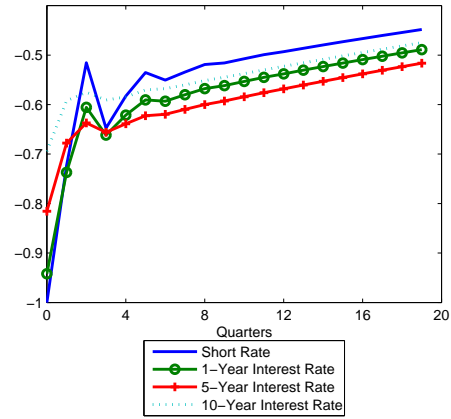
(b) Yield curve
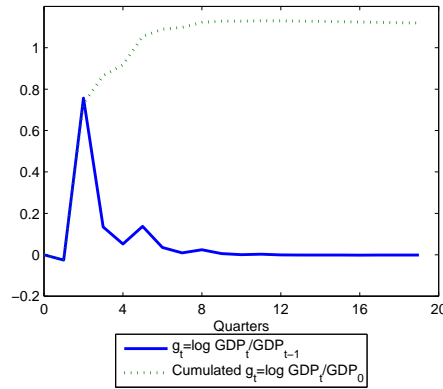
(c) GDP growth

Figure 21: Responses to a (negative) shock on the short term interest rate

be driven by its expectation part [see figure 21(a)]. This result confirms the intuition according to which the EXS shock can be viewed as a monetary policy shock.

However, some slight difference can be noted. Looking at figure 21(a), we observe that the response at $t = 0$ of the term premium to a short rate shock is negative. In the case of a EXS shock, the response of the term premium becomes negative after three quarters (recall that we controlled it to be zero at $t = 0$). In addition, at $t = 0$, the amplitude of the fall in the expectation part of the long term interest rate, $EX_0(40)$, is comparable to the one observed after an EXS shock (for the short rate shock: $EX_0(40)/r_0 = 0.45$ in quarterly basis; for the EXS shock: $EX_0(40)/r_0 = 0.44$ in quarterly basis). Therefore, recalling that $R_t(40) = EX_t(40) + TP_t(40)$, the long rate also decreases after the short term rate shock, but the fall is relatively higher in absolute value than the one obtained after an EXS shock (for the short rate shock: $R_t(40)/r_0 = 0.7$; for the EXS shock: $R_0(40)/r_0 = 0.44$ in quarterly basis). In other words, the increase in the spread is smaller than after

an EXS shock (for the short rate shock: $S_0/r_0 = 0.3$; for the EXS shock: $S_0/r_0 = 0.55$ in quarterly basis). More generally the yield curve tends to steepen, but the steepening is less pronounced than after an EXS shock (compare figures 19(b) and 21(b)).

Looking at figure 21(c), we see that the real GDP tends to increase after a negative shock on the short rate, but the long run impact is much smaller than the one associated to an EXS shock or a spread shock. Indeed, the immediate reduction in the long rate is, in that case, much larger and therefore the immediate rise in the spread is only 0.3. Here again, the positive relationship between the spread and future activity is verified.

8 Conclusions and Further Developments

In this paper we have used and developed both econometric tools and asset pricing models to study various problems concerned with the dynamic relationships between economic activity, yields and term premia on long-term bonds. The econometric tools we have used are mainly, Kullback causality measures, unit root and cointegration tests, information criteria, local-to-unit root and near-cointegration analysis. Moreover, we have developed the notion of New Information Response Function. As far as asset pricing models are concerned, we have used the theory of no-arbitrage discrete-time affine term structure models to build the yield curve, and we have introduced a notion of unbiased term premia. In addition, this notion of term premia is decomposed in various forward term premia over different horizons and in various risk premia attached to one-period holdings of bonds at different maturities.

The results obtained are promising in terms of fitting and prediction properties of our Near-Cointegrated VAR(p) term structure model, as well as in terms of evaluating term premia and disentangling the dynamic impact on the GDP growth of shocks on the expectation part and on the term premium part of the spread. Our starting point was the model proposed by APW (2006), but the various methodologies proposed here could clearly be used in different contexts, and there are obvious possible extensions of our approach. On the econometric side we could, for instance, consider the introduction of stochastic volatilities, switching regimes or fractionally integrated processes. On the macroeconomic side, it would be useful to extend the state vector in order to introduce other variables and, in particular, inflation. These are the objectives of ongoing and future research works.

Appendix 1: Further details about the unit root analysis.

The number of lags in the ADF test is selected minimizing the Akaike Information Criterion (AIC). In the (heteroskedastic-consistent) PP test, the Bartlett spectral kernel is used to estimate the spectrum, and the Newey-West (1994) procedure is used to determine the number of auto-covariance terms used. In the efficient unit root tests, we use GLS detrended data to estimate the spectral density at frequency zero, and the lag length is selected minimizing the Modified AIC (MAIC), as suggested by Ng and Perron (2001)[13]. In each test, the minimization of the information criterion is applied over lags $p \in \{0, \ldots, p_{max}\}$, with $p_{max} = [12(T/100)^{1/4}]$, where $[x]$ denotes the integer part of x, and where T denotes the sample size (in our case, $p_{max} = 13$).

In the ADF tests, we use MacKinnon (1996) critical values for the t statistic (to test the null hypothesis $\xi_0 = 0$ under $c = 0$, with ξ_0 denoting the parameter multiplying the lagged value of the process in the test regressions), while we consider Dickey and Fuller (1981, Tables IV-VI) critical values for the F statistics [to test the null joint hypothesis $(c, \xi_0)' = (0, 0)'$ or $(b, \xi_0)' = (0, 0)'$, c and b respectively denoting the constant term and the parameter of the linear time trend in the test regressions]. In the PP tests, we use MacKinnon (1996) critical values. In the Ng-Perron test, critical values are taken from their original paper (Table 1). With regard to the Dickey-Fuller GLS test, if only a constant is included in the test regression, we use MacKinnon (1996) critical values, while, if we include also a linear time trend, we apply critical values taken from Elliott, Rothenberg and Stock (1996, Table 1). Indeed, in the first case only, their t-statistic follows a Dickey-Fuller distribution. In the Point-Optimal test, critical values are provided by Elliott, Rothenberg and Stock (1996, Table 1).

In all unit root tests we have considered, the null hypothesis (presence of a unit root in the scalar time series) is rejected when the value of the test statistic is lower than the critical value (critical region). On the contrary, in the F test, the null hypothesis is rejected when the value of the test statistic is bigger than the critical value.

---

[13]Ng and Perron (2001) show that, starting from the findings of Elliott, Rothenberg and Stock (1996) and Dufour and King (1991), the use in conjonction of the MAIC and GLS detrended data, lead to tests with size and power gains with respect to the tests proposed by Ng and Perron (1996).

Appendix 2: Proof of Proposition 1

Assuming that (19) is true for $h - 1$, we get:

$$
\begin{aligned}
B_t(h) &= \exp(c_h' \tilde{X}_t + d_h) \\[2mm]
&= E_t[M_{t,t+1} \cdots M_{t+H-1,t+H}] \\[2mm]
&= E_t[M_{t,t+1} B_{t+1}(h-1)] \\[2mm]
&= \exp\left[-\beta - \alpha' \tilde{X}_t - \tfrac{1}{2}\Gamma_t'\Gamma_t + d_{h-1}\right] \times E_t[\exp\left(\Gamma_t'\eta_{t+1} + c_{h-1}'\tilde{X}_{t+1}\right)] \\[2mm]
&= \exp\left[-\beta - \alpha'\tilde{X}_t - \tfrac{1}{2}\Gamma_t'\Gamma_t + d_{h-1} + c_{h-1}'\tilde{\Phi}\tilde{X}_t + c_{1,h-1}'\tilde{\nu}\right] \\[2mm]
&\qquad \times E_t[\exp\left(\Gamma_t + \Sigma' c_{1,h-1}\right)'\eta_{t+1})] \\[2mm]
&= \exp\left[\left(-\alpha + \tilde{\Phi}' c_{h-1} + (\Sigma\gamma)' c_{1,h-1}\right)' \tilde{X}_t \right. \\[2mm]
&\qquad \left. + \left(-\beta + c_{1,h-1}'(\tilde{\nu} + \Sigma\gamma_o) + \tfrac{1}{2}c_{1,h-1}'\Sigma\Sigma' c_{1,h-1} + d_{h-1}\right)\right],
\end{aligned}
$$

(A.1)

and by identifying the coefficients we find the recursive relation presented in Proposition 1. $\square$

Appendix 3: Tables and Graphs.

Unit Root test	r_t $p-1$	test value	1 %	5 %	10 %	R_t $p-1$	test value	1 %	5 %	10 %
ADF t-stat	12	-1.90	-3.47	-2.88	-2.58	0	-1.83	-3.47	-2.88	-2.58
ADF F-stat		1.81	6.70	4.71	3.86		1.60	6.70	4.71	3.86
PP Adj. t-stat		-2.66	-3.47	-2.88	-2.58		-1.83	-3.47	-2.88	-2.58
Ng-Perron MZ_α^{GLS} stat	8	-7.89	-13.8	-8.10	-5.70	5	-2.27	-13.8	-8.10	-5.70
MZ_t^{GLS} stat	8	-1.99	-2.58	-1.98	-1.62	5	-1.06	-2.58	-1.98	-1.62
MSB^{GLS} stat	8	0.25	0.17	0.23	0.27	5	0.47	0.17	0.23	0.27
MP_T^{GLS} stat	8	3.11	1.78	3.17	4.45	5	10.79	1.78	3.17	4.45
D-F GLS t-stat	8	-1.78	-2.58	-1.94	-1.62	5	-1.02	-2.58	-1.94	-1.62
Point-Opt P_T-stat	8	3.19	1.92	3.15	4.29	5	12.25	1.92	3.15	4.29

Table A. 1: Unit root tests for the (one-quarter) short rate r_t (left panel) and for the (40-quarters) long rate R_t (right panel) (a constant is included in test regressions). In the ADF unit root test, based on the OLS regression $\Delta x_t = c + \xi_0 x_{t-1} + \sum_{j=1}^{p-1} \xi_j \Delta x_{t-j} + \varepsilon_t$ [with $\varepsilon_t \sim i.i.d.(0,\sigma^2)$ and with p denoting the AR order], we consider both the t-statistic, to test the null hypothesis $\xi_0 = 0$, and the F-statistic to test the joint hypothesis $(c, \xi_0)' = (0,0)'$.

Unit Root test	G_t $p-1$	test value	1 %	5 %	10 %	G_t $p-1$	test value	1 %	5 %	10 %
ADF t-stat	2	-0.91	-3.47	-2.88	-2.58	2	-4.39	-4.01	-3.44	-3.14
ADF F-stat		13.29	6.70	4.71	3.86		9.89	8.73	6.49	5.47
PP Adj. t-stat		-1.13	-3.47	-2.88	-2.58		-3.80	-4.01	-3.44	-3.14
Ng-Perron MZ_α^{GLS} stat	11	1.46	-13.80	-8.10	-5.70	1	-9.83	-23.80	-17.30	-14.20
MZ_t^{GLS} stat	11	2.49	-2.58	-1.98	-1.62	1	-2.16	-3.42	-2.91	-2.62
MSB^{GLS} stat	11	1.70	0.17	0.23	0.27	1	0.22	0.14	0.17	0.18
MP_T^{GLS} stat	11	207.15	1.78	3.17	4.45	1	9.54	4.03	5.48	6.67
D-F GLS t-stat	11	1.58	-2.58	-1.94	-1.62	1	-2.16	-3.49	-2.96	-2.67
Point-Optimal P_T-stat	11	251.67	1.92	3.15	4.29	1	11.31	4.10	5.65	6.84

Table A. 2: Left Panel: unit root tests for the log-GDP G_t (a constant is included in test regressions). In the ADF unit root test, based on the OLS regression $\Delta G_t = c + \xi_0 G_{t-1} + \sum_{j=1}^{p-1} \xi_j \Delta G_{t-j} + \varepsilon_t$ [with $\varepsilon_t \sim i.i.d.(0,\sigma^2)$ and with p denoting the AR order], we consider both the t-statistic, to test the null hypothesis $\xi_0 = 0$, and the F-statistic to test the joint hypothesis $(c, \xi_0)' = (0,0)'$. Right Panel: Unit root tests for the log-GDP G_t (a constant and a linear time trend are included in test regressions). In the ADF unit root test, based on the OLS regression $\Delta G_t = c + bt + \xi_0 G_{t-1} + \sum_{j=1}^{p-1} \xi_j \Delta G_{t-j} + \varepsilon_t$ [with $\varepsilon_t \sim i.i.d.(0,\sigma^2)$ and with p denoting the AR order], we consider both the t-statistic, to test the null hypothesis $\xi_0 = 0$, and the F-statistic to test the joint hypothesis $(b, \xi_0)' = (0,0)'$.

Lag p	LR	FPE	AIC	SIC	HQ
0	N.A.	6.15e-11	-15.00	-14.94	-14.98
1	1885.28	7.99e-16	-26.25	-26.03*	-26.16
2	39.84	6.96e-16	-26.39	-26.00	-26.23*
3	22.43*	6.72e-16*	-26.42*	-25.87	-26.20
4	10.90	6.98e-16	-26.39	-25.67	-26.09

Table A. 3: Criteria for VAR order selection. Given a sample period of size T, and a n-dimensional Gaussian VAR(p) process with empirical white noise covariance matrix $\hat{\Omega}(p)$, $LR = (T-m)[log|\hat{\Omega}(p-1)| - log|\hat{\Omega}(p)|]$ denotes, for each lag p, the sequential modified [Sims (1980)] likelihood ratio (LR) test statistic, where m is the number of parameters per equation under the alternative. The modified LR statistics are compared to the 5% critical values. $FPE = [(T+np+1)/(T-np-1)]^n \det(\hat{\Omega}(p))$ denotes, for each lag p, the final prediction error criterion. If we denote by log-L $= -(Tn/2)\log(2\pi) + (T/2)\log(|\hat{\Omega}(p)^{-1}|) - (Tn/2)$ the maximum value of the log-likelihood function associated to the VAR(p) model, $AIC = -2\text{log-L}/T + 2pn^2/T$, $SIC = -2\text{log-L}/T + (\log(T)/T)pn^2$ and $HQ = -2\text{log-L}/T + (2\log(\log(T))/T)pn^2$ denote, respectively and for each lag p, the Akaike, Schwarz and Hannan-Quinn information criteria. For each criterion, and starting from a maximum lag of $p=4$, (*) denotes the optimal number of lags.

	ν	Φ_1			Φ_2			Φ_3		
r_t	0.0013	0.6068	0.1431	0.0350	0.1023	-0.1896	0.0492	0.3239	-0.0570	-0.0843
	[0.2289]	[6.4211]	[0.7662]	[1.1271]	[0.9346]	[-0.8029]	[1.0878]	[3.1638]	[-0.3204]	[-2.9603]
R_t	0.0016	0.0198	0.8108	0.0120	0.0111	0.0855	-0.0030	0.0797	-0.0462	-0.0091
	[0.5235]	[0.3886]	[8.0530]	[0.7187]	[0.1889]	[0.6715]	[-0.1235]	[1.4430]	[-0.4818]	[-0.5931]
G_t	0.04404	0.2289	-0.3115	1.1425	-1.0742	0.4420	0.0129	0.2328	0.3175	-0.1597
	[2.8522]	[0.9121]	[-0.6282]	[13.8549]	[-3.6955]	[0.7048]	[0.1075]	[0.8563]	[0.6721]	[-2.1117]

$\Omega \times 10^3$			Corr.		log-L	$|\psi|$
0.0079	0.0026	0.0048	ρ_{12}	0.6060	2289.43	0.9968
[8.9722]	[6.5758]	[2.8558]			AIC	0.9315
.	0.0023	0.0036	ρ_{13}	0.2310	-26.4261	0.8647
	[8.9722]	[3.8719]			SIC	0.6170
.	.	0.0555	ρ_{23}	0.3205	-25.8749	0.5595$(^c)$
		[8.9722]			FPE	0.2891
					6.70e-16	0.0840$(^c)$

Table A. 4: Parameter estimates of the state dynamics $Y_t = \nu + \sum_{j=1}^3 \Phi_j Y_{t-j} + \varepsilon_t$, with $Y_t = (r_t, R_t, G_t)'$ [Gurkaynak-Sack-Wright (2007) data base; sample period : 1964:Q1 - 2007:Q2]. t-values are in brackets. ρ_{ij} denotes the (empirical) correlation between (ε_{it}) and (ε_{jt}). log-L denotes the maximum value of the log-Likelihood function. $|\psi|$ indicates the modulus of the roots of equation $|\tilde{\Phi}(\psi)| = 0$, with $\tilde{\Phi}(\psi) = (I_{3\times3}\psi^3 - \Phi_1\psi^2 - \Phi_2\psi - \Phi_3)$ denoting the characteristic polynomial; $(^c)$ indicates a pair of complex conjugate roots.

r	Eigen. λ_i	Trace Statistic	5 % Crit. Value	p-value	Max-Eigen Statistic	5 % Crit. Value	p-value
0	0.1413	34.4260	29.7971	0.0136	26.0462	21.1316	0.0094
1	0.0398	8.3798	15.4947	0.4257	6.9429	14.2646	0.4959
2	0.0084	1.4369	3.8415	0.2306	1.4369	3.8415	0.2306
α	0.0419	0.1178	-0.5778	β	1.0000	-1.0029	0.0031
	[0.5551]	[2.8829]	[-2.8754]			$\langle 0.00004 \rangle$	$\langle 0.0877 \rangle$

Table A. 5: Johansen cointegration tests for the variables (r_t, R_t, G_t) observed quarterly from 1964:Q1 to 2007:Q2 [Gurkaynak-Sack-Wright (2007) data base]. The null hypothesis is for both tests $H_0 : \text{rank}(\Pi) = r$. In the Trace test, the alternative hypothesis is $H_A : \text{rank}(\Pi) = 3$, and the associated statistic is given by $2(\log\text{-L}_A - \log\text{-L}_0) = -T\sum_{i=r+1}^3 \log(1 - \lambda_i)$, where $\log\text{-L}_A$ and $\log\text{-L}_0$ denote, respectively, the maximum value of the log-Likelihood function (of model (13)) under the case of 3 and $r < 3$ cointegrating relations. In the Maximum Eigenvalue test, $H_A : \text{rank}(\Pi) = r + 1$, and $2(\log\text{-L}_A - \log\text{-L}_0) = -T\log(1 - \lambda_{r+1})$. Both test statistics accept at 5 % the hypothesis $\text{rank}(\Pi) = 1$ [we use MacKinnon, Haug, and Michelis (1999) p-values]. Under the restriction $r = 1$, the second half of the table provides the estimates of the adjusement parameters $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ (t-values are in brackets) and the cointegrating vector $\beta = (1, \beta_2, \beta_3)'$. For parameters β_2 and β_3 we report in angled brackets, respectively, the p-value of the $\chi^2(1)$-distributed likelihood ratio statistic associated to the test $H_0 : \beta = (1, 0, \beta_3)'$ and $H_0 : \beta = (1, \beta_2, 0)'$. The alternative hypothesis is $H_A : \beta = (1, \beta_2, \beta_3)'$ in both cases, and the 5% and 1% critical values for a $\chi^2(1)$ are, respectively, 3.84 and 6.63.

	γ	Γ_1			Γ_2			α	μ
Δr_t	-0.0010	-0.4154	0.1941	0.0393	-0.3152	0.0110	0.0893	-0.0349	-0.0029
	[-2.7075]	[-3.9779]	[1.1085]	[1.2985]	[-3.1595]	[0.0626]	[3.1587]	[-0.5074]	
ΔR_t	-0.0002	-0.0862	-0.0675	0.0148	-0.0761	0.0216	0.0120	-0.1132	
	[-0.9910]	[-1.5306]	[-0.7148]	[0.9048]	[-1.4135]	[0.2270]	[0.7891]	[-3.0492]	
ΔG_t	0.0048	0.6854	-0.7554	0.1964	-0.3700	-0.3038	0.1912	0.4086	
	[4.8954]	[2.4406]	[-1.6043]	[2.4138]	[-1.3792]	[-0.6415]	[2.5147]	[2.2074]	

$\Omega \times 10^3$			log-L	$\mid\psi\mid$
0.0079	0.0026	0.0051	2283.60	1.0000**
[9.0277]	[6.6826]	[2.9613]	AIC	0.8478
.	0.0023	0.0038	-26.3930	0.6271
	[9.0277]	[3.9628]	SIC	0.5489(^{c})
.	.	0.0573	-25.8969	0.2468
		[9.0277]	FPE	0.1050(^{c})
			7.18e-16	

Table A. 6: Parameter estimates of the model $\Delta Y_t = \alpha(\beta' Y_{t-1} + \mu) + \sum_{j=1}^{2} \Gamma_j \Delta Y_{t-j} + \gamma + \varepsilon_t$, with $\Delta Y_t = (\Delta r_t, \Delta R_t, \Delta G_t)'$, when $\text{rank}(\alpha\beta') = 1$ and $\beta = (-1, 1, 0)'$ [Gurkaynak-Sack-Wright (2007) data base; sample period : 1964:Q1 - 2007:Q2]. t-values are in brackets. log-L denotes the maximum value of the log-Likelihood function. $\mid\psi\mid$ indicates the modulus of the roots of equation $|\tilde{\Phi}(\psi)| = 0$, with $\tilde{\Phi}(\psi) = (I_{3\times3}\psi^3 - \Phi_1\psi^2 - \Phi_2\psi - \Phi_3)$ denoting the characteristic polynomial; (^{c}) indicates a pair of complex conjugate roots, while (**) denote a root with multiplicity two.

	$\tilde{\nu}$	$\tilde{\Phi}_1$			$\tilde{\Phi}_2$			$\tilde{\Phi}_3$		
r_t	-0.0009	0.7787	0.1592	0.0393	-0.0829	-0.1831	0.0893	0.3042	-0.0110	0.0000
S_t	0.0010	0.0675	0.6600	-0.0245	0.1822	0.2722	-0.0773	-0.2497	-0.0106	0.0000
g_t	0.0036	-0.0701	-0.3469	0.1964	-0.6038	0.4516	0.1912	0.6739	0.3038	0.0000

$\tilde{\Omega} \times 10^3$			Corr.	
0.0079	-0.0053	0.0051	ρ_{12}	-0.8435
.	0.0050	-0.0013	ρ_{13}	0.2385
.	.	0.0573	ρ_{23}	-0.0786

Table A. 7: Parameter estimates of the CVAR(3) state dynamics $X_t = \tilde{\nu} + \sum_{j=1}^{3} \tilde{\Phi}_j X_{t-j} + \varepsilon_t$, with $X_t = (r_t, S_t, g_t)'$ [Gurkaynak-Sack-Wright (2007) data base; sample period : 1964:Q1 - 2007:Q2]. ρ_{ij} denotes the (empirical) correlation between (ε_{it}) and (ε_{jt}).

	ν	Φ			$\Omega \times 10^3$		
r_t	0.0009	0.9307	0.0694	0.0022	0.0088	-0.0061	0.0049
	[1.0679]	[23.8278]	[0.9523]	[0.0785]	[9.1378]	[-8.4468]	[2.6368]
S_t	0.00008	0.0296	0.8223	-0.0043	.	0.0056	-0.0014
	[0.1275]	[0.9490]	[14.1173]	[-0.1970]		[9.1378]	[-0.9733]
g_t	0.0079	-0.1626	0.1898	0.2414	.	.	0.0626
	[3.6441]	[-1.5640]	[0.9782]	[3.2890]			[9.1378]

Corr.		log-L	$|\psi|$
ρ_{12}	-0.8637	2258.76	0.9472
		AIC	0.8037
ρ_{13}	0.2084	-26.2779	0.2435
		SIC	
ρ_{23}	-0.0755	-26.0574	
		FPE	
		7.77e-16	

Table A. 8: Parameter estimates of the model $X_t = \nu + \Phi X_{t-1} + \varepsilon_t$, with $X_t = (r_t, S_t, g_t)'$ [Gurkaynak-Sack-Wright (2007) data base; sample period : 1964:Q1 - 2007:Q2]. t-values are in brackets. ρ_{ij} denotes the (empirical) correlation between (ε_{it}) and (ε_{jt}). log-L denotes the maximum value of the log-Likelihood function. $|\psi|$ indicates the modulus of the roots of equation $|\tilde{\Phi}(\psi)| = 0$, with $\tilde{\Phi}(\psi) = (I_{3\times3}\psi - \Phi)$ denoting the characteristic polynomial.

	ν	Φ_1			Φ_2			Φ_3		
r_t	0.0002	0.7588	0.1777	0.0295	-0.0739	-0.1960	0.0743	0.2486	-0.0783	0.0428
	[0.2043]	[4.9693]	[0.9530]	[0.9736]	[-0.3727]	[-0.8361]	[2.5760]	[1.6408]	[-0.4438]	[1.5634]
S_t	0.0006	0.0737	0.6416	-0.0182	0.1745	0.2798	-0.0676	-0.2192	0.0250	-0.0312
	[0.9415]	[0.6043]	[4.3039]	[-0.7544]	[1.1009]	[1.4931]	[-2.9311]	[-1.8104]	[0.1771]	[-1.4249]
g_t	0.0061	-0.1316	-0.4212	0.1994	-0.6263	0.4486	0.1969	0.6386	0.2473	-0.0532
	[2.7461]	[-0.3165]	[-0.8295]	[2.4207]	[-1.1596]	[0.7028]	[2.5077]	[1.5481]	[0.5148]	[-0.7140]

$\Omega \times 10^3$			Corr.		log-L	$|\psi|$
0.0077	-0.0052	0.0050	ρ_{12}	-0.8412	2287.93	0.9299
[8.9722]	[-8.1680]	[2.9381]			AIC	0.8888
.	0.0049	-0.0013	ρ_{13}	0.2380	-26.4085	0.6015(^{c})
	[8.9722]	[-0.9965]			SIC	0.49188
.	.	0.0574	ρ_{23}	-0.0788	-25.8574	0.4283(^{c})
		[8.9722]			FPE	0.2449
					6.82e-16	0.1598

Table A. 9: Parameter estimates of the unconstrained VAR(3) state dynamics $X_t = \nu + \sum_{j=1}^{3} \Phi_j X_{t-j} + \varepsilon_t$, with $X_t = (r_t, S_t, g_t)'$ [Gurkaynak-Sack-Wright (2007) data base; sample period : 1964:Q1 - 2007:Q2]. t-values are in brackets. ρ_{ij} denotes the (empirical) correlation between (ε_{it}) and (ε_{jt}). log-L denotes the maximum value of the log-Likelihood function. $|\psi|$ indicates the modulus of the roots of equation $|\tilde{\Phi}(\psi)| = 0$, with $\tilde{\Phi}(\psi) = (I_{3\times3}\psi^3 - \Phi_1\psi^2 - \Phi_2\psi - \Phi_3)$ denoting the characteristic polynomial; (^{c}) indicates a pair of complex conjugate roots.

	γ_o	γ_1			γ_2			γ_3		
r_t	-0.176	105.592	105.983	-17.555	30.179	24.127	-23.555	-111.760	-10.126	-0.942
	[-3.499]	[7.290]	[5.773]	[-5.776]	[1.302]	[0.893]	[-7.760]	[-8.712]	[-1.309]	[-0.386]
S_t	-0.012	45.465	88.585	2.057	-106.482	-91.175	9.348	47.026	40.574	-1.870
	[-0.132]	[1.646]	[2.524]	[0.358]	[-2.423]	[-1.771]	[1.664]	[2.091]	[1.521]	[-0.426]
g_t	2.613	-113.000	49.931	-39.328	-120.313	-29.174	-38.817	104.875	-267.899	64.851
	[1.024]	[-0.152]	[0.053]	[-0.327]	[-0.108]	[-0.022]	[-0.328]	[0.175]	[-0.375]	[0.637]

Table A. 10: Risk sensitivity parameter estimates for the NCVAR(3) factor-based term structure model [Gurkaynak-Sack-Wright (2007) data base; sample period : 1964:Q1 - 2007:Q2]. t-values are in brackets.

I	γ_o	γ_1			γ_2			γ_3		
r_t	-0.137	100.691	102.269	-14.896	19.314	8.438	-20.437	-100.213	-15.560	0
	[-2.844]	[6.450]	[5.324]	[-4.821]	[0.800]	[0.304]	[-7.622]	[-7.337]	[-0.978]	
S_t	0.092	47.176	93.102	-0.956	-97.518	-79.733	5.604	32.176	30.835	0
	[1.450]	[1.743]	[2.743]	[-0.166]	[-2.225]	[-1.573]	[1.233]	[1.467]	[1.188]	
g_t	1.540	47.995	99.149	-41.930	-232.010	77.676	62.374	91.746	-349.378	0
	[1.648]	[0.126]	[0.204]	[-0.587]	[-0.391]	[0.111]	[1.057]	[0.304]	[-0.968]	

II	γ_o	γ_1			γ_2			γ_3		
r_t	-0.525	108.229	98.626	-11.656	19.073	15.907	-18.090	-83.383	4.561	-11.029
	[-10.560]	[7.004]	[5.143]	[-3.828]	[0.792]	[0.577]	[-6.043]	[-6.107]	[0.286]	[-4.488]
S_t	-0.162	54.490	97.186	-0.681	-97.839	-80.545	7.869	38.467	41.643	-1.424
	[-2.253]	[2.001]	[2.831]	[-0.112]	[-2.210]	[-1.585]	[1.401]	[1.747]	[1.565]	[-0.325]
g_t	1.714	43.853	140.023	-61.069	-250.282	108.642	4.151	109.366	-416.392	61.739
	[1.374]	[0.095]	[0.242]	[-0.646]	[-0.340]	[0.130]	[0.045]	[0.300]	[-0.955]	[0.912]

III	γ_o	γ_1		
r_t	-0.478	34.789	56.968	-4.557
	[-6.629]	[11.523]	[9.066]	[-1.777]
S_t	-0.301	2.384	66.540	5.373
	[-4.304]	[1.115]	[13.899]	[1.771]
g_t	1.588	0.648	4.256	-5.401
	[0.539]	[0.006]	[0.018]	[-0.057]

Table A. 11: Risk sensitivity parameter estimates for the Cointegrated VAR(3) (panel I), the unconstrained VAR(3) (panel II) and the (unconstrained) VAR(1) (panel III) factor-based term structure models [Gurkaynak-Sack-Wright (2007) data base; sample period : 1964:Q1 - 2007:Q2]. t-values are in brackets.
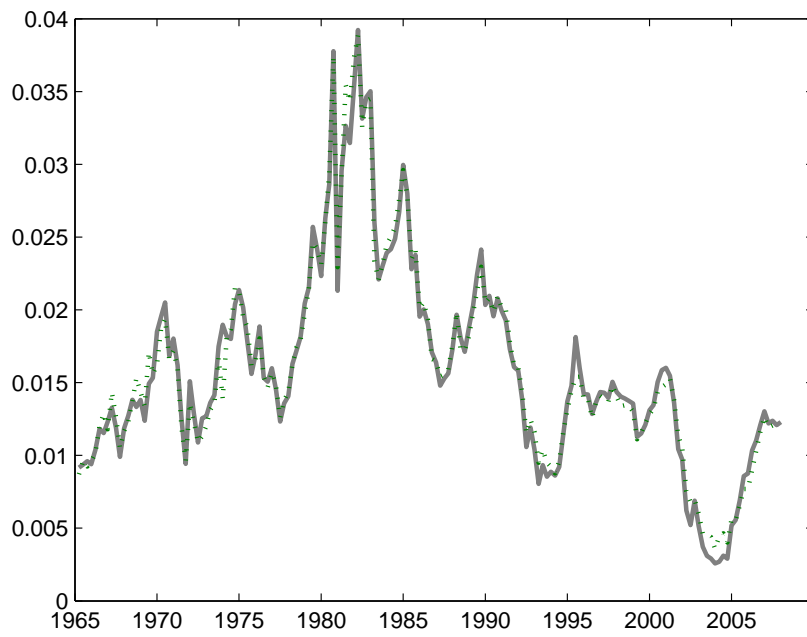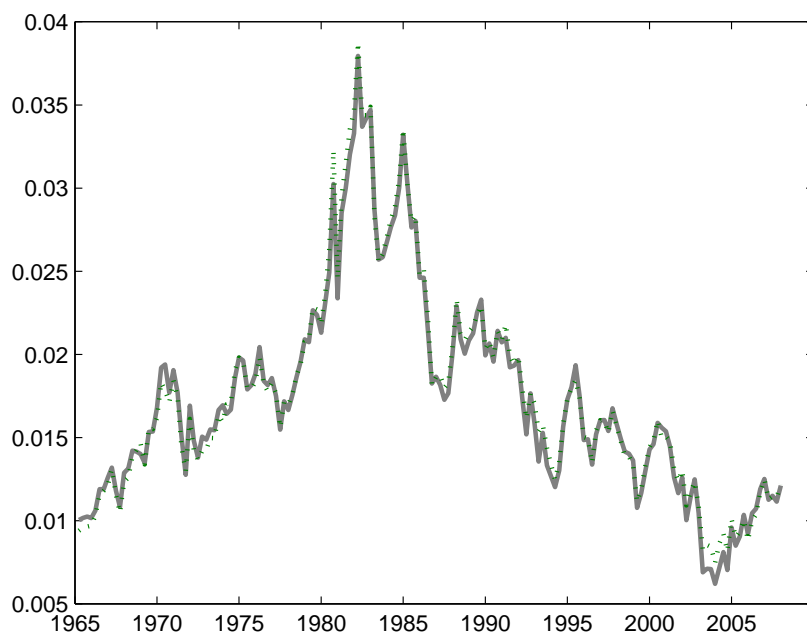
Figure A. 1:

Figure A. 2:
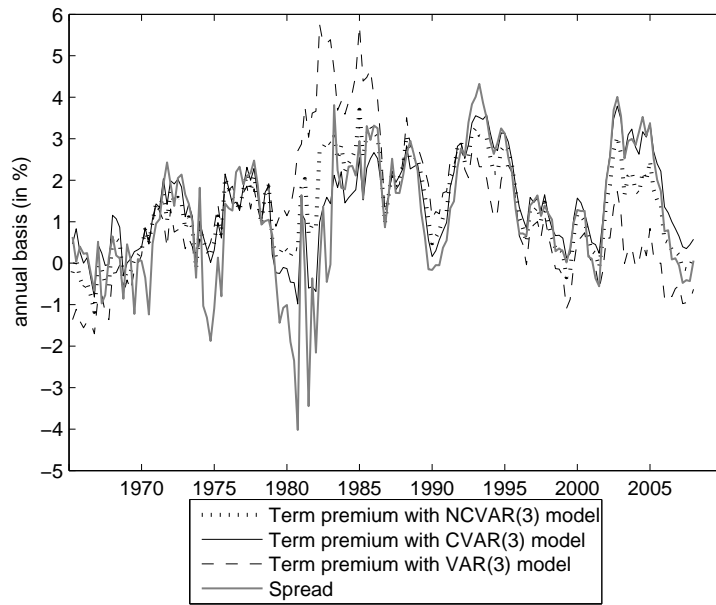5-Year interest rate, fitted (dashed) and observed (solid)
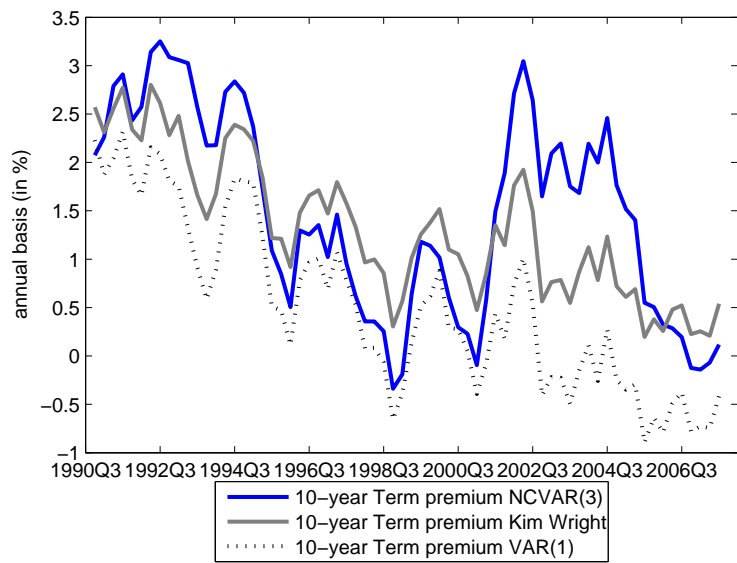
Figure A. 3: Term premia

Figure A. 4:
NCVAR(3), Kim Wright and Ang, Piazzesi, Wei (VAR(1))
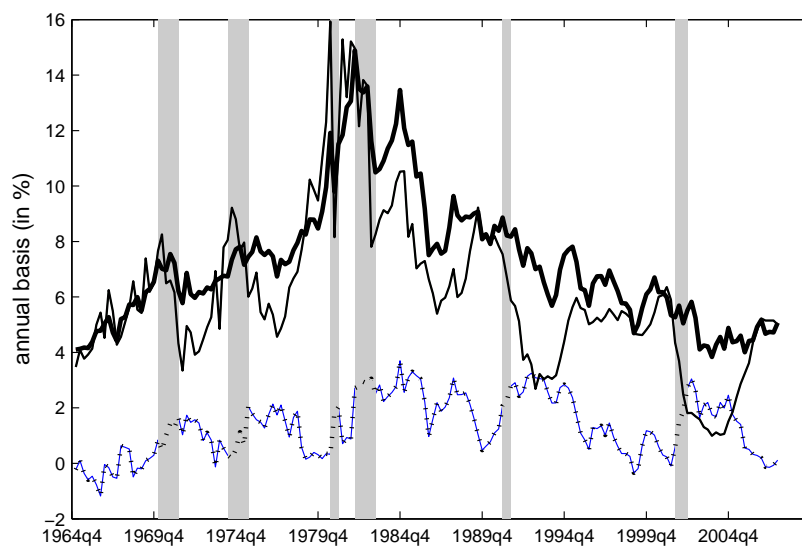10-year Term premia

Figure A. 5: NCVAR(3) term premium
Bold line: 10-year interest rate
Thin line: short term interest rate
Dotted line: 10-year term premium
Shaded areas: recession dates (NBER)

R E F E R E N C E S

Agiakloglou, C., Newbold, P., Wohar, M., (1993): "Bias in an Estimator of the Fractional Difference Parameter", Journal of Time Series Analysis, 14, 235-246.

Andrews, D.W.K. (1993): "Exactly median-unbiased estimation of first order autoregressive/unit root models", Econometrica, 61, 1, 139-165.

Andrews, D.W.K., Guggenberger, P. (2007): "Asymptotics for stationary very nearly unit root processes", Cowles Foundation D.P. n. 167.

Ang, A., Bekaert, G., (2002): "Short Rate Nonlinearities and Regime Switches", Journal of Economic Dynamics and Control, 26, 1243-1274.

Ang, A., Bekaert, G., Wei, M. (2008): "Term Structure of Real Rates and Expected Inflation", The Journal of Finance, 63(2), 797-849.

Ang, A., Piazzesi, M., Wei, M., (2006): "What does the Yield Curve tell us about GDP Growth", Journal of Econometrics, 131, 359-403.

Bansal, R., Zhou, H., (2002): "Term Structure of Interest Rates with Regime Shifts", Journal of Finance, 57, 1997-2043.

Bansal, R., Tauchen, G., Zhou, H., (2004): "Regime Shifts, Risk Premiums in the Term Structure, and the Business Cycle", Journal of Business and Economic Statistics, 22(4), 396-409.

Bernanke, B.S., Reinhart, V.R., and Sack, B. (2004): "Monetary Policy Alternatives at the Zero Bound: An Empirical Assessment", Board of Governors of the Federal Reserve System, Working Paper 2004-48.

Bikbov, R., and Chernov, M., (2006): "No-Arbitrage Macroeconomic Determinants of the Yield Curve", London Business School, Working Paper.

Campbell, J.Y., Shiller, R.J. (1987): "Cointegration and tests of present value models", Journal of Political Economy, 95 (5), 1062-1088.

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

Cochrane, J., and Piazzesi, M., (2008): "Decomposing the Yield Curve", Graduate School of Business, University of Chicago, Working Paper.

Chan, N.H., Wei, C.Z., (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes", Annals of Statistics, 15, 1050-1063.

Chernov, M., Mueller, P., (2008): "The Term Structure of Inflation Expectations", London Business School, Working Paper.

Dai, Q., and Singleton, K., (2002): "Expectations Puzzles, Time-varying Risk Premia, and Affine Model of the Term Structure", Journal of Financial Economics, 63, 415-441.

Dai, Q., and Singleton, K., (2003): "Term Structure Dynamics in Theory and Reality", Review of Financial Studies, 16, 631-678.

Dai, Q., Singleton, K., and W. Yang, (2007) : "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields", Review of Financial Studies, 20(5), 1669-1706.

DeJong, D.N., Nankervis, J.C., Savin, N.E., Whiteman, C.H., (1992a): "The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors", Journal of Econometrics, 53, 323-343.

DeJong, D.N., Nankervis, J.C., Savin, N.E., Whiteman, C.H., (1992b): "Integration Versus Trend Stationary in Time Series", Econometrica, 60(2), 423-433.

Dewachter, H., Lyrio, M., (2006): "Macro Factors and the Term Structure of Interest Rates", Journal of Money, Credit and Banking, 38(1), 119-140.

Dewachter, H., Lyrio, M., (2008): "Learning, Macroeconomic Dynamics and the Term Structure of Interest Rates", Asset Prices and Monetary Policy, NBER Conference Report, University of Chicago Press.

Dewachter, H., Lyrio, M., Maes, K., (2006): "A Joint Model for the Term Structure of Interest Rates and the Macroeconomy", Journal of Applied Econometrics, 21(4), 439-462.

Dickey, D.A., Fuller, W.A., (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with Unit Roots", Econometrica, 49(4), 1057-1072.

Diebold, F.X., Kilian, L., (2000): "Unit-root tests are useful for selecting forecasting models", Journal of Business and Economic Statistics, 18, 265-273.

Dotsey, M., (1998): "The predictive content of the interest rate term spread for future economic growth", Federal Reserve Bank of Richmond Economic Quarterly 84 (3), 31-51.

Driffill, J., Kenc, T., Sola, M., (2003): "An Empirical Examination of Term Structure Models with Regime Shifts", Working Paper.

Dufour J-M., King, M.L. (1991): "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors," Journal of Econometrics, 47, 115-143.

Elliott, G. (1999): "Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution", International Economic Review, 40, 767-783.

Elliott, G., (2006): "Unit root pre-testing and forecasting", Working Paper, UCSD.

Elliott, G., and Muller, U. (2003): "Tests for unit roots and the initial conditions", Econometrica, 71, 1269-1286.

Elliott, G., Stock, J.H. (2001): "Confidence intervals for autoregressive coefficients near one", Journal of Econometrics, 103, 155-181.

Elliott, G., Rothenberg, T.J., Stock, J.H., (1996): "Efficient Tests for an Autoregressive Unit Root", Econometrica, 64(4), 813-836.

Engle, R. F., Granger, C. W. J., (1987): "Cointegration and Error Correction: Representation, Estimation and Testing", Econometrica, 55, 251-276.

Estrella, A., Hardouvelis, G.A., (1991): "The term structure as predictor of real economic activity", Journal of Finance 46, 555-576.

Estrella, A., Mishkin, F.S., (1998): "Predicting U.S. recessions: financial variables as leading indicators", Review of Economics and Statistics 1, 45-61.

Evans, G.B.A., Savin, E. (1981): "Testing for unit roots I", Econometrica, 49, 3, 753-779.

Evans, M., (2003): "Real Risk, Inflation Risk, and the Term Structure", The Economic Journal, 113, 345-389.

Fama, E.F., Bliss, R.R., (1987): "The Information in Long-Maturity Forward Rates". American Economic Review 77, 680-692.

Favero, C., Kaminska, I., Södeström, U., (2005): "The Predictive Power of the Yield Spread: Further Evidence and a Structural Interpretation", Working Paper, Università Bocconi.

Gallant, R.A., Rossi, P.E., and Tauchen, G., (1993): "Nonlinear Dynamic Structures", Econometrica, 61(4), 871-907.

Garcia, R., Luger, R., (2007): "Risk Aversion, Intertemporal Substitution, and the Term Structure of Interest Rates", Working Paper, Université de Montreal.

Garcia, R., Perron, P., (1996): "An Analysis of Real Interest Rate Under Regime Shifts", Reviews of Economics and Statistics, 1, 111-125.

Geweke, J., Porter-Hudak, S., (1983): "The Estimation and Application of Long Memory Time Series Models", Journal of Time Series Analysis, 4(4), 221-238.

Giraitis, L., Phillips, P.C.B. (2006): "Uniform limit theory for stationary autoregressions", Journal of Time Series Analysis, 27, 51-60.

Gourieroux, C., Jasiak, J., (1999): "Nonlinear Innovations and Impulse Responses", Working Paper, CREST DP.

Gourieroux, C., Monfort, A. (1996): "Simulation based econometric methods", Oxford University Press.

Gourieroux, C., Monfort, A., (1997): "Time series and dynamic models", Cambridge University Press.

Gourieroux, C., Monfort, A., Renault, E., (1987): "Kullback causality measures", Annales d'économie et de Statistique, 6/7, 369-410.

Gourieroux, C., Monfort, A., Renault, E., (1993): "Indirect Inference", Journal of Applied Econometrics, 8, S85-S118.

Gourieroux, C., Phillips, P.C.A., Yu, J. (2007): "Indirect Inference of Dynamic Panel Models", Journal of Econometrics, forthcoming.

Gourieroux, C., Renault, E., Touzi, N. (2000): "Calibration by simulation for small sample's bias correction", in Simulation-Based Inference in Econometrics, Mariano, Schuermann, Weeks eds, Cambridge University Press.

Gray, S., (1996): "Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process", Journal of Financial Economics, 42, 27-62.

Greenwood, R., and Vayanos, D., (2008): "Bond Supply and Excess Bond Returns", Harvard Business School, working paper.

Gürkaynak, R., Sack, B., Wright, J., (2007), The U.S. Treasury Yield Curve: 1961 to the Present", Journal of Monetary Economics, 54, 2291-2304.

Hall, A.D., Anderson, H.M., Granger, C.W.J. (1992): "A Cointegration Analysis of Treasury Bill Yields", The Review of Economics and Statistics 74, 116-126.

Hall, P. (1997): "The Bootstrap and Edgeworth expansion", Springer Verlag.

Hamilton, J.D., (1988): "Rational-Expectations Econometric Analysis of Changes in Regimes - An Investigation of the Term Structure of Interest Rates", Journal of Economic Dynamics and Control, 12, 385-423.

Hamilton, J.D., Kim, D.H., (2002): "A re-examination of the predictability of the yield spread for real economic activity", Journal of Money, Credit, and Banking 34, 340-360.

Hansen, B., (2007): "Least Square Model Averaging", Econometrica, 75(4), 1175-1189.

Hansen, B., (2008): "Least Square Forecast Averaging", Journal of Econometrics, 146, 342-350.

Hansen, B., (2009): "Averaging Estimators for Autoregressions with a Near Unit Root", Journal of Econometrics, forthcoming.

Hansen, L.P., and Sargent, T.J., (2007): "Robustness", Princeton University Press.

Hansen, L.P., and Sargent, T.J., (2008): "Fragile Beliefs and the Price of Model Uncertainty", University of Chicago, working paper.

Harvey, C.R. (1989): "Forecasts of Economic Growth from the Bond and Stock Market", Financial Analysts Journal, 45, 38-45.

Harvey, C.R. (1993): "The Term Structure Forecasts Economic Growth", Financial Analysts Journal, 49, 6-8.

Jansson, M., Haldrup, N., (2002): "Regression Theory for Nearly Co-integrated Time Series", Econometric Theory, 18(6), 1309-1335.

Jardet, C., Monfort, A., and Pegoraro, F., (2009a): "Persistence, Bias, Prediction and Averaging Estimators", Banque de France Working Paper.

Jardet, C., Monfort, A., and Pegoraro, F., (2009b): "New Information Response Functions", Banque de France Working Paper.

Johansen, S., (1988): "Statistical Analysis of Cointegration Vectors", Journal of Economic Dynamics and Control, 12, 231-254.

Johansen, S., (1995): "Likelihood-Based Inference in Cointegrated Vector Autoregressive Models", Oxford University Press, Oxford.

Kemp, G.C.R., (1999): "The Behavior of Forecast Errors from a Nearly Integrated AR(1) Model as both Sample Size and Forecast Horizon become large", Econometric Theory, 15, 238-256.

Kendall, M.G. (1954): "A note on bias in the estimation of autocorrelation", Biometrika, 41, 403-404.

Kim, D.H., and Wright, J., (2005): "An Arbitrage-Free Three-Factor Term Structure Model and the Recent Behavior of Long-Term Yields and Distant-Horizon Forward Rates", Board of Governors

of the Federal Reserve System, Working Paper 2005-33.

Koop, G., Pesaran M.H., and Potter, S.M., (1996): "Impulse Response Analysis in Nonlinear Multivariate Models", Journal of Econometrics, 74(1), 119-147.

Kozicki, S., and Tinsley, P.A., (2001a): "Term Structure Views of Monetary Policy Under Alternative Models of Agent Expectations", Journal of Economic Dynamics and Control, 25, 149-184.

Kozicki, S., and Tinsley, P.A., (2001b): "Shifting Endpoints in the Term Structure of Interest Rates", Journal of Monetary Economics, 47, 613-652.

Kugler, P., (1990): "The Term Structure of Euro Interest Rates and Rational Expectations", Journal of International Money and Finance, 9, 234-244.

Lütkepohl, H., (2005): "New Introduction to Multiple Time Series Analysis", Springer-Verlag, Berlin.

MacDonald, R., Speight, A., (1991): "The Term Structure of Interest Rates Under Rational Expectations: Some International Evidence", Applied Financial Economics, 1, 211-221.

MacKinnon, J.G., (1996): "Numerical Distribution Functions for Unit Root and Cointegration Tests", Journal of Applied Econometrics, 11, 601-618.

MacKinnon, J.G., Haug, A.A., Michelis, L. (1999): "Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration", Journal of Applied Econometrics, 14, 563-577.

Magdalinos, T., Phillips, P.C.B. (2007): "Limit theory for moderate deviations from a unit root", Journal of Econometrics, 136, 115-130.

Marriott, F.H.C., Pope, J.A. (1954): "Bias in the estimation of autocorrelations", Biometrika, 41, 390-402.

McCulloch, J.H., Kwon, H-C., (1993): "U.S. Term Structure Data, 1947-1991", Working Paper 93-6, Ohio State University.

Monfort, A., Pegoraro, F., (2007): "Switching VARMA Term Structure Models", Journal of Financial Econometrics, 5(1), 105-153.

Newey, W.K., West, K.D., (1994): "Automatic Lag Selection in Covariance Matrix Estimation", Review of Economic Studies, 61, 631-653.

Ng, S., Perron, P., (2001): "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power", Econometrica, 69(6), 1519-1554.

Pesaran, M.H., Shin, Y., (1998): "Generalized Impulse Response Analysis in Linear Multivariate Models", Economic Letters, 58(1), 17-29.

Phillips, P.C.B., (1987): "Towards a Unified Asymptotic Theory for Autoregression", Biometrika, 74, 535-547.

Phillips, P.C.B., (1988): "Regression Theory for Near-Integrated Time Series", Econometrica, 56(5), 1021-1044.

Phillips, P.C.B., Magdalinos, T., (2007): "Limit Theory for Moderate Deviations from a Unit Root", Journal of Econometrics, 136, 115-130.

Phillips, P.C.B., Quintos, C., (1993): "Parameter Constancy in Cointegrating Regressions", Empirical Economics, 18, 675-708.

Phillips, P.C.B., Yu, J. (2009): "Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance", Handbook of Financial Time Series, 497-530.

Robinson, P. M., (1995): "Log Periodogram Regression of Time Series with Long Range Dependence", Annals of Statistics, 23, 1048-1072.

Rosenberg, J.V., Maurer, S., (2007): "Signal or Noise? Implications of the Term Premium for Recession Forecasting", Federal Reserve Bank of New York Economic Policy Review, forthcoming.

Rudebusch, G.D., Sack, B., Swanson, E., (2007): "Macroeconomic Implications of Changes in the Term Premium", Federal Reserve Bank of St. Louis Review, 89(4), 241-269.

Rudebusch, G.D., Swanson, E., (2008a): "Examining the Bond Premium Puzzle with a DSGE Model", Federal Reserve Bank of San Francisco, working paper.

Rudebusch, G.D., Swanson, E., (2008b): "The Bond Premium in a DSGE Model with Long-Run Real and Nominal Risks", Federal Reserve Bank of San Francisco, working paper.

Rudebusch, G.D., Swanson, E., Wu, T., (2006): "The Bond Yield Conundrum form a Macro-Finance Perspective", Monetary and Economic Studies, 24(S-1), 83-128.

Rudebusch, G.D., Williams, J.C. (2008): "Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve", Federal Reserve Bank of San Francisco, Working Paper 2007-16.

Rudebusch, G.D., Wu, T. (2007): "Accounting for a Shift in Term Structure Behavior with No-Arbitrage and Macro-Finance Models", Journal of Money, Credit and Banking, 39(2-3), 395-422.

Rudebusch, G.D., Wu, T. (2008): "A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy", Economic Journal, 118, 906-926.

Shaman, P., Stine, P. (1988): "The bias of autoregressive coefficient estimators", Journal of the American Statistical Association, 83, 403, 842-848.

Schwert, G.W., (1989): "Test for unit roots: A Monte Carlo investigation", Journal of Business and Economic Statistics, 7, 147-158.

Shea, G. S., (1992): "Benchmarking the Expectations Hypothesis of the Interest-Rate Term Structure: An Analysis of Cointegration Vectors", Journal of Business and Economic Statistics 10, 347-366.

Sims, C. A., Uhlig, H., (1991): "Understanding Unit Rooters: A Helicopter Tour", Econometrica, 59(6), 1591-1599.

Sowell, F., (1992): "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series", Journal of Econometrics, 53, 165-188.

Stock, J. H., (1996): "VAR, error-correction and pretest forecasts at long-horizons", Oxford Bulletin of Economics and Statistic, 58, 685-701.

Stock, J. H., Watson, M. W., (1989): "New Indexes of Coincident and Leading Indicators". In Blanchard, O., Fischer, S. (Eds.), 1989 NBER Macroeconomics Annual. MIT Press, Cambridge, MA.

Svensson, L.E.O., (1994): "Estimating and Interpreting Forward Rates: Sweden 1992-4", National Bureau of Economic Research Working Paper n. 4871.

Taylor, M.P., (1992): "Modelling the Yield Curve", Economic Journal, 102, 524-537.

Uhlig, H., (1994): "What Macroeconomists Should Know about Unit Roots: A Bayesian Perspective", Econometric Theory, 10(3/4), 645-671.