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

Caroline JARDET [(1)]
Banque de France

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

Fulvio PEGORARO[(3)]
Banque de France and CREST

First version : September, 2008
PRELIMINARY

Abstract

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

The present paper has five objectives. First, using dynamic regressions and Kullback causality measures, we build stylized facts about the dynamic links between the GDP growth and the spread between a long rate and a short rate, and between GDP growth and the short rate. The idea is to extend to a more realistic dynamic setting the usual results typically based on standard static regressions. Second, we carefully study the stationarity and the persistence properties of our variables of interest (the short rate, the spread between the long and the short rate, and the GDP growth), and we propose a modelling which focus on prediction performances in order to obtain, in particular, a reliable estimation of the term premia on the long-term bond. Third, this joint dynamics of the state variables is used to build a no-arbitrage macro-finance term structure model giving us the possibility to fit and to forecast the entire yield curve, and to extract term premia from any yield to maturity. Fourth, in order to try to conciliate the different viewpoints about the effects of the term premia on future GDP growth, we develop a dynamic analysis which studies and compares the respective roles of the expectation component and of the risk premium component of the spread. This analysis is based on a novel approach called "New Information Response Function". Fifth, in order to analyze more deeply the dynamic behavior of the term premia, we use a decomposition in terms of forward term premia at different horizons and a decomposition in terms of risk premia attached to the one-period holding of bonds of 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.

Keywords: 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, Economics and Finance Research Center [DGEI-DIR-RECFIN; E-mail: Caroline.JARDET@banque-france.fr].

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

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

We received helpful comments and suggestions from Sharon Kozicki, Monika Piazzesi, Mark Reesor, Eric T. Swanson and seminar participants at the September 2008 Bank of Canada Conference on Fixed Income Markets.

1 Introduction

One of the most important questions of the macroeconomics and finance literature on interest rates is the understanding of the dynamic relationships between economic activity, yields and term premia on long-term bonds, because of the related important implications for the conduct of monetary policy. The recent rise of federal funds rates (f.f.r.) of 425 basis points and the low and relative stable level of the long-term (10-years) interest rate, observed between June 2004 and June 2006 on the U.S. market, has further induced much interest in trying to detect the economic reasons behind this phenomenon [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] and, also, to well specify the links between financial and macro variables and risk premia.

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. Moreover, some of these 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) [APW(2006), hereafter], Rudebusch, Sack and Swanson (2007), and Rosenberg and Maurer (2007) find that the term premium has no predictive power for future GDP growth. Practitioner and private sector macroeconomic forecaster views agree on the decline of the term premium behind the conundrum but, in contrast, 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.

The present paper has five objectives. First, using dynamic regressions and Kullback causality measures, we build stylized facts about the dynamic links between the GDP growth and the spread between a long rate and a short rate, and between GDP growth and the short rate. The idea is to extend to a more realistic dynamic setting the usual results typically based on standard static regressions. Second, we carefully study the stationarity and the persistence properties of our variables of interest (the short rate, the spread between the long and the short rate, and the GDP growth), and we propose a modelling which focus on prediction performances in order to obtain, in particular, a reliable estimation of the term premia on the long-term bond. Third, this joint dynamics of the state variables is used to build a no-arbitrage macro-finance term structure model giving us the possibility to fit and to forecast the entire yield curve, and to extract term premia from any yield to maturity. Fourth, in order to analyze more deeply the dynamic behavior of the term premia, we use a decomposition in terms of forward term premia at different horizons and a decomposition in terms of risk premia attached to the one-period holding of bonds of different maturities. Fifth, in order to try to conciliate the different viewpoints about the effects of the term premia on future GDP growth, we develop a dynamic analysis which studies and compares the respective roles of the expectation component and of the risk premium component of the spread. This analysis is based on a novel approach called "New Information Response Function".

The specification of our model starts from the well known paper of APW (2006), in which the authors propose a no-arbitrage affine term structure model for joint GDP growth and yields dynamics. Their approach improves out-of-sample GDP forecasts of classical OLS regressions [see, among others, Harvey (1989, 1993), Stock and Watson (1989), Estrella and Hardouvelis (1991), Estrella and Mishkin (1998), Dotsey (1998), Hamilton and Kim (2002), Favero, Kaminska and Sodestrom (2005), Rudebusch and Williams (2008)] at all horizons and, contrary to previous results, they find that the short rate has more predictive power that any term spread. In their model, the yield curve factor (X_t, say) is given by the short rate (r_t), the spread between the long and the short rate (S_t), and the one-period real GDP growth (g_t), and its dynamics is described by a (unconstrained) Gaussian VAR(1) process.

We take a more general approach by leaving the data choose the number of lags, the number and the nature of possible cointegration relationships. In other words we check whether the joint dynamics is described by a Cointegrated VAR(p) model (CVAR(p)) in which the cointegrating relations is given by spreads [see Campbell and Shiller (1987), Kugler (1990), MacDonald and Speight (1991), Hall, Anderson and Granger (1992), Taylor (1992), and Shea (1992)].

Choosing to impose or not unit roots on interest rates joint dynamics [VAR(p) against CVAR(p) modelling] has important consequences. Indeed, it is well known that moving from a stationary environment to a unit root one, implies various types of discontinuity problems, in particular, in terms of asymptotic behavior of the estimation or testing procedure, or in terms of prediction. In the context of macro-finance modelling of interest rates, Cochrane and Piazzesi (2008) also noted very different long term predictions of yields, depending on whether unit roots are imposed or not. 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 polynomial matrix and that this behavior becomes very different as soon as at least one unit root is present.

This "discontinuity problem" is tackled in the literature by using different approaches, based on fractionally integrated processes, switching regimes, time-varying parameters, bayesian modelling or local-to-unity asymptotics. In the present work, we use the latter approach and we propose a no-arbitrage term structure model in which the dynamics of the factor $X_t = (r_t, S_t, g_t)'$ [the same as APW(2006)] is given by a Near-Cointegrated VAR(p) model [NCVAR(p)] which is able to improve out-of-sample forecasts and which is able to build a reliable measure of the term premia of interest rates. The specification of the NCVAR(p) factor dynamics that we propose is obtained as follows. First, on the basis of typical econometric procedures (lag order selection criteria, Johansen cointegration analysis) we select and estimate a VAR(p) and CVAR(p) model for the 3-dimensional vector $(r_t, S_t, g_t)'$. Then, we apply an average of the estimated VAR and CVAR parameters, in which the optimal weight $\lambda^* \in [0, 1]$ (say) is selected in order to minimize the prediction error of a variable of interest. The associated multivariate autoregressive model we obtain, called Near-Cointegrated VAR(p) model, will describe the dynamics of the factor driving term structure shapes in our yield-curve model. Since, in this paper, we are particularly interested in analyzing the dynamic relationship between the spread (between the long and short rate), its components (expectation part, and term premium) and the future economic activity, the optimal weight used to average the VAR(3) and CVAR(3) estimated parameters will be the one providing the best prediction of the expectation part of the spread. Our study is based on 174 quarterly observations of U.S. zero-coupon bond yields, for maturities from 1 to 40 quarters, and U.S. real GDP, covering the period from 1964:Q1 to 2007:Q2. In our model, the short rate (r_t) and the long rate (R_t) are, respectively, given by the 1-quarter and 40-quarter yields, and the one-quarter GDP growth at date t is denoted by g_t. These three variables constitute the information that investors use to price bonds. We select and estimate a VAR(3) and CVAR(3) model for the 3-dimensional vector X_t and, thus, the yield curve formula will be driven by a NCVAR(3) factor. In the estimated CVAR(3) model, we have one cointegrating relationship given by the spread, and an unrestricted constant.

Then, we study and evaluate the abilities of our model in several ways: a) out-of-sample predictions of yields, their expectation part, output growth, for various maturities and prediction horizons; b) ability to fit the yield curve (absolute pricing error); c) ability to match Campbell-Shiller regression coefficients. These performances are compared with those of other competing models, like the VAR(1) model of APW(2006), its generalization given by the VAR(3) model, and the CVAR(3) model which capture sources of non-stationarity. Two particularly important results are the following: i) in an interest rates forecast exercise, the NCVAR(3) term structure model reduces the out-of-sample root-mean-square forecast error (RMSFE), of the competing models, up to 45% (for long forecasting horizons); ii) in addition, our preferred model forecast the expectation part yields, for any time to maturity, better than the VAR(1), VAR(3) and CVAR(3) models. Consequently, our methodology seems to be a promising one to extract a reliable measure of term premia from long term bonds. The methodology of estimation of the term premia and of their different decompositions is used to analyze the recent period of the "conundrum" and to compare it with other periods showing an increase of the short rate.

Finally, we are interested in measuring the effects of a shock hitting a given factor, or a filtered transformation of the factors, on the output growth, the yield curve, the spread and its components. More precisely, our aim is to provide a dynamic analysis of the relationship between the spread and future 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). This approach allow us to measure the dynamic effects of a new information at date $t = 0$ (an unexpected increase of the spread or one of its component, for instance) on the variables of our model. Similar to the results found in the literature, we find that an increase of the spread implies a rise of 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 induce two effects on the 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 comes 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 introduces the data base used in the present work, and the stylized facts about the dynamic links between interest rates and GDP growth. Section 3 describes the Near-cointegration methodology, and presents some empirical performances of the

NCVAR(3) model in terms of out-of-sample forecast of short rate, long rate and GDP growth. Section 4 shows how the Near-cointegrated model can be completed by a no-arbitrage affine term structure model. In Section 5, we present decompositions of the term premia in terms of forward and risk term premia, and we show how these measures can be used to analyse more accurately the recent "conundrum" episode. Section 6 presents the impulse response analysis and, in particular, introduces the notion of New Information Response Function. Section 7 concludes, Appendix 1 gives further details about unit root analysis, Appendix 2 derives the yield-to-maturity formula and Appendix 3 gives details about impulses responses and definition of shocks. In Appendix 4 we gather additional tables and graphs.

2 Data and Stylized Facts

2.1 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 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, 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 (Trasury 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).

Yields	1-Q	4-Q	8-Q	12-Q	16-Q	20-Q	40-Q	log-GDP	1Q GDP growth
Mean	0.0152	0.0158	0.0164	0.0168	0.0170	0.0173	0.0180	8.7094	0.0080
Std. Dev.	0.0071	0.0070	0.0067	0.0065	0.0064	0.0063	0.0059	0.3854	0.0083
Skewness	1.0292	0.8455	0.8422	0.8626	0.8860	0.9087	0.9817	-0.0248	-0.0849
Kurtosis	4.6910	4.1357	4.0080	3.8800	3.7824	3.7196	3.6426	1.8298	4.4873
Minimum	0.0025	0.0026	0.0033	0.0043	0.0053	0.0062	0.0096	7.9897	-0.0204
Maximum	0.0398	0.0392	0.0395	0.0389	0.0384	0.0379	0.0372	9.3518	0.0387
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 [given by $\log(GDP_t)$] and one-quarter GDP growth rate [given by $\log(GDP_t/GDP_{t-1})$] 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.

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.

2.2 Dynamic Regressions, Bivariate Causality Measures and Impulse Response Functions

Since the work by Stock and Watson (1989) on leading economic indicators, many studies documented that the spreads between the ten or five years yields and three months interest rate are useful predictors of the real output growth [see e.g. Hamilton and Kim (2002), Ang, Piazzesi and Wei (2006), Rudebusch, Sack and Swanson (2006)]. Most of these studies consider static regressions of future (mean) GDP growth for the next k quarters on present spread and, possibly, present one-quarter GDP growth. In particular, for $k = 1$, the regressions are:

$$g_t = a_0 + b_1 S_{t-1} + \varepsilon_t \,,$$

$$\text{and} \qquad g_t = a_0 + a_1 g_{t-1} + b_1 S_{t-1} + \varepsilon_t \,.$$

In order to analyze more precisely the dynamic links between $g_t = \log(GPD_t/GDP_{t-1})$ and S_t (the spread between the 10-years yield and the short rate) it is important to introduce higher order lags in these regressions. These dynamic regressions can be viewed as an intermediate step between the static approach and the more comprehensive study presented in the following sections, which could provide useful stylized facts. More precisely, we have considered four lags in each variable and we have estimated sequentially the impact of these lags. Table 2 shows the estimations and the t-values of the a_i and b_j (for $i \in \{1, \ldots, 4\}$) coefficients in the regression:

$$g_t = a_0 + \textstyle\sum_{i=1}^4 a_i g_{t-i} + \sum_{j=1}^4 b_j S_{t-j} + \varepsilon_t \,, \qquad (1)$$

Number of lags	Panel A					Panel B		
(p,q)	(0,1)	(1,1)	(2,2)	(3,3)	(4,4)	(4,1)	(4,2)	(4,3)
a_1		0.25	0.17	0.20	0.20	0.21	0.17	0.21
		[3.4]	[2.2]	[2.6]	[2.6]	[2.7]	[2.3]	[2.6]
a_2			0.15	0.17	0.17	0.14	0.14	0.15
			[2.0]	[2.2]	[2.1]	[1.8]	[1.9]	[2.0]
a_3				-0.03	-0.04	-0.007	-0.05	-0.045
				[-0.4]	[-0.5]	[-0.08]	[-0.6]	[-0.6]
a_4					0.09	0.07	0.07	0.09
					[1.2]	[0.9]	[0.9]	[1.2]
b_1	0.42	0.376	-0.35	-0.29	-0.21	0.37	-0.35	-0.25
	[2.5]	[2.2]	[-1.4]	[-1.1]	[-0.8]	[2.3]	[-1.4]	[-0.9]
b_2			0.91	1.14	1.189		0.92	1.17
			[3.6]	[3.8]	[4.0]		[3.6]	[3.9]
b_3				-0.38	-0.34			-0.44
				[-1.4]	[-1.1]			[-1.6]
b_4					-0.19			
					[-0.7]			

Table 2: Parameter estimates of the dynamic regressions $g_t = a_o + \sum_{i=1}^{p} a_i g_{t-i} + \sum_{j=1}^{q} b_j S_{t-j} + \varepsilon_t$ (t-values are in brackets). In Panel A we first regress g_t on S_{t-1} only (first column), and then we regress the same variable on $(g_{t-1}, \dots, g_{t-p})$ and $(S_{t-1}, \dots, S_{t-q})$, with $p = q$ and $p \in \{1, \dots, 4\}$. In Panel B, we regress g_t on $(g_{t-1}, \dots, g_{t-p})$ and $(S_{t-1}, \dots, S_{t-q})$, with $p = 4$ and $q \in \{1, 2, 3\}$.

in which the lags are introduced progressively. When S_{t-1} only is introduced (first column in Panel A of Table 2), we find a significant coefficient equal to 0.42. This coefficient is smaller than the one found by APW (2006), namely, 0.65 from data between 1964:Q1 to 2001:Q4 (and with the spread based on the five years yield to maturity). This result confirms the decreasing impact of the spread in recent years. Interestingly, the introduction of S_{t-2} has a strong and very significant impact, which is robust to the introduction of additional lags both in S_t and g_t. This lagged effect is obviously missed in the static regressions mentioned above.

To go further in this dynamic analysis, it is worthwhile to introduce the notions of causality measures and of their decomposition (see Gourieroux Monfort (1997), chapter 10, and Gourieroux, Monfort and Renault (1987)). The global causality measure from S_t to g_t, based on a maximal number of lags equal to 4 (additional lags are not significant), is defined as Kullback discrepancy between the conditional models:

$$g_t = a_0 + \sum_{i=1}^{4} a_i g_{t-i} + \sum_{i=1}^{4} b_i S_{t-i} + \varepsilon_t\,, \quad \varepsilon_t \sim N(0, \sigma^2)$$

$$\text{and} \qquad g_t = \tilde{a}_0 + \sum_{i=1}^{4} \tilde{a}_i g_{t-i} + \tilde{\varepsilon}_t\,, \quad \tilde{\varepsilon}_t \sim N(0, \sigma_0^2)\,.$$

This measure is equal to $\frac{1}{2}\log(\sigma_0^2/\sigma^2)$, and can be consistently estimated by $\frac{1}{2}\log(\hat{\sigma}_0^2/\hat{\sigma}^2)$, where $\hat{\sigma}^2$ and $\hat{\sigma}_0^2$ are the standard estimators of σ^2 and σ_0^2. When we compare different causality measures for different variables or different lags, any multiplicative constant can obviously be introduced and, for statistical reasons, it is convenient to retain the measure $C = T\log(\hat{\sigma}_0^2/\hat{\sigma}^2)$, with T denoting the sample size, because this measure is the likelihood ratio statistic for the null hypothesis of no causality and is asymptotically distributed as $\chi^2(4)$ under the null. This global measure can be decomposed into:

$$C = \sum_{i=1}^{4} C_i$$

$$\text{with} \qquad C_i = T \log(\hat{\sigma}^2_{i-1}/\hat{\sigma}^2_i)\,,$$

where $\hat{\sigma}^2_i$ is the estimator of the variance of the error in the regression of g_t on $(g_{t-1}, g_{t-2}, g_{t-3}, g_{t-4}, S_{t-1}, \ldots, S_{t-i})$, and where $\hat{\sigma}^2_4 = \hat{\sigma}^2$.

Under the null hypothesis of no causality, the marginal causality measures C_i are independently distributed as $\chi^2(1)$. We can also define the cumulated causality measures $C^{(j)} = \sum_{i=1}^{j} C_i$, which is distributed as a $\chi^2(j)$ under the null. In Figure[4] A.1 we present these cumulative measures $C^{(j)}$, for $j \in \{1, \ldots, 4\}$, and the benchmark curves $\chi^2_{0.90}(j)$, $\chi^2_{0.95}(j)$ and $\chi^2_{0.99}(j)$, for $j \in \{1, \ldots, 4\}$.

From this figure we clearly see that the bulk of the causality intensity appears at lag 2 and that the cumulative causality remains strongly significant at higher lags. The global causality measure is equal to 21.6. Note that, we did not consider the instantaneous causality corresponding to $i = 0$, because S_t is measured at the end of the quarter and, therefore, such a causality cannot go from S_t to g_t.

A further dynamic analysis is based on the impulse response function deduced from the bivariate modelling:

$$\begin{cases} g_t &= a_0 + \sum_{i=1}^{4} a_i g_{t-i} + \sum_{j=1}^{4} b_j S_{t-j} + \varepsilon_t\,, \;\; \varepsilon_t \sim N(0, \sigma^2_\varepsilon) \\ S_t &= \bar{a}_0 + \sum_{i=0}^{4} \bar{a}_i g_{t-i} + \sum_{j=1}^{4} \bar{b}_j S_{t-j} + \eta_t\,, \;\; \eta_t \sim N(0, \sigma^2_\eta)\,. \end{cases} \tag{2}$$

Note that this is a recursive representation since g_t appears in the second equation, which implies that ε_t and η_t are independent and can be shocked independently. We choose this recursive form because it implies no instantaneous effect on g_t of a shock on S_t, in agreement with the fact already mentioned that S_t is measured at the end of the quarter.

Figure A.2 represents the responses of a positive shock on η_t. In particular, the solid line shows the propagation of this shock on g_t, in terms of proportions of the initial shock. We see that, after a small negative impact at horizon 1, the response jumps to a more than proportional positive impact at horizon 2. For instance, a shock on the spread (measured on a quarterly basis) equal to 40bp (annual basis) has an impact on the (one-quarter) GDP growth slightly larger than 0.1%, two quarters later. Figure A.3 gives the cumulated impact on the (one-quarter) GDP growth, that is to say the impact on the long run GDP growth, which converges to 3.5 after 5 years. This means, for instance, that a shock of 10^{-3} on the quarterly spread has a long run effect of 0.35% on the GDP growth.

The cumulated causality measures and the impulse response functions are also given when replacing the spread by the one-quarter short rate r_t [see figures A.4, A.5 and A.6]. The cumulated causality measure from r_t to the quarterly GDP growth is always above all the khi-square lines [see figure A.4]. However, the shape is different from the corresponding curve for the spread: the jump at horizon 2 is smaller, the curve is smoother and it reaches a higher level for the global causality, namely 26.9 (instead of 21.6 for the spread).

[4] In the following, prefix A. before the number of a figure or a table indicates that it is presented in Appendix 4.

The impulse response functions to a negative shock on the short rate are given in figures A.5 and A.6. We see a positive response on the GDP growth at horizon 2 which is more than proportional, and the long run response of the GDP growth is equal to +3.2.

All these results could be viewed as advanced stylized facts which must be confirmed and developed by a more precise multivariate analysis. In particular, it would be important to analyze the long run properties of the variables to separate the spread into an expectation part and a term premium part and to disentangle their specific effects on the growth of the GDP.

3 Near-Cointegration Analysis

The specification of the state variable dynamics is based on the following steps. First, in Section 3.1, 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 [Section 3.1.1]. 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 3.1.2].

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, Ang and Bekaert (2002), and the references therein]. Second, as indicated by Cochrane and Piazzesi (2008), interest rate forecasts over long horizons, coming from the alternative CVAR(3) and VAR(3) specifications, have very different behaviors [see Section 3.2.1] because of the discontinuity problem induced by the presence or not of a unit root in a VAR dynamics. As a consequence, important differences are found about the term premia extraction. In order to propose a state dynamics able to explain the observed serial dependence, and in order to propose a solution to the discontinuity problem, we assume X_t given by a Near-Cointegrated VAR(3) model, as defined in Section 3.2.2.

3.1 A Vector Error Correction Model of the State Variable

3.1.1 Unit Root Tests

The first step of our modelling studies 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 Elliot, 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.

3.1.2 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:

$$Y_t = \nu + \sum_{j=1}^{p} \Phi_j Y_{t-j} + \varepsilon_t \,, \tag{3}$$

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 :

$$\begin{aligned} &\Delta Y_t = \Pi Y_{t-1} + \sum_{j=1}^{2} \Gamma_j \Delta Y_{t-j} + \nu + \varepsilon_t \,, \\ &\text{with} \quad \Pi = -\left[I_{3\times 3} - \textstyle\sum_{j=1}^{3} \Phi_j \right] \text{ and } \Gamma_j = -\sum_{i=j+1}^{3} \Phi_i \,, \end{aligned} \tag{4}$$

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 (4), 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 and γ a (3×1) vector orthogonal to α), we have tested the null hypothesis $\mathrm{H}_0 : \nu = \alpha\mu$ (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 $\mathrm{H}_0 : \beta = (-1, 1, 0)'$ (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 (4), when $\Pi = \alpha\beta' = 1$, 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 (4) in terms of the 3-dimensional state process $X_t = (r_t, S_t, g_t)'$, with $S_t = R_t - r_t$ and $g_t = G_t - G_{t-1}$:

$$X_t = \tilde{\nu} + \sum_{j=1}^{3} \tilde{\Phi}_j X_{t-j} + \tilde{\eta}_t \,, \tag{5}$$

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

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

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

[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 (4). 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 (4) 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.

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

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

$$\text{where} \quad \tilde{\Phi} = \begin{bmatrix} \tilde{\Phi}_1 & \tilde{\Phi}_2 & \tilde{\Phi}_3 \\ I_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I_{3\times3} & 0_{3\times3} \end{bmatrix}, \quad \tilde{X}_t = (X_t', X_{t-1}', X_{t-2}')'\,, \tag{6}$$

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

3.2 Near-Cointegrated VAR(p) Dynamics

3.2.1 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 (2006)) or in term of prediction (see e.g. Stock (1996), Kemp (1999), Diebold and Kilian (2000), Elliot (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 1 and 2). 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 K-step ahead short rate forecasts obtained from the CVAR(3) and VAR(3) models (figures 1 and 2 respectively) for K = 1, 4, 8, 12, 16, 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.

3.2.2 Averaging estimations

The discontinuity problems can be tackled in different ways based on fractionally integrated processes, switching regimes, time-varying parameters, bayesian methods or local-to-unity asymptotics. For instance, Kozicki and Tinsley (2001a, 2001b, 2005), introducing shifting endpoint-based time-varying parameters in the short rate dynamics, significantly improve yield predictions. Here we adopt the local-to-unity approach. More precisely, we start from the results by Hansen (2007, 2008a, 2008b) which shows, among other results, that using a local-to-unity approach, the optimal weight averaging an unconstrained and a unit root (constrained) estimator, in terms of forecast error minimization, is strictly between 0 and 1. So, the idea is to consider an average of the VAR(3)

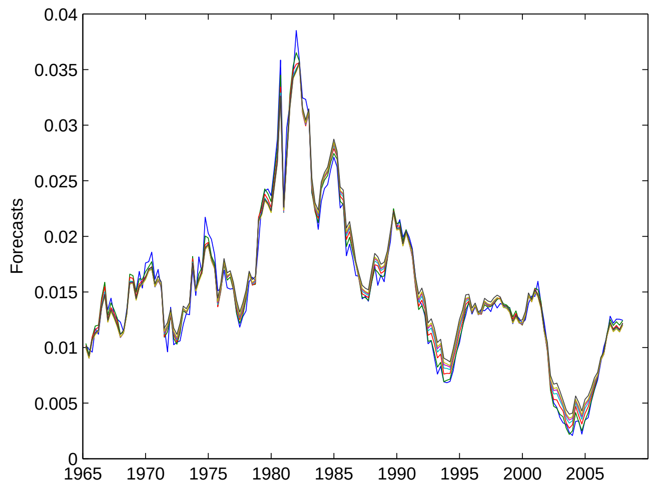


Figure 1:
K-step ahead short rate forecasts from the CVAR(3) model
K = 1, 4, 8, 12, 16, 20, 40 quarters.

and CVAR(3) parameters and to find the optimal weight in terms of prediction of a variable of interest.

The Near-Cointegrated VAR(3) model for the state vector X_t is obtained in the following way: once we have estimated 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 (Table A.7), the vector of parameters θ_{nc} specifying the Near-Cointegrated VAR(3) model is given by:

$$\theta_{nc} = \theta_{nc}(\lambda) = \lambda\theta_{var} + (1-\lambda)\theta_{cvar} \,, \tag{7}$$

with $\lambda \in [0,1]$ a free parameter selected to minimize a criterion of interest. In this paper we focus on minimizing the prediction error, measured by the root mean squared forecast error (RMSFE thereafter), of a variable of interest, and given our aim to provide a reliable measure of the term premia on long term bonds, we focus on the best estimation of the associated expectation part.

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 5 for a detailed presentation). For a given maturity h, the parameter $\lambda = \lambda(h)$ (say) is selected as the solution of the following problem:

$$\lambda^*(h) = arg \min_{\lambda(h)\in[0,1]} \sum_{t=1}^{T} [\tilde{B}_t^*(h) - \hat{B}_t^*(h)]^2 \tag{8}$$

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 $\hat{B}_t^*(h)$ 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 increasing size window for the estimation of models VAR(3) and CVAR(3). More

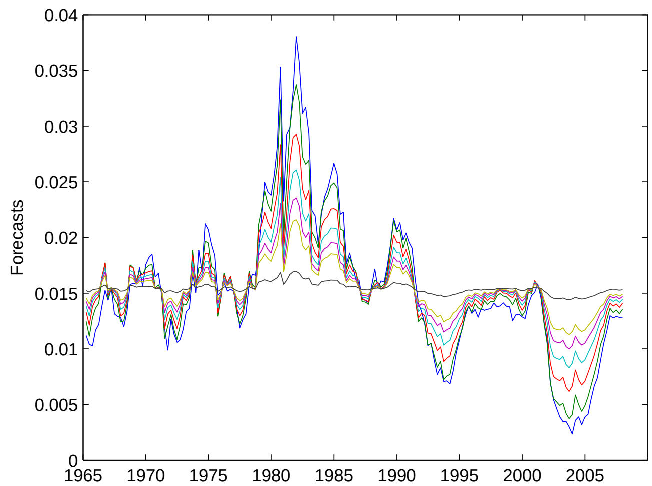


Figure 2:
K-step ahead short rate forecasts from the VAR(3) model
K = 1, 4, 8, 12, 16, 20, 40 quarters.

precisely, we first estimate the parameters θ_{var} and θ_{cvar} over the period 1964:Q1 to 1989:Q4 and we calculate $\hat{B}_t^*(h)$ 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 3 we compare, for h ranging from 2 to 40 quarters, the RMFSE obtained from the NCVAR(3) model, with $\lambda^*(h)$ solution of (8), 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. More precisely, 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 3]. 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, our model reduces their out-of-sample RMSFE of 20-30%.

Since, in this work, one of the main objectives is to extract the term premium from the 40-quarter long term bonds, we will assume that the NCVAR(3) state dynamics, driving term structure shapes over time and maturities, is specified by a $\lambda^*(40) = 0.2624$. Nevertheless, 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 3.2.3 and 4.3].

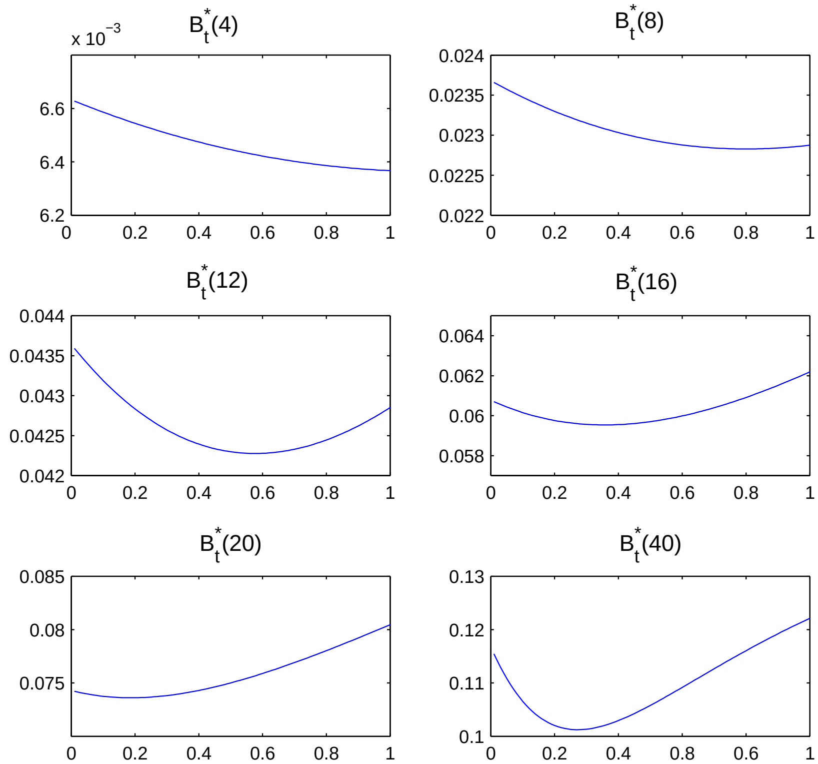


Figure 3:
X axis: λ; Y axis: RMSFE of $B_t^*(h)$ obtained from the NCVAR(3) model with vector of parameters $\theta_{nc} = \lambda\theta_{var} + (1-\lambda)\theta_{cvar}$. $\lambda = 0$ corresponds to the CVAR(3) case, while $\lambda = 1$ corresponds to the VAR(3) case.

3.2.3 Out-of-Sample Forecasts with Near-Cointegrated VAR(3) State Dynamics

In Section 3.2.2 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 averaging estimations 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 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, the 10-years long rate and the one-quarter GDP growth 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

	h	AR(1) (Vasicek)	$\lambda^*(h)$	NCVAR(3)	CVAR(3)	VAR(3)	VAR(1)
	2	0.00144198	1.0000	0.00149067	0.00154129	0.00149067	0.00157565
	4	0.00713912	1.0000	0.00636612	0.00662780	0.00636612	0.00754511
	8	0.02572808	0.7911	0.02282756	0.02365905	0.02287877	0.02609197
	12	0.04756626	0.5672	0.04227720	0.04359307	0.04287781	0.04691679
$\tilde{B}_t^*(h)$	16	0.06890693	0.3471	0.05953802	0.06070613	0.06226470	0.06674954
	20	0.08817411	0.1772	0.07361418	0.07421954	0.08058584	0.08475891
	28	0.11514307	0.1782	0.09627664	0.09791432	0.10841901	0.11136596
	32	0.12290695	0.2320	0.09910924	0.10372194	0.11350414	0.11963159
	36	0.13260412	0.2598	0.10338498	0.11207893	0.11972531	0.13043550
	40	0.14107400	0.2624	0.10123857	0.11547323	0.12239660	0.14055551

Table 3: Out-of-sample forecasts of $\tilde{B}_t^*(h) = \exp(-r_t - ... - r_{t+h-1})$. Table entries are 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.

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 4.3 we will concentrate on the forecast of yields with maturities between 3 and 40 quarters]. As in Section 3.2.2, the out-of-sample forecast exercise is performed using an increasing-size 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 4, 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 $\lambda^*(q) = 0$, for the short and long rate, while it remains equal to zero in the case of the GDP growth rate. This result indicates that, for interest rates, 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. With regard to GDP growth forecasts, we have a complete preference ($\lambda^*(q) = 0$ for each q) for the CVAR(3) component. Second, as far as the short and long rate forecasts are concerned, 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 25% when $q = 32$ quarters, and 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. Third, as far as the models' forecast of g_t are concerned, for $h = 16$ and $h = 20$, the CVAR(3) (and NCVAR(3)) slightly outperforms the AR(1) and VAR(1) model while, for shorter maturities, the AR(1) specification proposes the best performances.

b) If we consider the forecast of the state variables obtained by the NCVAR(3) process with

	q	AR(1)	$\lambda^*(q)$	NCVAR(3)	NCVAR(3) $[\lambda = 0.2624]$	CVAR(3)	VAR(3)	VAR(1)
r_t	1	0.00147585	1.0000	0.00152379	0.00155154	0.00157312	0.00152379	0.00161267
	4	0.00442546	0.7585	0.00388532	0.00392834	0.00398789	0.00389483	0.00449470
	8	0.00659594	0.5993	0.00607209	0.00612820	0.00625836	0.00613987	0.00647230
	12	0.00745883	0.4329	0.00694911	0.00697577	0.00713955	0.00716155	0.00716000
	16	0.00779739	0.2698	0.00708623	0.00708631	0.00721580	0.00755620	0.00745265
	20	0.00811347	0.0378	0.00671292	0.00684157	0.00671763	0.00782944	0.00782010
	28	0.00908244	0.0000	0.00749682	0.00778469	0.00749682	0.00884670	0.00902846
	32	0.00953839	0.0000	0.00723228	0.00783472	0.00723228	0.00924450	0.00956114
	36	0.01003454	0.0000	0.00650092	0.00778389	0.00650092	0.00967746	0.01012154
	40	0.01055810	0.0000	0.00582954	0.00789679	0.00582954	0.01015693	0.01070457
R_t	1	0.01129684	1.0000	0.00119449	0.00120936	0.00121914	0.00119449	0.00118238
	4	0.01013233	0.8894	0.00230502	0.00235331	0.00240402	0.00230642	0.00232552
	8	0.00900139	0.6469	0.00300762	0.00307379	0.00320254	0.00305507	0.00312799
	12	0.00823473	0.4707	0.00392395	0.00396178	0.00413364	0.00410495	0.00427545
	16	0.00770173	0.2201	0.00416466	0.00416755	0.00425506	0.00476777	0.00507070
	20	0.00695753	0.0000	0.00424712	0.00445743	0.00424712	0.00563220	0.00594186
	28	0.00607529	0.0000	0.00537884	0.00577048	0.00537884	0.00706849	0.00732260
	32	0.00570298	0.0000	0.00591024	0.00629019	0.00591024	0.00758718	0.00789753
	36	0.00548394	0.0000	0.00573424	0.00632595	0.00573424	0.00781734	0.00821376
	40	0.00551052	0.0000	0.00547662	0.00637219	0.00547662	0.00812138	0.00860646
g_t	1	0.00499098	0.0000	0.00549931	0.00560079	0.00549931	0.00597070	0.00573126
	4	0.00518494	0.0000	0.00631169	0.00643978	0.00631169	0.00680449	0.00555359
	8	0.00464024	0.0000	0.00521692	0.00531731	0.00521692	0.00547284	0.00486811
	12	0.00476664	0.0000	0.00491617	0.00496746	0.00491617	0.00498269	0.00491073
	16	0.00476048	0.0000	0.00472333	0.00475241	0.00472333	0.00474189	0.00483858
	20	0.00483377	0.0000	0.00473452	0.00475326	0.00473452	0.00477318	0.00483837

Table 4: Out-of-sample forecasts of state variables. Table entries are RMSFEs. r_t denotes the (one-quarter) short rate, R_t is the 10-years interest rate, and g_t is the one-quarter GDP growth. AR(1) denotes a Gaussian scalar autoregressive of order one process used to forecast, respectively, the r_t, R_t and g_t. The forecasting horizons (q) are measured in quarters.

$\lambda = 0.2624$, the results we obtain are the following. With regard to the short rate, even if the averaging parameter is selected using the expectation term 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 $q = 32$ quarters, our selected factor reduces the RMSFE of the AR(1) and VAR(1) models of almost 20%, and this percentage increases to 25% when $q = 40$ quarters. 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 $q = 32$, 36 and 40 quarters, they get slightly worse than the AR(1) forecasts, but still better than those obtained by the VAR(1) process. Finally, the forecasts of the one-quarter GDP growth remain almost the same as in case a).

4 Affine Term Structure Models

4.1 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:

$$M_{t,t+1} \quad = \quad \exp\left[-\mu_0 - \mu_1'\tilde{X}_t + \Gamma_t'\,\eta_{t+1} - \tfrac{1}{2}\Gamma_t'\Gamma_t\right] \,, \tag{9}$$

where $\tilde{\eta}_{t+1} = \Sigma\eta_{t+1}$.

$\Gamma_t = \gamma_0 + \gamma\tilde{X}_t = \gamma_0 + \gamma_1 X_t + \gamma_2 X_{t-1} + \gamma_3 X_{t-2}$ is the affine (multiple lags) stochastic risk sensitivity vector, γ_0 is a (3×1) vector and $\gamma = [\gamma_1 : \gamma_2 : \gamma_3]$ is a (3×9) matrix (in the case of CVAR(3) term structure model, the third column of γ_3 is a vector of zeros). γ_0, γ_1, γ_2, γ_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 $M_{t,t+1} = \exp\left[-r_t + \Gamma_t'\eta_{t+1} - \frac{1}{2}\Gamma_t'\Gamma_t\right]$. 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:

$$B_t(h) = \exp(c_h'\tilde{X}_t + d_h)\,, \tag{10}$$

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

$$\left\{\begin{array}{lcl} c_h & = & -\tilde{e}_1 + \tilde{\Phi}'c_{h-1} + (\Sigma\gamma)'c_{1,h-1} \\ & & \\ & = & -\tilde{e}_1 + \tilde{\Phi}^{*\prime}c_{h-1}\,, \\ & & \\ d_h & = & c_{1,h-1}'(\tilde{\nu} + \Sigma\gamma_o) + \frac{1}{2}c_{1,h-1}'\Sigma\Sigma'c_{1,h-1} + d_{h-1}\,, \end{array}\right. \tag{11}$$

and where :

$$\tilde{\Phi}^* = \left[\begin{array}{ccc} \tilde{\Phi}_1 + \Sigma\gamma_1 & \tilde{\Phi}_2 + \Sigma\gamma_2 & \tilde{\Phi}_3 + \Sigma\gamma_3 \\ I_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I_{3\times3} & 0_{3\times3} \end{array}\right] \text{ is a } (9\times9) \text{ matrix}\,,$$

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 :

$$\begin{aligned} R_t(h) &= -\frac{1}{h} \log B_t(h) \\ &= -\frac{c_h'}{h} \tilde{X}_t - \frac{d_h}{h}, \quad h \geq 1 \,. \end{aligned} \tag{12}$$

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}.

4.2 Risk Sensitivity Parameter Estimates

The estimation of historical (VAR, CVAR and NCVAR) and risk sensitivity paramaters follow a consistent two-step procedure, as adopted, among the others, by APW(2006), Monfort and Pegoraro(2007), and Garcia and Luger (2007). In Section 3 we have presented the (first step) least squares estimates of θ_{var}, θ_{cvar}, and θ_{ncvar}, thanks to the observations of the 1-quarter and 40-quarters yields, and of the real GDP. Given these parameter estimates, 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:

$$\begin{cases} \hat{\theta}_\gamma = Arg\min_{\theta_\gamma} S^2(\theta_\gamma) \\ S^2(\theta_\gamma) = \sum\limits_t \sum\limits_h (\tilde{R}_t(h) - R_t(h))^2, \\ \text{s. t. } \sum\limits_t (\tilde{R}_t(40) - R_t(40))^2 = 0 \,, \end{cases} \tag{13}$$

where, for each date t and maturity h, $R_t(h)$ is the theoretical yield determined by formula (12), and $\tilde{R}_t(h)$ indicate the observed one.

Risk sensitivity parameter estimates of the CVAR(3), VAR(3), NCVAR(3) and VAR(1) factor-based term structure models are presented, respectively, in Tables A.10 and A.11.

4.3 Empirical Comparisons

4.3.1 Pricing Errors

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 4.3.2), we will study the performances of our model to forecast out-of-sample these interest rates. Moreover, in Section 4.3.3, 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 5, 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 θ_γ, the pricing error in the following way:

$$PE_t = \frac{\sum_h |\tilde{R}_t(h) - R_t(h)|}{H}\,, \tag{14}$$

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 5: 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 (13), 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. The reason of this result is shown in the second column of Table 5, 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 4.2. We observe that the optimal value of the averaging parameter, for the criterion (13), is $\lambda^* = 0.2583$, which is very close to $\lambda^*(40) = 0.2624$.

4.3.2 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 followed in Section 3, 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 6, 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 4, 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 $q = 36$ and $q = 40$, $\lambda^*(h, q) = 0$ 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 6 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 ($q = 36$ and $q = 40$), 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 6, is that our model still 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

[9] We have also performed some forecast exercise 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.

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.

4.3.3 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 as far as h increases from three to forty quarters (see the second column of Table 7).

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:

$$\phi_h = \frac{Cov[R_{t+1}(h-1) - R_t(h), (R_t(h) - r_t)/(h-1)]}{Var[(R_t(h) - r_t)/(h-1)]}, \tag{15}$$

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 7, 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 $h = 8$).

	q	AR(1)	$\lambda^*(q)$	NCVAR(3)	NCVAR(3) $[\lambda = 0.2624]$	CVAR(3)	VAR(3)	VAR(1)
$R_t(4)$	1	0.00150457	0.9171	0.00173200	0.00179392	0.00187151	0.00226702	0.00212967
	4	0.00410245	0.7542	0.00388512	0.00396777	0.00406611	0.00467780	0.00459689
	8	0.00618828	0.6578	0.00595052	0.00605602	0.00621890	0.00705335	0.00644976
	12	0.00712904	0.5120	0.00687839	0.00694667	0.00714956	0.00835300	0.00725270
	16	0.00753698	0.3051	0.00695090	0.00695414	0.00709709	0.00874877	0.00764111
	20	0.00803245	0.0520	0.00664430	0.00677684	0.00661315	0.00918641	0.00830445
	28	0.00926663	0.0061	0.00760501	0.00790518	0.00755524	0.01011159	0.00974535
	32	0.00977851	0.0000	0.00737861	0.00800062	0.00732978	0.01035500	0.01030866
	36	0.01026247	0.0000	0.00656263	0.00789903	0.00652089	0.01061580	0.01083760
	40	0.01080059	0.0000	0.00577706	0.00796956	0.00574673	0.01094494	0.01138883
$R_t(8)$	1	0.00154723	0.6245	0.00175921	0.00181942	0.00196628	0.00258628	0.00221387
	4	0.00378185	0.6838	0.00357272	0.00368763	0.00381515	0.00462851	0.00420143
	8	0.00563986	0.6414	0.00543996	0.00558128	0.00575273	0.00685439	0.00584279
	12	0.00662399	0.5290	0.00641456	0.00651503	0.00671885	0.00818448	0.00667537
	16	0.00712138	0.3149	0.00651845	0.00652452	0.00664151	0.00865068	0.00710065
	20	0.00774318	0.0680	0.00631748	0.00645255	0.00622776	0.00925017	0.00788360
	28	0.00908665	0.0122	0.00734640	0.00767379	0.00721745	0.01021373	0.00945996
	32	0.00963692	0.0000	0.00730348	0.00788722	0.00717071	0.01048004	0.01010022
	36	0.01010066	0.0000	0.00656615	0.00779931	0.00644671	0.01066727	0.01062787
	40	0.01061729	0.0000	0.00574192	0.00780954	0.00564300	0.01092711	0.01115301
$R_t(12)$	1	0.00153930	0.5399	0.00175446	0.00182192	0.00203052	0.00287847	0.00230031
	4	0.00348641	0.6364	0.00331522	0.00345525	0.00359999	0.00464893	0.00392987
	8	0.00508890	0.6124	0.00496925	0.00513329	0.00529701	0.00669243	0.00536835
	12	0.00609559	0.5213	0.00596558	0.00608394	0.00626409	0.00802816	0.00621587
	16	0.00667039	0.3160	0.00613023	0.00613781	0.00620965	0.00858206	0.00667918
	20	0.00737411	0.0811	0.00602085	0.00615635	0.00586493	0.00928250	0.00752741
	28	0.00880667	0.0192	0.00709789	0.00744032	0.00687984	0.01024790	0.00918380
	32	0.00939450	0.0022	0.00722974	0.00775906	0.00699564	0.01052059	0.00987538
	36	0.00985087	0.0000	0.00661061	0.00770813	0.00638871	0.01064768	0.01039711
	40	0.01035066	0.0000	0.00583726	0.00769212	0.00563554	0.01083864	0.01089231
$R_t(20)$	1	0.00145368	0.4892	0.00179187	0.00188328	0.00218170	0.00345216	0.00267997
	4	0.00306038	0.5905	0.00300003	0.00318839	0.00336102	0.00484218	0.00378866
	8	0.00425306	0.5681	0.00425957	0.00445899	0.00460784	0.00654362	0.00482962
	12	0.00527566	0.4990	0.00525419	0.00539504	0.00552664	0.00784873	0.00564470
	16	0.00594500	0.3176	0.00552177	0.00553244	0.00552459	0.00852924	0.00612972
	20	0.00673015	0.1060	0.00556158	0.00568689	0.00528903	0.00934363	0.00701994
	28	0.00824621	0.0408	0.00673454	0.00706000	0.00635259	0.01027103	0.00874448
	32	0.00888492	0.0386	0.00710840	0.00751844	0.00668822	0.01052619	0.00948993
	36	0.00933858	0.0000	0.00670731	0.00752999	0.00626099	0.01055783	0.00999968
	40	0.00982848	0.0000	0.00608401	0.00750527	0.00563944	0.01063770	0.01045906

Table 6: 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.

h	sample $\phi_{h,T}$	NCVAR(3) ϕ_h	Confidence Bands	VAR(3) ϕ_h	Confidence Bands	VAR(1) ϕ_h	Confidence Bands
3	-0.494 [0.278]	-0.204	(-0.722, 0.880)	0.195	(-0.536, 1.058)	0.025	(-0.678, 1.182)
4	-0.745 [0.398]	-0.308	(-0.918, 0.760)	-0.0002	(-0.773, 0.933)	-0.052	(-0.778, 1.061)
8	-0.976 [0.684]	-0.581	(-0.904, 0.860)	-0.485	(-0.904, 0.889)	-0.377	(-0.572, 1.204)
12	-1.196 [0.822]	-0.877	(-1.460, 0.564)	-0.882	(-1.433, 0.609)	-0.706	(-1.158, 0.839)
20	-1.546 [0.934]	-1.457	(-2.364, 0.106)	-1.516	(-2.348, 0.057)	-1.351	(-2.236, 0.180)
40	-2.567 [1.186]	-2.743	(-4.330, -0.695)	-2.754	(-4.377, -0.880)	-2.698	(-4.442, -0.595)

Table 7: 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.

5 Unbiased Term Premia

5.1 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:

$$\widetilde{TP}_t(h) = R_t(h) - \frac{1}{h} E_t \sum_{j=0}^{h-1} r_{t+j}\,, \tag{16}$$

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

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

and a similar decomposition for the spread gives:

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

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:

$$EX_t(h) = -\frac{1}{h}\log B_t^*(h) \tag{17}$$

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:

$$TP_t(h) = R_t(h) - EX_t(h)\,. \tag{18}$$

Note that $EX_t(h)$ is easily computed using the recursive equations (11), with $\gamma_0 = 0$ and $\gamma = 0$.

5.2 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:

$$TP_t^f(h) = f_t(h) - EX_t^f(h)\,. \tag{19}$$

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

$$EX_t(h) = \frac{1}{h}\sum_{j=1}^{h} EX_t^f(j)\,, \tag{20}$$

we get:

$$TP_t(h) = \frac{1}{h}\sum_{j=1}^{h} TP_t^f(j)\,. \tag{21}$$

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

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

$$S_t(h) = \tfrac{1}{h}\sum_{j=1}^{h} EXS_t^f(j) + \frac{1}{h}\sum_{j=1}^{h} TP_t^f(j)\,, \tag{22}$$

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

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 (11) we have:

$$\begin{aligned} f_t(h) &= \log B_t(h-1) - \log B_t(h) \\ &= (c_{h-1}-c_h)'\tilde{X}_t + d_{h-1} - d_h \\ &= (\tilde{e}_1 + (I-\tilde{\Phi}')c_{h-1} - (\Sigma\gamma)'c_{1,h-1})'\tilde{X}_t - c'_{1,h-1}(\tilde{\gamma}+\Sigma\gamma_0) - \frac{1}{2}c'_{1,h-1}\Sigma\Sigma'c_{1,h-1} \\ &= r_t + c'_{h-1}(I-\tilde{\Phi})\tilde{X}_t - c'_{1,h-1}\tilde{\nu} - \frac{1}{2}c'_{1,h-1}\Sigma\Sigma'c_{1,h-1} - c'_{1,h-1}\Sigma\Gamma_t \end{aligned} \tag{23}$$

Similarly

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

where c^*_{h-1} is obtained from (11) with $\gamma=0$, and

$$\begin{aligned} TP_t^f(h) &= (c_{h-1}-c^*_{h-1})'(I-\tilde{\Phi})\tilde{X}_t - (c_{1,h-1}-c^*_{1,h-1})'\tilde{\nu} \\ &\quad -\frac{1}{2}c'_{1,h-1}\Sigma\Sigma'c_{1,h-1} + \frac{1}{2}c^{*'}_{1,h-1}\Sigma\Sigma'c^{*}_{1,h-1} - c'_{1,h-1}\Sigma\Gamma_t \end{aligned} \tag{24}$$

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

$$\begin{aligned} TP_t^f(h) &= (c_{h-1}-c^*_{h-1})'(I-\tilde{\Phi})\tilde{X}_t - (c_{1,h-1}-c^*_{1,h-1})'\tilde{\nu} \\ &\quad -\frac{1}{2}c'_{1,h-1}\tilde{\Omega}c_{1,h-1} + \frac{1}{2}c^{*'}_{1,h-1}\tilde{\Omega}c^{*}_{1,h-1} - c'_{1,h-1}\tilde{\Omega}\tilde{\Gamma}_t \end{aligned} \tag{25}$$

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 $c_{h-1} = c^*_{h-1}$, $TP^f_t(h) = -c'_{1,h-1}\Sigma\gamma_0$ $(= -c'_{1,h-1}\tilde{\Omega}\tilde{\gamma}_0)$, and

$$\begin{aligned} TP_t(h) &= \frac{1}{h}\sum_{j=1}^{h} TP^f_t(j) \\ &= -\frac{1}{h}\sum_{j=1}^{h} c'_{1,j-1}\Sigma\gamma_0 \end{aligned} \tag{26}$$

5.3 Yield term premia, risk premia and risk sensitivities

Formula (21) gives a decomposition of the yield term premium $TP_t(h)$ in terms of the forward term premia $TP^f_t(j)$. Anoher interesting decomposition of $TP_t(h)$ is based on the expected term premia attached to future one-period holdings of the zero-coupon bond with residual maturity h at time t. Indeed we have:

$$\begin{aligned} R_t(h) &= -\frac{1}{h}\sum_{j=1}^{h} \log \frac{B_{t+j-1}(h-j+1)}{B_{t+j}(h-j)} \\ &= \frac{1}{h}\sum_{j=1}^{h} \rho_{t+j}(h-j+1) \end{aligned} \tag{27}$$

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:

$$\begin{aligned} EX_t(h) &= -\frac{1}{h}\sum_{j=1}^{h} \log \frac{B^*_{t+j-1}(h-j+1)}{B^*_{t+j}(h-j)} \\ &= \frac{1}{h}\sum_{j=1}^{h} \rho^*_{t+j}(h-j+1)\,. \end{aligned} \tag{28}$$

So:

$$\begin{aligned} TP_t(h) &= R_t(h) - EX_t(h) \\ &= \frac{1}{h}\sum_{j=1}^{h} \left[\rho_{t+j}(h-j+1) - \rho^*_{t+j}(h-j+1)\right]\,. \end{aligned} \tag{29}$$

Note that in equations (27), (28) and (29) 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:

$$\begin{aligned} TP_t(h) &= \frac{1}{h}\sum_{j=1}^{h} E_t\left[\rho_{t+j}(h-j+1)-\rho^*_{t+j}(h-j+1)\right] \\ &= \frac{1}{h}\sum_{j=1}^{h} E_t E_{t+j-1}\left[\rho_{t+j}(h-j+1)-\rho^*_{t+j}(h-j+1)\right] \\ &= \frac{1}{h}\sum_{j=1}^{h} E_t RP_{t+j-1}(h-j+1)\,, \end{aligned} \tag{30}$$

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 (11) we have:

$$\begin{aligned} \rho_{t+j}(h-j+1) &= d_{h-j}+c'_{h-j}\tilde{X}_{t+j}-d_{h-j+1}-c'_{h-j+1}\tilde{X}_{t+j-1} \\ &= c'_{h-j}\tilde{X}_{t+j}-(-\tilde{e}_1-c'_{h-j}\tilde{\Phi}-c'_{1,h-j}\Sigma\gamma)\tilde{X}_{t+j-1} \\ &\quad -c'_{1,h-j}(\tilde{\nu}+\Sigma\gamma_0)-\frac{1}{2}c'_{1,h-j}\Sigma\Sigma' c_{1,h-j} \\ &= c'_{h-j}(\tilde{X}_{t+j}-\tilde{\Phi}\tilde{X}_{t+j-1}-e_1\tilde{\nu})+\tilde{e}'_1\tilde{X}_{t+j-1} \\ &\quad -c'_{1,h-j}\Sigma\gamma\tilde{X}_{t+j-1}-c'_{1,h-j}\Sigma\gamma_0-\frac{1}{2}c'_{1,h-j}\Sigma\Sigma' c_{1,h-j} \\ &= c'_{1,h-j}\tilde{\eta}_{t+j}+r_{t+j-1}-c'_{1,h-j}\Sigma\Gamma_{t+j-1}-\frac{1}{2}c'_{1,h-j}\Sigma\Sigma' c_{1,h-j} \end{aligned} \tag{31}$$

Similarly:

$$\rho^*_{t+j}(h-j+1)=c^{*'}_{1,h-j}\tilde{\eta}_{t+j}+r_{t+j-1}-\frac{1}{2}c^{*'}_{1,h-j}\Sigma\Sigma' c^{*'}_{1,h-j} \tag{32}$$

where the $c^*_{1,h-j}$ are obtained from (11) with $\gamma=0$. Finally

$$RP_{t+j-1}(h-j+1) = -c'_{1,h-j}\Sigma\Gamma_{t+j-1}-\frac{1}{2}c'_{1,h-j}\Sigma\Sigma' c_{1,h-j}+\frac{1}{2}c'^{*}_{1,h-j}\Sigma\Sigma' c^*_{1,h-j} \tag{33}$$

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

$$RP_{t+j-1}(h-j+1) = -c'_{1,h-j}\tilde{\Omega}\tilde{\Gamma}_{t+j-1}-\frac{1}{2}c'_{1,h-j}\tilde{\Omega} c_{1,h-j}+\frac{1}{2}c^{*'}_{1,h-j}\tilde{\Omega} c^*_{1,h-j} \tag{34}$$

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 (30) becomes:

$$\begin{aligned} TP_t(h) &= \frac{1}{h}\sum_{j=1}^{h} E_t RP_{t+j-1}(h-j+1) \\ &= \frac{1}{h}\sum_{j=1}^{h} -c'_{1,h-j}\Sigma E_t\Gamma_{t+j-1}-\frac{1}{2}c'_{1,h-j}\Sigma\Sigma' c_{1,h-j}+\frac{1}{2}c^{*'}_{1,h-j}\Sigma\Sigma' c^*_{1,h-j}\,, \end{aligned} \tag{35}$$

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 (26) we get $-c'_{1,i-1}\Sigma\gamma_0 = TP^f_t(i) = E_t RP_{t+h-i}(i)$, $\forall t,\ i,\ h > i$, 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 (35) is equal to zero since $RP_{t+h-1}(1) = 0$.

5.4 Computation of the term premia

Figure A.9 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.10 and A.11 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.10 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

[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,

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.10, 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.11 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.

5.5 An Application to the *Conundrum*

Figure 4 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 present in figure 5 the term premium decomposition in risk premia as it is described in equation (35). 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). 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 5 we see that the decreasing trend of the term premium observed during the rise of short term interest rate is mainly driven by expected risk premia over the two following years. During the 1994 tightening, the risk premia decreases but remains positive, except for one quarter. During the 1999 episode, negative expected risk premia are more frequent but do not exceed three quarters. In contrast, during the 2004 episode, period of negative expected risk premia lasts at least six quarters. The observed increasing periods of negative 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 of the 2004 episode and argue that

such as the Rudebusch and Wu (2008) measure (see Rudebusch, Sack and Swanson (2007) for further details).

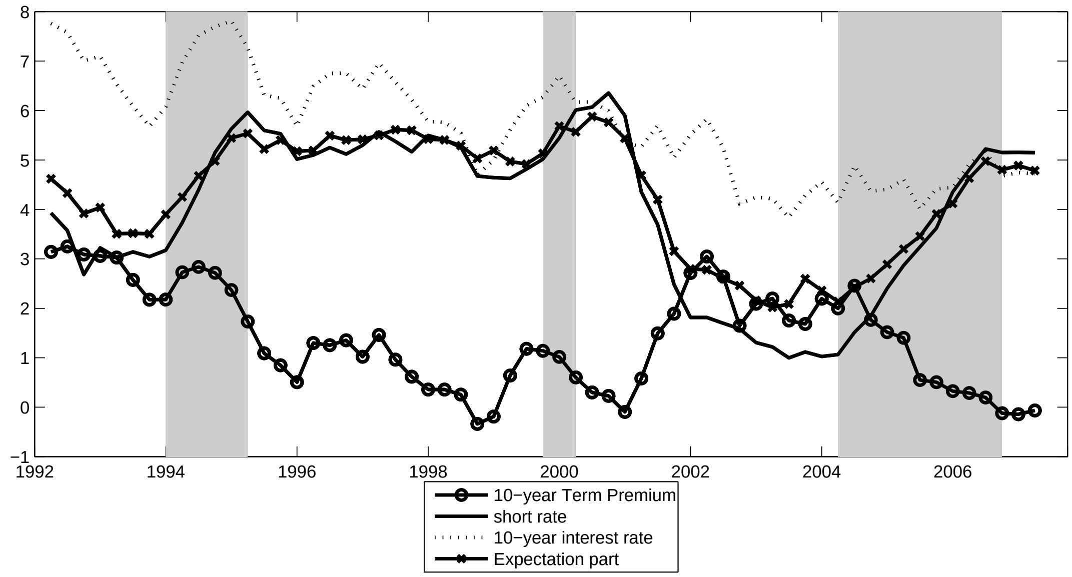


Figure 4:
Short rate, 10-year interest rates and its components over the three previous monetary policy tightening episodes
Shaded areas: monetary policy tightening episodes.

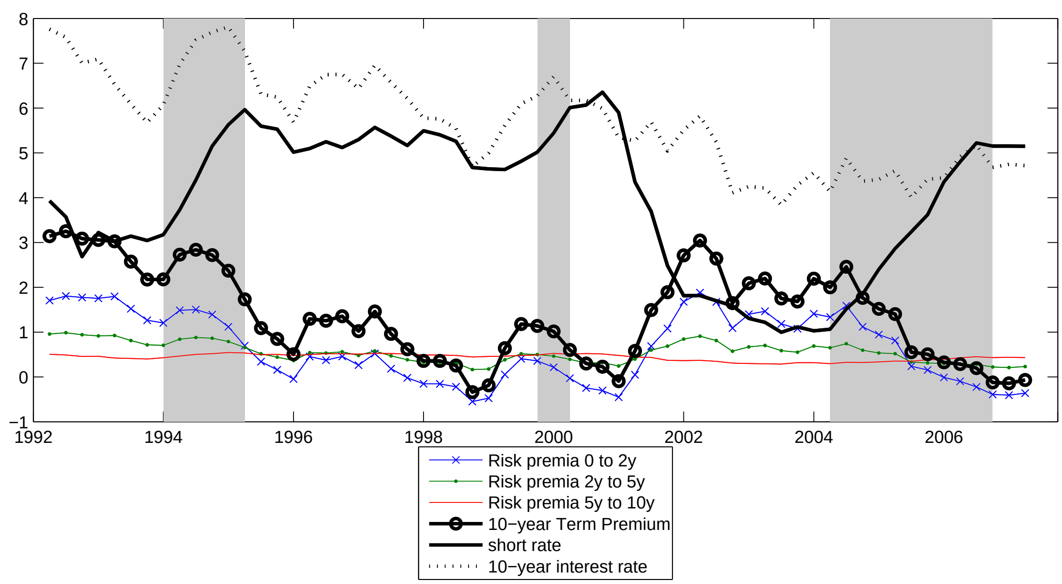


Figure 5:
Expected risk premia and 10-year term premium
Shaded areas: monetary policy tightening episodes.

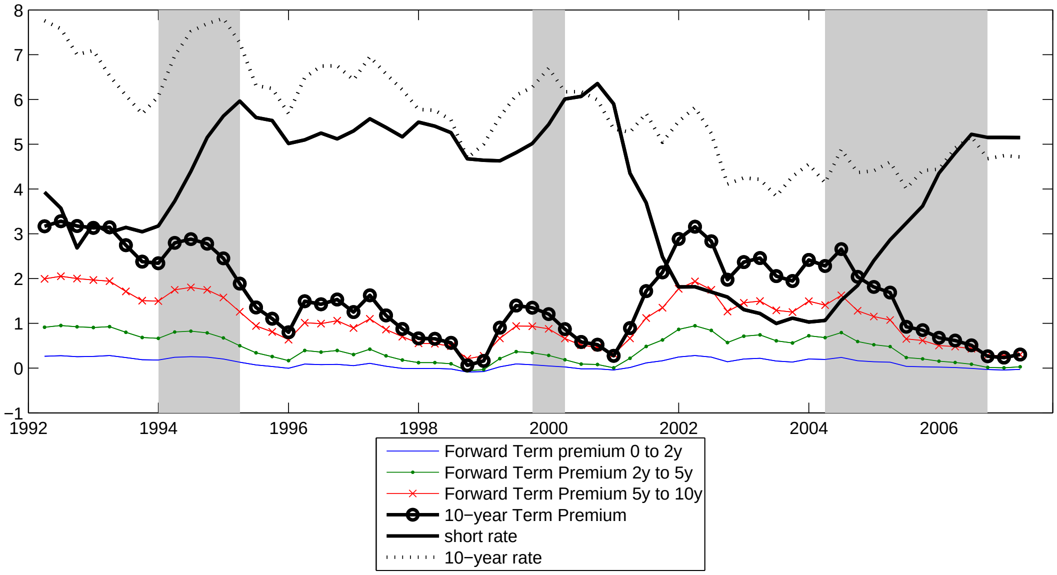


Figure 6:
Forward term premia and 10-year term premium
Shaded areas: monetary policy tightening episodes.

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 6 the decomposition of the 10-year term premium, in terms of forward term premium (see equation (26)) 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 6, 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.

6 Impulse 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 propose in this section a new approach based on a generalization of the Impulse Response Function, called *New Information Response Function.* The first two subsections present 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.

6.1 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:

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

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

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

and:

$$NIRF(t) = D_t \delta \tag{36}$$

with $\delta = E(\eta_0 | h(\eta_0) = h_0)$, and D_t is the t^{th} Markov matrix coefficient of the MA representation of y_t (see Appendix 3).

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 j^{th} 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 j^{th} equal to 1). Second, if $V(\eta_0)$ $=\Sigma$, it is usual to consider a shock of 1 on the j^{th} 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 j^{th} column of P ($P^{(j)}$ can also be normalized in order to have its j^{th} component equal to 1; see Appendix 3). Third, Pesaran and Shin (1998) also considered a "generalized" IRF, in which the new information is $\eta_{0j} = 1$ and therefore, in formula (36), $\delta = E(\eta_0 | \eta_{0j} = 1) = Cov(\eta_0, \eta_{0j}) / Var(\eta_{0j})$ (in the gaussian case).

But the New Information Response Function is useful in a much more general context, in particular when considering shocks on filtered variables.

6.2 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.

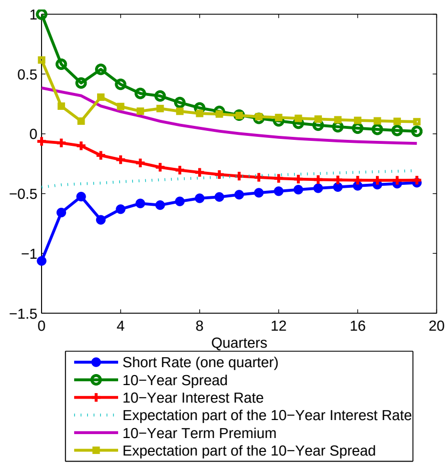
6.3 Impulse responses to a shock on the 10-year spread

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 $t = 0$.

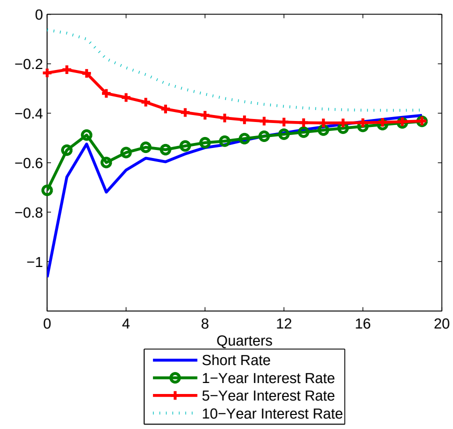
For that purpose and following previous notations, we need to determine the value of the (3×1) vector δ such that $\delta = E(\eta_0 | I_0)$, where η_0 is the innovation of the vector (r_t, S_t, g_t) and I_0 is the new information at date $t = 0$. Here, the new information I_0 includes, first of all, $\eta_{0,2} = 1$, 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 7 present the responses over 20 quarters of the real GDP, interest rates, term premia and expectation components of yield as defined by equation (17) and(18). Figure 7(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 7(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.

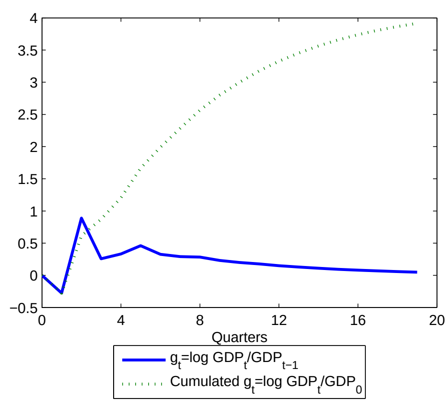
In figure 7(c) we observe that after a slight decrease that does not exceed 1 quarter, the real GDP tends to increase until to reach its new steady state level. After 20 quarters, the real GPD has increased by 4%, corresponding to an average annual growth rate equal to 0.8%. This result confirms the well documented results in the literature that emphasizes the positive relationship between the slope of the yield curve and future activity. Note that the response of real GDP to a shock on the spread are very close to the one obtained in the bivariate analysis of section 2 (see figures A.3 and 7(c)). In particular the long run effect on the real GDP growth displays the same amplitude.



(a) Spread and components



(b) Yield curve



(c) GDP growth

Figure 7: Responses to a shock on the spread

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:

$$S_t = TP_t(40) + EX_t(40) - r_t \tag{37}$$

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 (18) and (17) 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:

$$EXS_t(40) = EX_t(40) - r_t \tag{38}$$

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 (37) 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.

6.4 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 $y_t = (r_t, S_t, g_t)'$:

$$\begin{aligned} EXS_t(40) &= F_{1,1}(L)r_t + F_{1,2}(L)S_t + F_{1,3}(L)g_t && (39) \\ TP_t(40) &= F_{2,1}(L)r_t + F_{2,2}(L)S_t + F_{2,3}(L)g_t && (40) \end{aligned}$$

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:

$$\begin{aligned} \tilde{\eta}_{0,1} &= F_{1,1}(0)\eta_{0,1} + F_{1,2}(0)\eta_{0,2} + F_{1,3}(0)\eta_{0,3} && (41) \\ \tilde{\eta}_{0,2} &= F_{2,1}(0)\eta_{0,1} + F_{2,2}(0)\eta_{0,2} + F_{2,3}(0)\eta_{0,3} && (42) \end{aligned}$$

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]:

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

6.4.1 Shock on the expectation part

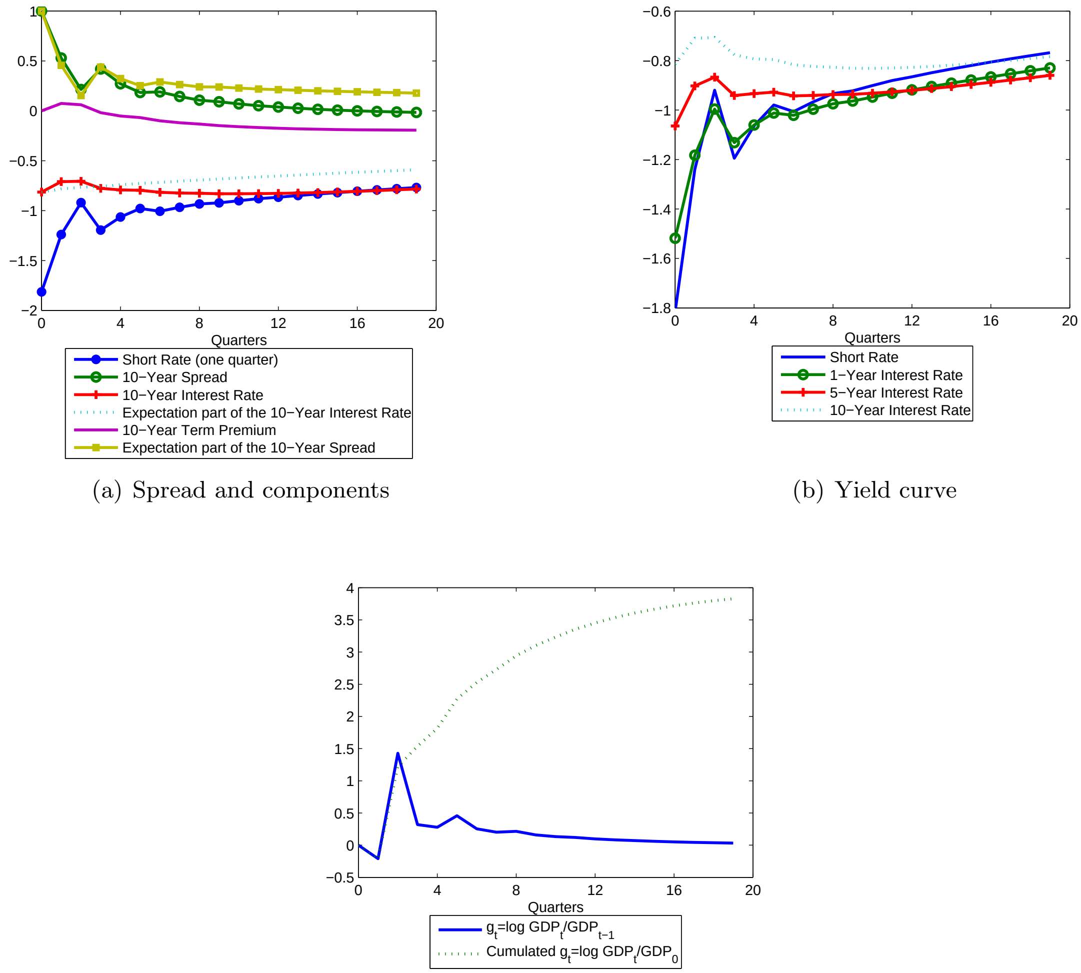
We are interested in the dynamic effects of 1 percentage point increase in the spread that would be completely due to a 1 percentage point increase in the expectation part of the spread. More precisely, the new information I_0 includes $\eta_{0,2} = 1$, $\tilde{\eta}_{0,1} = 1$ and $\tilde{\eta}_{0,2} = 0$. 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 (41) 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 8 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 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 8(a)]. More generally, figure 8(b) shows that this shock principally affects the short run of the yield curve (steepening of the yield curve). Figure 8(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.

[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 $F_{1,3}(0) + F_{2,3}(0) = 0$



(a) Spread and components

(b) Yield curve

(c) GDP growth

Figure 8: Responses to a shock on the Expectation part of the spread

6.4.2 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 (42) 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 9 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 (9(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 (9(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 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.

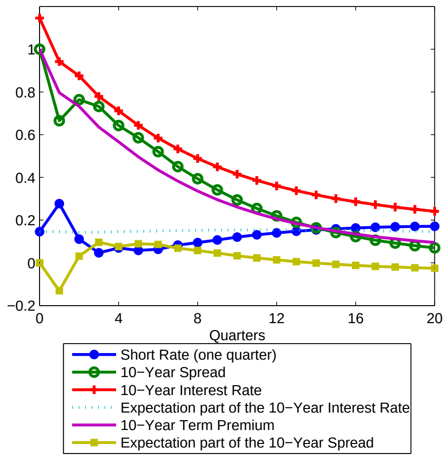
6.5 Shock on the short term interest rate

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:

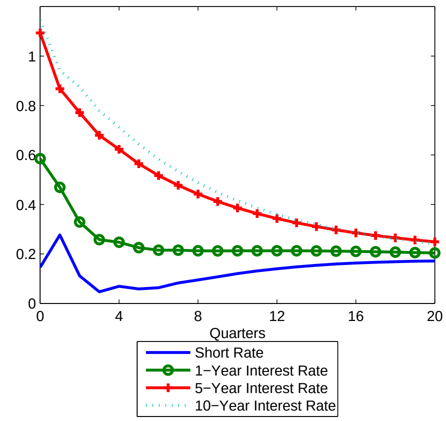
$$\delta = (-1, -\zeta, 0)'$$

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 10 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 be driven by its expectation part [see figure 10(a)]. This result confirms the intution according to which the EXS shock can be viewed as a monetary policy shock.

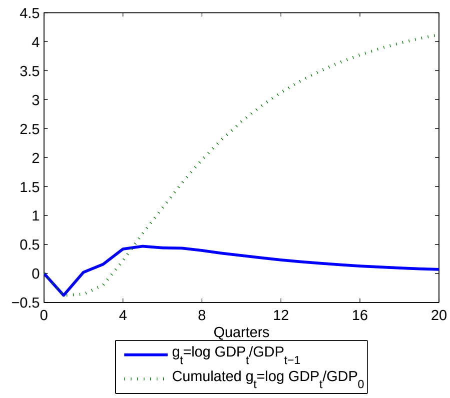
However, some slight difference can be noted. Looking at figure 10(a), we observe that the response at $t = 0$ of the term premium to a short rate shock is negative. In the case of a EXS shock, the response of the term premium becomes negative after three quarters (recall that we controlled



(a) Spread and components



(b) Yield curve

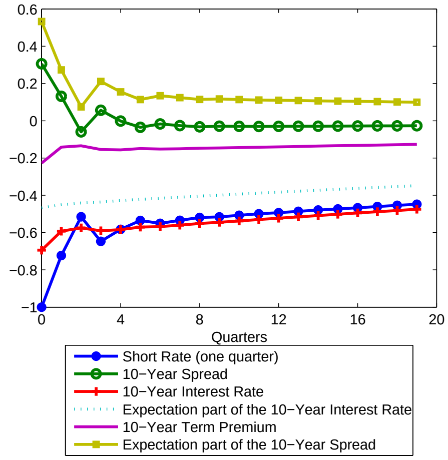


(c) GDP growth

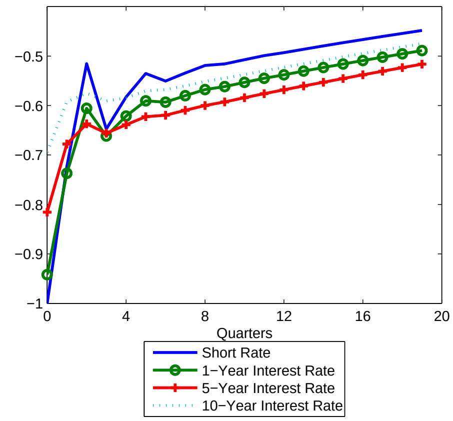
Figure 9: Responses to a shock on the Term Premium

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 8(b) and 10(b)).

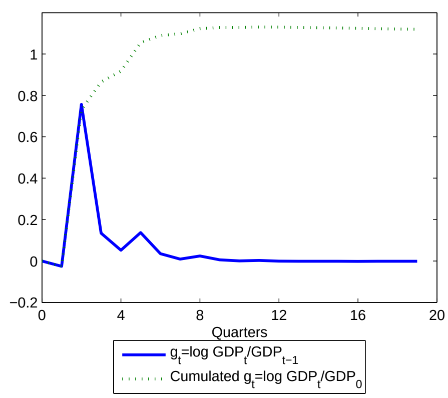
Looking at figure 10(c), we see that the real GDP tends to increase after a negative shock on the short rate, but the long run impact is much smaller than the one associated to an EXS shock



(a) Spread and components



(b) Yield curve



(c) GDP growth

Figure 10: Responses to a (negative) shock on the short term interest rate

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.

7 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 autocovariance 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 Elliot, 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 Elliot, 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 Elliot, 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 (10) is true for $h-1$, we get:

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

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

Appendix 3: Impulse Responses and Definition of Shocks

Let us consider a VARMA model defined by:

$$B(L)y_t = \nu + C(L)\eta_t\,, \tag{A.2}$$

where y_t is a vector of size n, ν is a vector of constant terms, η_t is a white noise with mean 0 and variance-covariance matrix Σ, $B(L)$ and $C(L)$ are matrices of lag polynomials of maximal degree p for B and q for C. We also assume that η_t is the innovation of y_t and, therefore, $B(0) = I_{n\times n}$, $C(0) = I_{n\times n}$. The process y_t may be non stationary, so $B(1)$ is not necessarily invertible.

Impulse Responses

We want to compute the impact on y_t, $t \geq 0$, of a shock δ on η_0. Let us introduce the following notations: $y_{\underline{-p}} = (y'_{-1}, y'_{-2}, \ldots, y'_{-p})'$, $\eta_{\underline{-q}} = (\eta'_{-1}, \eta'_{-2}, \ldots, \eta'_{-q})'$, $\underline{y_t} = (y'_t, y'_{t-1}, \ldots, y'_0)'$, and $\underline{\eta_t} = (\eta'_t, \eta'_{t-1}, \ldots, \eta'_0)'$.

Given that y_t is a linear function $\underline{\eta_t}$, $\eta_{\underline{-q}}$ and $y_{\underline{-p}}$ and, since we want to evaluate the differential impact on $\underline{y_t}$, with respect to a given path, of a shock on η_0, we can assume that $\eta_{\underline{-q}} = 0$ and $y_{\underline{-p}} = 0$. Therefore, we have the following moving average representation of y_t:

$$y_t = \mu_t + \sum_{i=0}^{t} D_i \eta_{t-i}\,, \tag{A.3}$$

where μ_t is deterministic. Replacing relation (A.3) in (A.2) we get, in particular

$$\left(\sum_{i=0}^{p} B_i L^i\right)\left(\sum_{i=0}^{t} D_i L^i\right)\eta_t = \left(\sum_{i=0}^{q} C_i L^i\right)\eta_t\,,$$

and, therefore,

$$\begin{aligned} & D_h = C_h - \sum_{i=1}^{h} B_i D_{h-i}\,,\ h \geq 1\,, \\ & \text{with}\ \ B_0 = I_{n\times n}, C_0 = I_{n\times n}, D_0 = I_{n\times n}, D_j = 0_{n\times n}\ \text{ if }\ j < 0 \\ & \text{and}\ \ B_j = 0_{n\times n}\ \text{ if }\ j > p\ \ C_j = 0_{n\times n}\ \text{ if }\ j > q. \end{aligned} \tag{A.4}$$

So, the matrices D_h's can be computed recursively from (A.4). The impact on y_t of a shock δ on η_0 is $D_t\delta$. In particular, the impact on y_0 is δ.

Unit Variance Orthogonalized Errors

If Σ is diagonal, the components of any η_t, in particular η_0, are uncorrelated and we can shock a component of η_0 independently of the others. If Σ is not diagonal, we can decompose it in $\Sigma = PP'$, where P is lower triangular. In other words, the white noise ξ_t defined by $\eta_t = P\xi_t$ has a variance-covariance matrix equal to $I_{n\times n}$. A shock of 1 on the j^{th} component of ξ_0 has an immediate impact on η_0, or on y_0, equal to $\delta_1 = P^{(j)}$, where $P^{(j)}$ is the j^{th} column of P. In particular, there is no impact on y_{0i}, for $i < j$, and the impact on y_{0j} is P_{jj}. The impact on y_t is $D_t P^{(j)}$. These impacts obviously depend on the ordering of the components of y_t. Moreover, the immediate impact on y_{0j} is not equal to 1.

Unit Impact Orthogonalized Errors

If we want to obtain an immediate impact on y_{0j}, or η_{0j}, equal to 1, we have to write η_t as:

$$\begin{aligned} & \eta_t = P\Delta^{-1}\Delta\xi_t\,, \\ \text{or} \qquad & \eta_t = \tilde{P}\zeta_t\,, \end{aligned} \tag{A.5}$$

where Δ is the diagonal matrix with diagonal terms P_{jj}, $\tilde{P} = P\Delta^{-1}$, $\zeta_t = \Delta\xi_t$, and $Var(\zeta_t) = \Delta\Delta' = \Delta^2$. Note that $\tilde{P}$ is a lower triangular matrix whose diagonal terms are equal to 1. The same is true for $\tilde{P}^{-1}$ and, therefore, $\zeta_t = \tilde{P}^{-1}\eta_t$ can be interpreted as the vector of the residuals of the regression of each component of η_t on the previous ones. A shock of 1 on the j^{th} component of ζ_0 has an immediate impact on y_0, or η_0, equal to $\delta_2 = \tilde{P}^{(j)}$; in particular, the impact on y_{0j} is 1. The impact on y_t is $D_t\tilde{P}^{(j)}$.

Moreover, observe that relation (A.2) can be rewritten in the following way:

$$\begin{aligned} \tilde{P}^{-1}B(L)y_t &= \tilde{P}^{-1}\nu + \tilde{P}^{-1}C(L)\tilde{P}\zeta_t\,, \\ \text{or} \ \ \tilde{B}(L)y_t &= \tilde{\nu} + \tilde{C}(L)\zeta_t\,, \\ \text{with} \ \ \tilde{B}(0) &= \tilde{P}^{-1}\,, \ \ \tilde{C}(0) = I_{n\times n}\,. \end{aligned} \tag{A.6}$$

In other words, relation (A.6) is a recursive version of (A.2). Note that, if $C(L) = c(L)I_{n\times n}$, we have $\tilde{C}(L) = c(L)$. The impact on the y_t's of a shock equal to 1 on ζ_{0j} can be computed in different ways. A first way is to compute $\delta_2 = \tilde{P}^{(j)}$, to compute the D_t's using (A.4) and get $D_t\tilde{P}^{(j)}$. A second way is to set $\underline{y_{-p}} = 0$, $\underline{\eta_{-q}} = 0$, $\eta_0 = \tilde{P}^{(j)}$, $\eta_t = 0 \ \forall \ \ t > 0$, and compute recursively the y_t's from (A.2). A third way is to set $\underline{y_{-p}} = 0$, $\zeta_0 = e_j$, $\zeta_t = 0$ for $t \neq 0$ and to compute the y_t's recursively from (A.6).

Non Orthogonalized Errors

A third approach has been proposed by Pesaran and Shin (2006). They suggested to define the impact on y_t, of a unitary shock on η_{0j} by:

$$E(y_t \,|\, \underline{y_{-p}}, \underline{\eta_{-q}}, \eta_{0j} = 1) - E(y_t \,|\, \underline{y_{-p}}, \underline{\eta_{-q}})\,. \tag{A.7}$$

The linearity of the model implies that this impact is simply $D_tE(\eta_0 \,|\, \eta_{0j} = 1)$. Under normality we know that:

$$\delta_3 = E(\eta_0 \,|\, \eta_{0j} = 1) = \tfrac{\Sigma^{(j)}}{\Sigma_{jj}}\,, \tag{A.8}$$

where $\Sigma^{(j)}$ is the j^{th} column of Σ. The impact on η_0 is δ_3, and in particular the impact on η_{0j} and y_{0j} is obviously 1 like in the unit impact orthogonalization. However, the impact on η_0 is in general different except for $j = 1$. Indeed, in this case, the impact on η_0 of a unitary shock on η_{0j}, in the unit impact case, is:

$$\tilde{P}^{(1)} = \tfrac{P^{(1)}}{P_{11}}\,,$$

and, using the identity $PP' = \Sigma$, we have $P^{(1)}P_{11} = \Sigma^{(1)}$, $P_{11}^2 = \Sigma_{11}$, and finally $\tilde{P}^{(1)} = \Sigma^{(1)}/\Sigma_{11}$.

New Information Response Function (NIRF)

Let us assume that we have the new information I_0 at $t = 0$. The differential response of y_t, induced by I_0, is:

$$E(y_t \,|\, \underline{y_{-p}}, \underline{\eta_{-q}}, I_0) - E(y_t \,|\, \underline{y_{-p}}, \underline{\eta_{-q}})\,. \tag{A.9}$$

The linearity implies that this differential response is:

$$E(y_t \,|\, \underline{y_{-p}} = 0, \underline{\eta_{-q}} = 0, \eta_t = 0 \;\; \forall t > 0, \eta_0 = E(\eta_0|I_0))\,, \tag{A.10}$$

which is easily obtained recursively putting $\underline{y_{-p}} = 0, \underline{\eta_{-q}} = 0, \eta_0 = E(\eta_0|I_0)$, and $\eta_t = 0 \;\; \forall t > 0$.

The examples of impulse responses presented above, are particular cases of this New Information Response Function (NIRF) framework. In the case of unit variance and unit impact orthogonalized errors, I_0 is a specific value of η_0, and $E(\eta_0|I_0)$ is equal to this value. In the Pesaran and Shin (2006) case, I_0 is made of a unit value for a component η_{0j} of η_0 and, therefore, $E(\eta_0|I_0) = Cov(\eta_{0i}, \eta_{0j})/Var(\eta_{0j})$, $i \in \{1, \ldots, n\}$.

We could also consider an information I_0 given by $\eta_{0j} = 1$ and $\eta_{0k} = 0$, with $k \neq j$. In this case we have $E(\eta_{0i}|I_0) = \beta_i$, $i \in \{1, \ldots, n\}$, where β_i is the coefficient of η_{0j} in the theoretical regression of η_{0i} on η_{0j} and η_{0k} (in particular, we have $\beta_j = 1$ and $\beta_k = 0$).

Response to a shock on a filtered variable

Let us suppose that $\widetilde{y}_t = F(L)y_t$, where $F(L) = (F_1(L), \ldots, F_n(L))$ is a row vector of polynomials in L. The innovation of $\widetilde{y}_t$ at $t = 0$ is $\tilde{\eta}_0 = F(0)\eta_0$.

An information I_0 based on $\tilde{\eta}_0$ and some component of η_0 is summarized in $E(\eta_0|I_0)$, which can take different values. For instance:

- if I_0 is given by $\tilde{\eta}_0 = 1$ and $\eta_{0i} = 0$, for $i \in \{1, \ldots, n-1\}$, we have $E(\eta_0|I_0) = (0, \ldots, 0, 1/F_n(0))'$;
- if I_0 is $\tilde{\eta}_0 = 1$, we have $E(\eta_0|I_0) = Cov(\eta_{0i}, \tilde{\eta}_0)/Var(\tilde{\eta}_0)$, for $i \in \{1, \ldots, n\}$, and, moreover, $\sum_{i=1}^{n} F_i(0)\frac{Cov(\eta_{0i}, \tilde{\eta}_0)}{Var(\tilde{\eta}_0)} = 1$;
- if I_0 is $\tilde{\eta}_0 = 1$ and $\eta_{0j} = 0$, we have $E(\eta_0|I_0) = \beta_i$, for $i \in \{1, \ldots, n\}$, where β_i is the coefficient of $\tilde{\eta}_0$ in the theoretical regression of η_{0i} on $\tilde{\eta}_0$ and η_{0j} (in particular, $\beta_j = 0$).

Appendix 4: 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	-1.8972	-3.4712	-2.8794	-2.5764	0	-1.8341	-3.4683	-2.8781	-2.5757
F-stat		1.8053	6.70	4.71	3.86		1.5972	6.70	4.71	3.86
PP										
Adj. t-stat		-2.6572	-3.4683	-2.8781	-2.5757		-1.8332	-3.4683	-2.8781	-2.5757
Ng-Perron										
MZ_{α}^{GLS} stat	8	-7.8890	-13.800	-8.1000	-5.7000	5	-2.2698	-13.800	-8.1000	-5.7000
MZ_{t}^{GLS} stat	8	-1.9861	-2.5800	-1.9800	-1.6200	5	-1.0649	-2.5800	-1.9800	-1.6200
MSB^{GLS} stat	8	0.2518	0.1740	0.2330	0.2750	5	0.4692	0.1740	0.2330	0.2750
MP_{T}^{GLS} stat	8	3.1056	1.7800	3.1700	4.4500	5	10.7914	1.7800	3.1700	4.4500
D-F GLS										
t-stat	8	-1.7829	-2.5791	-1.9428	-1.6154	5	-1.0227	-2.5788	-1.9427	-1.6154
Point-Opt										
P_T-stat	8	3.1911	1.9204	3.1544	4.2884	5	12.2549	1.9204	3.1544	4.2884

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	-0.9076	-3.4687	-2.8783	-2.5758	2	-4.3909	-4.0126	-3.4363	-3.1423
F-stat		13.2907	6.70	4.71	3.86		9.8924	8.73	6.49	5.47
PP										
Adj. t-stat		-1.1299	-3.4683	-2.8781	-2.5757		-3.7962	-4.0120	-3.4360	-3.1421
Ng-Perron										
$\mathrm{MZ}_{\alpha}^{GLS}$ stat	11	1.4637	-13.800	-8.1000	-5.7000	1	-9.8256	-23.800	-17.300	-14.200
MZ_{t}^{GLS} stat	11	2.4864	-2.5800	-1.9800	-1.6200	1	-2.1582	-3.4200	-2.9100	-2.6200
MSB^{GLS} stat	11	1.6987	0.1740	0.2330	0.2750	1	0.2196	0.1430	0.1680	0.1850
MP_{T}^{GLS} stat	11	207.147	1.7800	3.1700	4.4500	1	9.5412	4.0300	5.4800	6.6700
D-F GLS										
t-stat	11	1.5767	-2.5793	-1.9428	-1.6154	1	-2.1560	-3.4936	-2.9580	-2.6680
Point-Optimal										
P_T-stat	11	251.67	1.9204	3.1544	4.2884	1	11.3128	4.1046	5.6548	6.8418

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	-14.99847	-14.94313	-14.97601
1	1885.283	7.99e-16	-26.24971	-26.02836*	-26.15989
2	39.83684	6.96e-16	-26.38823	-26.00087	-26.23104*
3	22.42721*	6.72e-16*	-26.42252*	-25.86914	-26.19796
4	10.89691	6.98e-16	-26.38604	-25.66665	-26.09412

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	0.001331	0.606831	0.143068	0.035002	0.102311	-0.189622	0.049223	0.323890	-0.057000	-0.084323
	[0.22889]	[6.42110]	[0.76623]	[1.12708]	[0.93464]	[-0.80287]	[1.08776]	[3.16383]	[-0.32044]	[-2.96027]
R_t	0.001642	0.019803	0.810821	0.012036	0.011148	0.085518	-0.003013	0.079661	-0.046213	-0.009111
	[0.52354]	[0.38857]	[8.05297]	[0.71875]	[0.18886]	[0.67147]	[-0.12347]	[1.44302]	[-0.48177]	[-0.59314]
G_t	0.044041	0.228884	-0.311457	1.142524	-1.074207	0.442041	0.012913	0.232781	0.317451	-0.159731
	[2.85216]	[0.91207]	[-0.62819]	[13.8549]	[-3.69555]	[0.70484]	[0.10747]	[0.85632]	[0.67207]	[-2.11175]

$\Omega \times 10^3$			Correlations		log-L	$\vert\lambda\vert$
0.00786	0.00257	0.00482	ρ_{12}	0.6060	2289.428	0.996757
[8.97218]	[6.57583]	[2.85580]			AIC	0.931549
.	0.00229	0.00361	ρ_{13}	0.2310	-26.42606	0.864719
	[8.97218]	[3.87193]			SIC	0.616957
.	.	0.05548	ρ_{23}	0.3205	-25.87489	0.559512(^{c})
		[8.97218]			FPE	0.289147
					6.70e-16	0.083990(^{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. $\vert\lambda\vert$ indicates the modulus of the roots of equation $\vert\tilde{\Phi}(\lambda)\vert = 0$, with $\tilde{\Phi}(\lambda) = (I_{3\times 3}\lambda^3 - \Phi_1\lambda^2 - \Phi_2\lambda - \Phi_3)$ denoting the characteristic polynomial; (^{c}) indicates a pair of complex conjugate roots.

r	Eigenvalue	Trace Statistic	5 % Critical Value	p-value	Max-Eigen Statistic	5 % Critical Value	p-value
0	0.141284	34.42603	29.79707	0.0136	26.04625	21.13162	0.0094
1	0.039789	8.379788	15.49471	0.4257	6.942923	14.26460	0.4959
2	0.008368	1.436865	3.841466	0.2306	1.436865	3.841466	0.2306
α	0.041894	0.117752	-0.577805	β	1.000000	-1.002883	0.003063
	[0.55511]	[2.88295]	[-2.87536]			$\langle 0.00004 \rangle$	$\langle 0.08771 \rangle$

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 : rank$(\Pi) = r$. In the Trace test, the alternative hypothesis is H_A : rank$(\Pi) = 3$, and the associated statistic is given by 2(log-L_A - log-L_0) $= -T \sum_{i=r+1}^{3} \log(1-\lambda_i)$, where log-$\mathrm{L}_A$ and log-L_0 denote, respectively, the maximum value of the log-Likelihood function (of model (4)) under the case of 3 and $r < 3$ cointegrating relations. In the Maximum Eigenvalue test, H_A : rank$(\Pi) = r+1$, and 2(log-L_A - log-L_0) $= -T \log(1-\lambda_{r+1})$. Both test statistics accept at 5 % the hypothesis 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 adjustement 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 $\mathrm{H}_0 : \beta = (1, 0, \beta_3)'$ and $\mathrm{H}_0 : \beta = (1, \beta_2, 0)'$. 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	-0.000979	-0.415423	0.194112	0.039291	-0.315243	0.011025	0.089297	-0.034924	-0.002868
	[-2.70748]	[-3.97793]	[1.10849]	[1.29848]	[-3.15953]	[0.06260]	[3.15866]	[-0.50737]	
ΔR_t	-0.000193	-0.086245	-0.067537	0.014772	-0.076095	0.021575	0.012037	-0.113247	
	[-0.99102]	[-1.53059]	[-0.71479]	[0.90479]	[-1.41348]	[0.22705]	[0.78913]	[-3.04918]	
ΔG_t	0.004761	0.685382	-0.755436	0.196407	-0.370046	-0.303811	0.191169	0.408580	
	[4.89545]	[2.44064]	[-1.60429]	[2.41381]	[-1.37923]	[-0.64152]	[2.51471]	[2.20741]	

$\Omega \times 10^3$			log-L	$\mid\lambda\mid$
0.00793	0.00263	0.00509	2283.600	1.00000**
[9.02773]	[6.68263]	[2.96130]	AIC	0.847825
.	0.00231	0.00376	-26.39299	0.627061
	[9.02773]	[3.96284]	SIC	0.548880(^{c})
.	.	0.05735	-25.89693	0.246769
		[9.02773]	FPE	0.104975(^{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 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. $\mid\lambda\mid$ indicates the modulus of the roots of equation $|\tilde{\Phi}(\lambda)| = 0$, with $\tilde{\Phi}(\lambda) = (I_{3\times3}\lambda^3 - \Phi_1\lambda^2 - \Phi_2\lambda - \Phi_3)$ 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	-0.000879	0.778688	0.159188	0.039291	-0.082906	-0.183087	0.089297	0.304218	-0.011025	0.000000
S_t	0.001010	0.067530	0.660029	-0.024519	0.182168	0.272199	-0.077260	-0.249698	-0.010550	0.000000
g_t	0.003588	-0.070055	-0.346856	0.196407	-0.603802	0.451625	0.191169	0.673856	0.303811	0.000000

$\tilde{\Omega} \times 10^3$			Correlations	
0.00793	-0.00530	0.00509	ρ_{12}	-0.8435
.	0.00498	-0.00133	ρ_{13}	0.2385
.	.	0.05735	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	0.000866	0.930676	0.069426	0.002164	0.00884	-0.00610	0.00490
	[1.06788]	[23.8278]	[0.95226]	[0.07846]	[9.13783]	[-8.44682]	[2.63677]
S_t	0.000083	0.029614	0.822297	-0.004342	.	0.00564	-0.00142
	[0.12750]	[0.94903]	[14.1173]	[-0.19706]		[9.13783]	[-0.97331]
g_t	0.007867	-0.162556	0.189790	0.241397	.	.	0.06259
	[3.64410]	[-1.56400]	[0.97825]	[3.28905]			[9.13783]

Correlations		log-L	$\vert\lambda\vert$
ρ_{12}	-0.8637	2258.762	0.947180
		AIC	0.803692
ρ_{13}	0.2084	-26.27792	0.243498
		SIC	
ρ_{23}	-0.0755	-26.05745	
		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. $|\lambda|$ indicates the modulus of the roots of equation $|\tilde{\Phi}(\lambda)| = 0$, with $\tilde{\Phi}(\lambda) = (I_{3\times 3}\lambda - \Phi)$ denoting the characteristic polynomial.

	ν	Φ_1			Φ_2			Φ_3		
r_t	0.000168	0.758766	0.177748	0.029460	-0.073947	-0.196031	0.074309	0.248621	-0.078296	0.042825
	[0.20432]	[4.96926]	[0.95300]	[0.97364]	[-0.37275]	[-0.83610]	[2.57596]	[1.64085]	[-0.44379]	[1.56340]
S_t	0.000619	0.073745	0.641578	-0.018244	0.174550	0.279790	-0.067578	-0.219235	0.024977	-0.031195
	[0.94150]	[0.60429]	[4.30389]	[-0.75443]	[1.10089]	[1.49310]	[-2.93113]	[-1.81038]	[0.17713]	[-1.42490]
g_t	0.006153	-0.131586	-0.421217	0.199406	-0.626298	0.448599	0.196940	0.638584	0.247280	-0.053245
	[2.74610]	[-0.31655]	[-0.82954]	[2.42072]	[-1.15963]	[0.70280]	[2.50771]	[1.54808]	[0.51484]	[-0.71399]

$\Omega \times 10^3$			Correlations		log-L	$\lvert\lambda\rvert$
0.00775	-0.00521	0.00502	ρ_{12}	-0.8412	2287.931	0.929918
[8.97218]	[-8.16798]	[2.93811]			AIC	0.888804
.	0.00495	-0.00133	ρ_{13}	0.2380	-26.40855	0.601542(^{c})
	[8.97218]	[-0.99649]			SIC	0.491884
.	.	0.05743	ρ_{23}	-0.0788	-25.85738	0.428283(^{c})
		[8.97218]			FPE	0.244932
					6.82e-16	0.159802

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. $\lvert\lambda\rvert$ indicates the modulus of the roots of equation $|\tilde{\Phi}(\lambda)| = 0$, with $\tilde{\Phi}(\lambda) = (I_{3\times3}\lambda^3 - \Phi_1\lambda^2 - \Phi_2\lambda - \Phi_3)$ denoting the characteristic polynomial; (^{c}) indicates a pair of complex conjugate roots.

	γ_o	γ_1			γ_2			γ_3		
r_t	-0.136652	100.691518	102.269258	-14.896053	19.314151	8.437605	-20.437526	-100.212879	-15.560201	0
	[-2.844382]	[6.450301]	[5.323759]	[-4.821233]	[0.799806]	[0.304243]	[-7.621634]	[-7.337268]	[-0.978338]	
S_t	0.092360	47.175696	93.102439	-0.955629	-97.518067	-79.733476	5.603760	32.175722	30.834653	0
	[1.450239]	[1.743514]	[2.742617]	[-0.165901]	[-2.225377]	[-1.573247]	[1.232679]	[1.467221]	[1.187849]	
g_t	1.540296	47.994907	99.149410	-41.930044	-232.010236	77.676090	62.374147	91.745662	-349.377840	0
	[1.648148]	[0.126182]	[0.203772]	[-0.587535]	[-0.390555]	[0.110982]	[1.057419]	[0.304262]	[-0.968137]	

	γ_o	γ_1			γ_2			γ_3		
r_t	-0.524919	108.228826	98.625876	-11.656333	19.073429	15.906897	-18.089892	-83.382967	4.560793	-11.029529
	[-10.559728]	[7.003605]	[5.143045]	[-3.828487]	[0.792034]	[0.576711]	[-6.042908]	[-6.106854]	[0.285957]	[-4.487555]
S_t	-0.161976	54.490247	97.185702	-0.680821	-97.839133	-80.544700	7.868587	38.467362	41.642690	-1.423593
	[-2.252678]	[2.001090]	[2.830724]	[-0.111710]	[-2.210141]	[-1.584640]	[1.401269]	[1.746867]	[1.565135]	[-0.325354]
g_t	1.714034	43.853199	140.022996	-61.068570	-250.281939	108.641615	4.151276	109.366549	-416.391956	61.739192
	[1.374114]	[0.094733]	[0.242222]	[-0.645551]	[-0.340414]	[0.129669]	[0.045088]	[0.300037]	[-0.955248]	[0.912193]

Table A. 10: Risk sensitivity parameter estimates for the Cointegrated VAR(3) (top panel) and the unconstrained VAR(3) (bottom panel) factor-based term structure models [Gurkaynak-Sack-Wright (2007) data base; sample period : 1964:Q1 - 2007:Q2]. t-values are in brackets.

	γ_o	γ_1			γ_2			γ_3		
r_t	-0.175976	105.591876	105.983133	-17.554761	30.179033	24.126697	-23.555232	-111.759561	-10.126020	-0.941943
	[-3.499081]	[7.290341]	[5.772738]	[-5.775648]	[1.302222]	[0.892926]	[-7.760015]	[-8.712175]	[-1.308643]	[-0.385633]
S_t	-0.012362	45.465498	88.585328	2.057232	-106.482043	-91.174578	9.348216	47.025998	40.573734	-1.870025
	[-0.131771]	[1.645902]	[2.524100]	[0.358244]	[-2.422960]	[-1.770751]	[1.663735]	[2.090681]	[1.521398]	[-0.426417]
g_t	2.613361	-112.999796	49.931272	-39.328490	-120.312695	-29.173913	-38.816691	104.875009	-267.898853	64.850553
	[1.024150]	[-0.151638]	[0.053316]	[-0.327007]	[-0.107655]	[-0.022382]	[-0.328538]	[0.174723]	[-0.375226]	[0.637249]

	γ_o	γ_1		
r_t	-0.478376	34.789443	56.968201	-4.557311
	[-6.629479]	[11.523107]	[9.066444]	[-1.776612]
S_t	-0.301354	2.383774	66.540381	5.373331
	[-4.304486]	[1.114791]	[13.898731]	[1.771460]
g_t	1.587789	0.648518	4.255942	-5.401042
	[0.539493]	[0.006127]	[0.018443]	[-0.056872]

Table A. 11: Risk sensitivity parameter estimates for the NCVAR(3) factor-based term structure model with $\lambda = 0.2624$ (top panel), and the (unconstrained) VAR(1) (bottom panel) factor-based term structure model [Gurkaynak-Sack-Wright (2007) data base; sample period : 1964:Q1 - 2007:Q2]. t-values are in brackets.

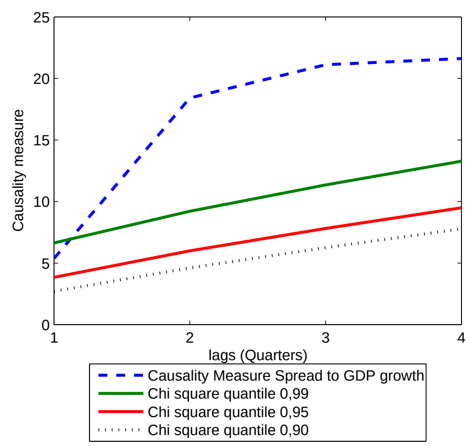


Figure A. 1:
Causality Measures from Spread to GDP Growth (Dashed: Chi Square Lines)

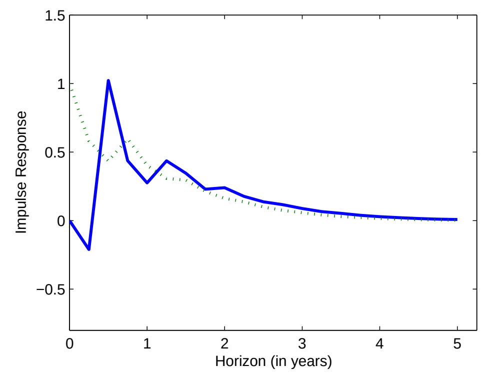


Figure A. 2:
Impulse Responses Functions to a Shock on the Spread
Solide line: GDP growth
Dashed line: Spread

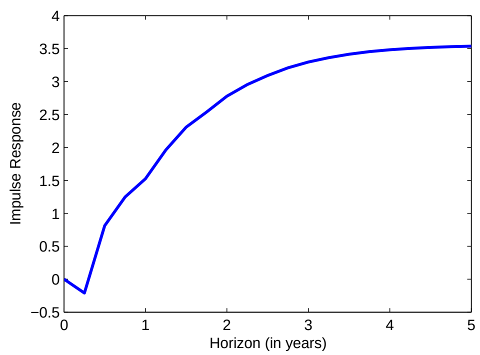


Figure A. 3:
Impulse Response Functions to a Shock on the Spread
Solide line: cumulated GDP growth

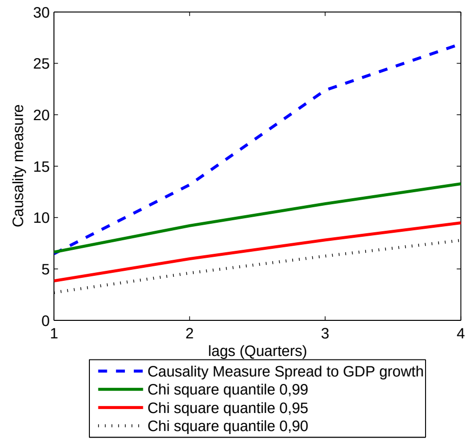


Figure A. 4:
Causality Measures from Short Rate to GDP growth (Dashed: Chi Square Lines)

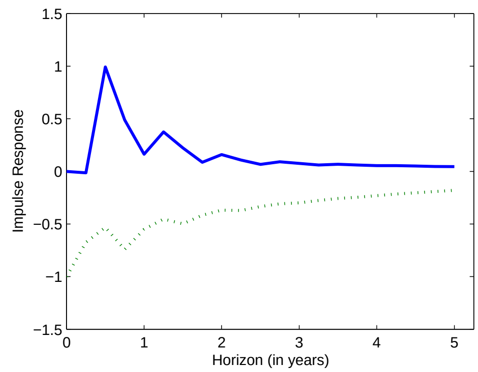


Figure A. 5:
Impulse Response Functions to a (negative) Shock on the Short Rate
Solid Line: GDP growth
Dashed line: Short rate

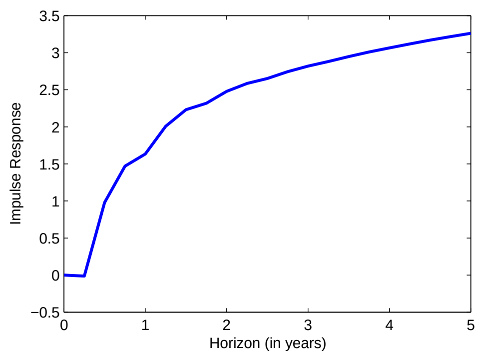


Figure A. 6:
Impulse Responses Functions to a (negative) Shock on the Short Rate
Solid Line: Cumulated GDP growth

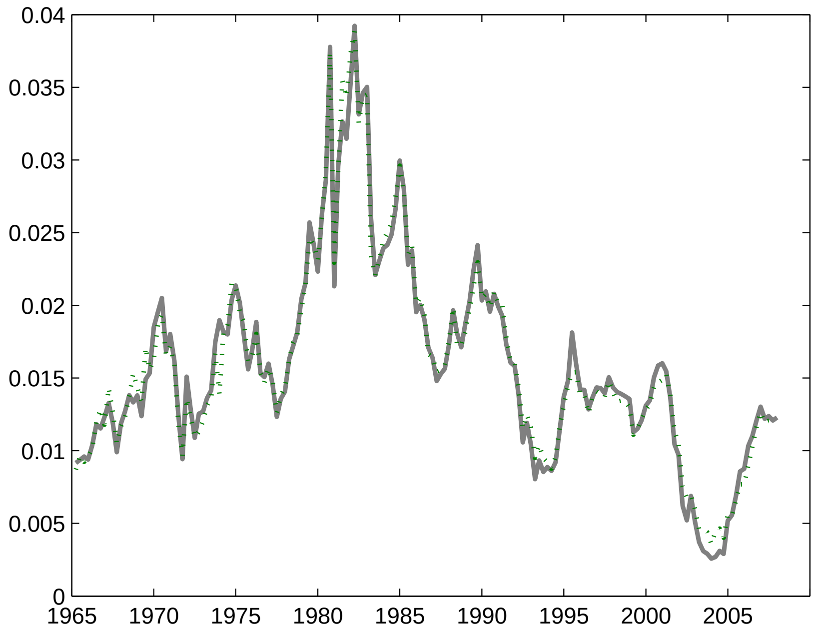


Figure A. 7:
1-Year interest rate, fitted (dashed) and observed (solid)

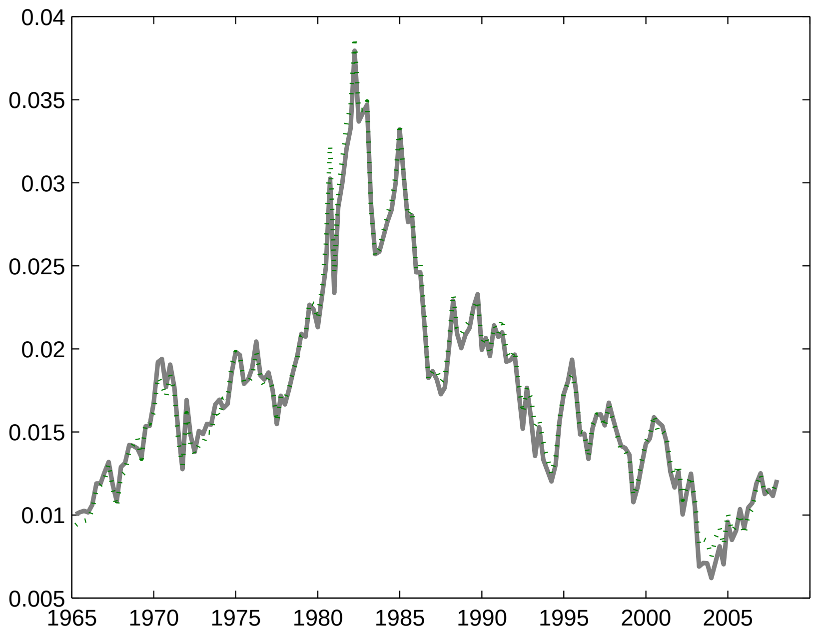


Figure A. 8:
5-Year interest rate, fitted (dashed) and observed (solid)

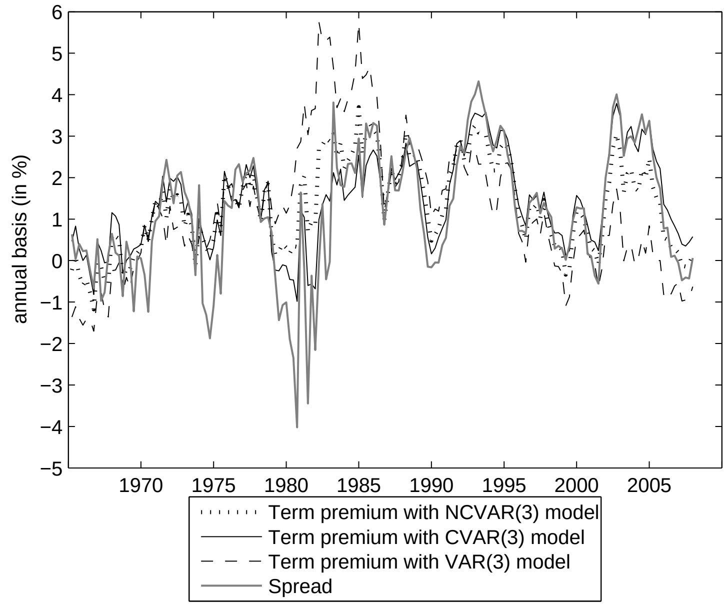


Figure A. 9: Term premia

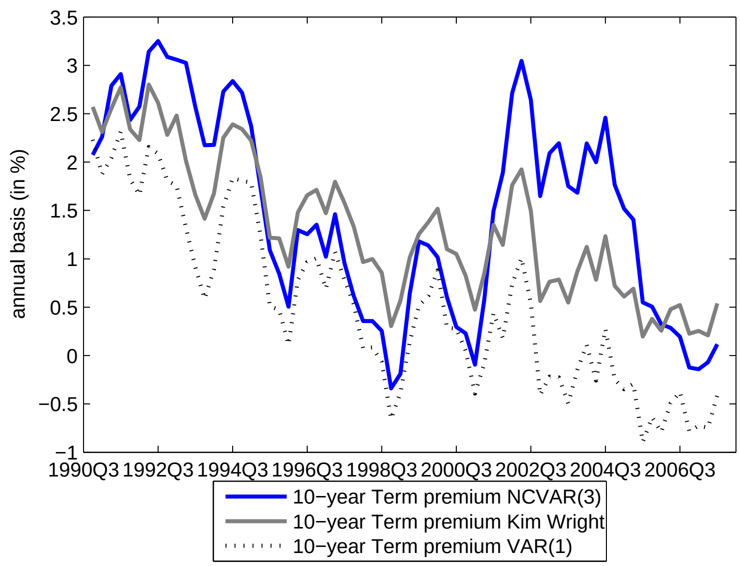


Figure A. 10:
NCVAR(3), Kim Wright and Ang, Wei, Piazzesi (VAR(1))
10-year Term premia

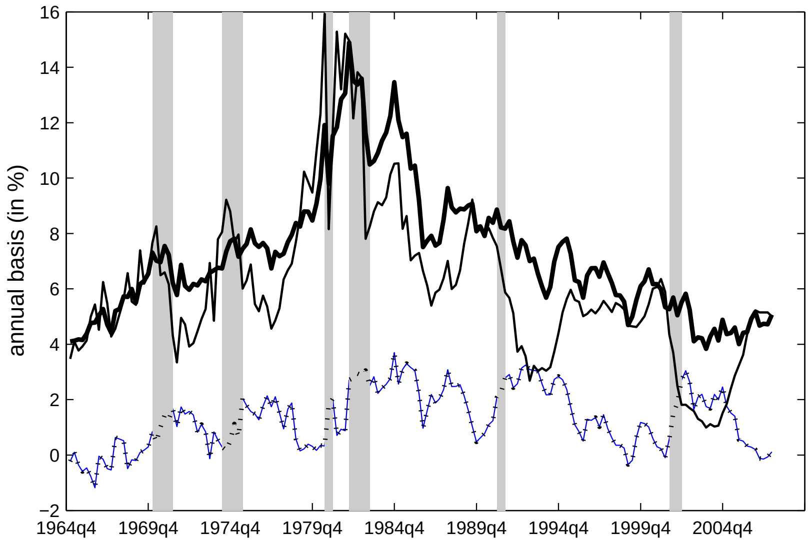


Figure A. 11: NCVAR(3) term premium
Bold line: 10-year interest rate
Thin line: short term interest rate
Shaded areas: recession dates (NBER)

R E F E R E N C E S

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. (2007): "Term Structure of Real Rates and Expected Inflation", The Journal of Finance, forthcoming.

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.

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., 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.

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

Elliot, 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, 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.

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

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.

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.

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., (2008a): "Least Square Forecast Averaging", Journal of Econometrics, forthcoming.

Hansen, B., (2008b): "Averaging Estimators for Autoregressions with a Near Unit Root", Journal of Econometrics, forthcoming.

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.

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.

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.

Kozicki, S., and Tinsley, P.A., (2005): "What Do You Expect? Imperfect Policy Credibility and Tests of the Expectation Hypothesis", Journal of Monetary Economics, 52, 421-447.

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.

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., (2006): "Limit Theory for Moderate Deviations from a Unit Root", , Journal of Econometrics, forthcoming.

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.

Schwert, G.W., (1989): ”Test for unit roots: A Monte Carlo inverstigation”, 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.

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.