Online Appendix for

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

March, 2012.

I Introduction

The purpose of this online Appendix is to report additional details and results about state dynamics specification, in-sample fit performances and new information response functions associated to the class of affine term structure models that we have introduced in the paper "No-arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth". In Section II we present further results about the NCVAR modelling, about the reliability of the selected averaging parameter, and we present historical and risk-neutral parameter estimates of the competing VAR and CVAR models. Section III presents the interest rates forecast performances provided by the NCVAR(3) term structure model, as well as the above mentioned competing models, in Section IV we show the ability of our NCVAR(3) yield curve model to explain the violation of the Expectation Hypothesis, while in Section V we study the responses of economic activity to a shock on the h-period spread, and its components, exploiting the concept of normalized shock for *NIRFs*.

II Further Details about Near-Cointegration Analysis

II.1 Data

In the data base provided by Gurkaynak, Sack and Wright (2007) [GSW (2007), hereafter], the authors 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 Rudebusch, Sack and Swanson (2007, footnote 26), for the 10-years yield-to-maturity] by the fact that (even if there were few bond observations with these maturities), the reconstructed time series are highly correlated with other well known and widely used time series [like, for instance, the FRED interest rates data base (Treasury Constant Maturity interest rates), or the McCulloch and Kwon (1993) data base]. Moreover, in order to be coherent with the literature and, in particular, with the majority of the papers concerned with the predictive ability of the term spread for GDP [see, for instance, Fama and Bliss (1987), and Ang, Piazzesi and Wei (2006)], we have decided to start the sample period in 1964.

Summary statistics about the yields (expressed on a quarterly basis), the real log-GDP and its first difference are presented in Table $a.i$). 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 increasing with maturity for any given lag and, obviously, decreasing with lag for any given maturity. 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	G_t	g_t
Mean	0.015	0.016	0.016	0.017	0.017	0.017	0.018	8.709	0.008
Std. Dev.	0.007	0.007	0.007	0.006	0.006	0.006	0.006	0.385	0.008
Skewness	1.029	0.846	0.842	0.863	0.886	0.909	0.982	-0.025	-0.085
Kurtosis	4.691	4.136	4.008	3.880	3.782	3.720	3.643	1.830	4.487
Minimum	0.002	0.003	0.003	0.004	0.005	0.006	0.009	7.990	-0.020
Maximum	0.040	0.039	0.039	0.039	0.038	0.038	0.037	9.352	0.039
ACF(1)	0.910	0.932	0.940	0.946	0.951	0.955	0.959	0.981	0.268
ACF(4)	0.760	0.788	0.805	0.817	0.826	0.831	0.842	0.925	0.093
ACF(8)	0.513	0.581	0.627	0.658	0.679	0.693	0.717	0.853	-0.167
ACF(12)	0.335	0.426	0.494	0.538	0.566	0.585	0.616	0.785	-0.170
ACF(16)	0.240	0.307	0.365	0.404	0.430	0.448	0.482	0.718	0.004
ACF(20)	0.224	0.252	0.283	0.308	0.325	0.336	0.356	0.655	0.127

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

II.2 State Dynamics Specification

II.2.1 Unit Root Analysis

The first step of our modelling is the study of 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 by Elliott, Rothenberg and Stock (1996) [Dickey-Fuller test with GLS detrending (denoted Dickey-Fuller GLS), and Point-Optimal test], and in the work of Ng and Perron (2001) (denoted Ng-Perron). It is well known that ADF and PP tests have size distortion and low power against various alternatives, and against trend-stationary alternatives when conventional sample size are considered [see, for instance, De Jong, Nankervis, Savin and Whiteman (1992a, 1992b)]. 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.

The results [see tables from *a.ii*) and *a.iii*)] 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. The purpose of the next section is to search for long-run equilibrium relationships among the components of Y_t, using cointegration techniques.

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)[1]. 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 and PP tests, we use MacKinnon (1996) critical values for the t statistic while we consider Dickey and Fuller (1981, Tables IV-VI) critical values for the F statistics. In the Ng-Perron test, critical values are taken from their original paper (Table 1). With regard to the

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

Dickey-Fuller GLS test, if only a constant is included in the test regression, we use MacKinnon (1996) critical values, while, if we include also a linear time trend, we apply critical values taken from Elliott, Rothenberg and Stock (1996, Table 1). Indeed, in the first case only, their t-statistic follows a Dickey-Fuller distribution. In the Point-Optimal test, critical values are provided by Elliott, Rothenberg and Stock (1996, Table 1).

In the ADF tests of table *a.ii*) and in the left panel of table *a.iii*), when we test the null hypothesis of unit root, we assume the following test regression with a non null constant:

$$\Delta x_t = c + \xi_0 \, x_{t-1} + \sum_{j=1}^{p-1} \xi_j \, \Delta x_{t-j} + \varepsilon_t \,, \quad \varepsilon_t \sim i.i.d.(0, \sigma^2) \,, \quad x_t \in \{r_t, R_t\} \,, \qquad (1)$$

but we assume that $c = 0$ in the true process, that is we test the presence of a unit root in x_t ($\xi_0 = 0$) assuming $c = 0$ under the null. The same assumption is made about the true process and the test regression (the non-augmented version of (1)) used for the PP test. The F statistics test the null joint hypothesis $(c, \xi_0)' = (0, 0)'$. In the right panel of table *a.iii*), the test regression of the ADF test is assumed to be:

$$\Delta x_t = c + b\,t + \xi_0 \, x_{t-1} + \sum_{j=1}^{p-1} \xi_j \, \Delta x_{t-j} + \varepsilon_t \,, \quad \varepsilon_t \sim i.i.d.(0, \sigma^2) \,, \quad x_t = \log(GDP_t) = G_t \,, \qquad (2)$$

while the true process assumes $c \neq 0$, $b = 0$ and $\xi_0 = 0$, that is we test the null hypothesis of unit root ($\xi_0 = 0$) assuming $c \neq 0$ and $b = 0$ under the null. The same assumptions apply to the true process and the test regression (the non-augmented version of (2)) behind the PP test. The F statistics test the null joint hypothesis $(b, \xi_0)' = (0, 0)'$, assuming $c \neq 0$.

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.

The results can be summarized as follows. With regard to the short rate and the long rate, the

F statistic of Table $a.ii$) accepts (at 5% or 10% level) for both series the joint hypothesis of unit root and no constant. This result is confirmed by ADF and PP tests as well as efficient unit root tests. As far as the log-GDP level is concerned, the F statistic rejects the above mentioned joint hypothesis [see left panel of Table $a.iii$)] and, once we assume the presence of a constant under the null, we accept the hypothesis of unit root (t-value = -0.907) with a constant term consistently found significantly different from zero (t-value = 5.08). When also a linear time trend is included in the test regression (but not in the true process; see Table $a.iii$), right panel), the F statistic rejects the joint hypothesis of unit root and no linear time trend. Nevertheless, if we assume the presence of a unit root, the regression test leads to statistically accept the presence of a constant and no linear trend. When we consider the efficient unit root tests (in which we typically have the same deterministic terms in the test regression and in the true process), 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.

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

Table $a.ii$): 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).

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

Table *a.iii*): Left Panel: unit root tests for the log-GDP G_t (a constant is included in test regressions). Right Panel: Unit root tests for the log-GDP G_t (a constant and a linear time trend are included in test regressions).

II.2.2 Order Selection and Parameter Estimates of the VAR(p) Model for $Y_t = (r_t, R_t, G_t)'$

We present in Tables *a.iv*) and *a.v*) the VAR order selection and parameter estimates of the model:

$$Y_t = \nu + \sum_{j=1}^{p} \Phi_j \, Y_{t-j} + \varepsilon_t \, , \;\; \varepsilon_t \sim IIN(0, \Omega) \, , \tag{3}$$

describing the joint dynamics of $Y_t = (r_t, R_t, G_t)'$. As far as Table *a.iv*) is concerned, 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. The results in Table *a.iv*) lead us to select $p = 3$ and, thus, in Table *a.v*) we provide parameter estimates of a VAR(3) model for Y_t.

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

Table *a.iv*): Criteria for VAR order selection.

	ν	Φ_1			Φ_2			Φ_3		
r_t	0.0013	0.6068	0.1431	0.0350	0.1023	-0.1896	0.0492	0.3239	-0.0570	-0.0843
	[0.2289]	[6.4211]	[0.7662]	[1.1271]	[0.9346]	[-0.8029]	[1.0878]	[3.1638]	[-0.3204]	[-2.9603]
R_t	0.0016	0.0198	0.8108	0.0120	0.0111	0.0855	-0.0030	0.0797	-0.0462	-0.0091
	[0.5235]	[0.3886]	[8.0530]	[0.7187]	[0.1889]	[0.6715]	[-0.1235]	[1.4430]	[-0.4818]	[-0.5931]
G_t	0.04404	0.2289	-0.3115	1.1425	-1.0742	0.4420	0.0129	0.2328	0.3175	-0.1597
	[2.8522]	[0.9121]	[-0.6282]	[13.8549]	[-3.6955]	[0.7048]	[0.1075]	[0.8563]	[0.6721]	[-2.1117]
	$\Omega \times 10^3$			Corr.		log-L	$\mid\psi\mid$			
	0.0079	0.0026	0.0048	ρ_{12}	0.6060	2289.43	0.9968			
	[8.9722]	[6.5758]	[2.8558]			AIC	0.9315			
	.	0.0023	0.0036	ρ_{13}	0.2310	-26.4261	0.8647			
		[8.9722]	[3.8719]			SIC	0.6170			
	.	.	0.0555	ρ_{23}	0.3205	-25.8749	0.5595(^{c})			
			[8.9722]			FPE	0.2891			
						6.70e-16	0.0840(^{c})			

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

II.2.3 Cointegration Analysis

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:

$$Y_t = \nu + \sum_{j=1}^{p} \Phi_j Y_{t-j} + \varepsilon_t \,, \tag{4}$$

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, the lag length is selected to be $p = 3$ [see tables *a.iv)* and *a.v)*]. Then, we write the Gaussian VAR(3) model in the equivalent vector error correction model (VECM) representation:

$$\begin{aligned} &\Delta Y_t = \Pi Y_{t-1} + \sum_{j=1}^{2} \Gamma_j \Delta Y_{t-j} + \nu + \varepsilon_t \,, \\ &\text{with} \quad \Pi = -\left[I_{3\times 3} - \textstyle\sum_{j=1}^{3} \Phi_j \right] \text{ and } \Gamma_j = -\sum_{i=j+1}^{3} \Phi_i \,, \end{aligned} \tag{5}$$

and we determine the rank $r \in \{0, 1, 2, 3\}$ of the matrix Π using the trace and maximum eigenvalue tests. The results indicate that both tests accept the presence of one cointegrating relation at 5 % level, and, thus, they decide for the presence of two unit roots. Consequently, we can write $\Pi = \alpha\beta'$, where α and β are (3×1) vectors [see Table *a.vi)*], 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 (5), in which the unrestricted constant term ν induces a linear trend in Y_t. Given the decomposition $\nu = \alpha\mu + \gamma$ (with μ a scalar determined so that the error correction term has a sample mean of zero, and γ a (3×1) vector), we have tested the null hypothesis $\mathrm{H}_0 : \nu = \alpha\mu$ (the intercept is restricted to lie in the α

direction) using the $\chi^2(2)$-distributed (under H_0) likelihood ratio statistic 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[2].

Moreover, in order to achieve economic interpretability of the cointegrating relation, we have tested the null hypothesis $H_0 : \beta = (-1, 1, 0)'$ (the spread between the long and the short rate is the cointegrating relation). The likelihood ratio statistic taking the value[3] 3.276, which is smaller than the chi-square 5 % quantile (with two degrees of freedom) $\chi^2_{0.05}(2) = 5.99$, the null hypothesis is accepted, and, therefore, the spread provides the long-run equilibrium relationship [see Table *a.vii*) and Campbell and Shiller (1987), Engle and Granger (1987), Hall, Anderson and Granger (1992)][4].

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 (5) in terms of the 3-dimensional state process $X_t = (r_t, S_t, g_t)'$, with $S_t = R_t - r_t$ and $g_t = G_t - G_{t-1}$:

$$X_t = \widetilde{\nu} + \sum_{j=1}^{3} \widetilde{\Phi}_j X_{t-j} + \eta_t\,, \text{ with } \widetilde{\nu} = A\nu\,, \ A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\,, \ \widetilde{\Phi}_1 = \widetilde{\Gamma}_1 + \widetilde{\alpha}\,(0,1,0) + B\,,$$

$$\widetilde{\Phi}_2 = \widetilde{\Gamma}_2 - \widetilde{\Gamma}_1 B\,, \ \widetilde{\Phi}_3 = -\widetilde{\Gamma}_2 B\,, \ B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\,, \ \widetilde{\Gamma}_i = A\Gamma_i A^{-1} \text{ for } i \in \{1,2\}\,, \ \tilde{\alpha} = A\alpha\,, \tag{6}$$

and where η_t is a 3-dimensional Gaussian white noise with $\mathcal{N}(0, \widetilde{\Omega})$ distribution and $\widetilde{\Omega}$[5]. Note that

[2]The likelihood ratio statistic is $\widetilde{lr} = -T\sum_{k=2}^{3} \log[(1-\tilde{\lambda}_k)/(1-\lambda_k)]$, 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 (5). The estimation of the two models leads to $(\tilde{\lambda}_2, \tilde{\lambda}_3) = (0.0962431, 0.032958)$ and $(\lambda_2, \lambda_3) = (0.039789, 0.008368)$.

[3]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 (5) under H_0.

[4]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 as in APW (2006) (we are very grateful to Andrew Ang, Monika Piazzesi and Min Wei for providing us the data set).

[5]The parameter estimates are presented in Table *a.viii*), while the estimates of the VAR(3) and VAR(1) (*à la*

the third column of $\tilde{\Phi}_3$ is a vector of zeros.

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

Table *a.vi*): 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 (5)) 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 H_0: $\beta = (1, 0, \beta_3)'$ and H_0 : $\beta = (1, \beta_2, 0)'$. The alternative hypothesis is H_A : $\beta = (1, \beta_2, \beta_3)'$ in both cases, and the 5% and 1% critical values for a $\chi^2(1)$ are, respectively, 3.84 and 6.63.

APW (2006)) state dynamics are organized, respectively, in Tables *a.ix*) and *a.x*).

II.2.4 Historical Parameter Estimates of VAR and CVAR Models for $X_t = (r_t, S_t, g_t)'$

	γ	Γ_1			Γ_2			α	μ
Δr_t	-0.0010	-0.4154	0.1941	0.0393	-0.3152	0.0110	0.0893	-0.0349	-0.0029
	[-2.7075]	[-3.9779]	[1.1085]	[1.2985]	[-3.1595]	[0.0626]	[3.1587]	[-0.5074]	
ΔR_t	-0.0002	-0.0862	-0.0675	0.0148	-0.0761	0.0216	0.0120	-0.1132	
	[-0.9910]	[-1.5306]	[-0.7148]	[0.9048]	[-1.4135]	[0.2270]	[0.7891]	[-3.0492]	
ΔG_t	0.0048	0.6854	-0.7554	0.1964	-0.3700	-0.3038	0.1912	0.4086	
	[4.8954]	[2.4406]	[-1.6043]	[2.4138]	[-1.3792]	[-0.6415]	[2.5147]	[2.2074]	
	$\Omega \times 10^3$			log-L	$\lvert \psi \rvert$				
	0.0079	0.0026	0.0051	2283.60	1.0000**				
	[9.0277]	[6.6826]	[2.9613]	AIC	0.8478				
	.	0.0023	0.0038	-26.3930	0.6271				
		[9.0277]	[3.9628]	SIC	0.5489(^{c})				
	.	.	0.0573	-25.8969	0.2468				
			[9.0277]	FPE	0.1050(^{c})				
				7.18e-16					

Table $a.vii$): 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. $\lvert \psi \rvert$ indicates the modulus of the roots of equation $|\tilde{\Phi}(\psi)| = 0$, with $\tilde{\Phi}(\psi) = (I_{3\times 3}\psi^3 - \Phi_1\psi^2 - \Phi_2\psi - \Phi_3)$ denoting the characteristic polynomial; (^{c}) indicates a pair of complex conjugate roots, while (**) denote a root with multiplicity two.

	$\tilde{\nu}$	$\tilde{\Phi}_1$			$\tilde{\Phi}_2$			$\tilde{\Phi}_3$		
r_t	-0.0009	0.7787	0.1592	0.0393	-0.0829	-0.1831	0.0893	0.3042	-0.0110	0.0000
S_t	0.0010	0.0675	0.6600	-0.0245	0.1822	0.2722	-0.0773	-0.2497	-0.0106	0.0000
g_t	0.0036	-0.0701	-0.3469	0.1964	-0.6038	0.4516	0.1912	0.6739	0.3038	0.0000
	$\tilde{\Omega} \times 10^3$			Corr.						
	0.0079	-0.0053	0.0051	ρ_{12}	-0.8435					
	.	0.0050	-0.0013	ρ_{13}	0.2385					
	.	.	0.0573	ρ_{23}	-0.0786					

Table $a.viii$): Parameter estimates from the CVAR(3) model $X_t = \tilde{\nu} + \sum_{j=1}^{3} \tilde{\Phi}_j X_{t-j} + \eta_t$, with $X_t = (r_t, S_t, g_t)'$ and $\eta_t \sim IIN(0, \tilde{\Omega})$ [Gurkaynak-Sack-Wright (2007) data base; sample period: 1964:Q1 - 2007:Q2]. ρ_{ij} denotes the (empirical) correlation between (η_{it}) and (η_{jt}).

	$\bar{\nu}$	$\bar{\Phi}_1$			$\bar{\Phi}_2$			$\bar{\Phi}_3$		
r_t	0.0002	0.7588	0.1777	0.0295	-0.0739	-0.1960	0.0743	0.2486	-0.0783	0.0428
	[0.2043]	[4.9693]	[0.9530]	[0.9736]	[-0.3727]	[-0.8361]	[2.5760]	[1.6408]	[-0.4438]	[1.5634]
S_t	0.0006	0.0737	0.6416	-0.0182	0.1745	0.2798	-0.0676	-0.2192	0.0250	-0.0312
	[0.9415]	[0.6043]	[4.3039]	[-0.7544]	[1.1009]	[1.4931]	[-2.9311]	[-1.8104]	[0.1771]	[-1.4249]
g_t	0.0062	-0.1316	-0.4212	0.1994	-0.6263	0.4486	0.1969	0.6386	0.2473	-0.0532
	[2.7461]	[-0.3165]	[-0.8295]	[2.4207]	[-1.1596]	[0.7028]	[2.5077]	[1.5481]	[0.5148]	[-0.7140]
	$\Omega \times 10^3$			Corr.		log-L	$\lvert \psi \rvert$			
	0.0077	-0.0052	0.0050	ρ_{12}	-0.8412	2287.93	0.9299			
	[8.9722]	[-8.1680]	[2.9381]			AIC	0.8888			
	.	0.0049	-0.0013	ρ_{13}	0.2380	-26.4085	0.6015(^{c})			
		[8.9722]	[-0.9965]			SIC	0.49188			
	.	.	0.0574	ρ_{23}	-0.0788	-25.8574	0.4283(^{c})			
			[8.9722]			FPE	0.2449			
						6.82e-16	0.1598			

Table $a.ix$): Parameter estimates from the unconstrained VAR(3) model $X_t = \bar{\nu} + \sum_{j=1}^{3} \bar{\Phi}_j X_{t-j} + \xi_t$, with $X_t = (r_t, S_t, g_t)'$ and $\xi_t \sim IIN(0, \bar{\Omega})$ [Gurkaynak-Sack-Wright (2007) data base; sample period: 1964:Q1 - 2007:Q2]. t-values are in brackets. ρ_{ij} denotes the (empirical) correlation between (ξ_{it}) and (ξ_{jt}). log-L denotes the maximum value of the log-Likelihood function. $\lvert \psi \rvert$ indicates the modulus of the roots of equation $|\bar{\Phi}(\psi)| = 0$, with $\bar{\Phi}(\psi) = (I_{3\times3}\psi^3 - \bar{\Phi}_1\psi^2 - \bar{\Phi}_2\psi - \bar{\Phi}_3)$ denoting the characteristic polynomial; (^{c}) indicates a pair of complex conjugate roots.

	$\bar{\nu}$	$\bar{\Phi}$			$\bar{\Omega} \times 10^3$		
r_t	0.0008	0.9347	0.0800	-0.0020	0.0088	-0.0061	0.0049
	[0.9836]	[24.1968]	[1.1122]	[-0.0749]	[9.1924]	[-8.5031]	[2.6382]
S_t	0.0001	0.0270	0.8148	-0.0012	.	0.0056	-0.0014
	[0.2058]	[0.8759]	[14.1675]	[-0.0558]		[9.1924]	[-0.9778]
g_t	0.0080	-0.1680	0.1820	0.2428	.	.	0.0619
	[3.7540]	[-1.6401]	[0.9533]	[3.3550]			[9.1924]
	Corr.		log-L	$\lvert \psi \rvert$			
	ρ_{12}	-0.8647	2287.21	0.9510			
			AIC	0.7985			
	ρ_{13}	0.2072	-26.3377	0.2427			
			SIC				
	ρ_{23}	-0.0754	-26.1737				
			FPE				
			7.57e-16				

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

II.2.5 Risk-Neutral Parameter Estimates of VAR and CVAR Yield Curve Models

A	γ_o	γ_1			γ_2			γ_3		
r_t	-0.137	100.691	102.269	-14.896	19.314	8.438	-20.437	-100.213	-15.560	0
	[-2.844]	[6.450]	[5.324]	[-4.821]	[0.800]	[0.304]	[-7.622]	[-7.337]	[-0.978]	
S_t	0.092	47.176	93.102	-0.956	-97.518	-79.733	5.604	32.176	30.835	0
	[1.450]	[1.743]	[2.743]	[-0.166]	[-2.225]	[-1.573]	[1.233]	[1.467]	[1.188]	
g_t	1.540	47.995	99.149	-41.930	-232.010	77.676	62.374	91.746	-349.378	0
	[1.648]	[0.126]	[0.204]	[-0.587]	[-0.391]	[0.111]	[1.057]	[0.304]	[-0.968]	
B	γ_o	γ_1			γ_2			γ_3		
r_t	-0.525	108.229	98.626	-11.656	19.073	15.907	-18.090	-83.383	4.561	-11.029
	[-10.560]	[7.004]	[5.143]	[-3.828]	[0.792]	[0.577]	[-6.043]	[-6.107]	[0.286]	[-4.488]
S_t	-0.162	54.490	97.186	-0.681	-97.839	-80.545	7.869	38.467	41.643	-1.424
	[-2.253]	[2.001]	[2.831]	[-0.112]	[-2.210]	[-1.585]	[1.401]	[1.747]	[1.565]	[-0.325]
g_t	1.714	43.853	140.023	-61.069	-250.282	108.642	4.151	109.366	-416.392	61.739
	[1.374]	[0.095]	[0.242]	[-0.646]	[-0.340]	[0.130]	[0.045]	[0.300]	[-0.955]	[0.912]
C	γ_o	γ_1								
r_t	-0.478	34.789	56.968	-4.557						
	[-6.629]	[11.523]	[9.066]	[-1.777]						
S_t	-0.301	2.384	66.540	5.373						
	[-4.304]	[1.115]	[13.899]	[1.771]						
g_t	1.588	0.648	4.256	-5.401						
	[0.539]	[0.006]	[0.018]	[-0.057]						

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

II.3 Reliability of $\lambda^*(40) \approx 0.26$

II.3.1 Reliability over the Time Series Dimension

The purpose of this section is to empirically highlight, in the time series dimension, the reliability of the averaging parameter $\lambda^*(40)$ determined in Jardet, Monfort and Pegoraro (2012) to optimally extract the expectation term from the 10-year bond. The value $\lambda^*(40)$ is obtained by solving the following problem:

$$\lambda^*(40) = arg \min_{\lambda \in [0,1]} \sum_t [\widetilde{B}_t^{obs}(40) - \widetilde{B}_t(40, \lambda)]^2 \tag{7}$$

where, for each date t, $\widetilde{B}_t^{obs}(40)$ is the observed realization of $\exp(-r_t - ... - r_{t+39})$ while $\widetilde{B}_t(40, \lambda)$ is the NCVAR(3) model's forecast of $\exp(-r_t - ... - r_{t+39})$. In Jardet, Monfort and Pegoraro (2012), the out-of-sample forecasts are performed during the period 1990:Q1 - 2007:Q2, using an expanding window for the estimation of models VAR(3) and CVAR(3), and the minimization of criterion (7) is obtained for $\lambda^*(40) \approx 0.26$.

What we show in Table *a.xii*) is that this value remains around 0.26 even if the window of observations [1964:Q1, t], starting the above mentioned out-of-sample forecast exercise (in which we have taken $t = 1989 : Q4$), is specified for t varying from $t =$ 1986:Q4 to $t =$ 1992:Q4. Moreover, if we consider a rolling window instead of an expanding one, we obtain $\lambda^*(40) \approx 0.23$, with the NCVAR(3) model still being the best one in extracting term premia from long-term bonds [see Table *a.xiii*)].

t	1986:Q4	1987:Q4	1988:Q4	1989:Q4	1990:Q4	1991:Q4	1992:Q4
$\lambda^*(40)$	0.28	0.29	0.29	0.26	0.25	0.28	0.25
RMSFE	0.1010	0.1053	0.1090	0.1012	0.0911	0.0935	0.1011

Table *a.xii*): Table entries are $\lambda^*(40)$ and associated RMSFEs. The time to maturity ($h = 40$) is measured in quarters.

	h	AR(1) (Vasicek)	$\lambda^*(h)$	NCVAR(3)	CVAR(3)	VAR(3)	VAR(1)
$\widetilde{B}_t^{obs}(h)$	2	0.0015	0.1554	0.0016	0.0017	0.0017	0.0016
	4	0.0072	0.1802	0.0070	0.0071	0.0075	0.0076
	8	0.0261	0.1709	0.0246	0.0247	0.0264	0.0264
	12	0.0483	0.1438	0.0448	0.0450	0.0492	0.0470
	16	0.0700	0.0981	0.0617	0.0619	0.0709	0.0664
	20	0.0904	0.0341	0.0743	0.0744	0.0917	0.0850
	28	0.1205	0.0791	0.0978	0.0985	0.1238	0.1150
	32	0.1294	0.1549	0.1025	0.1067	0.1292	0.1245
	36	0.1405	0.1926	0.1088	0.1179	0.1364	0.1371
	40	0.1483	0.2318	0.1062	0.1248	0.1360	0.1460

Table *a.xiii*): Out-of-sample forecasts of $\widetilde{B}_t^{obs}(h) = \exp(-r_t - ... - r_{t+h-1})$ with rolling window.

II.3.2 Reliability over the Cross-Sectional Dimension

The purpose of this section is to further testify the reliability of the value $\lambda^* = \lambda^*(40) \approx 0.26$, and also to show the ability of our yield-to-maturity formula to explain the observed interest rates variability in terms of fitting performance.

The reliability check of the estimated averaging parameter is done in the following way: instead of using criterion (4) in Jardet, Monfort and Pegoraro (2012) to estimate λ, we estimate it at the same time as the risk sensitivity parameters $\theta_\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$, by minimizing the fitting error of the whole yield curve, that is by using the CNLLS methodology of Section 4.2. This value is found to be $\lambda^{**} = 0.256$, very close to $\lambda^*(40) \approx 0.26$.

	NCVAR(3) $[\lambda^{**} = 0.256]$	NCVAR(3) $[\lambda^* = 0.262]$	CVAR(3)	VAR(3)	VAR(1)
Mean	16.79	16.81	16.91	16.86	18.76
Median	12.11	12.22	12.89	12.55	16.02
Std. Dev.	13.80	13.88	14.02	13.96	15.23

Table *a.xiv*): Annualized Absolute Pricing Errors (Basis Points).

As far as the in-sample fit is concerned, we compare in Table *a.xiv*) the annualized absolute yield-to-maturity errors of our selected NCVAR(3) factor-based term structure model with the performances of the competing CVAR(3), VAR(3) and VAR(1) yield curve models. For each date t and each estimated model, we compute, over the maturities used to estimate the risk sensitivity parameters $\theta_\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$, the pricing error in the following way:

$$PE_t = \frac{\sum_h |R_t^{obs}(h) - R_t(h)|}{H}\,, \tag{8}$$

where $R_t^{obs}(h)$ 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 and standard deviation. We see from table *a.xiv*) that the fitting based on the NCVAR(3) estimates is slightly better than the one based on the CVAR(3) and VAR(3) estimates and much better than the one based on the VAR(1) model [see table $a.xiv$) for further details].

We also have considered the case of a weighting parameter λ optimally selected on the basis of a criterion of interest like the forecast of yields over several horizons, and found that our model is again able to provide a reduction of the associated root mean squared forecast errors [see Section III]. We have also verified, working with Campbell-Shiller regressions, the ability of our approach to explain the observed violation of the Expectation Hypothesis theory over the maturity spectrum [see Section IV].

III Interest Rates Forecasts with the NCVAR(3) Model

In Section 3.2.2 we have seen that the extraction of the expectation term of the long-term bond $B_t(40)$ is more precise when performed by our NCVAR(3) model. The purpose of the present section is to analyze the interest rates out-of-sample forecast performances that the NCVAR(3) affine model is able to produce. The forecast exercise considers two cases: i) when λ is selected to minimize, for each forecasting horizon q (say) and for each variable, the associated RMSFE; in this context λ is considered as a free parameter which gives a further degree of freedom in order to improve model's performances like, in this case, the forecast of a variable of interest over a certain horizon; ii) when the averaging parameter is fixed to $\lambda = 0.262$. As in Section 3.2.2, the out-of-sample forecast exercise is performed using an expanding window: we first compute the estimates $\widehat{\theta}_{t,var}$ and $\widehat{\theta}_{t,cvar}$ over the period 1964:Q1 - 1989:Q4 and then, at each later point in time t, we recompute $\widehat{\theta}_{t,var}$ and $\widehat{\theta}_{t,cvar}$ taking into account the new observation. The results, organized in Table $a.xv$), are the following.

First, regarding to the optimal value of $\lambda = \lambda(h, q)$ (say) in the NCVAR(3) specification, we observe that for any residual maturity h, as far as q increases, $\lambda^*(\cdot, q)$ decreases from values (rather) close to one towards zero. This result indicates that the minimization of the forecast error, when the forecasting horizon increases, gives an increasing weight to the CVAR(3) component and, thus, it indicates how important it is for obtaining reliable long-run forecasts. Second, our NCVAR(3) model

	q	AR(1)	$\lambda^*(h,q)$	NCVAR(3)	NCVAR(3) $[\lambda = 0.262]$	CVAR(3)	VAR(3)	VAR(1)
$R_t(1) = r_t$	1	0.0015	1.0000	0.0015	0.0016	0.0016	0.0015	0.0016
	4	0.0044	0.7585	0.0039	0.0039	0.0040	0.0039	0.0045
	8	0.0066	0.5993	0.0061	0.0061	0.0063	0.0061	0.0065
	12	0.0075	0.4329	0.0069	0.0070	0.0071	0.0072	0.0072
	20	0.0081	0.0378	0.0067	0.0068	0.0067	0.0078	0.0078
	40	0.0106	0.0000	0.0058	0.0079	0.0058	0.0102	0.0107
$R_t(4)$	1	0.0015	0.8781	0.0017	0.0018	0.0019	0.0023	0.0021
	4	0.0041	0.7316	0.0039	0.0040	0.0041	0.0047	0.0046
	8	0.0062	0.6468	0.0060	0.0061	0.0062	0.0071	0.0064
	12	0.0071	0.5086	0.0069	0.0070	0.0071	0.0084	0.0073
	20	0.0080	0.0566	0.0067	0.0068	0.0066	0.0091	0.0083
	40	0.0108	0.0000	0.0059	0.0080	0.0057	0.0109	0.0114
$R_t(8)$	1	0.0015	0.5851	0.0017	0.0018	0.0020	0.0026	0.0022
	4	0.0038	0.6630	0.0036	0.0037	0.0038	0.0046	0.0042
	8	0.0056	0.6340	0.0055	0.0057	0.0058	0.0069	0.0058
	12	0.0066	0.5307	0.0065	0.0066	0.0067	0.0082	0.0067
	20	0.0077	0.0803	0.0064	0.0065	0.0062	0.0093	0.0079
	40	0.0106	0.0000	0.0059	0.0080	0.0056	0.0109	0.0112
$R_t(12)$	1	0.0015	0.5026	0.0017	0.0018	0.0020	0.0029	0.0023
	4	0.0035	0.6235	0.0034	0.0035	0.0036	0.0046	0.0039
	8	0.0050	0.6137	0.0051	0.0052	0.0053	0.0067	0.0054
	12	0.0061	0.5320	0.0061	0.0062	0.0063	0.0081	0.0062
	20	0.0074	0.1006	0.0062	0.0063	0.0059	0.0093	0.0075
	40	0.0104	0.0000	0.0062	0.0079	0.0056	0.0108	0.0109
$R_t(20)$	1	0.0015	0.4543	0.0017	0.0018	0.0022	0.0035	0.0027
	4	0.0031	0.5908	0.0031	0.0033	0.0034	0.0048	0.0038
	8	0.0043	0.5848	0.0044	0.0046	0.0046	0.0065	0.0048
	12	0.0053	0.5273	0.0055	0.0056	0.0055	0.0078	0.0056
	20	0.0067	0.1380	0.0058	0.0059	0.0053	0.0093	0.0070
	40	0.0098	0.0000	0.0067	0.0079	0.0056	0.0106	0.0105
$R_t(40) = R_t$	1	0.0113	1.0000	0.0012	0.0012	0.0012	0.0012	0.0012
	4	0.0101	0.8894	0.0023	0.0024	0.0024	0.0023	0.0023
	8	0.0090	0.6469	0.0030	0.0031	0.0032	0.0031	0.0031
	12	0.0082	0.4707	0.0039	0.0040	0.0041	0.0041	0.0043
	20	0.0070	0.0000	0.0042	0.0045	0.0042	0.0056	0.0059
	40	0.0055	0.0000	0.0055	0.0064	0.0055	0.0081	0.0086

Table $a.xv$): Out-of-sample forecasts of $R_t(h)$, with $h \in \{1,4,8,12,20,40\}$ 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 $h \in \{1,4,8,12,20,40\}$. Forecasting horizons (q) are measured in quarters. r_t denotes the (one-quarter) short rate and R_t is the 10-year interest rate.

outperforms, over both short and long forecasting horizons, the AR(1) and VAR(1) specifications, and over short and medium horizons the CVAR(3) model.

If we consider the forecast of the state variables obtained by the NCVAR(3) process with $\lambda = 0.262$, the results we obtain are the following. Concerning the short rate, even if the averaging parameter is selected using the expectation term criterion, the RMSFEs produced by our selected state process remain lower than or equal to those obtained by the AR(1) and VAR models. Its performances are slightly better or in line with the CVAR(3) model for short and medium horizons. If we consider the long rate, the average forecast errors remain quite close to those obtained in the previous case, for short and medium forecasting horizons, while, for $q = 40$ quarters, the AR(1) model turns out to work better. As far as the yields with residual maturity between 4 and 20 quarters are concerned, we see that the NCVAR(3) model still outperforms the AR(1) and VAR specifications over long horizons, and works better that the CVAR(3) model over short and medium horizons.

IV Campbell-Shiller Regressions

Let us now study the ability of our NCVAR(3) term structure model (with $\lambda = 0.262$) 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, 2003), Monfort and Pegoraro (2007)]. We find confirmation of this stylized fact also in the GSW (2007) data base considered in our empirical analysis; indeed, the estimated slope coefficients $\phi_{h,T}$ (say) obtained from the above mentioned regression is always negative and moves from -0.494 to -2.567 when h increases from three to forty quarters (see the second column of Table $a.xvi$)).

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:

$$\phi_h = \frac{Cov[R_{t+1}(h-1) - R_t(h), (R_t(h) - r_t)/(h-1)]}{Var[(R_t(h) - r_t)/(h-1)]}, \tag{9}$$

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 $a.xvi$), 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 $h = 8$).

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

Table $a.xvi)$: 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.

V Building Normalized Shocks for *NIRFs*

The purpose of this section is to transpose the analysis of Sections 6.3 and 6.4 in Jardet, Monfort and Pegoraro (2012) to the case of a shock on the 1-year spread $S_t(4)$ (say) and 5-year spread $S_t(20)$ (say) as well as their components (i.e., with obvious notation, $(EXS_t(4), TP_t(4))$ and $(EXS_t(20), TP_t(20))$ respectively).

In order to generate impulse responses comparable with those presented in the two above mentioned sections, we propose to specify any shock in such way that the instantaneous reaction of the 10-year spread is equal to one. In other words, the family of shocks we define share the common feature of generating the same instantaneous variation in the slope of the yield curve (the unitary shock on S_t). We will name this shock a normalized shock.

More formally, and following the same assumptions as in Sections 6.3 and 6.4 of Jardet, Monfort and Pegoraro (2012), we specify the shock at date t on the variable $\vartheta_t^{(h)}$, with $\vartheta_t^{(h)} \in \{S_t(h), EXS_t(h), TP_t(h)\}$ for $h = 1$ and $h = 5$ years, in the following way. Given the affine structure of our model, the h-year spread $S_t(h)$, its expectation part $EXS_t(h)$ and its term premium $TP_t(h)$ are obtained by applying a linear filter on $y_t = (r_t, S_t, g_t)'$:

$$S_t(h) = F_{1,1}^{(h)}(L)r_t + F_{1,2}^{(h)}(L)S_t + F_{1,3}^{(h)}(L)g_t \tag{10}$$

$$EXS_t(h) = F_{2,1}^{(h)}(L)r_t + F_{2,2}^{(h)}(L)S_t + F_{2,3}(L)g_t \tag{11}$$

$$TP_t(h) = F_{3,1}^{(h)}(L)r_t + F_{3,2}^{(h)}(L)S_t + F_{3,3}^{(h)}(L)g_t\,. \tag{12}$$

Hence, innovations at $t = 0$ of $S_t(h)$, $EXS_t(h)$ and $TP_t(h)$, respectively denoted by $\tilde{\eta}_{0,1}^{(h)}$, $\tilde{\eta}_{0,2}^{(h)}$ and $\tilde{\eta}_{0,3}^{(h)}$, are:

$$\tilde{\eta}_{0,1}^{(h)} = F_{1,1}^{(h)}(0)\eta_{0,1} + F_{1,2}^{(h)}(0)\eta_{0,2} + F_{1,3}^{(h)}(0)\eta_{0,3} \tag{13}$$

$$\tilde{\eta}_{0,2}^{(h)} = F_{2,1}^{(h)}(0)\eta_{0,1} + F_{2,2}^{(h)}(0)\eta_{0,2} + F_{2,3}^{(h)}(0)\eta_{0,3} \tag{14}$$

$$\tilde{\eta}_{0,3}^{(h)} = F_{3,1}^{(h)}(0)\eta_{0,1} + F_{3,2}^{(h)}(0)\eta_{0,2} + F_{3,3}^{(h)}(0)\eta_{0,3}\,, \tag{15}$$

where $\eta_{0,1}$, $\eta_{0,2}$ and $\eta_{0,3}$ are innovations at $t = 0$ of r_t, S_t and g_t respectively.

Normalized Shock on the h-year spread $S_t(h)$

We have to determine the value of the vector $\delta^{(S)}(hy) := E(\eta_0 \,|\, I_0) = E(\eta_0 \,|\, \tilde{\eta}_{0,1}^{(h)} = \tilde{\delta}_1, \eta_{0,3} = 0)$ with $\tilde{\delta}_1$ such that $E(\eta_{0,2} \,|\, \tilde{\eta}_{0,1}^{(h)} = \tilde{\delta}_1, \eta_{0,3} = 0) = 1$. Hence, the normalized shock is given by:

$$\delta^{(S)}(h\,y) := E(\eta_0 \,|\, I_0) = \left(\beta \frac{F_{1,2}^{(h)}(0)}{1 - F_{1,1}^{(h)}(0)\beta}, 1, 0\right)',$$

where β is the coefficient of $\tilde{\eta}_{0,1}^{(h)}$ in the theoretical regression of $\eta_{0,1}$ on $\tilde{\eta}_{0,1}^{(h)}$ and $\eta_{0,3}$.

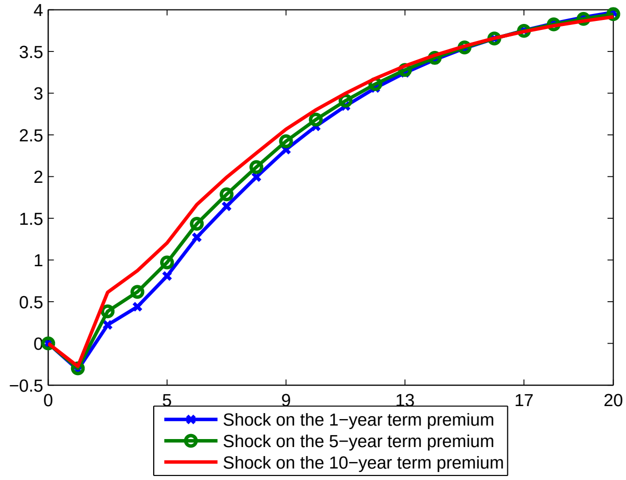

Figure a.1): Response of cumulated GDP after a shock on 1-y, 5-y and 10-y Spread.

Normalized Shock on $EXS_t(h)$

From (15), considering $\tilde{\eta}_{0,3}^{(h)} = 0$, $\eta_{0,2} = 1$ and $\eta_{0,3} = 0$ we obtain :

$$\delta^{(EXS)}(h\,y) := E(\eta_0 \,|\, I_0) = \left(-\frac{F_{3,2}^{(h)}(0)}{F_{3,1}^{(h)}(0)}, 1, 0 \right)'.$$

Normalized Shock on $TP_t(h)$

From (14), considering $\tilde{\eta}_{0,2}^{(h)} = 0$, $\eta_{0,2} = 1$ and $\eta_{0,3} = 0$ we obtain :

$$\delta^{(TP)}(h\,y) := E(\eta_0 \,|\, I_0) = \left(-\frac{F_{2,2}^{(h)}(0)}{F_{2,1}^{(h)}(0)}, 1, 0 \right)'.$$

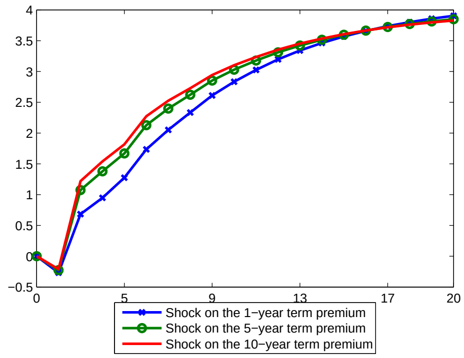

Figure $a.2$):
Response of cumulated GDP after a shock on the expectation part of the 1-y, 5-y and 10-y spread.

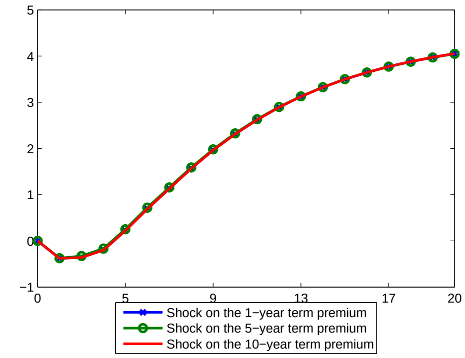

Figure $a.3$):
Response of cumulated GDP after a shock on 1-y, 5-y and 10-y Term Premium.

REFERENCES

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.

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.

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.

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.

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

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

Elliott, G., 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.

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

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.

Jardet C., Monfort A., and F. Pegoraro, (2012): "No-arbitrage Near-Cointegrated VAR(p) term structure models, term premia and GDP growth", Working Paper available at http://www.crest.fr/pageperso/pegoraro/pegoraro.htm.

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.

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., (1996): "Useful Modifications to Unit Root Tests with Dependent Errors and their Local Asymptotic Properties", Review of Economic Studies, 63, 435-463.

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.

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.

Sims, C., (1980): "Macroeconomics and Reality", Econometrica, 48(1), 1-48.

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