Canonical term-stucture models with observable factors and the dynamics of bond risk premia*

Marcello Pericoli† and Marco Taboga‡

JEL classification: C5, G1.

Keywords: term structure models, yield curve, risk premium.

March 21, 2008

*Any views expressed in this article are the authors' and do not necessarily represent those of the Bank of Italy. We are extremely grateful to Fabio Panetta, who encouraged us to undertake this research project, and to an anonymous referee for the many improvements she (he) suggested. We thank for helpful discussion Paolo Angelini, Fabio Fornari, Giuseppe Grande, Marco Protopapa, Glenn Rudebusch, Oreste Tristani, Paolo Zaffaroni and seminar participants and discussants at the Bank of Italy, at the ECB, at the FFM 2005, EEA-ESEM 2006 and EFA 2006 Conferences and at the IX Quantitative Finance Workshop.

†Marcello Pericoli (marcello.pericoli@bancaditalia.it), Economist, Economic Outlook and Monetary Policy Department, Bank of Italy.

‡Marco Taboga (marco.taboga@bancaditalia.it), Economist, Economic Outlook and Monetary Policy Department, Bank of Italy.

Abstract

We derive a canonical representation for the no-arbitrage discrete-time term structure models with both observable and unobservable state variables, popularized by Ang and Piazzesi (2003). We conduct a specification analysis based on this canonical representation and we analyze how alternative parametrizations affect estimated risk premia, impulse response functions and variance decompositions. We find a trade-off between the need to obtain parsimonious parametrizations and the ability of the models to match observed patterns of variation in risk premia. We also find that more richly parametrized models uncover a greater influence of macroeconomic fundamentals on the long-end of the yield curve.

1 Introduction

We derive a canonical representation for a class of affine models with both observable and unobservable variables, which includes as special cases the models of Ang and Piazzesi (2003), Ang, Dong and Piazzesi (2004), Ang, Piazzesi and Wei (2006), Hördal, Tristani and Vestin (2006) and Rudebusch and Wu (2004). Since the seminal paper of Ang and Piazzesi (2003), it has been acknowledged that identification schemes provided by Dai and Singleton (2000) for affine term structure models cannot be applied to models with both observable and unobservable variables. In an affine setting with only unobservable variables, equivalent representations of a model can be obtained by any rotation and translation of the state vector; Dai and Singleton (2000) derived a set of restrictions which permit to identify one and only one representation of the model (the canonical representation) for each class of equivalent representations. However, when the set of state variables also comprises some observables, the identifying restrictions provided by Dai and Singleton (2000) are not applicable, because equivalent representations can be obtained only by rotations and translations of the state vector which leave the observables unchanged. The canonical representation we provide takes this fact into account.

An important feature of our canonical representation is that it does not require the often-made assumption of statistical independence between observable and unobservable variables. The importance of relaxing such assumption, from both a theoretical and an empirical standpoint, has been stressed, among others, by Rudebusch, Sack and Swanson (2006) and Diebold, Rudebusch and Aruoba (2006). As far as our dataset is concerned, the statistical tests we perform on a battery of models reject the overidentifying restriction of independence both under the historical probability measure and the risk-neutral (pricing) one. However, selection criteria which tend to put a strong emphasis on parsimony indicate that overidentified models might be preferable in some applications, for example in out of sample forecasting exercises (on this point see Favero, Niu and Sala 2007). We also find that imposing less overidentifying restrictions helps to better match observed patterns of variation in risk premia, in particular some of the empirical findings reported by Fama and Bliss (1987), Campbell and Shiller (1991) and Dai and Singleton (2002): the lpy(i) and lpy(ii) tests first proposed by Dai and Singleton (2002) suggest that more richly parametrized models are able to better account for the predictability in bond returns found by running simple regressions of bond returns on the slope of the yield curve. Furthermore, we find that relaxing some overidentifying restrictions increases the explanatory power of macroeconomic fundamentals as regards the variability of the long-end of the yield curve.

We use our canonical representation to carry out a specification analysis in the spirit of that conducted by Dai and Singleton (2000). Besides evaluating the consequences of imposing the aforementioned overidentifying restrictions, we also conduct a statistical analysis to find the optimal number of unobservable factors and lags of the observable macro variables. Standard selection criteria prefer a model where the state vector includes three unobservable variables and only contemporaneous values of the macro variables. The choice of parametrization depends on the selection criterion.

The seminal paper by Ang and Piazzesi (2003) has inaugurated a proliferous literature which uses modern no-arbitrage pricing models to analyze the relation between the yield curve and macroeconomic fundamentals; some examples are: Ang, Dong and Piazzesi (2004), Ang, Piazzesi and Wei (2006), Chabi-Yo and Yang (2007), Gallmeyer, Hollifield and Zin (2005), Hördal, Tristani and Vestin (2006) and Rudebusch and Wu (2005). For a survey, we refer the reader to Diebold, Piazzesi and Rudebusch (2005). Earlier studies investigating the relation between the yield curve and macroeconomic variables, like Fama (1990), Mishkin (1990), Estrella and Mishkin (1995) and Evans and Marshall (2002) do not consider no-arbitrage relations among yields and do not model bond pricing. As a consequence, they are able to make predictions only about the yields explicitly analyzed (typically no more than three), they do not rule out theoretical inconsistencies due to the presence of arbitrage opportunities along the yield curve and they make no predictions about risk premia and their evolution over time. For these reasons, the more recent studies we mentioned above have proposed to enrich macro-finance models with rigorous asset pricing relations, imposing no-arbitrage constraints on bond prices. All these studies employ Gaussian affine term-structure models where risk premia are allowed to vary over time. Our contribution to this literature is two-fold: we enrich its theoretical foundations, by deriving the most general identified formulation of the Gaussian affine models with observable macro-factors, and we perform a thorough empirical analysis aimed at understanding the differences between alternative specifications.

The paper is organized as follows: Section 2 presents the class of affine models we are going to estimate and gives the minimal identifying conditions; Section 3 describes our dataset; Section 4 discusses the empirical evidence; Section 5 concludes.

2 The model

2.1 The baseline model

Our model of the term structure is a standard Gaussian affine model, set in discrete time, as in the majority of the recent literature about macro term structure models. The model consists of three equations. The first equation describes the dynamics of the vector of state variables X_t (a k-dimensional vector, $k \in \mathbb{N}$):

$$X_t = \mu + \rho X_{t-1} + \Sigma \varepsilon_t \tag{1}$$

where $\varepsilon_t \sim N(0, I_k)$, μ is a $k \times 1$ vector and ρ and Σ are $k \times k$ matrices. Without loss of generality, it can be assumed that Σ is lower triangular. The (historical) probability measure associated to the above specification of X_t will be denoted by P.

The second equation relates the one-period interest rate r_t to the state variables (positing that it be an affine function of the state variables):

$$r_t = a + b^{\intercal} X_t \tag{2}$$

where a is a scalar and b is a $k \times 1$ vector.

The third equation is related to bond pricing in an arbitrage-free market.

A sufficient condition for the absence of arbitrage on the bond market is that there exists a risk-neutral measure Q, equivalent to P, under which the process X_t follows the dynamics:

$$X_t = \overline{\mu} + \overline{\rho} X_{t-1} + \Sigma \eta_t \tag{3}$$

where $\eta_t \sim N(0, I_k)$ under Q and such that the price at time t of a bond paying a unitary amount of cash at time $t+n$ (denoted by p_t^n) equals:

$$p_t^n = \mathrm{E}_t^Q \left[\exp(-r_t) \, p_{t+1}^{n-1} \right] \tag{4}$$

where E_t^Q denotes expectation under the probability measure Q, conditional upon the information available at time t.

The vector $\overline{\mu}$ and the matrix $\overline{\rho}$ are in general different from μ and ρ, while equivalence of P and Q guarantees that Σ is left unchanged. The link between the risk-neutral distribution Q and the historical distribution P is given by the prices of risk, denoted by $\lambda_0 = \Sigma^{-1}(\mu - \overline{\mu})$ and $\lambda_1 = \Sigma^{-1}(\rho - \overline{\rho})$:

$$\begin{aligned} \left. \frac{dQ}{dP} \right|_t &= \xi_{t+1} / \mathrm{E}_t \left[\xi_{t+1} \right] \\ \xi_{t+1} &= \prod_{j=1}^{\infty} \exp\left[-\left(\lambda_0 + \lambda_1 X_{t+j-1} \right) \varepsilon_{t+j} \right] \end{aligned}$$

Multifactor affine models of the term structure, such as the one just described, are very popular in the finance literature and their properties have long been studied by many researchers. Thorough specification analyses of these models have been conducted (e.g. Dai and Singleton 2000) and their properties are now well-known. A distinguishing feature of these models is that they are able to describe the dynamics of yields in terms of a small set of unobservable state variables: typically three variables are deemed a sufficient number to describe the whole yield curve and this is supported also by empirical studies, such as the seminal paper by Litterman and Scheinkman (1991). Although such models are capable of describing accurately and parsimoniously the evolution of interest rates over time, the factors they identify as the driving forces of interest rates often lack economic intuition and are difficult to relate to relevant economic variables. This is one of the reasons why recent studies have proposed to augment the usual set of unobservable state variables with some observable variables. Typically, inflation and a measure of the output gap are the two observable variables, while a small number of unobservable factors, ranging from one to three, are included into the models: recent examples are Ang and Piazzesi (2003), Rudebusch and Wu (2004), Hördal, Tristani and Vestin (2006) and Ang, Piazzesi and Wei (2006). All these works impose some set of restrictions on the system of equations (1-3) and, after estimating the coefficients, derive bond prices using equation (4).

We take the same approach, adding inflation and output gap to the unobservable factors, but rather than imposing parameter restrictions derived from economic theory or from reduced-form models of the economy, we derive a set of minimal identifying restrictions and we perform a specification analysis to understand the consequences of some of the overidentifying restrictions previously imposed in the literature and to assess the advantages and disadvantages of alternative specifications.

Our minimal set of identifying restrictions is not the standard set of restrictions usually imposed for identification of affine term structure models (e.g.: Dai and Singleton 2000). Standard models of the term structure include only unobservable factors and equivalent representations of the factor dynamics can be obtained by performing any rotation and translation of the factors. On the contrary, our set of identifying restrictions takes into account the fact that in a model with both observable and unobservable factors equivalent representations can be obtained only with rotations and translations which leave the observable factors unchanged.

Suppose that the first k^o variables included in the model are observable and the remaining $k^u = k - k^o$ are unobservable. Collect their values at time t into the $k^o \times 1$ vector X_t^o and the $k^u \times 1$ vector X_t^u respectively. Equations (1-3) can be written as follows:

$$
\begin{array}{ll}
\text{Short-rate process} & \left\{ r_t = a + b^{o\top} X_t^o + b^{u\top} X_t^u \right. \\[2ex]
\text{Law of motion under } P & \left\{ \begin{array}{l} X_t^o = \mu^o + \rho^{oo} X_{t-1}^o + \rho^{ou} X_{t-1}^u + \Sigma^{oo} \varepsilon_t^o \\ X_t^u = \mu^u + \rho^{uo} X_{t-1}^o + \rho^{uu} X_{t-1}^u + \Sigma^{uo} \varepsilon_t^o + \Sigma^{uu} \varepsilon_t^u \end{array} \right. \\[2ex]
\text{Law of motion under } Q & \left\{ \begin{array}{l} X_t^o = \overline{\mu}^o + \overline{\rho}^{oo} X_{t-1}^o + \overline{\rho}^{ou} X_{t-1}^u + \Sigma^{oo} \eta_t^o \\ X_t^u = \overline{\mu}^u + \overline{\rho}^{uo} X_{t-1}^o + \overline{\rho}^{uu} X_{t-1}^u + \Sigma^{uo} \eta_t^o + \Sigma^{uu} \eta_t^u \end{array} \right.
\end{array}
\tag{5}
$$

where all the matrices are obtained by separating into blocks the matrices in equations (1-3).

The following proposition, proved in the Appendix, gives the minimal set of restrictions to be imposed in order to identify the model:

Proposition 1 *Model (5) always admits a unique equivalent representation (eventually after renaming the unobservable factors and the error terms) satisfying the following restrictions:*

- Σ^{oo} *is lower triangular*
- $\Sigma^{uo} = 0$
- $\Sigma^{uu} = I$
- $b^u \geq 0$
- $X_0^u = 0$

Further restrictions found in the literature are: $\rho^{uo} = 0$, $\rho^{ou} = 0$, ρ^{uu} is lower triangular, $\overline{\rho}^{uo} = 0$ and $\overline{\rho}^{ou} = 0$. For example, a set of restrictions equivalent to these is imposed by Ang and Piazzesi (2003) and Favero, Niu and Sala (2007). However, as clarified by Proposition 1, these further restrictions are overidentifying, i.e. not necessary to identify the model. These overidentifying restrictions, together with those in Proposition 1, imply that the observable factors are statistically independent from the unobservable factors, both under the historical and the risk-neutral measure. This is a strong assumption, as it is tantamount to saying that there are no interactions between macroeconomic variables and other factors related to the shape of the term-structure (for a discussion of this point, see Rudebusch, Sack and Swanson 2006). Instead, the minimal set of restrictions in Proposition 1 allows for a lagged response of macroeconomic variables to changes in the unobservable factors (and viceversa). As some recent studies confirm (e.g. Diebold, Rudebusch and Aruoba 2006) the hypothesis of no interactions between macroeconomic variables and other factors which determine the shape of the yield curve is challenged by formal statistical tests.

The restriction that the unobservable variables be equal to zero at time zero ($X_0^u = 0$) replaces the restriction $\mu^u = 0$ usually found in the literature. However, while the latter can be derived only assuming that the process X_t be stationary, such an assumption is not needed to derive the former. Hence, the restriction we propose is more general: for example, it allows for the possibility that the process X_t has one ore more unit roots.

Within this Gaussian framework, bond yields are affine functions of the state variables:

$$y_t^n = -\frac{1}{n}\ln\left(p_t^n\right) = A_n + B_n^\intercal X_t$$

where y_t^n is the yield at time t of a bond maturing in n periods and A_n and B_n are coefficients obeying the following simple system of Riccati equations, derived from (4):

$$\begin{aligned}
&A_1 = a \qquad\qquad (6)\\
&B_1 = b\\
&\ldots\\
&A_n = \frac{1}{n}\left[a + (n-1)\left(A_{n-1} + B_{n-1}^\intercal \overline{\mu} - \frac{n-1}{2} B_{n-1}^\intercal \Sigma\Sigma^\intercal B_{n-1}\right)\right]\\
&B_n = \frac{1}{n}\left[b + (n-1)\,\overline{\rho}^\intercal B_{n-1}\right]
\end{aligned}$$

2.2 An extension

In this subsection we extend the results of the previous section to the case where the set of state variables includes also some lags of the observable variables. Let the state variables be ordered in such a way that the vector X_t can be partitioned as follows:

$$X_t = \left[\begin{array}{ccc} X_t^{o\top} & X_t^{u\top} & X_t^{l\top} \end{array} \right]^\top$$

where X_t^o is the $k^o \times 1$ vector of observable variables, X_t^u is the $k^u \times 1$ vector of unobservable variables and X_t^l is the $k^l \times 1$ vector of lags of the observable variables. Equations (1-3) can be written as follows:

$$\begin{array}{ll}
\text{Short-rate process} & \left\{ r_t = a + b^{o\top} X_t^o + b^{u\top} X_t^u + b^{l\top} X_t^l \right. \\
\text{Law of motion under } P & \left\{ \begin{array}{l} X_t^o = \mu^o + \rho^{oo} X_{t-1}^o + \rho^{ou} X_{t-1}^u + \rho^{ol} X_{t-1}^l + \Sigma^{oo} \varepsilon_t^o \\ X_t^u = \mu^u + \rho^{uo} X_{t-1}^o + \rho^{uu} X_{t-1}^u + \rho^{ul} X_{t-1}^l + \Sigma^{uo} \varepsilon_t^o + \Sigma^{uu} \varepsilon_t^u \\ X_t^l = \rho^{lo} X_{t-1}^o + \rho^{ll} X_{t-1}^l \end{array} \right. \\
\text{Law of motion under } Q & \left\{ \begin{array}{l} X_t^o = \overline{\mu}^o + \overline{\rho}^{oo} X_{t-1}^o + \overline{\rho}^{ou} X_{t-1}^u + \overline{\rho}^{ol} X_{t-1}^l + \Sigma^{oo} \eta_t^o \\ X_t^u = \overline{\mu}^u + \overline{\rho}^{uo} X_{t-1}^o + \overline{\rho}^{uu} X_{t-1}^u + \overline{\rho}^{ul} X_{t-1}^l + \Sigma^{uo} \eta_t^o + \Sigma^{uu} \eta_t^u \\ X_t^l = \overline{\rho}^{lo} X_{t-1}^o + \overline{\rho}^{ll} X_{t-1}^l \end{array} \right.
\end{array} \qquad (7)$$

where ρ^{lo}, ρ^{ll} are two matrices whose entries are either equal to zero or to one. For example, when $X_t^l = X_{t-1}^o$, ρ^{lo} is the identity matrix and ρ^{ll} is the zero matrix. Besides, $\overline{\rho}^{lo} = \rho^{lo}$ and $\overline{\rho}^{ll} = \rho^{ll}$, since each lagged variable is defined to be equal to itself also under the risk neutral measure. Notice that the dimension of ε_t and η_t is not equal to the number of state variables, as in the baseline case, but is equal to $k^o + k^u$. Furthermore, (7) obviously imply zero restrictions on the lower blocks of μ, ρ, $\overline{\mu}$, $\overline{\rho}$.

The following proposition, proved in the Appendix, extends Proposition 1 to the case when the state variables include some lags of the observable variables:

Proposition 2 *Model (7) always admits a unique equivalent representation (eventually after renaming the unobservable factors and the error terms) satisfying the following restrictions:*

- Σ^{oo} *is lower triangular*
- $\Sigma^{uo} = 0$
- $\Sigma^{uu} = I$
- $b^u \geq 0$
- $X_0^u = 0$

Hence, the inclusion of some lags of the observable variables in the state vector does not change the identification conditions found for the baseline case.

3 The data

For our empirical analysis of the term structure we rely on a dataset of zero coupon rates extracted from US government bond yields and recorded at a quarterly frequency, provided by the Federal Reserve: the yield curve consists of ten equally spaced maturities, from 1 to 10 years. The sample goes from the first quarter of 1960 to the last of 2006 and the yields are registered on the last trading day of each quarter. We utilize all the ten maturities to carry out estimation of the models. In this respect our paper differs from most existing studies, which select only small subsets of the available maturities and typically do not employ yields of maturities longer than five years. We prefer not to exclude a priori any maturity from our sample, because we are also interested in understanding the capability of the models to fit the entire yield curve.

We include two macroeconomic variables in our model: an inflation rate and a measure of the output gap. The inflation rate is the twelve-month growth rate of the consumer price index. The output gap is HP-filtered real GDP. The empirical results we present are robust to inclusion of other measures of the output gap, for example Baxter and King (1995) bandpass filtered GDP at different frequency ranges (2-4, 3-5 and 2-8 years).

4 Empirical evidence

We use the canonical representations given in Section 2 to carry out a specification analysis, in order to find the best model, according to statistical criteria, as regards the number of unobservable variables and lags of the observables, and to analyze the consequences of overidentifying restrictions.

To simplify the exposition, we denote a model by $M(i, j, r)$, where i is the number of unobservable variables, j is the number of lags of the observable variables included in the state vector[1] and r specifies which overidentifying restrictions are imposed:

$r = U$	no overidentifying restrictions
$r = R1$	$\rho^{uo} = 0$, $\rho^{ou} = 0$, ρ^{uu} is lower triangular
$r = R2$	$\overline{\rho}^{uo} = 0$, $\overline{\rho}^{ou} = 0$
$r = R3$	$R1 + R2$

As previously explained, the above restrictions have been imposed in the literature, together with those in Proposition 1 and 2, and they imply either independence of observables and unobservables under the historical probability measure ($r = R1$), or independence under the risk-neutral probability measure ($r = R2$), or both ($r = R3$). All the models have inflation and the output gap as observable variables.

We carry out the specification analysis simultaneously along the three dimensions i, j and r, estimating a total of 64 models $M(i, j, r)$: we let i range from 1 to 4, j from 0 to 3 and estimate for each of the 16 models thus obtained both the unrestricted version and the three restricted versions. We

[1]Note that when j lags of X_t^o are included in X_t (the state vector), X_t is a function of $j+1$ lags of X_t^o; for example, when $j = 1$:

$$X_t = \left[\begin{array}{ccc} X_t^{o\top} & X_t^{u\top} & X_{t-1}^{o\top} \end{array}\right]^\top = \mu + \rho \left[\begin{array}{ccc} X_{t-1}^{o\top} & X_{t-1}^{u\top} & X_{t-2}^{o\top} \end{array}\right]^\top + \Sigma \varepsilon_t$$

also re-estimate all the models with $j > 1$, imposing the further restrictions:

$$\begin{aligned} \rho^{ul} &= 0 \\ \overline{\rho}^{ol} &= 0 \\ \overline{\rho}^{ul} &= 0 \end{aligned} \tag{8}$$

These restrictions on lagged state variables (imposed also by Ang and Piazzesi 2003) turn out to be generally not rejected by statistical tests in our sample and help to avoid computational difficulties generated by overparametrization. Furthermore, without imposing these restrictions, adding lags to the state vector causes an explosion in the number of parameters, hence selection criteria based on parameter numerosity, like AIC and BIC, tend to overwhelmingly reject specifications with $j > 1$. For these reasons, we present the results obtained after re-estimating the models with the restrictions in (8).

The models are estimated by maximum likelihood, using Chen and Scott's (1993) methodology: given a set of parameters, observed bond prices are used to infer the values of the unobservable variables. In order to do so, one has to assume that a number of bonds equal to the number of unobservable factors are exactly priced and their prices are measured without error: we choose the 1-year bond for the model with one unobservable variable ($i = 1$) and we add the 5-year, the 10-year and the 3-year when we increase the number of unobservable variables to two, three and four, respectively ($i = 2, 3, 4$).

The overidentifying restrictions $R1$, $R2$ and $R3$ are rejected by χ^2-tests at conventional levels of significance and for any choice of i and j (see Table 1). Also the AIC criterion (Table 2) always selects the unrestricted models over those with restrictions. However, the BIC criterion, which tends to penalize overparametrization more heavily, always selects $R3$ models. This suggests that, in spite of the rejection by χ^2-tests, overidentified models might be better than unrestricted ones for some applications, for example out-of-sample forecasting exercises (see, for example, Favero, Niu and Sala 2007).

Moving along the i dimension, we find that the models with three unobservable variables are unanimously selected by all criteria: hence, the classical finding that multifactor models with three unobservable factors provide the best balance between parsimony and statistical fit (e.g. Litterman and Scheinkman 1991 and Knez, Litterman and Scheinkman 1994) is not altered by the inclusion of observable state variables.

As far as the number of lags of the observable variables is concerned, the evidence is more mixed (Table 1 and 2). When $M(i, 3, r)$ is the encompassing model, the restriction $j = 2$ is in most cases not rejected by χ^2-tests. Further restrictions on the number of lags ($j = 1$ or $j = 0$) are rejected at the 5% significance level, but not at the 1% level for the $M(i, j, U)$ models, while they are rejected at both levels of significance for the overidentified models. Furthermore, the more overidentifying restrictions are imposed, the stronger is the rejection of a lesser number of lags. Also according to the AIC criterion, a model with no lags ($j = 0$) is best when there are no overidentifying restrictions ($r = U$), while a model with two or three lags is preferred in conjunction with overidentifying restrictions ($r = R1, R2, R3$). According to the BIC criterion, on the other hand, a model with no lags is preferred in any case. Models with more than three lags of the observable variables included in the state vector ($j > 3$) are rejected by both the AIC and BIC criterion. Portmanteau tests of residual autocorrelation indicate that autocorrelation in the estimates of ε_t^u is not significant, irrespective of the model chosen; instead, residual autocorrelation for ε_t^o remains significant (at the 5% level) unless one increases the number of lags of the observable variables included in the VAR equations to four ($j = 3$). Hence, although selection criteria suggest to limit the number of lags, a more complex autoregressive structure is needed to eliminate residual autocorrelation.

Overall, the AIC criterion picks $M(3,0,U)$ as the best model, while BIC selects $M(3,0,R3)$, despite the χ^2-test rejection of the restrictions in $R3$. We further investigate the properties of these two models (parameter estimates for $M(3,0,U)$ are reported in Table 3; parameter estimates for the other models are available upon request and in previous versions of the paper), in order to better understand their differences.

We compute risk-premia as model-implied expected excess holding-period returns. More precisely, when a bond maturing in n periods is bought at time t and held for $h < n$ periods, the risk premium $e_t^{n,h}$ is:

$$e_t^{n,h} = \mathrm{E}_t^P \left[\ln \left(p_{t+h}^{n-h} \right) - \ln \left(p_t^n \right) \right] + \ln \left(p_t^h \right)$$

In Figure 1 we plot estimated risk premia on the 10-year bond for a holding-period of 1 year, obtained with the $M(3,0,U)$ and the $M(3,0,R3)$ models. To interpret the differences between the estimates of the risk premia produced by the two models, we run the LPY(i) and LPY(ii) diagnostics first proposed by Dai and Singleton (2002):

- LPY(i) model-implied population coefficients from the projections of $y_{t+h}^{n-h} - y_t^n$ onto $h\left(y_t^n - y_t^h\right)/(n-h)$ are compared with the coefficients estimated from the data;
- LPY(ii) estimated coefficients from the projections of $y_{t+h}^{n-h} - y_t^n + e_t^{n,h}/(n-h)$ onto $h\left(y_t^n - y_t^h\right)/(n-h)$ ($e_t^{n,h}$ is derived from the models, everything else comes from the data) are compared with the model-implied population values of 1.

LPY(i) essentialy assesses the ability of a model to reproduce the stylized fact that the slope of the yield curve is a good predictor of holding-period returns on long-term bonds (e.g.: Fama and Bliss 1987, Campbell and Shiller 1991). LPY(ii) checks whether the risk premia produced by the model are able to explain the stylized fact.

In Figure 2 and 3 we plot the results obtained running the above diagnostics for a holding-period of one-year ($h = 4$ in our quarterly model). LPY(i) results indicate that $M(3,0,U)$ produces model-implied population coefficients that are significantly closer to those found in the data. LPY(ii) results show that the risk-adjustment produced by $M(3,0,U)$ helps to recover projection coefficients which are closer to their theoretical value of 1 than those produced by $M(3,0,R3)$. As explained by Dai and Singleton (2002), a closer match between estimated and model-implied coefficients in LPY(i) indicates a better performance in capturing the P-dynamics of yields, while closeness of estimated and theoretical coefficients in LPY(ii) reveals a good performance in reproducing the Q-dynamics.

The differences between the two models seem to be consequential also for impulse-response analysis, especially as far as longer maturities are concerned. In particular, according to $M(3, 0, R3)$, the response of the long-end of the yield curve (plots are available in previous versions of the paper) to changes in inflation and output is weaker than suggested by $M(3, 0, U)$.

As far as variance decompositions are concerned, relaxing overidentifying restrictions increases the proportion of variance of yields explained by shocks to the macroeconomic variables, especially at longer maturities. For example, according to $M(3, 0, U)$, the proportion of variance of the 10-year yield explained by macro factors is low at short time horizons (around 10 per cent), but then increases as time elapses and, already after four years, macroeconomic variables explain more than half of the variation; according to $M(3, 0, R3)$, instead, the proportion explained by macro factors remains around 10 per cent at any time horizon.

5 Concluding remarks

Ang and Piazzesi's (2003) seminal paper has inaugurated a proliferous literature which analyzes the relation between the yield curve and macroeconomic fundamentals using modern no-arbitrage pricing models. We have derived a canonical representation (i.e. the most general identified parametrization) for the class of Gaussian affine models with both observable and unobservable factors utilized in these studies. The specification analysis we have conducted, guided by the theoretical results on canonical forms of these models, provides interesting insights on the advantages and disadvantages of different modelling strategies. On the one hand, we have found that very rich parametrizations are needed to capture important empirical properties of bond returns (as highlighted by LPY(i) and LPY(ii) tests) and the autocorrelation structure of the state variables driving bond yields. On the other hand, statistical selection criteria suggest that a need for parsimonous models might well justify overidentifying restrictions previously imposed in the literature. The theoretical foundations laid in this paper, as well as the insights from the empirical analysis, might hopefully be of help for future research efforts aimed at building structural models capable of capturing the dynamics of bond risk premia in a parsimonious manner.

6 Appendix

Proof of Proposition 1. Given the process X_t defined as in (5), whose

law of motion under P is:

$$X_t = \mu + \rho X_{t-1} + \Sigma \varepsilon_t \tag{9}$$

we can obtain an equivalent representation Y_t by rotating and translating X_t in such a way that the observable variables are left unchanged:

$$Y_t := m + C X_t$$

where m is any $(k^o + k^u) \times 1$ vector whose first k^o entries are equal to zero and C is any invertible $(k^o + k^u) \times (k^o + k^u)$ matrix whose first k^o rows are the first k^o vectors of the Euclidean basis of $\mathbb{R}^{k^o+k^u}$, i.e.:

$$C = \begin{bmatrix} e_1 & \dots & e_{k^o} & v_1 & \dots & v_{k^u} \end{bmatrix}^\top$$

where:

$$\begin{aligned} e_1 &= \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \end{bmatrix}^\top \\ e_2 &= \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \end{bmatrix}^\top \\ &\dots \end{aligned}$$

are the first k^o vectors of the Euclidean basis of $\mathbb{R}^{k^o+k^u}$ and $v_1, \dots, v_{k^u}$ are k^u vectors such that C is invertible.

The equivalent representation Y_t has law of motion:

$$Y_t = \mu_* + \rho_* Y_{t-1} + \Sigma_* \varepsilon_t$$

where

$$\begin{aligned} \mu_* &= \left(I - C\rho C^{-1}\right) m + C\mu \\ \rho_* &= C\rho C^{-1} \\ \Sigma_* &= C\Sigma \end{aligned}$$

A set of restrictions on μ_*, ρ_* and Σ_* is a set of minimal identifying restrictions (and Y_t is a canonical representation of X_t), if there exists a unique couple (m, C) such that the equivalent representation Y_t satisfies the restrictions (this must be true for any initial choice of X_t).

We first prove existence. The set of restrictions on Σ_* is:

$$\begin{aligned}&\Sigma_*^{oo} \text{ is lower triangular}\\&\Sigma_*^{uo} = 0\\&\Sigma_*^{uu} = I\end{aligned}$$

Since the first k^o rows of C are the first k^o vectors of the Euclidean basis of $\mathbb{R}^{k^o+k^u}$ and Σ^{oo} is already lower triangular, the requirement that Σ_*^{oo} be lower triangular is trivially satisfied.

The restrictions $\Sigma_*^{uo} = 0$ and $\Sigma_*^{uu} = I$ are satisfied if:

$$e_{k^o+i}^\top = v_i^\top \Sigma \quad i = 1, \ldots, k^u$$

Since Σ is invertible, the restrictions are satisfied with:

$$v_i^\top = e_{k^o+i}^\top \Sigma^{-1}$$

Since the distribution of any component of ε_t does not change when you multiply it by -1, you can always change the sign of an unobservable component of Y_t leaving Σ_* unchanged, in order to satisfy the restrictions $b^u \geq 0$. The restriction $X_0^u = 0$ can be satisfied only by subtracting from the unobservable components of Y_t their respective values at $t = 0$.

Uniqueness of the equivalent representation is guaranteed by the uniqueness of Σ^{-1} and of the changes of sign which are necessary to get $b^u \geq 0$.

Finally, note that redefining the unobservable factors also affects the law of Y_t under Q, so that in general no restriction can be imposed on the Q-dynamics. ■

Proof of Proposition 2. The proof of Proposition 2 is a trivial extension of that of Proposition 1. The rotation matrix C is defined as follows:

$$C = \left[\begin{array}{cccccccccc} e_1 & \dots & e_{k^o} & v_1^\top & \dots & v_{k^u}^\top & e_{k^o+k^u+1} & \dots & e_{k^o+k^u+k^l} \end{array} \right]^\top$$

where the vectors v_i are defined exactly as in the previous proof.

It suffices to note that the rotation leaves the equation

$$X_t^l = \rho^{lo} X_{t-1}^o + \rho^{ll} X_{t-1}^l$$

unchanged in (7). The vector m is also obtained as in the previous proof, with the only difference that you must adjoin a vector of zeros of length k^l.

■

References

[1] Ang, A., S. Dong and M. Piazzesi. (2004) "No-arbitrage Taylor Rules." mimeo.

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

[3] Ang, A., M. Piazzesi and M. Wei. (2006) "What Does the Yield Curve Tell us about GDP Growth?" *Journal of Econometrics*, 131, 359-403.

[4] Baxter, M., and R. G. King. (1995) "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series." NBER Working Paper No. 5022.

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

[6] Chabi-Yo, F., and J. Yang. (2007) "A No-Arbitrage Analysis of Macroeconomic Determinants of Term Structures and the Exchange Rate." *Bank of Canada Working Papers*, 2007-21.

[7] Chen, R. R., and L. Scott. (1993) "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates." *Journal of Fixed Income*, 3, 14-31.

[8] Dai, Q., and K. J. Singleton. (2000) "Specification Analysis of Affine Term Structure Models." *Journal of Finance*, 55, 1943-78.

[9] Dai, Q., and K. J. Singleton. (2002) "Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure." *Journal of Financial Economics*, 63, 415-441.

[10] Diebold, X. F., G. D. Rudebusch and S. B. Aruoba. (2006) "The Macroeconomy and the yield curve: a Dynamic Latent Factor Approach." *Journal of Econometrics*, 127, 309-338.

[11] Diebold, X. F., M. Piazzesi and G. D. Rudebusch. (2005) ”Modelling Bond Yields in Finance and Macroeconomics.” *American Economic Review, Papers and Proceedings*, 95, 415-420.

[12] Estrella, A., and F. S. Mishkin. (1995) ”The Term Structure of Interest Rates and its Role in Monetary Policy for the European Central Bank.” NBER Working Paper No. 5279, September.

[13] Evans, C. L., and D. A. Marshall. (2002) ”Economic Determinants of the Nominal Treasury Yield Curve.” Working Paper 2001-16, Federal Reserve Bank of Chicago.

[14] Fama, E. (1990) ”Term-structure Forecasts of Interest Rates, Inflation and Real Returns.” *Journal of Monetary Economics*, 25, 59-76.

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

[16] Favero, C. A., L. Niu and L. Sala. (2007) "Term Structure Forecasting: No-Arbitrage Restrictions vs. Large Information Sets." IGIER Working Papers, 318.

[17] Gallmeyer, M., B. Hollifield and S. Zin. (2005) "Taylor Rules, McCallum Rules and the Term Structure of Interest Rates." *Journal of Monetary Economics*, 52, 921-950.

[18] Hördal, P., O. Tristani and D. Vestin. (2006) ”A Joint Econometric Model of Macroeconomic and Term Structure Dynamics.” *Journal of Econometrics*, 127, 405-444.

[19] Knez, P. K., R. Litterman and J.A. Scheinkman. (1994) "Explorations into factors explaining money market returns." *Journal of Finance*, 49, 1861-1882.

[20] Litterman, R. and J. A. Scheinkman. (1991) ”Common factors affecting bond returns.” *Journal of Fixed Income*, 1, 54-61.

[21] Mishkin, F. S. (1990) ”What Does the Term-structure Tell Us About Future Inflation?” *Journal of Monetary Economics*, 25, 76-95.

[22] Rudebusch, G. D., B. P. Sack and E. T. Swanson. (2006) "Macroeconomic implications of changes in the term premium." Working Paper Series 2006-46, Federal Reserve Bank of San Francisco.

[23] Rudebusch, G. D. and T. Wu. (2004) ”A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy.” *Proceedings of the Federal Reserve Bank of San Francisco*, March.

7 Tables

Table 1

Goodness of Fit - χ^2 tests

	Encompassing model: M(3,3,U)		
	M(3,3,R1)	M(3,3,R2)	M(3,3,R3)
χ^2	35.4	36.6	65.4
d.f.	15	12	27
p-value	0.21%	0.02%	0.00%
	Encompassing model: M(3,0,U)		
	M(3,0,R1)	M(3,0,R2)	M(3,0,R3)
χ^2	40.6	47.0	82.2
d.f.	15	12	27
p-value	0.03%	0.00%	0.00%
	Encompassing model: M(3,3,U)		
	M(3,2,U)	M(3,1,U)	M(3,0,U)
χ^2	12.0	21.8	31.4
d.f.	6	12	18
p-value	6.19%	3.98%	2.58%
	Encompassing model: M(3,3,R3)		
	M(3,2,R3)	M(3,1,R3)	M(3,0,R3)
χ^2	10.6	39.4	48.2
d.f.	6	12	18
p-value	10.16%	0.01%	0.01%

Table 2

Goodness of Fit - Selection Criteria

	M(3,3,U)	M(3,3,R1)	M(3,3,R2)	M(3,3,R3)
Log-likelihood	1374.9	1357.2	1356.6	1342.2
AIC	-2575.7	-2570.3	-2563.1	-2564.5
BIC	-2297.9	-2340.4	-2323.7	-2372.9
	M(3,2,U)	M(3,2,R1)	M(3,2,R2)	M(3,2,R3)
Log-likelihood	1368.9	1351.0	1344.5	1336.9
AIC	-2575.8	-2569.9	-2551.0	-2565.7
BIC	-2317.1	-2359.2	-2330.7	-2393.3
	M(3,1,U)	M(3,1,R1)	M(3,1,R2)	M(3,1,R3)
Log-likelihood	1364.0	1343.2	1340.1	1322.5
AIC	-2578.1	-2566.4	-2554.2	-2549.0
BIC	-2338.6	-2374.8	-2353.0	-2395.8
	M(3,0,U)	M(3,0,R1)	M(3,0,R2)	M(3,0,R3)
Log-likelihood	1359.2	1338.9	1335.7	1318.1
AIC	-2580.6	-2569.7	-2557.5	-2552.2
BIC	-2360.2	-2397.3	-2375.5	-2418.1

Table 3 - $M(3,0,U)$ model - Parameter estimates

(continued on the next page)

	a_1	a_2	a_3	a_4	a_5
	0.2500	0.4271	0.4348	0.6879	0.5050
	(0.0332)	(0.0586)	(0.0526)	(0.2675)	(0.1340)
	μ_1	μ_2	μ_3	μ_4	μ_5
	0.1955	0.1821	-0.1168	-0.1776	0.3984
	(0.0661)	(0.0705)	(0.1270)	(0.1442)	(0.1922)
	$\overline{\mu}_1$	$\overline{\mu}_2$	$\overline{\mu}_3$	$\overline{\mu}_4$	$\overline{\mu}_5$
	0.6865	-0.2511	0.0001	0.2660	0.0481
	(0.1405)	(0.1692)	(0.0559)	(0.0629)	(0.0808)
	ρ_{i1}	ρ_{i2}	ρ_{i3}	ρ_{i4}	ρ_{i5}
ρ_{1j}	0.9759	0.2011	-0.0201	-0.0038	0.0451
	(0.0043)	(0.0233)	(0.0108)	(0.0204)	(0.0243)
ρ_{2j}	-0.0467	0.8663	0.0000	-0.0319	0.0252
	(0.0167)	(0.0248)	(0.0136)	(0.0233)	(0.0318)
ρ_{3j}	0.1139	-0.0169	0.9339	0.0022	0.0003
	(0.0243)	(0.0431)	(0.0167)	(0.0418)	(0.0593)
ρ_{4j}	0.0843	0.0227	-0.0432	0.7652	0.0900
	(0.0226)	(0.0511)	(0.0213)	(0.0308)	(0.0545)
ρ_{5j}	-0.0630	0.1917	0.0390	0.0007	0.5211
	(0.0314)	(0.0430)	(0.0415)	(0.0462)	(0.0430)

	$\overline{\rho}_{i1}$	$\overline{\rho}_{i2}$	$\overline{\rho}_{i3}$	$\overline{\rho}_{i4}$	$\overline{\rho}_{i5}$
$\overline{\rho}_{1j}$	0.6763	0.7031	0.4095	0.6441	-1.2805
	(0.0450)	(0.0562)	(0.0737)	(0.0376)	(0.1500)
$\overline{\rho}_{2j}$	-0.0244	0.5147	-0.1950	-0.1551	0.9127
	(0.0286)	(0.0246)	(0.0333)	(0.0204)	(0.0844)
$\overline{\rho}_{3j}$	0.0825	0.0002	0.9682	-0.1239	0.0000
	(0.0189)	(0.0069)	(0.0272)	(0.0151)	(0.0216)
$\overline{\rho}_{4j}$	-0.0204	0.1401	0.0925	1.0152	-0.4309
	(0.0145)	(0.0096)	(0.0200)	(0.0212)	(0.0451)
$\overline{\rho}_{5j}$	0.0444	-0.0654	-0.0238	-0.1362	1.001
	(0.0171)	(0.0333)	(0.0211)	(0.0285)	(0.0386)

	Σ_{i1}	Σ_{i2}
Σ_{1j}	0.5695	0
	(0.1619)	-
Σ_{2j}	0.1572	0.7077
	(0.0410)	(0.3920)

a_0
2.0506
(0.1196)

Standard deviations of pricing errors

2y	3y	4y	6y
0.0726	0.0375	0.0360	0.0707
(0.0427)	(0.0340)	(0.0081)	(0.0210)

7y	8y	9y
0.0592	0.0845	0.0422
(0.0619)	(0.0319)	(0.0212)

Figure 1 - Expected excess holding-period returns

(10-year bond; 1-year holding period)

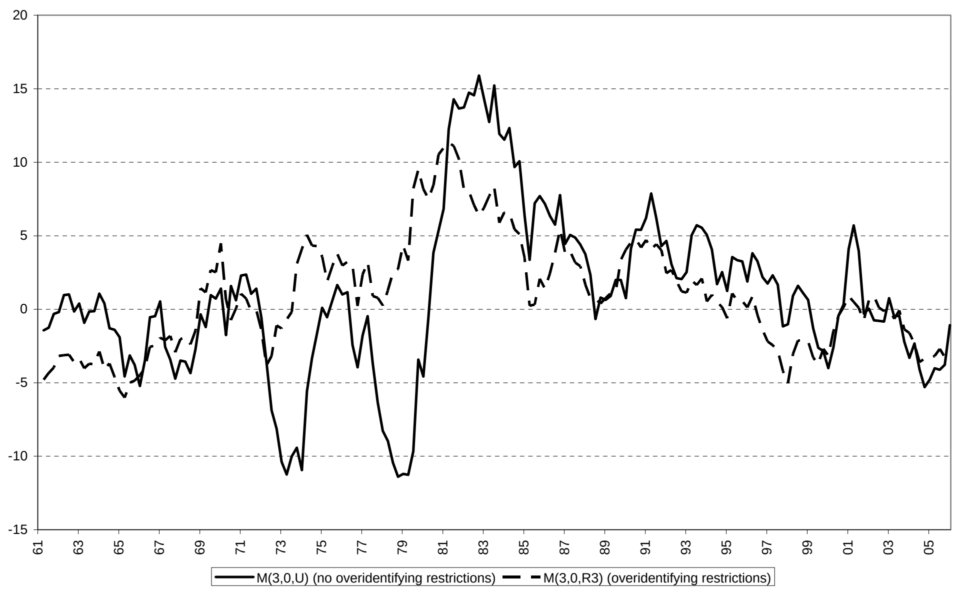

Figure 2 - LPY(i) projection coefficients

(maturities on the x-axis)

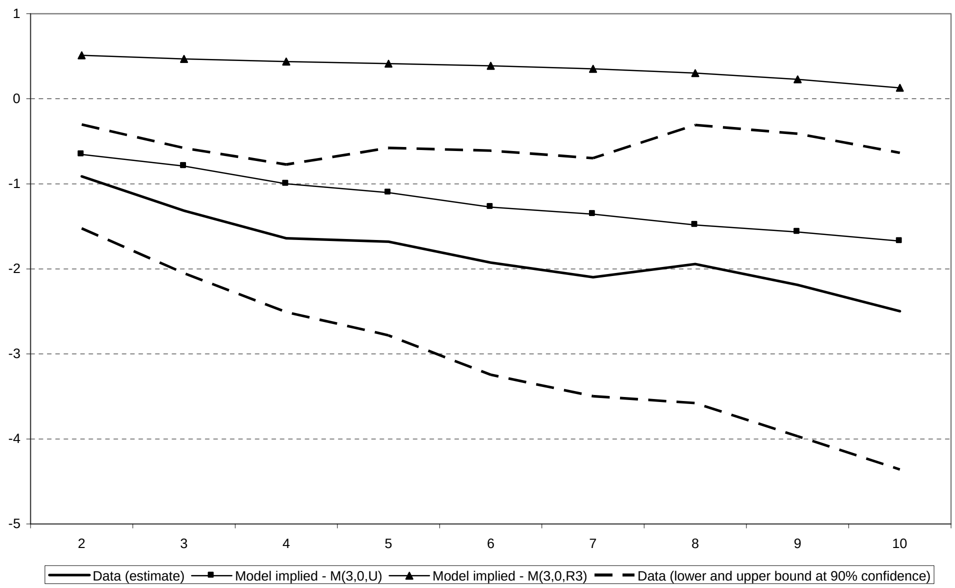

Figure 3 - LPY(ii) projection coefficients

(maturities on the x-axis)