The Role of No-Arbitrage on Forecasting: Lessons from a Parametric Term Structure Model*

Caio Almeida

Graduate School of Economics
Getulio Vargas Foundation
Praia de Botafogo 190, 11th Floor,
Botafogo, Rio de Janeiro, Brazil

Phone: 5521-2559-5828, Fax: 5521-2553-8821,

`calmeida@fgv.br`

José Vicente

Research Department
Central Bank of Brazil
Av. Presidente Vargas 730, 7th Floor,
Centro, Rio de Janeiro, Brazil

Phone: 5521-21895762, Fax: 5521-21895092,

`jose.valentim@bcb.gov.br`

October 30, 2007

1

Abstract

Parametric term structure models have been successfully applied to innumerous problems in fixed income markets, including pricing, hedging, managing risk, as well as studying monetary policy implications. On their turn, dynamic term structure models, equipped with stronger economic structure, have been mainly adopted to price derivatives and explain empirical stylized facts. In this paper, we combine flavors of those two classes of models to test if no-arbitrage affects forecasting. We construct cross section (allowing arbitrages) and arbitrage-free versions of a parametric polynomial model to analyze how well they predict out-of-sample interest rates. Based on U.S. Treasury yield data, we find that no-arbitrage restrictions significantly improve forecasts. Arbitrage-free versions achieve overall smaller biases and Root Mean Square Errors for most maturities and forecasting horizons. Furthermore, a decomposition of forecasts into forward-rates and holding return premia indicates that the superior performance of no-arbitrage versions is due to a better identification of bond risk premium.

Keywords: Dynamic term structure models, parametric functions, factor loadings, time series analysis, time-varying bond risk premia

1 Introduction

Fixed income portfolio managers, central bankers, and market participants are in a continuous search for econometric models to better capture the evolution of interest rates. As the term structure of interest rates carries out important information about monetary policy and market risk factors, those models might be seen as useful decision-orienting tools. In fact, in a quest to better understand the behavior of interest rates, a large literature on excess returns predictability and interest rates forecasting has emerged[1]. In particular, some models are not consistent inter-temporally while others impose no-arbitrage restrictions, and so far the importance of such restrictions on the forecasting context has not been established yet.

Testing the importance of no-arbitrage on interest rate forecasts should be relevant for at least two reasons. First, since imposing no-arbitrage implies stronger economic structure, testing how it will affect model ability to capture risk premium dynamics should be of direct concern to researchers. In principle, although we could expect that a more theoretically-sound model would better capture risk premiums, only careful empirical analysis might manage to answer such question. On the other hand, from a practitioner's viewpoint, testing how no-arbitrage affects forecasting will objectivelly informe managers if it is worth to implement more complex interest rate models or not. Since latent factor models with no economic restrictions usually represent a simpler alternative to be implemented, if no-arbitrage restrictions don't aggregate practical gains, they do not necessarily have to be enforced.

In this paper, we address the above mentioned points by testing how no-arbitrage restrictions affect the forecasting ability and risk premium structure of a parametric term structure model[2]. We argue that parametric models

---

[1]Fama (1984), Fama and Bliss (1987), Campbell and Shiller (1991), Dai and Singleton (2002), Duffee (2002), and Cochrane and Piazzesi (2005) analyze the failure of the expectation hypothesis and the importance of time-varying risk premia. Kargin and Onatski (2007), Bali et al. (2006), Diebold and Li (2006), and Bowsher and Meeks (2006) study different model specifications in a search for adequate forecasting candidates. Ang and Piazessi (2003), Hordahl et al. (2006), Huse (2007), Favero et al. (2007), and Mönch (2007) relate interest rates and macroeconomic variables through term structure models.

[2]In parametric term structure models, the term structure is a linear combination of predetermined parametric functions, such as polynomials, exponentials, or trigonometric functions among others. For examples, see for instance, McCulloch (1971), Vasicek and Fong (1982), Chambers et al. (1984), Nelson and Siegel (1987), and Svenson (1994), among others.

are particularly appropriate to test the effects of no-arbitrage on forecasting, since they keep a **fixed** factor-loading structure that is **independent** of the underlying factors' dynamics. This invariant loading structure implies that across different versions of the model, bond risk premia relate to a common set of underlying factors, i.e. term structure movements. Based on this fixed set of factors, it should be possible to perform a careful analysis of how each model version and no-arbitrage restrictions affect risk premium.

We parameterize the term structure of interest rates as a linear combination of Legendre polynomials. This framework supports flexible factors' dynamics, including versions that allow for arbitrage opportunities and others that are arbitrage-free. Focusing the analysis on three-factor models[3], we compare a cross section (CS) version, which allows for the existence of arbitrages, to two affine arbitrage-free versions, one Gaussian (AFG) and the other with one factor driving stochastic volatility (AFSV).

The CS polynomial version is similar to the exponential model adopted by Diebold and Li (2006) to forecast the U.S. term structure of Treasury bonds, i.e. they are both parametric models that don't rule out arbitrages. On their turn, the arbitrage-free versions of the Legendre model share many characteristics with the class of affine models proposed by Duffie and Kan (1996). No-arbitrage restrictions are imposed through the inclusion of conditionally deterministic factors of small magnitude that guarantee the existence of an equivalent martingale probability measure (Almeida 2005). Each arbitrage-free version is implemented with six latent factors: three stochastic, and three conditionally deterministic. Interestingly, by affecting the dynamics of the three basic stochastic factors ("level", "slope" and "curvature"), the conditionally deterministic factors directly affect bond risk premium structure.

More general arbitrage-free versions of the polynomial model exist and could also be analyzed[4]. However, priming for objectivity and transparency, a more concise analysis was favored, with choices of Gaussian (AFG) and Stochastic Volatility (AFSV) affine versions motivated by Dai and Singleton

---

[3]Litterman and Scheinkman (1991) show that most of the variability of the U.S. term structure of Treasury bonds can be captured by three factors: level, slope and curvature. Many subsequent more recent works have confirmed their findings. An exception is Cochrane and Piazzesi (2005) who find that a fourth latent factor improves forecasting ability.

[4]For instance, versions with more than one factor driving stochastic volatility within the affine family, or even models with a non-affine diffusion structure. For examples, see Almeida (2005).

(2002), Duffee (2002), and Tang and Xia (2007). Duffee (2002) elects the three-factor affine Gaussian model as the best (within affine) to predict U.S. bond excess returns. Dai and Singleton (2002) identify that the same Gaussian model correctly reproduces the failures of the expectation hypothesis documented by Fama and Bliss (1987) for U.S. Treasury bonds. In contrast, Tang and Xia (2007) show that a three-factor affine model with one factor driving stochastic volatility generates bond risk premium patterns compatible with data from five major fixed income markets (Canada, Japan, UK, US, and Germany). A key ingredient to all these findings is the flexible essentially affine parameterization of the market prices of risk (Duffee 2002), which we also adopt in our work.

Based on monthly U.S. zero-coupon Treasury data, we analyze the out-of-sample behavior of the three proposed versions under different forecasting horizons (1-month, 6-month, and 12-month). Forecasting results indicate that dynamic arbitrage-free versions of the model achieve overall lower bias and root mean square errors for most maturities, with stronger results holding for longer forecasting horizons. Diebold and Mariano (1995) tests confirm the statistical significance of obtained results.

In order to analyze the effects of no-arbitrage in the risk premium structure, we decompose yield forecasts into forward rates and risk premium components. The decomposition allows us to identify that the superior forecasting performance of arbitrage-free versions is primarily due to a better identification of bond risk premium dynamics. This result represents an important effort in the direction of understanding *how* no-arbitrage affects forecasting. It also indicates that further analysis with other classes of parametric models should be seriously considered.

Related works include the papers by Duffee (2002), Ang and Piazzesi (2003), Favero et al. (2007), and Christensen et al. (2007). Duffee (2002) tests the ability of affine models on forecasts of interest rates, concluding that completely affine models fail to reproduce U.S. term structure stylized facts, while essentially affine models do a better job due to a richer risk premium structure. While Duffee (2002) analyzes how different market prices of risk specifications affect forecasting in *arbitrage-free models*, we study how no-arbitrage affects forecasting, what stands for including models that allow for arbitrages in our analysis.

Ang and Piazzesi (2003) show that imposing no-arbitrage restrictions to a VAR with macroeconomic variables improves its forecasting ability. Similarly, Favero et al. (2007) test how macroeconomic variables and no-arbitrage

restrictions affect interest rate forecasting, finding that no-arbitrage models, when supplemented with macro data, are more effective in forecasting. Both papers model factor dynamics with a Gaussian VAR structure, while we include stochastic volatility in our analysis, finding it to be relevant to improve forecasting. In addition, both allow for changes in term structure loadings when comparing no-arbitrage models to models allowing for arbitrages. Those changes in factors and bond risk premiums make it harder to isolate the pure effects of no-arbitrage on forecasting. In contrast, the parametric term structure polynomial model adopted in our work avoids this issue due to its fixed factor-loading structure.

Christensen et al. (2007) obtain a Gaussian arbitrage-free version of the parametric exponential model proposed by Diebold and Li (2006). They empirically test their arbitrage-free version and identify that it offers predictive gains for moderate to long maturities and forecasting horizons. Although in this case they keep a fixed factor loading structure as we do, there are interesting differences between the two papers. First, the two papers analyze distinct parametric families, each offering interesting insights on their own. Second, the technique used to derive arbitrage-free versions is quite distinct. While we base our derivations on Filipovic's (2001) consistency work, which is not attached to the class of affine models, they make use of Duffie and Kan's (1996) arguments, which are valid only under affine models. Third, they present a Gaussian arbitrage-free version while we also include the important case where volatility is stochastic. Last, in addition to the forecasting analysis, we propose a careful analysis of the risk premium structure, which should be particularly interesting for portfolio managers and risk managers, as a complementing tool.

Our results should be important to managers and practitioners in general. They suggest it should be worth constructing arbitrage-free versions of other parametric models to test their performances as practical forecasting/hedging tools. The techniques adopted to construct arbitrage-free versions of the polynomial model can be found in Filipovic (2001), and can be readily applied to other parametric families, such as variations of Nelson and Siegel (1987), and Svenson (1994) models[5], and splines models with fixed knots, among

---

[5]Filipovic (1999) showed that there is no non-trivial arbitrage-free version of the original Nelson and Siegel (1987) model. Nevertheless, it is possible to construct arbitrage-free versions of variations of the Nelson and Siegel and Svenson (1994) models, as shown for instance, by Christensen et al. (2007). On this matter, see Sharef and Filipovic (2005) for theoretical results, and De Rossi (2004) for an implementation of a Gaussian exponential

others.

The paper is organized as follows. Section 2 introduces the polynomial model, presenting its CS and arbitrage-free versions. Section 3 explains the dataset adopted, and presents empirical results, including an interesting discussion relating bond risk premium to model forecasting ability. Section 4 offers concluding remarks and possibly extending topics. The Appendix presents details on the arbitrage-free versions of the polynomial model.

2   The Legendre Polynomial Model

Almeida et al. (1998) proposed modeling the term structure of interest rates $R(.)$ as a linear combination of Legendre polynomials[6]:

$$R(t, \tau) = \sum_{n \geq 1} Y_{t,n} P_{n-1}(\frac{2\tau}{\ell} - 1), \tag{1}$$

where τ denotes time to maturity, P_n is the Legendre polynomial of degree n and ℓ is the longest maturity in the bond market. In this model, each Legendre polynomial represents a term structure movement, providing an intuitive generalization of the principal components analysis proposed by Litterman and Scheinkman (1991). The constant polynomial is related to parallel shifts, the linear polynomial is related to changes in the slope, and the quadratic polynomial is related to changes in the curvature. Naturally, higher-order polynomials are interpreted as loadings of different types of curvatures. For illustration purposes, Figure 1 depicts the first four Legendre polynomials[7]. This model has been applied to problems involving scenario-based portfolio allocation, risk management, and hedging with non-paralell movements (see, for instance, Almeida et al. 2000, 2003).

---

arbitrage-free model.

[6]A parametric term structure model based on the power series as opposed to the Legendre polynomial basis, appeared before in Chambers et al. (1984). The advantage of Legendre polynomials is that they form an orthogonal basis, being less subject to multicolinearity problems.

[7]They are respectively $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = \frac{1}{2}(3x^2 - 1)$, and $P_3(x) = \frac{1}{2}(5x^3 - 3x)$, defined within the interval [-1,1]. The Legendre polynomials of degrees four and five, $P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$ and $P_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x)$, are also of interest, since they will be adopted to build arbitrage-free versions of the Legendre model.

On the estimation process, the number of Legendre polynomials is fixed according to some statistical criterion[8]. When considering zero-coupon yields, on each date, the model is estimated by running a linear regression of the corresponding vector of observed yields into the set of Legendre polynomials ($P_n(\cdot)$'s) previously selected. The cross section version (CS) of the model is characterized by repeatedly running this linear regression at different instants of time, to extract a time series of term structure movements $\{Y_t\}_{t=1,...,T}$. Equipped with those time series one can choose any arbitrary time-series process to fit their joint dynamics. It is important to note, however, that the time-series extraction step imposes **no inter-temporal restrictions** to term structure movements, consequently allowing for the **existence of arbitrages** within the model[9].

From an economic point of view, it would be interesting to add enough structure to our model so as to enforce absence of arbitrages. To that end, we begin by assuming the following dynamics for the stochastic factors driving term structure movements:

$$\mathrm{d}Y_t = \mu(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}W_t, \tag{2}$$

where W is a N-dimensional independent standard Brownian motion under the objective probability measure $\mathcal{P}$ and $\mu(\cdot)$ and $\sigma(\cdot)$ are progressively measurable processes with values in $\mathbb{R}^N$ and in $\mathbb{R}^{N \times N}$, respectively, such that the differential system above is well-defined.

How do we impose no-arbitrage conditions to the polynomial model? From finance theory, it suffices to guarantee the existence of a martingale measure equivalent to $\mathcal{P}$ (see Duffie 2001). More specifically, in order to rule out arbitrage opportunities, and to keep the polynomial term structure form, the following conditions (hereafter denominated AF conditions) must hold

1. The time t price of a bond with time to maturity $\tau = T - t$, $B(t,T)$, should be given by:
$$B(t,T) = \mathrm{e}^{-\tau G(\tau)'Y_t}, \tag{3}$$

---

[8]Almeida et al. (1998) suggest the use of a stepwise regression, Akaike or Bayesian information criteria.

[9]This is the same approach chosen by Diebold and Li (2006) to extract time-series of term structure movements implied by a parametric exponential model to forecast U.S. Treasury interest rates.

where $G(\tau)$ is a vector containing the first N Legendre polynomials evaluated at maturity τ:

$$G(\tau) = \left[P_0 \left(\frac{2\tau}{\ell} - 1 \right) \quad P_1 \left(\frac{2\tau}{\ell} - 1 \right) \quad \cdots \quad P_{N-1} \left(\frac{2\tau}{\ell} - 1 \right) \right]'. \quad (4)$$

2. There should exist a probability measure $\mathcal{Q}$ equivalent to $\mathcal{P}$ such that, under $\mathcal{Q}$, discounted bond prices are martingales.

The next theorem establishes restrictions (hereafter denominated AF restrictions[10]) that will provide arbitrage-free versions of the polynomial model.

Theorem 1 *Assume Y_t-dynamics under a probability measure $\mathcal{Q}$ equivalent to $\mathcal{P}$ given by:*

$$\mathrm{d}Y_t = \mu^{\mathcal{Q}}(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}W_t^*, \quad (5)$$

where W^* is a Browian motion under $\mathcal{Q}$.

If $\mu^{\mathcal{Q}}(Y_t)$ satisfies the restriction expressed in Equation 6, $\mathcal{Q}$ is an equivalent martingale measure and the AF conditions hold[11].

(6.1) $\quad \sum_{j=2}^{N}(j-1)L_j Y_{t,j}\tau^{j-2} = \sum_{j=1}^{N} L_j \mu_j^{\mathcal{Q}}(Y_t)\tau^{j-1} - \sum_{j=1}^{[\frac{N}{2}]} \sum_{k=1}^{[\frac{N}{2}]} \Gamma_{jk}(Y_t)\frac{\tau^{j+k-1}}{k}$

(6.2) $\quad \Gamma_{jk}(Y_t) = 0$ *for* $j > [\frac{N}{2}]$ *or* $k > [\frac{N}{2}]$

$$(6)$$

with $\Gamma(Y_t) = L\sigma(Y_t)\sigma(Y_t)L'$, L_j standing for the j_{th}-line of an upper triangular matrix that depends only on ℓ, and $[\cdot]$ representing the integer part of a number.

Proof and technical details are provided in the Appendix.

The AF restriction has a fundamental implication for any AF version of the Legendre polynomial model: for each stochastic term structure movement there must exist a corresponding conditionally deterministic movement whose

---

[10]The AF restriction is equivalent to imposing the Heath et al. (1992) forward rate drift restriction that ensures absence of arbitrages in the market.

[11]In addition to the drift restriction, $\sigma(Y_t)$ should present enough regularity to guarantee that discounted bond prices that are local martingales, also become martingales. In practical problems, a bounded or a square-affine $\sigma(Y_t)$ is enough to enforce the martingale condition.

drift will compensate the diffusion of the former. If we adopt, for instance, a CS version with N factors driving movements of the term structure, the corresponding arbitrage-free versions should present $2N$ latent factors in order to become stochastically compatible with CS: N stochastic factors with non-null diffusion coefficients, and N conditionally deterministic factors. Observe that although the AF restriction is enforced to the drift of the risk neutral dynamics (5), in principle, we can work with any general drift (for the first N factors) under the objective dynamics (2) by taking general market prices of risk processes. However, the restriction that imposes the existence of conditionally deterministic factors must hold under both the risk neutral and the objective measures, and this is what enforces no-arbitrage, and distinguishes AF versions from CS.

In this paper, we focus our analysis on AF versions whose dynamics belong to the class of affine models (Duffie and Kan 1996). This is implemented by restricting the diffusion coefficient of the state vector Y to be within the affine class, simplifying the SDEs for Y to[12]:

$$dY_t = \kappa^{\mathcal{Q}}(\theta - Y_t)dt + \Sigma\sqrt{S_t(Y_t)}dW_t^*, \tag{7}$$

where the matrix S_t is diagonal with elements $S_t^{ii} = \alpha_i + \beta_i'Y_t$ for some scalar α_i and some $\mathbb{R}^N$-vector β_i.

In the empirical section, we compare a three factor CS version with two AF versions that present three stochastic factors with non-null diffusions. We have seen before that this implies arbitrage-free versions with six factors (three stochastic, three conditionally deterministic). The first AF version is a Gaussian model ($\beta_i = 0, \quad \forall i$) and the second is a stochastic volatility model with only one factor driving the volatility. In the Appendix, we show in details how to translate the AF restriction to the affine framework, and further specialize the results to the Gaussian and stochastic volatility AF versions.

Following Duffee (2002) we specify the connection between risk neutral probability measure $\mathcal{Q}$ and objective probability measure $\mathcal{P}$ through an essentially affine market price of risk

$$\Lambda_t = \sqrt{S_t}\lambda_0 + \sqrt{S_t^-}\lambda_Y Y_t, \tag{8}$$

---

[12]Note that although bond prices are exponential affine functions of the state space vector Y (see (3)), in general the dynamics of Y is not restricted to be that of an affine model. For instance, if we choose $\sigma(Y)$ not to be the square root of an affine function of Y, the dynamics of Y will be non-affine.

where λ_0 is a $N \times 1$ vector, λ_Y is a $N \times N$ matrix, S_t appears in Equation 7, and S_t^- is defined by:

$$S_t^{ii-} = \begin{cases} \frac{1}{S_t^{ii}} & \text{if } \inf(\alpha_i + \beta_i^t Y_t) > 0 \\ \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The market prices of risk turn out to be of fundamental importance since the dependence of bond expected excess returns $e_{t,\tau}^i$ on term structure movements Y is what moves the model away from the Expectation Hypothesis Theory:

$$e_{t,\tau}^i = -\tau G\left(\tau\right) \Sigma \left(S_t \lambda_0 + I^- \lambda_Y Y_t\right). \tag{10}$$

Equation 10 indicates that zero coupon bond instantaneous expected excess return is a linear combination of model factors, with weights depending on matrices λ_Y, and Σ, and on a predetermined vector of maturity-dependent Legendre polynomial terms.

Finally, to estimate the parameters of the two AF versions we use a Quasi-Maximum Likelihood procedure since, within the class of affine models, both first and second conditional moments of latent factors are known in closed-form formulas (see Appendix for details).

2.1 Forecasting with the Polynomial Model

Within the sub-class of affine polynomial models with essentially affine market prices of risk, any arbitrage-free version will correspond to a continuous time vector autoregressive model of order 1 (possibly with stochastic volatility). In order to provide fair comparisons, we match the lagging structure of the time series processes describing arbitrage-free and CS versions, therefore, specializing the CS version to forecast with a VAR(1) process.

The procedure to forecast under the CS version is divided in two steps: First extract the time series $Y_t^{\mathtt{CS}}$ of term structure movements by running cross section regressions and then to fit a VAR(1) process to those series of term structure movements:

$$Y_t^{\mathtt{CS}} = c + \phi Y_{t-1}^{\mathtt{CS}} + \epsilon_t. \tag{11}$$

Given a fixed maturity τ and a fixed forecasting horizon (h-step horizon), forecasts are produced by calculating the conditional expectation of CS fac-

tors under the VAR(1) structure:

$$E_t \left(Y_{t+h}^{\mathtt{CS}} \right) = c \sum_{j=0}^{h-1} \phi^j + \phi^h Y_t^{\mathtt{CS}}. \qquad (12)$$

The conditional expectation of the τ-maturity yield is obtained by substituting factor forecasts in (1):

$$E_t \left(R(t+h, \tau) \right) = G(\tau)' E_t \left(Y_{t+h}^{\mathtt{CS}} \right). \qquad (13)$$

Similarly, for the arbitrage-free affine versions, interest rate forecasts can be produced by using the closed form structure of conditional factor means. As under the affine sub-class the drift of latent factors $Y^{\mathtt{arb.free}}$ can be written as $\mu^{\mathcal{Q}}(Y_t^{\mathtt{arb.free}}) = \kappa^{\mathcal{Q}}(\theta - Y_t^{\mathtt{arb.free}})$, the time t conditional expectation of $Y_{t+h}^{\mathtt{arb.free}}$ is given by (Duffee 2002):

$$E_t \left(Y_{t+h}^{\mathtt{arb.free}} \right) = (I_{2N} - e^{-\kappa^{\mathcal{Q}}h})\theta + e^{-\kappa^{\mathcal{Q}}h} Y_t^{\mathtt{arb.free}} \qquad (14)$$

where I_{2N} is the identity matrix of order $2N$. Finally, for any fixed maturity τ, the term structure formula in (1) should be used to forecast:

$$E_t \left(R(t+h, \tau) \right) = G(\tau)' E_t \left(Y_{t+h}^{\mathtt{arb.free}} \right) \qquad (15)$$

Under both CS and arbitrage-free versions, forecasts considering horizons longer than the sampling frequency are produced under a multi-step prediction structure, as opposed to re-estimating the models under each horizon frequency.

3 Empirical Results

3.1 Data Description

Data consists of 324 monthly observations of bootstrapped U.S. Treasury zero-coupon yields (2-, 3-, 5-, 7-, and 10-year maturities) observed from January, 1972 to December, 1998. Based on a sub-sample of 276 observations from January, 1972 to December, 1994, we estimate three distinct versions of the Legendre polynomial model: The CS version that allows for arbitrages, a Gaussian arbitrage-free version (AFG), and a stochastic volatility arbitrage-free version with one variable driving volatility (AFSV). The following subsequent four years of monthly data (from 1995 to 1998) not included in the estimation process, are used to measure models' forecasting ability, and to study their risk premium structure.

3.2 Estimation

The two AF versions were estimated using a Quasi-Maximum Likelihood procedure, explicitly exploring the fact that the conditional first and second moments of latent variables are known analytically. Adopting Chen and Scott's (1993) methodology, a subset of zero-rates (2-, 5- and 10-year maturities) was priced without errors, while the remaining rates were priced with i.i.d zero-mean errors. Parameters that identify the stochastic discount factor appear in Table 1. Σ's and β's are parameters related to volatility, λ's are related to factors' risk premia, and Y_0's define initial conditions for conditionally deterministic factors. Standard deviations from residual fits of 3- and 7-year zeros, indicate that the AFSV version presents a better in-sample cross section fitting than the AFG version (13.6 and 26.0 bps under AFG versus 9.3 and 16.0 bps under AFSV).

Figures 2 and 3 present time-series of factors capturing term structure movements, for respectively the AFG and AFSV versions. Left-hand side graphs present "level", "slope" and "curvature" factors. Right-hand side graphs depict the three conditionally deterministic factors. As yields have intrinsic stochastic behavior, it is natural to expect that conditionally deterministic factors will have their in-sample values minimized by the QML optimization procedure. Indeed, factors five and six, are practically negligible under both arbitrage-free versions. However, factor four, relating to the cubic Legendre polynomial (dashed blue line) gets up to 75 bps under the Gaussian version (in-sample), and gets up to 20 bps under the stochastic volatility version (in-sample). It doesn't vanish like the other two conditionally deterministic factors because it represents the "price" that the polynomial model has to pay in order to become arbitrage-free. The three higher order factors change the time-series of lower order movements ("level", "slope" and "curvature") in a way to guarantee no-arbitrage under each arbitrage-free version.

The small magnitude of conditionally deterministic factors explains why the three lower order movements present similar time series across different versions of the model (see Figures 2 and 3). Note that the two arbitrage-free versions present the same term structure parametric form, a linear combination of the first six Legendre polynomials, implying that any differences on the time series of the lower order movements should come from differences on the higher order conditionally deterministic factors across versions.

The CS version is a three-factor model estimated by running monthly

separate cross sectional regressions. While arbitrage-free versions were estimated under QML explicitly considering the dynamics of the six polynomial factors, the CS version, in contrast, assumes complete time-independence for factors dynamics, and is based on only the three lower order factors, "level", "slope" and "curvature", since conditionally deterministic factors are not necessary in this case, given that no-arbitrage restrictions are not imposed.

Figure 4 presents time-series of the differences between each factor in the CS version ("level", "slope" and "curvature"), and the corresponding factor on each dynamic version (AFG and AFSV). Those distances are small in magnitude and again, come predominantly from the conditionally deterministic factor due to the cubic Legendre polynomial. In fact, for each arbitrage-free version, the shape of the fourth factor time-series is carried out to Figure 4[13].

3.3   Forecast Comparisons

We proceed as in Section 2.1 to produce, for each version, forecasts based on fixed parameters estimated with the sample ranging from January 1972 to December 1994[14]. We argue that keeping fixed estimated parameters, as opposed to recursively re-estimating models out-of-sample (like performed in other studies), is an appropriate choice: With fixed estimated parameters, better out-of-sample forecasting suggests higher ability to capture the underlynd dynamics of interest rates. This choice is consistent with our goal of further analyzing the risk premium structure of the polynomial model.

Table 2 presents yield forecast biases and Root Mean Square Errors (RMSE) for the out-of-sample period, from January of 1995 to December of 1998. For each maturity and forecasting horizon h, a total of 49-h forecasts

---

[13]Favero et al. (2007) also compare time-series of term structure movements coming from models with and without no-arbitrage restrictions. They compare movements coming from a Gaussian arbitrage-free model to corresponding movements coming from the Diebold and Li (2006) model, finding that, across models, level factors are more homogenous, while slope and curvature present higher distances.

[14]In order to further check and validate our results, we performed a number of robustness tests: i) changed the number of factors in the CS version from three to six, ii) changed the in-sample estimation period to (1972-1996) and corresponding out-of-sample period to (1997-2000), iii) changed the estimation method of the CS version to invert from three bonds, similarly to the arbitrage-free versions. The two arbitrage-free versions continue to outperform the CS version, with stronger results in i), and with slightly weaker results but still statistically significant in ii) and iii). Those robustness test results are available upon request.

is produced, with h-month ahead forecasts beginning in the h_{th} month of 1995, and ending in December of 1998. Bias and RMSE are measured in basis points, and bold values indicate the lowest absolute value of bias/RMSE under a fixed maturity and forecasting horizon. We first concentrate our analysis on the bias results.

From a total of 15 entries appearing in the table (three forecasting horizons and five observed maturities), the CS version presents the lowest absolute bias in 4 of them, AFG version in 4, and AFSV in 7. In other words, in more than 70% of the entries the arbitrage-free models present significantly lower biases. Interestingly, the CS version is superior only on the shortest forecasting-horizon (1-month), indicating that no-arbitrage restrictions improve longer-horizon forecasts. A more appropriate comparison is proposed by separately comparing CS to each arbitrage-free version. In this case, the AFG version presents absolute bias lower than CS in 9 out of 15 entries, and the AFSV version presents absolute bias lower than CS in 11 out of 15 entries. In summary, from a bias perspective, no-arbitrage tremendously improves results, specially for longer forecasting horizons.

Bias results are pictured in Figure 5, where out-of-sample averaged observed and averaged model implied term structures appear. For instance, for a 1-month forecasting horizon, the solid blue line represents an average of the 48 curves that were observed between January 1995 and December of 1998. Correspondingly, the red dotted, the cyan dash-dotted, and the black dashed lines, represent the average of the 48 forecasts produced respectively by CS, AFG, and AFSV versions. The bias is simply the difference between averaged observed and model implied curves. Note how, due to the conditionally deterministic factors, arbitrage-free versions present much higher curvature than CS. This higher curvature produces two antagonistic effects: it makes arbitrage-free versions to get much closer to observed yields for most maturities, but also generates strong bias for a few cases[15].

Now observing RMSE results in Table 2, it is clear that arbitrage-free versions are again superior. When compared by pairs CS x AFSV and CS x AFG, AFSV is superior to CS in 11 out of 15 entries, and AFG is superior to CS in 9 out of 15 entries. For short-horizon forecasts, the AFSV version presents the best performance, under the RMSE criterion, among the three competitors, and for long-horizon forecasts, AFG takes its place. On its

---

[15]The AFG presents high bias at the 7-, and 10-year maturities, and the AFSV, at the 10-year maturity.

turn, CS version is only better on the 10-year maturity, where arbitrage-free versions are biased due to the conditionally deterministic factors (as mentioned above), and on the short-term forecast of the 7-year yield.

We check the statistical significance of our results by means of the Diebold and Mariano (1995) test. Under a Mean Absolute Error loss function (MAE), Table 3 compares forecasting errors produced with the arbitrage-free versions to corresponding CS forecasting errors[16]. Negative values of the statistics (S_1 or S_2) indicate that no-arbitrage improves forecasts. According to S_2, which is robust to small samples, from a total of 15 table entries, AFSV has forecasting ability superior to CS in 8 of them at a 99% confidence level (bi-caudal test) (in 9 entries at a 95% confidence level). On the other hand, in only 2 entries CS would be superior to AFSV, at both 95% or 99% confidence level. On comparisons between AFG and CS versions, results are more balanced but still in favor of no-arbitrage, with 6 entries in favor of AFG, significant at a 95% confidence level (5 entries at 99%), and 5 entries in favor of CS, at a 99% confidence level. Interestingly, against AFG, CS is strong on short-horizon forecasts and on forecasts for the 10-year yield. Against AFSV, CS is strong only on forecasts for the 10-year yield.

3.4 Discussion

3.4.1 The Effects of Bond Risk Premium in Bias.

In order to better understand the differences in forecasting ability across the three distinct versions of the polynomial model analyzed in this paper, we are interested in decomposing the conditional expectations of yields as the difference of a forward rate component and a bond risk premium component. The bond risk premium component is defined as a holding-return premium, similarly to Hordahl et al. (2006)[17].

Suppose we want to analyze model forecasting behavior for a fixed maturity of τ years, and forecasting horizon of h months, where one month is our basic time slot. The idea is to consider, at time t, the return of buying a zero-coupon bond with time to maturity $\tau + \frac{h}{12}$ and selling it h months in the future, leading to the following excess return expression with respect to

---

[16]Significance of results is not affected when we tested with a quadratic loss function.

[17]See Kim and Orphanides (2007) for a careful explanation about the term premium.

the time t short-term yield with maturity $\frac{h}{12}$, $R(t, \frac{h}{12})$:

$$BP(\tau, h) = E_t \left[\log \left(\frac{B(t + \frac{h}{12}, \tau)}{B(t, \tau + \frac{h}{12})} \right) - R \left(t, \frac{h}{12} \right) \right] \qquad (16)$$

We define this holding period return BP to be the bond premium. Now, defining the t_1-maturity forward rate, t_2 years in the future to be $f(t, t_1, t_2)$, the relation between bond premium, corresponding forward rate, and yield conditional expectation is given by:

$$E_t \left[R \left(t + \frac{h}{12}, \tau \right) \right] = f \left(t, \tau, \frac{h}{12} \right) - \left(\frac{\frac{h}{12}}{\tau} \right) BP(\tau, h) \qquad (17)$$

Equation 17 says that the h-month ahead forecast for the yield with maturity τ can be directly decomposed as the forward rate of a τ-maturity yield seen h months in the future, subtracted by a normalized risk premium (normalized by forecasting horizon over time-to-maturity).

This way, adopting Equation 17, conditional yields are decomposed in a forward rate, and a holding-return premium component. These decomposed forecasts might be useful for managers as an accessing tool to extract risk premium, since there is large interest in obtaining bond premiums from term structure data, and since they are hard to estimate (Kim and Orphanides 2007).

Tables 4, 5, and 6 respectively present out-of-sample averaged yields, averaged forward rates, and averaged bond premium. By looking at the first two tables, with a few exceptions, we note that forward rates are higher than average yields, directly indicating that models should present positive risk premium in order to compensate this difference, and to decrease bias. Interestingly, Table 6 indicates that both arbitrage-free versions indeed generate positive risk premiums, while in contrast, the CS version generates negative premiums. In other words, under a vector autoregressive structure of lag one, the version that allows arbitrages does not capture risk premium correctly[18]. For instance, the behavior of the 5-year yield under short/medium term forecasting horizons (1- and 6- month) is of particular interest to our

---

[18]It is important to say that the lack of CS ability to reproduce risk premiums can not be attributed to instability in the estimated VAR. In fact, the vector autoregressive model estimated under the CS version is stable, with all roots from the characteristic polynomial lying within the unit circle.

risk premium analysis. The short-term horizon is a good example because forward rates under the three versions of the model are close to each other (see Table 5) implying that differences in bias across versions come predominantly from differences in their implied risk premiums. For a 1-month forecasting horizon, Table 4 shows an averaged observed out-of-sample yield for the 5-year maturity equal to 5.648%[19]. From Table 5, the 1-month ahead 5-year forward rates are respectively 5.709%, 5.723%, and 5.723%, for CS, AFG, and AFSV versions, with roughly a difference of 1.5 bps between CS and arbitrage-free versions. On the other hand, from Table 6, the averaged risk premiums implied by CS, AFG, and AFSV versions are respectively -1.6, 7.8, and 6.0 bps, indicating that CS misses bond premium even when forward rates are all similar across versions, that is, when we control for differences in forward rates across versions. Similarly, considering the 6-month forecasting horizon, the 6-month ahead 5 year forward rates for the CS and AFSV versions are very similar, respectively, 5.906% and 5.895% (Table 5), but their implied risk-premiums are very distinct, respectively -18.6 and 25.1 bps (Table 6). It is clear that the forward rates coming from the two versions are overestimating future 5-year yields, but while the positive risk premium implied by the AFSV version corrects this overestimation, the negative risk premium implied by the CS version worsens.

3.4.2 What is the Contribution of No-arbitrage?

Why imposing no-arbitrage leads to better forecasts? The mechanics of the problem can be directly explained by the conditionally deterministic factors. Once they are included in the term structure parameterization, they change the original time series of "level", "slope" and "curvature" factors, consequently affecting the behavior of bond risk premium.

Further appreciation of the no-arbitrage effect on risk premium can be obtained from Table 7. It presents, for each model version, the ratio of the bias generated by assuming a zero bond risk premium (no model implied risk-premium effect), over the true bias generated when model implied bond risk premium is fully incorporated. Whenever risk premium has a positive effect on forecasting, we should immediately observe values higher than 1 for this ratio. For values lower than 1, the model is not correctly capturing

---

[19]The average of observed yields is depending on the forecasting horizon because the horizon defines the beginning of the averaging window. See the description of Table 4 for further explanations.

the risk premium dynamics. It is particularly interesting to observe that CS presents values lower than 1 in all table entries, indeed confirming that it is not correctly capturing risk premium dynamics. In sparkling contrast, arbitrage-free versions not only present (for most table entries) values higher than 1, but in addition, some entries have values much higher than 1[20], indicating that no-arbitrage tremendously increase model ability to correctly capture risk premium dynamics.

A dynamic picture of the risk premium effect described on the paragraph above can be readily observed in Figure 6. For a fixed 12-month forecasting horizon, it presents time-series of observed out-of-sample 2-year yields, with corresponding forward rates, and model implied bond risk premiums[21]. On each graph, the dotted line represents observed yields, the dashed line represents the 12-month ahead 2-year forward rate, and the solid line represents the risk premium corrected forward rate, that is, the yield forecast produced with 17. Once risk premium is included, it clearly improves forecasts under the two arbitrage-free versions: the solid line is much closer to the dotted line than the dashed line is. However, under the CS version, risk premium degrades its performance. The dashed line (the one with zero-premium) is much closer to the true observed yield than the solid line (the one including risk premium).

Figure 7 presents examples of risk premium dynamics along the 27 years, from 1972 to 1998, for different maturities and forecasting horizons. The goal of this picture is to show similarities and differences among risk premiums implied by each model version, both in- and out-of-sample. It presents the 1-month holding period return premium for the 5-year bond, the 6-month premium for the 10-year bond, and the 12-month premium for the 2-year bond. Those three maturities give pretty much an idea of the risk premium behavior across the U.S. Treasury term structure for maturities up to 10-years. For the three forecasting horizons, the less volatile premium comes from the AFSV arbitrage-free version. Despite presenting a smaller volatility, it has a very strong effect on improving forecasts as previously observed

---

[20]Under the AFG version, 7 ratio values are higher than 3, and under the AFSV version, 6 ratio values are higher than 3. A ratio value higher than 3 indicates that once model implied risk premium is considered in forecasting (and not only forward rates), bias decreases for less than one third of the bias value with zero-premium.

[21]The choice of a 12-month forecasting horizon is justified by our interest in making explicit the role of risk premium, since its importance is an increasing function of forecasting horizon.

in Table 7. Risk premiums coming from the other two versions (CS and AFG) have more similar in-sample behavior, but clearly get apart out-of-sample, with the AFG version generating positive premiums, and the CS version generating negative ones. This out-of-sample separation of premiums indicates that while CS might be doing a good job when fitting in-sample data, it is probably overfitting data and missing the true dynamics of yields.

The second picture in Figure 7 presents the premium behavior of the 10-year yield under a 6-month forecasting horizon. We have intentionally included this particular maturity to show that even the best arbitrage-free version of the polynomial model (analyzed in this paper) can not capture all features of data, ending up missing the risk premium for this particular maturity. Observe that in the out-of-sample period the AFSV premium converges to approximately the same negative values produced by the CS version, when both should be producing positive premiums. This is a first indication that the polynomial family, at least under its affine subclass, might not be the best candidate to simultaneously describe the behavior of **the whole** cross section of yields, and to guarantee inter-temporal consistency of the underlying term structure factors.

The third picture in Figure 7 presents the dynamic premium behavior of the 2-year yield under a 12-month forecasting horizon. Note how the out-of-sample behavior of the premium implied under the three versions is tremendously different, with the AFG premium highly positive, AFSV premium slightly positive, and CS premium highly negative. This distinct dynamic behavior translates into rather different implications for bias. For instance, the AFG excellent performance when forecasting the 2-year yield 12-months in future (-2.8 bps of bias) can be explained by it risk premium out-of-sample behavior. Picture 2 in Figure 6 indicates that its forward rates are exaggerated with respect to realized yields. However, its out-of-sample risk premium is positive and high, thus compensating those exaggerated forward rates, and bringing forecasts to values close to observed yields. On the other hand, AFSV version presents a positive bias of 35.2 basis points, indicating that it should have produced higher risk premium values to decrease bias. CS version clearly misses the premium as it should have been positive (see picture 3 in Figure 6), while it is negative during the whole out-of-sample period.

Finally, rather than looking for the best forecasting candidate, our specific interest was to identify if no-arbitrage improves or degrades the forecasting ability of a given parametric term structure model. However, with the in-

tention of putting the polynomial model among credible benchmarks, we present in Table 8, bias and Root Mean Square Errors coming from the best polynomial version AFSV, the established Random Walk (RW) benchmark, and the recently proposed Diebold and Li (2006) model (DL). Forecasting horizons (1-,6-, and 12- month) and maturities (2-, 3-, 5-, 7-, and 10-year) are the same as presented in previous tables. The polynomial model achieves smaller bias and RMSE in 9 out of 15 entries, and, interestingly 7 among those 9 entries are related to longer forecasting horizons (6- and 12-month).

4   Conclusion

We tested the effect of no-arbitrage restrictions on out-of-sample interest rate forecasts. This was implemented with the use of a parametric term structure model that expresses the term structure of interest rates as a linear combination of polynomials. We test this family by comparing forecasts of a model version which admits arbitrages, to two different arbitrage-free versions of the same model, concluding that absence of arbitrage decreases bias and RMSE, specially for longer forecasting horizons.

An important feature of performing this no-arbitrage effect test with a parametric family that presents closed-form formula for bond prices, is that it allows us to isolate the effects of no-arbitrage from other effects like changes in factor loadings under different model dynamic specifications. Fixed factor loadings not only put the forecasting comparison on a fixed basis, but also allow for a similar interpretation of bond risk premia across different model versions. By looking at model implied risk premia, we find that the different versions generate very distinct bond risk premium behavior, whose effect can be directly observed in the out-of-sample forecasting biases. The risk premium implied by arbitrage-free versions improves forward rates forecasting ability while the corresponding premium implied by the cross section version degrades forecasting ability.

Note that rather than proposing an isolated test of no-arbitrage effects on forecasting, the test is conditional to the Legendre polynomial term structure model. However, if something can be attributed to the particular polynomial structure, is that it is biased against no-arbitrage. This bias can be directly observed in Figures 2, 3, and 5, which show that for 7-, and 10-year maturities under the AFG, and 10-year maturity under the AFSV, out-of-sample forecasts are biased due to an explosion of the conditionally deterministic

factors, out-of-sample[22]. With this observation in mind, we could conjecture that under more flexible parametric families, the no-arbitrage restrictions might generate even more positive effects on forecasting. This way, it appears to be room for further evaluation of important parametric families such as the classical polynomial-exponential family whose models by Nelson and Siegel (1987), Diebold and Li (2006), and Svenson (1994) belong to, and also analysis of more complex families like "splines with fixed knots" (see Bowsher and Meeks 2006)[23]. Moreover, as the techniques used to generate arbitrage-free versions of parametric models readily allow for inclusion of extra variables in factor dynamics, tests including macroeconomic variables could possibly better identify bond risk premium behavior (see Ludvigson and Ng 2007). We leave those topics for future research.

---

[22]The explosion of these conditionally deterministic factors is exacerbated by the parametric polynomial structure of the yield curve. A test where all conditionally deterministic factors are kept at a constant value (their last in-sample value) during the whole out-of-sample period, considerably improves forecasts under both versions, at those "bad" maturities, while keeping the previous good results at other maturities. The results of this test are available upon request.

[23]Equipped with Filipovic's (2001) theoretical results on consistent term structure models, our tests can be readily extended to other parametric families, as long as they support at least one arbitrage-free version for the term structure model.

References

[1] Almeida C.I.R (2005). Affine Processes, Arbitrage-Free Term Structures of Legendre Polynomials, and Option Pricing, *International Journal of Theoretical and Applied Finance*, **8**, 2, 1-23.

[2] Almeida C.I.R., A.M. Duarte, and C.A.C. Fernandes (1998). Decomposing and Simulating the Movements of Term Structures of Interest Rates in Emerging Eurobonds Markets. *Journal of Fixed Income*, **8**, 1, 21-31.

[3] Almeida C.I.R., A.M. Duarte, and C.A.C. Fernandes (2000). Credit Spread Arbitrages in Emerging Eurobonds Markets. *Journal of Fixed Income*, **10**, 3, 100-111.

[4] Almeida C.I.R , A.M. Duarte, and C.A.C. Fernandes (2003) A Generalization of Principa Component Analysis for Non-observable Term Structures in Emerging Markets, *International Journal of Theoretical and Applied Finance*, **6**, 8, 885-903.

[5] 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.

[6] Bali T., M. Heidari, and L. Wu (2006). Predictability of Interest Rates and Interest Rate Portfolios. Working Paper, Baruch College.

[7] Bowsher C. and R. Meeks (2006). High Dimensional Yield Curves: Models and Forecasting. Working Paper, Nuffield College, University of Oxford.

[8] Campbell, J. Y. and R. Shiller (1991). Yield Spreads and Interest Rate Movements: A Birds Eye View. *Review of Economic Studies*, **58**, 495-514.

[9] Chambers D.R., W.T. Carleton, and D.W. Waldman (1984). A New Approach to Estimation of the Term Structure of Interest Rates. *Journal of Financial and Quantitative Analysis*, **19**, 3, 233-251.

[10] Cochrane J. and M. Piazzesi (2005). Bond Risk Premia. *American Economic Review*, **95**, 1, 138-160.

[11] 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.

[12] Christensen J.H.E., F. Diebold, and G.D. Rudebusch (2007). The Affine Arbitrage-free Class of the Nelson-Siegel Term Structure Models. Working Paper, Federal Reserve Bank of San Francisco.

[13] Dai Q. and K. Singleton (2000). Specification Analysis of Affine Term Structure Models. *Journal of Finance*, **LV**, 5, 1943-1977.

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

[15] De Rossi G. (2004). Kalman Filtering of Consistent Forward Rate Curves: A Tool to Estimate and Model Dynamically the Term Structure. *Journal of Empirical Finance*, **11**, 277-308.

[16] Diebold F.X. and C. Li (2006). Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics*, 130, 337-364.

[17] Diebold F.X. and R.S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, **13**, 253-263.

[18] Duffee G. R. (2002). Term Premia and Interest Rates Forecasts in Affine Models. *Journal of Finance*, **57**, 405-443.

[19] Duffie D. (2001). *Dynamic Asset Pricing Theory*. Princeton University Press.

[20] Duffie D. and Kan R. (1996). A Yield Factor Model of Interest Rates. *Mathematical Finance*, **6**, 4, 379-406.

[21] Fama E.F. (1984). The Information in the Term Structure of Interest Rates. *Journal of Financial Economics*, **13**, 2, 509-528.

[22] Fama, E. F. and Bliss R.R. (1987). The information in long-maturity forward rates. *American Economic Review*, **77**, 4, 680-692.

[23] Favero A.C., L. Niu, and L. Sala (2007). Term Structure Forecasting: No-Arbitrage Restrictions vs. Large Information Set. Working Paper, Bocconi University.

[24] Filipovic D. (1999). A Note on the Nelson and Siegel Family. *Mathematical Finance*, **9**, 4, 349-359.

[25] Filipovic D. (2001). Consistency Problems for Heath-Jarrow-Morton Interest Rate Models. *Lecture Notes in Mathematics*, **1760**, Springer-Verlag, Berlin.

[26] Heath D., R. Jarrow and A. Morton (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica*, **60**, 1, 77-105.

[27] Hordahl P., O. Tristani, and D. Vestin (2006). A Joint Econometric Model of Macroeconomic and Term Structure Dynamics. *Journal of Econometrics*, **131**, 405-440.

[28] Huse C. (2007). Term Structure Modelling with Observable State Variables. Working Paper, London School of Economics.

[29] Kargin V. and A. Onatski (2007). Curve Forecasting by Functional Autoregression. Working Paper, Columbia University.

[30] Kim D. and A. Orphanides (2007). The Bond Market Term Premium: What is it, and How can we Measure it?, *BIS Quarterly Review*, June.

[31] Litterman R. and Scheinkman J.A. (1991). Common Factors Affecting Bond Returns. *Journal of Fixed Income*, **1**, 54-61.

[32] Ludvigson S. and S. Ng (2007). Macro Factors in Bond Risk Premia. Working Paper, Department of Economics, New York University.

[33] McCulloch J.H. (1971). Measuring the Term Structure of interest Rates. *Journal of Business*, 44, 19-31.

[34] Mönch E. (2007). Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach. Working Paper, Humboldt Universitt zu Berlin.

[35] Nelson C. and A. Siegel (1987). Parsimonious Modeling of Yield Curves. *Journal of Business*, 60, 4, 473-489.

[36] Sharef E. and D. Filipovic (2004). Conditions for Consistent Exponential-Polynomial Forward Rate Processes with Multiple Nontrivial Factors. *International Journal of Theoretical and Applied Finance*, **7**, 685-700.

[37] Svensson L. (1994). Monetary Policy with Flexible Exchange Rates and Forward Interest Rates as Indicators. *Institute for International Economic Studies*, Stockholm University.

[38] Tang H. and Y. Xia (2007). An International Examination of Affine Term Structure Models and the Expectations Hypothesis. *Journal of Financial and Quantitative Analysis*, **42**, 1, 41-80.

[39] Vasicek O. and G. Fong (1982). Term Structure Modelling Using Exponential Splines. *Journal of Finance*, 37, 2, 339-48.

5 Appendix

5.1 Proof of Theorem 1.

Theorem 1.
Assume Y_t-dynamics under a probability measure $\mathcal{Q}$ equivalent to $\mathcal{P}$ given by:

$$dY_t = \mu^{\mathcal{Q}}(Y_t)dt + \sigma(Y_t)dW_t^*, \tag{18}$$

where W^* is a Browian motion under $\mathcal{Q}$.

If $\mu^{\mathcal{Q}}(Y_t)$ satisfies the restriction expressed in Equation (19), $\mathcal{Q}$ is an equivalent martingale measure and the AF conditions hold[24].

$$\sum_{j=2}^{N}(j-1)L_j Y_{t,j}\tau^{j-2} = \sum_{j=1}^{N} L_j \mu_j^{\mathcal{Q}}(Y_t)\tau^{j-1} - \sum_{j=1}^{[\frac{N}{2}]}\sum_{k=1}^{[\frac{N}{2}]}\Gamma_{jk}(Y_t)\frac{\tau^{j+k-1}}{k}$$

$$\Gamma_{jk}(Y_t) = 0 \text{ for } j > [\tfrac{N}{2}] \text{ or } k > [\tfrac{N}{2}]$$

$$\tag{19}$$

with $\Gamma(Y_t) = L\sigma(Y_t)\sigma(Y_t)L'$, L_j standing for the j_{th}-line of an upper triangular matrix that depends only on ℓ, and $[\cdot]$ representing the integer part of a number.

Proof of Theorem 1.
The term structure of the Legendre polynomial model is given by:

$$R(\tau, Y_t) = G(\tau)'Y_t = \sum_{n=1}^{N} Y_{t,n}P_{n-1}(\frac{2\tau}{\ell} - 1), \tag{20}$$

that is, the loadings of the term structure are Legendre polynomials. Therefore, the τ-maturity instantaneous forward rate is

$$f(\tau, Y_t) = \sum_{n=1}^{N} Y_{t,n}P_{n-1}(\frac{2\tau}{\ell} - 1) + \tau\left(\sum_{n=1}^{N} Y_{t,n}\frac{\partial P_{n-1}(\frac{2\tau}{\ell} - 1)}{\partial \tau}\right). \tag{21}$$

In the equation above, the forward rates are expressed as linear combinations of Legendre polynomials, which can be readily expressed as linear combina-

---

[24]In addition to the drift restriction, $\sigma(Y_t)$ should present enough regularity to guarantee that discounted bond prices that are local martingales, also become martingales. In practical problems, a bounded or a square-affine $\sigma(Y_t)$ is enough to enforce the martingale condition.

tions of powers of τ:

$$f(\tau, Y_t) = \sum_{n=1}^{N} L_n Y_t \tau^{n-1}, \tag{22}$$

where L_n is the nth row of the upper triangular matrix L. In fact, (22) defines matrix L. If $N = 6$ the matrix L is[25]

$$L = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ 0 & \frac{4}{\ell} & -\frac{12}{\ell} & \frac{24}{\ell} & -\frac{40}{\ell} & \frac{60}{\ell} \\ 0 & 0 & \frac{18}{\ell^2} & -\frac{90}{\ell^2} & \frac{270}{\ell^2} & -\frac{630}{\ell^2} \\ 0 & 0 & 0 & \frac{80}{\ell^3} & -\frac{560}{\ell^3} & \frac{2450}{\ell^3} \\ 0 & 0 & 0 & 0 & \frac{350}{\ell^4} & -\frac{3150}{\ell^4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1512}{\ell^5} \end{bmatrix}. \tag{24}$$

Proposition 3.2 of Filipovic (1999) presents conditions on $f(\tau, Y_t)$, which guarantee that discounted bond prices are martingales under any specific interest rate model[26]. Using these conditions, Almeida (2005) proves that if the AF restrictions (19) hold, then the Legendre polynomial model is arbitrage-free. $\diamond$

---

[25]Using the first six Legendre polynomials we have

$$f(\tau, Y_t) =$$
$$Y_{t,1} + Y_{t,2}x + \tfrac{Y_{t,3}}{2}(3x^2 - 1) + \tfrac{Y_{t,4}}{2}(5x^3 - 3x) +$$
$$\tfrac{Y_{t,5}}{8}(35x^4 - 30x^2 + 3) + \tfrac{Y_{t,6}}{8}(63x^5 - 70x^3 + 15x) +$$
$$\tfrac{2\tau}{\ell}\left[Y_{t,2} + 3Y_{t,3}x + \tfrac{Y_{t,4}}{2}(15x^2 - 3)\right] +$$
$$\tfrac{2\tau}{\ell}\left[\tfrac{Y_{t,5}}{8}(140x^3 - 60x) + \tfrac{Y_{t,6}}{8}(315x^4 - 210x^2 + 15)\right]. \tag{23}$$

where $x = \tfrac{2\tau}{\ell} - 1$. Collecting terms that are powers of τ in the expression above we obtain the upper triangular matrix L for $N = 6$.

[26]Basically, Proposition 3.2 of Filipovic (1999) imposes a specific relationship between the partial derivatives of $f(\tau, Y_t)$.

5.2 Technical Details about the Sub-Class of Arbitrage-Free Legendre Models with Affine Dynamics.

The affine class of dynamic term structure models is composed by processes whose state vector Y is an affine diffusion[27], and whose implied short term rate is affine in Y. Dai and Singleton (2000) proposed the following notation to describe the dynamics of canonical affine models under the risk neutral measure $\mathcal{Q}$:

$$dY_t = \mu^{\mathcal{Q}}(Y_t)dt + \Sigma\sqrt{S_t(Y_t)}dW_t^* = \kappa^{\mathcal{Q}}(\theta^{\mathcal{Q}} - Y_t)dt + \Sigma\sqrt{S_t(Y_t)}dW_t^* \quad (25)$$

where $\kappa^{\mathcal{Q}}$ and Σ are $N \times N$ matrices, $\theta^{\mathcal{Q}}$ is a $\mathbb{R}^N$-vector, and S_t is diagonal matrix with elements $S_t^{ii} = \alpha_i + \beta_i'Y_t$ for some scalar α_i and some $\mathbb{R}^N$-vector β_i.

Now suppose we want to equip the affine class of models with a loadings structure composed by Legendre polynomials[28]. To this end, we have to impose the AF restrictions of Theorem 1.

Consider the auxiliary state space vector $\tilde{Y}_t$ defined by

$$\tilde{Y}_t = LY_t, \quad (26)$$

where L is the upper triangular matrix of Theorem 1. This auxiliary process characterizes term structure movements when the loadings come from a power series in the maturity variable τ. It appears as an intermediate step in calculations.

The dynamics of $\tilde{Y}_t$ under probability measure $\mathcal{Q}$ is given by

$$d\tilde{Y}_t = \tilde{\mu}^{\mathcal{Q}}(\tilde{Y}_t)dt + \tilde{\Sigma}\sqrt{\tilde{S}_t(\tilde{Y}_t)}dW_t^*, \quad (27)$$

where the parameters of this stochastic differential equations system are defined in similar way to (25) (i.e., $\tilde{S}_t^{ii} = \tilde{\alpha}_i + \tilde{\beta}_i'\tilde{Y}_t$ for some scalar $\tilde{\alpha}_i$ and some $\mathbb{R}^N$-vector $\tilde{\beta}_i$ and so on) and are related through (26) with the corresponding parameters in (25). It should be clear that $\tilde{Y}_t$ is affine if, and only if, Y_t is affine, because L is invertible.

---

[27]This means that the drift and the squared diffusion terms of Y are affine functions of Y.

[28]Note that it is not possible to make use of Duffie and Kan (1996) separation arguments that lead to their pair of Ricatti equations since the Legendre polynomials do not satisfy one of the algebraic conditions stated in their main theorem.

Under this particular sub-class (affine plus polynomial loadings), the first requirement of AF restrictions becomes

$$\sum_{j=2}^{N}(j-1)\tilde{Y}_{t,j}\tau^{j-2} = \sum_{j=1}^{N}\mu_j^{\mathcal{Q}}(\tilde{Y}_t)\tau^{j-1} - \sum_{j=1}^{[\frac{N}{2}]}\sum_{k=1}^{[\frac{N}{2}]}(\tilde{H}_{0,jk} + \tilde{H}_{1,jk}\tilde{Y}_t)\frac{\tau^{j+k-1}}{k}, \quad (28)$$

where $\left(\tilde{\Sigma}\tilde{S}_t\tilde{\Sigma}'\right)_{ij} = \tilde{H}_{0ij} + \tilde{H}_{1ij}\tilde{Y}_t$, with $\tilde{H}_{0ij} \in \mathbb{R}$ and $\tilde{H}_{1ij} \in \mathbb{R}^N$.

In particular, by matching coefficients on the maturity variable τ in (28), we obtain an explicit expression for the drift of the auxiliary process:

$$\tilde{\mu}^{\mathcal{Q}}\left(\tilde{Y}_t\right)_i = i\tilde{Y}_{t,i+1} + \sum_{j=\text{Max}\left\{1,i-\left[\frac{N}{2}\right]\right\}}^{\text{Min}\left\{i-1,\left[\frac{N}{2}\right]\right\}} \frac{\tilde{H}_{0,j(i-j)} + \tilde{H}_{1,j(i-j)}\tilde{Y}_t}{i-j}. \quad (29)$$

This expression can be readily translated to a similar expression for the drift of the original state vector Y with the use of (26).

In the empirical section of our paper, we compare a three factor CS version with corresponding AF versions that present three stochastic factors with non-null diffusions. By Theorem 1, a natural way to implement this application, is to work with AF versions driven by six factors (three stochastic, three conditionally deterministic). In the next lines, we provide the restrictions that should be implemented to generate affine models with *polynomial* loadings, and how to translate those restrictions to generate affine models with *Legendre polynomial* loadings. After that, we explain in details the two AF versions chosen to be implemented in this work: the Arbitrage-Free Gaussian (AFG) version, in which the volatility of Y is deterministic and time independent, and the Arbitrage-Free Stochastic Volatility (AFSV) version, in which only one stochastic factor determines the volatility of Y.

When $N = 6$ the dynamics of $\tilde{Y}_t$ has the following form:

$$\tilde{S}_t^{ii}(\tilde{Y}_t) = \begin{cases} \tilde{\alpha}_i + \tilde{\beta}_i' \tilde{Y}_t & \text{if} \quad i \le 3 \\ 0 & \text{if} \quad i > 3, \end{cases}$$

$$\tilde{\Sigma}_{i,j} = 0 \quad i, j > 3,$$

$$\tilde{\mu}^{\mathcal{Q}}(\tilde{Y}_t)_1 = \tilde{Y}_{t,2},$$

$$\tilde{\mu}^{\mathcal{Q}}(\tilde{Y}_t)_2 = 2\tilde{Y}_{t,3} + \tilde{H}_{0,11} + \tilde{H}_{1,11}\tilde{Y}_t,$$

$$\tilde{\mu}^{\mathcal{Q}}(\tilde{Y}_t)_3 = 3\tilde{Y}_{t,4} + \frac{\tilde{H}_{0,12}}{2} + \frac{\tilde{H}_{1,12}}{2}\tilde{Y}_t + \tilde{H}_{0,21} + \tilde{H}_{1,21}\tilde{Y}_t,$$

$$\tilde{\mu}^{\mathcal{Q}}(\tilde{Y}_t)_4 = 4\tilde{Y}_{t,5} + \frac{\tilde{H}_{0,13}}{3} + \frac{\tilde{H}_{1,13}}{3}\tilde{Y}_t + \frac{\tilde{H}_{0,22}}{2} + \frac{\tilde{H}_{1,22}}{2}\tilde{Y}_t + \tilde{H}_{0,31} + \tilde{H}_{1,31}\tilde{Y}_t,$$

$$\tilde{\mu}^{\mathcal{Q}}(\tilde{Y}_t)_5 = 5\tilde{Y}_{t,6} + \frac{\tilde{H}_{0,23}}{3} + \frac{\tilde{H}_{1,23}}{3}\tilde{Y}_t + \frac{\tilde{H}_{0,32}}{2} + \frac{\tilde{H}_{1,32}}{2}\tilde{Y}_t,$$

$$\tilde{\mu}^{\mathcal{Q}}(\tilde{Y}_t)_6 = \frac{\tilde{H}_{0,33}}{3} + \frac{\tilde{H}_{1,33}}{3}\tilde{Y}_t.$$

$$(30)$$

The dynamics of the term structure movements Y under the original Legendre polynomial parameterization can then be obtained by solving (26). To that end, let us rewrite the drift $\tilde{\mu}^{\mathcal{Q}}$ in matrix notation, as an affine transformation of $\tilde{Y}$:

$$\tilde{\mu}^{\mathcal{Q}}(\tilde{Y}_t) = M + U\tilde{Y}_t, \tag{31}$$

where $U = U_1 + U_2$, and U_1, U_2 and M are given by:

$$
M = \begin{bmatrix} 0 \\[1em] \tilde{H}_{0,11} \\[1em] \frac{\tilde{H}_{0,12}}{2} + \tilde{H}_{0,21} \\[1em] \frac{\tilde{H}_{0,13}}{3} + \frac{\tilde{H}_{0,22}}{2} + \tilde{H}_{0,31} \\[1em] \frac{\tilde{H}_{0,23}}{3} + \frac{\tilde{H}_{0,32}}{2} \\[1em] \frac{\tilde{H}_{0,33}}{3} \end{bmatrix}, \tag{32}
$$

$$
U_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{33}
$$

$$
U_2 = \begin{bmatrix} 0_{1\times6} \\[1em] \tilde{H}_{1,11} \\[1em] \frac{\tilde{H}_{1,12}}{2} + \tilde{H}_{1,21} \\[1em] \frac{\tilde{H}_{1,13}}{3} + \frac{\tilde{H}_{1,22}}{2} + \tilde{H}_{1,31} \\[1em] \frac{\tilde{H}_{1,23}}{3} + \frac{\tilde{H}_{1,32}}{2} \\[1em] \frac{\tilde{H}_{1,33}}{3} \end{bmatrix}. \tag{34}
$$

Finally, the drift and diffusion of process Y are given by:

$$
\mu^{\mathcal{Q}}(Y_t) = L^{-1}\tilde{\mu}^{\mathcal{Q}}(\tilde{Y}_t) = L^{-1}\tilde{\mu}^{\mathcal{Q}}(LY_t) = L^{-1}M + L^{-1}ULY_t \tag{35}
$$

and

$$
\sigma(Y_t) = L^{-1}\tilde{\Sigma}\sqrt{\tilde{S}_t(LY_t)}. \tag{36}
$$

In our empirical application, the maximum maturity is equal to $\ell = 10$ years. Then, matrix L is given by:

$$
L = \begin{bmatrix}
1 & -1 & 1 & -1 & 1 & -1 \\
0 & 0.4 & -1.2 & 2.4 & -4 & 6 \\
0 & 0 & 0.180 & -0.9 & 2.70 & -6.3 \\
0 & 0 & 0 & 0.08 & -0.56 & 2.24 \\
0 & 0 & 0 & 0 & 0.035 & -0.3158 \\
0 & 0 & 0 & 0 & 0 & 0.0152
\end{bmatrix} . \tag{37}
$$

Now we are ready to specialize the drift restriction (30) to each particular AF version implemented in this paper (AFG and AFSV), and also to obtain the corresponding restrictions for the process of interest Y, the one that drives term structure movements within the Legendre polynomial model.

5.2.1 The AFG Version

Noting that in this version the matrix controlling the diffusion structure of vector $\tilde{Y}$, i.e. $\tilde{S}(.)$, is the identity matrix, we directly obtain $\tilde{\Sigma}\tilde{\Sigma}' = \tilde{H}_0$, and from (36) we obtain the relation between $\tilde{H}_0$ and Σ:

$$
\tilde{H}_0 = L\Sigma^2((L^{-1})')^{-1} = L\Sigma^2 L'. \tag{38}
$$

If we adopt a diagonal matrix representation for Σ[29], with Σ_{ii} as the ith-diagonal term, then, in order to match the second requirement of AF restrictions we must have $\Sigma_{ii} = 0$ for $i \geq 4$. Therefore, using transformation L between Y and $\tilde{Y}$, $\tilde{H}_0$ can be explicitly related to the non-null diagonal terms in Σ:

---

[29]This representation for Σ provides exactly the same identification structure of Dai and Singleton (2002).

$$
\tilde{H}_0 =
\begin{bmatrix}
\Sigma_{11}^2 + \Sigma_{22}^2 + \Sigma_{33}^2 & -0.4\Sigma_{22}^2 - 1.2\Sigma_{33}^2 & -0.18\Sigma_{33}^2 & 0 & 0 & 0 \\[2ex]
-0.4\Sigma_{22}^2 - 1.2\Sigma_{33}^2 & 0.16\Sigma_{22}^2 + 1.44\Sigma_{33}^2 & -0.216\Sigma_{33}^2 & 0 & 0 & 0 \\[2ex]
0.18\Sigma_{33}^2 & -0.216\Sigma_{33}^2 & 0.0324\Sigma_{33}^2 & 0 & 0 & 0 \\[2ex]
0 & 0 & 0 & 0 & 0 & 0 \\[2ex]
0 & 0 & 0 & 0 & 0 & 0 \\[2ex]
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}. \tag{39}
$$

Since U_2 is null under the Gaussian version, we learn from (35) that the two matrices ($L^{-1}M$ and $L^{-1}U_1L$) necessary to obtain an explicit expression for the drift $\mu^{\mathcal{Q}}(Y_t)$ are given by:

$$
L^{-1}M =
\begin{bmatrix}
\dfrac{5}{2}\Sigma_{11}^2 + \dfrac{5}{6}\Sigma_{22}^2 + \dfrac{1}{2}\Sigma_{33}^2 \\[2.5ex]
\dfrac{5}{2}\Sigma_{11}^2 + \dfrac{3}{2}\Sigma_{22}^2 + \dfrac{11}{14}\Sigma_{33}^2 \\[2.5ex]
\dfrac{5}{3}\Sigma_{22}^2 + \dfrac{5}{7}\Sigma_{33}^2 \\[2.5ex]
\Sigma_{22}^2 + \Sigma_{33}^2 \\[2.5ex]
\dfrac{9}{7}\Sigma_{33}^2 \\[2.5ex]
\dfrac{5}{7}\Sigma_{33}^2
\end{bmatrix} \tag{40}
$$

and

$$
L^{-1}U_1L =
\begin{bmatrix}
0 & 0.4 & -0.3 & 0.56667 & -0.41667 & 0.65667 \\
0 & 0 & 0.9 & -0.5 & 1.25 & -0.77 \\
0 & 0 & 0 & 1.3333 & -0.58333 & 1.7833 \\
0 & 0 & 0 & 0 & 1.75 & -0.63 \\
0 & 0 & 0 & 0 & 0 & 2.16 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \tag{41}
$$

Note that Y_4, Y_5, and Y_6 are deterministic factors under the Gaussian case. This is a consequence of two facts: (i) their dynamics do not depend on the Brownian motion vector, and (ii) their drifts do not depend on the first three components of the state vector. With matrices $L^{-1}M$ and $L^{-1}U_1L$ in hands, we obtain the drift of vector Y, and in particular, the drifts of the deterministic factors Y_4, Y_5, and Y_6:

$$\mu^{\mathcal{Q}}(Y_t)_4 = \Sigma_{22}^2 + \Sigma_{33}^2 + 1.75Y_{t,5} - 0.63Y_{t,6},$$

$$\mu^{\mathcal{Q}}(Y_t)_5 = \frac{9}{7}\Sigma_{33}^2 + 2.16Y_{t,6}, \tag{42}$$

$$\mu^{\mathcal{Q}}(Y_t)_6 = \frac{5}{7}\Sigma_{33}^2.$$

By explicitly solving the ordinary differential equations implied for these factors, we have

$$Y_{t,4} = Y_{0,4} + (\Sigma_{22}^2 + \Sigma_{33}^2 + 1.75Y_{0,5} - 0.63Y_{0,6})t + (0.9\Sigma_{33}^2 + 0.189Y_{0,6})t^2 + 0.45\Sigma_{33}^2 t^3, \tag{43}$$

$$Y_{t,5} = Y_{0,5} + (\frac{9}{7}\Sigma_{33}^2 + 2.16Y_{0,6})t + \frac{27}{35}\Sigma_{33}^2 t^2, \tag{44}$$

$$Y_{t,6} = Y_{0,6} + \frac{5}{7}\Sigma_{33}^2 t. \tag{45}$$

Note that, under this Gaussian version, the dynamics of the state variables $Y_{t,4}$, $Y_{t,5}$ and $Y_{t,6}$, in addition to being deterministic, are completely determined by parameters Σ_{22}, Σ_{33}, and the initial conditions $Y_{0,4}$, $Y_{0,5}$ and $Y_{0,6}$.

5.2.2 The AFSV Version

The AFSV version, presents one stochastic factor driving the stochastic volatility of the three stochastic factors (Y_1, Y_2 and Y_3). In order to keep the risk-neutral dynamics of both Y_t and $\tilde{Y}_t$ within the sub-class of affine models with only one factor determining the volatility, we choose factor Y_3

to drive the stochastic volatility[30]. Specifically we set

$$\beta_i' = \begin{bmatrix} 0 & 0 & \beta_{i3} & 0 & 0 & 0 \end{bmatrix},$$

what gives:

$$H_{1,ij} = \begin{bmatrix} 0 & 0 & h_{ij} & 0 & 0 & 0 \end{bmatrix} \quad 1 \leq i, j \leq 6;$$

where $(\Sigma S_t \Sigma')_{ij} = H_{0ij} + H_{1ij}Y_t$ with $H_{1ij} \in \mathbb{R}^N$. This specifications imply that

$$H_1 \cdot Y_t = Y_{t,3}H,$$

where in the right-hand side we have a tensor product, with H being a 6×6-matrix with elements h_{ij} (see Duffie (2001) for the tensorial notation).

From (36) and the relation $\tilde{Y}_t = LY_t$ we obtain

$$H_0 = \Sigma \left(\text{diag} \left[\alpha_1, \dots, \alpha_6 \right] \right) \Sigma',$$

$$\tilde{H}_0 = LH_0L',$$

$$H = \Sigma \left(\text{diag} \left[b_1, \dots, b_6 \right] \right) \Sigma'.$$

Since $\tilde{H}_1 \cdot \tilde{Y}_t = L \left(H_1 \cdot Y_t \right) L' = Y_{t,3}LHL'$ we have

$$\tilde{H}_{1,ij} = \begin{bmatrix} 0 & 0 & z_{ij} & 11.25z_{ij} & \dfrac{720}{7}z_{ij} & \dfrac{6250}{7}z_{ij} \end{bmatrix}, \tag{46}$$

with $z_{ij} = (LHL')_{ij}$.

Hence from (32), (33) and (34) we can express M and U as well as the drift of Y_t as functions of α_i, β_{i3} $(i = 1, \dots, 6)$ and Σ. Finally, for identification purposes, in the empirical implementation of this version, we fix Σ to be a diagonal matrix (with $\Sigma_{ii} = 0$ for $i \geq 4$ in order to match the second requirement of AF restrictions) and $\alpha = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$.

5.2.3 Estimation Procedures

How does one estimate CS and AF versions of the Legendre polynomial model?

For the CS version, we run cross-sectional independent regressions for each point t in time, within the sample period. In a market of zero coupon

---

[30]Choices of any of the two remaining factors to capture stochastic volatility could have been implemented but with higher computational costs.

bonds, assuming that we observe yields R_{obs} with measurement error, the model is estimated with the use of the following linear regression:

$$\hat{Y}_t = (F'F)^{-1} F' R_{obs}^t, \tag{47}$$

where R_{obs}^t is a vector containing observed yields, at time t, for different maturities $(\tau_1,...\tau_k)$, and F is the following matrix:

$$F = \begin{bmatrix} P_0(\frac{2\tau_1}{\ell} - 1) & P_1(\frac{2\tau_1}{\ell} - 1) & ... & P_{N-1}(\frac{2\tau_1}{\ell} - 1) \\ P_0(\frac{2\tau_2}{\ell} - 1) & P_1(\frac{2\tau_2}{\ell} - 1) & ... & P_{N-1}(\frac{2\tau_2}{\ell} - 1) \\ \vdots & \vdots & & \vdots \\ P_0(\frac{2\tau_k}{\ell} - 1) & P_1(\frac{2\tau_k}{\ell} - 1) & ... & P_{N-1}(\frac{2\tau_k}{\ell} - 1) \end{bmatrix}.$$

For both AF versions we use the Quasi-Maximum Likelihood (QML) procedure, adopting the methodology proposed by Chen and Scott (1993), with 2-,5-, and 10-year maturity zero-coupon bonds priced exactly and 3-, and 7-year maturity zero-coupon bonds priced with i.i.d zero-mean errors. The conditional transition densities are obtained with the use of closed-form formulas for the first and the second moments of Y within the affine framework (see for instance, Duffee (2002), Jacobs and Karoui (2006)). Observe that for the AFG version the QML is, in fact, a pure maximum likelihood procedure since the transitions densities come exactly from a normal distribution.

Parameter	AFG	AFSV
β_{13}	-	86.20 (13.96)
β_{23}	-	63.77 (21.72)
β_{33}	-	62.78 (7.45)
Σ_{11}	0.0206 (0.0005)	0.0218 (0.0005)
Σ_{22}	0.0094 (0.0004)	0.0099 (0.0004)
Σ_{33}	0.0023 (0.0000)	0.0031 (0.0001)
$\lambda_0(1)$	1.56 (1.02)	*
$\lambda_0(2)$	*	1.30 (0.92)
$\lambda_0(3)$	*	-0.31 (0.79)
$\lambda_Y(1,1)$	-17.65 (9.79)	-21.65 (68.72)
$\lambda_Y(1,2)$	*	*
$\lambda_Y(1,3)$	131.5 (111.48)	*
$\lambda_Y(2,1)$	1.70 (2.58)	14.08 (10.68)
$\lambda_Y(2,2)$	-164.81 (47.49)	186.89 (80.73)
$\lambda_Y(2,3)$	-268.89 (138.63)	480.15 (191.58)
$\lambda_Y(3,1)$	*	-4.42 (8.09)
$\lambda_Y(3,2)$	-82.53 (25.82)	149.31 (46.68)
$\lambda_Y(3,3)$	-144.66 (59.77)	521.53 (144.56)
$Y_{0,4}$	0.0082 (0.0006)	-0.0034 (0.0003)
$Y_{0,5}$	-0.0009 (0.0000)	0.0008 (0.0001)
$Y_{0,6}$	0.0000 (0.0000)	-0.0001 (0.0000)

Table 1: Estimated Parameters and Standard Errors for the AFG Model

Both models were estimated by QML adopting the methodology proposed by Chen and Scott (1993), with 2-,5-, and 10-year maturity zero-coupon bonds priced exactly and 3-, and 7-year maturity zero-coupon bonds priced with i.i.d zero-mean errors. Under AFSV model, for each i and $j \neq 3$, β_{ij} is fixed to zero (only the third factor drives stochastic volatility). Values with stars were not significant in a first QML estimation passage. Values with dashes do not apply to the specific model. Estimation sample ranges from January 1972 to December 1994. Standard errors were obtained by the BHHH method.

Maturity	2-Year	3-Year	5-Year	7-Year	10-Year
Model	**1-Month Forecasting Horizon**				
CS	13.8/20.5	**6.7**/20.1	7.8/24.6	**11.9/27.7**	**9.5/27.9**
AFG	-5.6/17.6	25.8/31.8	**-0.2**/23.8	-78.1/86.6	16.6/31.0
AFSV	**-0.9/15.2**	6.9/**19.8**	1.6/**23.7**	-20.2/34.4	25.8/37.6
Model	**6-Month Forecasting Horizon**				
CS	64.4/73.5	55.2/70.0	54.2/75.0	58.3/81.2	**59.6/84.0**
AFG	**-14.1**/46.4	21.4/**50.2**	**4.2/55.4**	-60.3/88.5	77.8/97.7
AFSV	15.1/**43.7**	**17.4**/52.7	9.5/60.1	**-2.5/65.8**	87.6/114.8
Model	**12-Month Forecasting Horizon**				
CS	109.5/116.5	98.6/109.1	96.7/111.8	100.9/117.9	102.7/**121.0**
AFG	**-2.8/52.8**	31.3/**58.7**	12.2/**57.9**	-49.3/79.8	120.0/133.9
AFSV	35.2/64.1	**26.8**/67.8	**4.1**/72.5	**-13.9/78.1**	**91.6**/127.5

Table 2: Bias and Root Mean Square Errors for Out-of-Sample Forecasts (in bps)

This table presents bias (first number in each cell) and RMSE (second number in each cell) for 1-month, 6-month and 12-month ahead out-of-sample forecasts, for the three versions of the polynomial model considered: Cross Sectional (CS), Arbitrage-free Gaussian (AFG), Arbitrage-free with Stochastic Volatility (AFSV). Out-of-sample period ranges from January 1995 to December 1998. Smaller absolute bias and RMSE across models appears in bold.

Maturity	2-Year	3-Year	5-Year	7-Year	10-Year
Model	**1-Month Forecasting Horizon**				
S1 AFSV x CS	-2.2^{**}	-0.13	-0.41	1.59	4.17***
S2 AFSV x CS	-3.17^{***}	0.0	0.29	0.87	3.17***
S1 AFG x CS	-0.80	5.61***	-0.16	8.70***	3.09***
S2 AFG x CS	-1.44	3.75***	0.29	5.48***	2.88***
Model	**6-Month Forecasting Horizon**				
S1 AFSV x CS	-3.76^{***}	-1.64^{*}	-1.15	-1.13	3.20***
S2 AFSV x CS	-4.42^{***}	-3.81^{***}	-2.90^{***}	-2.90^{***}	4.72***
S1 AFG x CS	-1.91^{**}	-2.61^{***}	-1.57	0.43	3.05***
S2 AFG x CS	-2.28^{**}	-4.11^{***}	-2.59^{***}	1.37	4.42***
Model	**12-Month Forecasting Horizon**				
S1 AFSV x CS	-35.04^{***}	-10.40^{***}	-1.78^{*}	-1.15	0.01
S2 AFSV x CS	-6.08^{***}	-5.10^{***}	-2.79^{***}	-2.46^{**}	1.15
S1 AFG x CS	-6.33^{***}	-3.95^{***}	-2.98^{***}	-0.96	3.75***
S2 AFG x CS	-5.42^{***}	-5.75^{***}	-4.77^{***}	-1.48	6.08***

Table 3: Statistical Comparison of Forecasts through the Diebold and Mariano (1995) Test

This table presents the Diebold and Mariano (1995) S1, and S2 statistics for 1-month, 6-month and 12-month ahead out-of-sample forecasts, comparing the AFSV and the AFG to the CS version. Comparisons are done as functions of Mean Absolute Errors (MAE). In-sample period ranges from January 1972 to December 1994. Out-of-sample period ranges from January 1995 to December 1998. Negative values are in favor of AFSV / AFG versions, and against the CS version. Small p-values indicate high probability of rejecting the null hypothesis of a zero difference in Mean Absolute Errors. Values with a star indicate significance at a 90% level, with two stars, significance at a 95% level, and three stars, significance at a 99% level, on a bi-caudal test.

Maturity	2-Year	3-Year	5-Year	7-Year	10-Year
	1-Month Forecasting Horizon				
Average Yields	5.337	5.502	5.648	5.717	5.822
	6-Month Forecasting Horizon				
Average Yields	5.245	5.411	5.550	5.614	5.717
	12-Month Forecasting Horizon				
Average Yields	5.208	5.389	5.536	5.601	5.706

Table 4: Observed Yields Averaged across the Out-of-Sample Period (in %)

This table presents observed yields averaged across the out-of-sample period, for the three different forecasting horizons. Out-of-sample period ranges from January 1995 to December 1998. For the h-month forecasting horizon, the average is performed with a window of data ranging from the hth month of 1995 up to December 1998.

Maturity	2-Year	3-Year	5-Year	7-Year	10-Year
Model	**1-Month Forecasting Horizon**				
CS	5.433	5.539	5.709	5.822	5.883
AFG	5.536	5.943	5.723	4.961	6.023
AFSV	5.453	5.680	5.723	5.510	5.943
Model	**6-Month Forecasting Horizon**				
CS	5.627	5.736	5.906	6.010	6.044
AFG	6.198	6.352	5.839	5.100	6.903
AFSV	5.823	5.960	5.895	5.683	6.381
Model	**12-Month Forecasting Horizon**				
CS	5.800	5.905	6.064	6.154	6.158
AFG	6.589	6.505	5.766	5.179	8.062
AFSV	6.082	6.130	5.974	5.806	6.896

Table 5: Model Implied Forward Rates Averaged Across the Out-of-Sample Period (in %)

This table presents model implied forward rates with maturities τ, and forward term equal to respectively 1-, 6-, and 12-month, averaged across the out-of-sample period, for the three versions of the polynomial model considered: Cross Sectional (CS), Arbitrage-free Gaussian (AFG), Arbitrage-free with Stochastic Volatility (AFSV). In-sample period ranges from January 1972 to December 1994. Out-of-sample period ranges from January 1995 to December 1998.

Maturity	2-Year	3-Year	5-Year	7-Year	10-Year
Model	δ_t=**1-Month**				
CS	-4.2	-3.1	-1.6	-1.4	-3.5
AFG	25.5	18.2	7.8	2.5	3.4
AFSV	12.5	10.7	6.0	-0.5	-13.8
Model	δ_t=**6-Month**				
CS	-26.3	-22.7	-18.6	-18.7	-27.0
AFG	109.3	72.7	24.8	8.9	40.8
AFSV	42.6	37.5	25.1	9.5	-21.2
Model	δ_t=**12-Month**				
CS	-50.3	-46.9	-43.8	-45.6	-57.5
AFG	140.9	80.4	10.9	7.1	115.6
AFSV	52.2	47.3	39.8	34.4	27.4

Table 6: Model Implied Bond Risk Premium Averaged Across the Out-of-Sample Period

This table presents model implied bond risk premium for 1-,6- and 12-month holding periods, averaged across the out-of-sample period, for the three versions of the polynomial model considered: Cross Sectional (CS), Arbitrage-free Gaussian (AFG), Arbitrage-free with Stochastic Volatility (AFSV). In-sample period ranges from January 1972 to December 1994. Out-of-sample period ranges from January 1995 to December 1998. Bond risk premium for maturity τ and holding period δ_t was normalized by a factor $\frac{\delta_t}{\tau}$.

Maturity	2-Year	3-Year	5-Year	7-Year	10-Year
Model	δ_t=1-Month				
CS	0.69	0.55	0.79	0.88	0.63
AFG	3.55	1.71	36.89	0.97	1.21
AFSV	13.71	2.55	4.80	1.03	0.47
Model	δ_t=6-Month				
CS	0.59	0.59	0.66	0.68	0.55
AFG	6.76	4.40	6.93	0.85	1.52
AFSV	3.82	3.15	3.65	2.73	0.76
Model	δ_t=12-Month				
CS	0.54	0.52	0.55	0.55	0.44
AFG	49.21	3.57	1.90	0.86	1.96
AFSV	2.48	2.77	10.81	1.48	1.30

Table 7: Effects of Bond Risk Premium on Forecasting Bias

This table presents ratios of the absolute value of forecasting bias imposing zero bond risk premium (using only forward rates) over the absolute value of the actual forecasting bias, for 1-, 6- and 12-month holding period intervals, for the three versions of the polynomial model considered: Cross Sectional (CS), Arbitrage-free Gaussian (AFG), Arbitrage-free with Stochastic Volatility (AFSV). In-sample period ranges from January 1972 to December 1994. Out-of-sample period ranges from January 1995 to December 1998. Ratio above one indicates that model implied risk premium contributes to decrease bias, and bellow one indicates that risk premium was not correctly estimated.

Maturity	2 Year	3 Year	5 Year	7 Year	10 Year
Model	**1-Month Forecasting Horizon**				
AFSV	**-0.9/15.2**	6.9/19.8	**1.6/23.7**	-20.2/34.4	25.8/37.6
RW	4.7/16.0	5.4/20.0	6.0/24.5	**6.3/26.4**	**6.4**/27.5
DL	3.7/15.9	**2.0/19.2**	5.8/24.2	8.9/26.6	6.7/**27.4**
Model	**6-Month Forecasting Horizon**				
AFSV	**15.1/43.7**	**17.4/52.7**	**9.5/60.1**	**-2.5/65.8**	87.6/114.8
RW	22.1/46.3	24.0/55.9	27.5/67.0	29.5/72.2	**30.2/74.7**
DL	39.1/56.9	37.9/62.6	44.2/73.4	49.6/80.0	49.2/82.1
Model	**12-Month Forecasting Horizon**				
AFSV	35.2/64.1	**26.8/67.8**	**4.1/72.5**	**-13.9/78.1**	91.6/127.5
RW	**29.4**/58.6	30.4/68.5	34.8/81.6	38.2/87.7	**39.8/90.4**
DL	76.7/90.1	73.8/92.4	80.6/103.7	87.2/111.5	87.9/113.7

Table 8: Bias and Root Mean Square Errors for Out-of-Sample Forecasting (in bps): Comparisons with the Random Walk and Diebold and Li (2006) models

This table presents bias (first number in each cell) and RMSE (second number in each cell) for 1-month, 6-month, and 12-month ahead out-of-sample forecasts for the RW, and DL models, and compare them to the AFSV polynomial model. In-sample period ranges from January 1972 to December 1994. Out-of-sample period ranges from January 1995 to December 1998. For a fixed forecasting horizon (1-month, 6-month, 12-month), smaller absolute bias and smaller RMSE across models appear in bold.
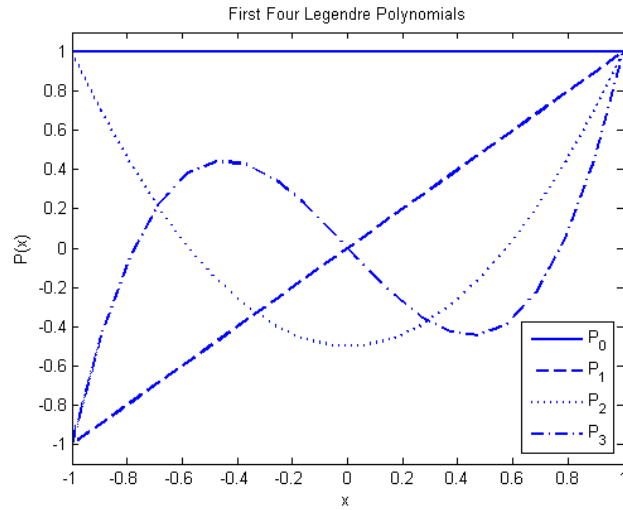
Figure 1: The First Four Legendre Polynomials

This picture depicts the first four Legendre polynomial, which are respectively $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = \frac{1}{2}(3x^2 - 1)$, and $P_3(x) = \frac{1}{2}(5x^3 - 3x)$, defined within the interval [-1,1].
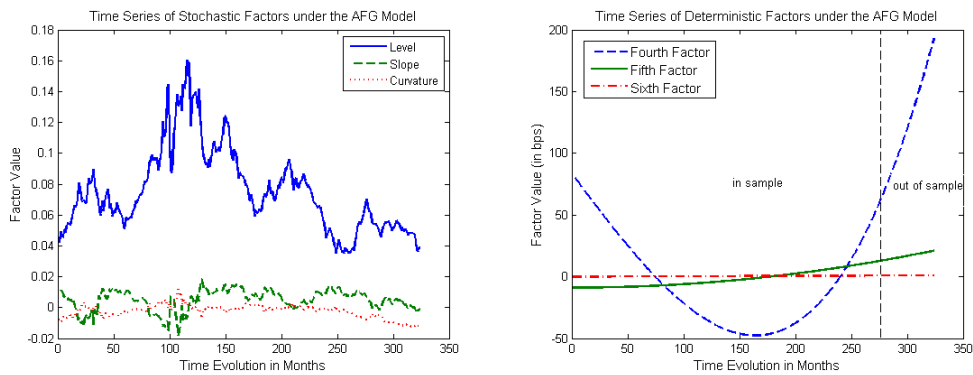
Figure 2: Dynamic Factors in the AFG Polynomial Model
This picture presents the time series of the six factors estimated under the AFG model version. The left-hand side factors are the three lower order factors with non-null diffusions, respectively capturing "level", "slope" and "curvature" movements. The right-hand side factors are the three conditionally deterministic higher order factors, respectively related to the Legendre polynomials of degree 3, 4 and 5. In-sample period ranges from January 1972 to December 1994. Out-of-sample period ranges from January 1995 to December 1998.
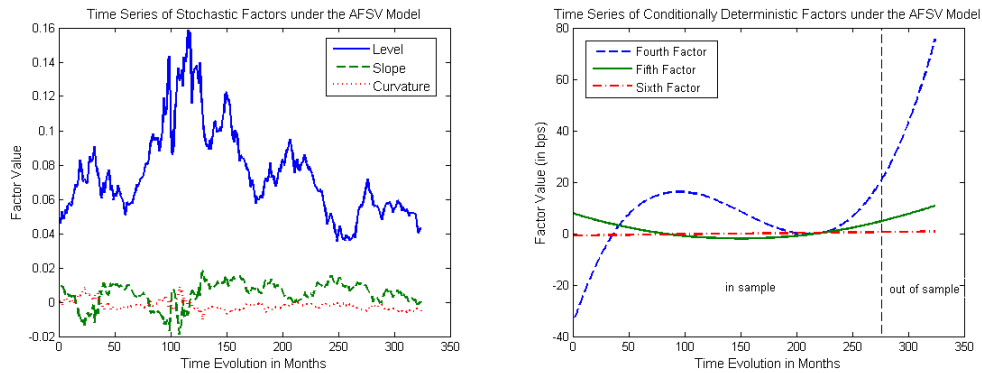
**Figure 3: Dynamic Factors in the AFSV Polynomial Model**

This picture presents the time series of the six factors estimated under the AFSV model. The left-hand side factors are the three lower order factors with non-null diffusions, respectively capturing "level", "slope" and "curvature" movements. The curvature (third) factor drives stochastic volatility of the three lower order factors. The right-hand side factors are the three conditionally deterministic higher order factors, respectively related to the Legendre polynomials of degree 3, 4 and 5. In-sample period ranges from January 1972 to December 1994. Out-of-sample period ranges from January 1995 to December 1998.

48
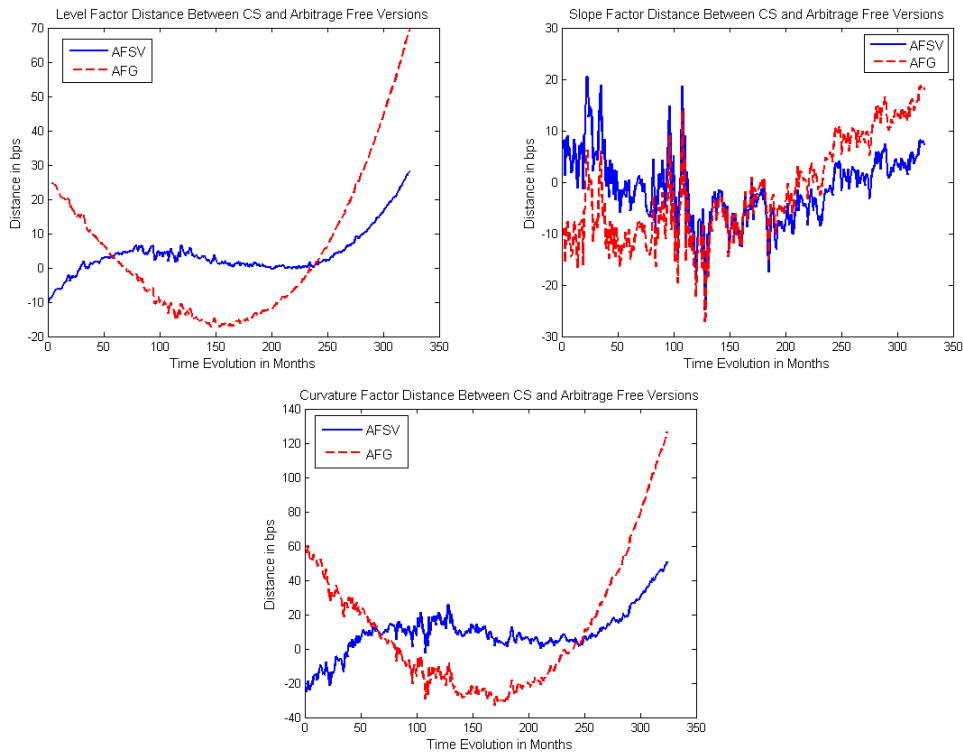
Figure 4: Distance Between Factors from CS and Arbitrage-Free Versions

This picture presents the distance between the CS "level", "slope" and "curvature" factors, and the same factors under each arbitrage-free version of the polynomial model. Blue full line captures the distance between a CS factor and the corresponding AFSV factor. Red dashed line captures the distance between a CS factor and the corresponding AFG factor. In-sample period ranges from January 1972 to December 1994.
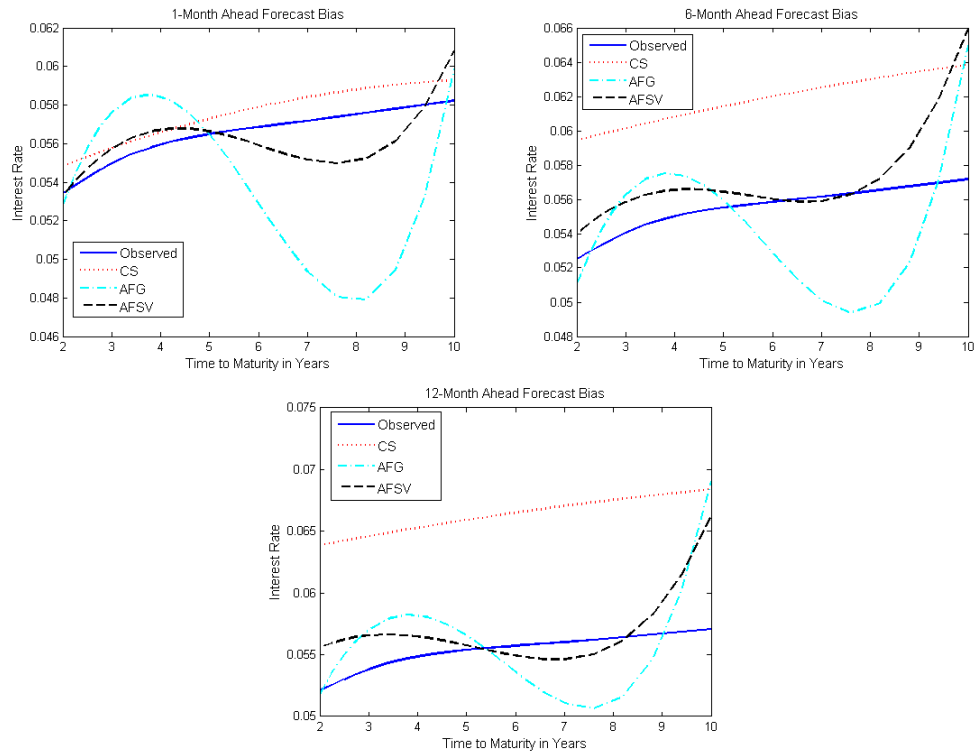
Figure 5: Out-of-Sample Averaged Forecasts and Observed Yield Curves

This picture presents a spline version of the observed yield curve (2-, 3-, 5-, 7-, and 10- year maturities) averaged across the out-of-sample period (from Jan. 95 to Dec. 98), and corresponding averaged yield curves implied by the different versions of the polynomial model. Blue solid line represents the observed yield curve, dotted red line represents the CS version, cyan dash-dotted line represents the AFG version, and black dashed line represents the AFSV version. In-sample period ranges from January 1972 to December 1994.
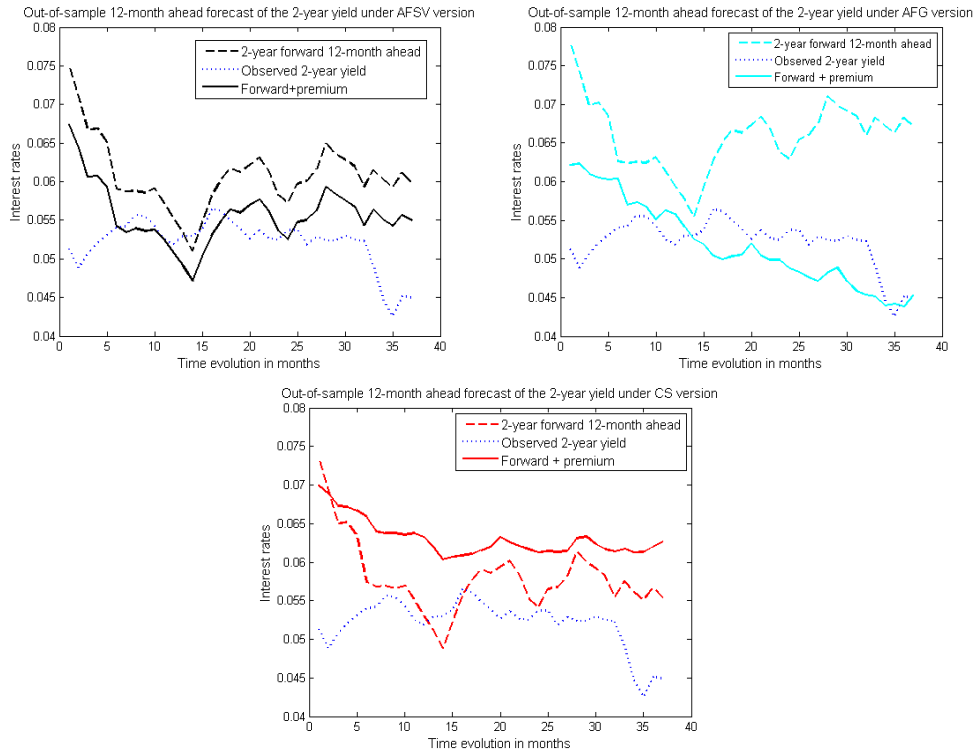
Figure 6: 12-Month Ahead Out-of-Sample Forecasting of the 2-Year Yield

This picture presents, out-of-sample time series of observed yields, model implied forward rates, and model implied bond risk premium, for different forecasting horizons. Dotted line represents observed yields. Solid line represents model forecast given by Equation (17). Dashed line represents model implied forward rate. In-sample period ranges from January 1972 to December 1994. For the h-month forecasting horizon, the out-of-sample period ranges from the h_{th} month of 1995 to December 1998.
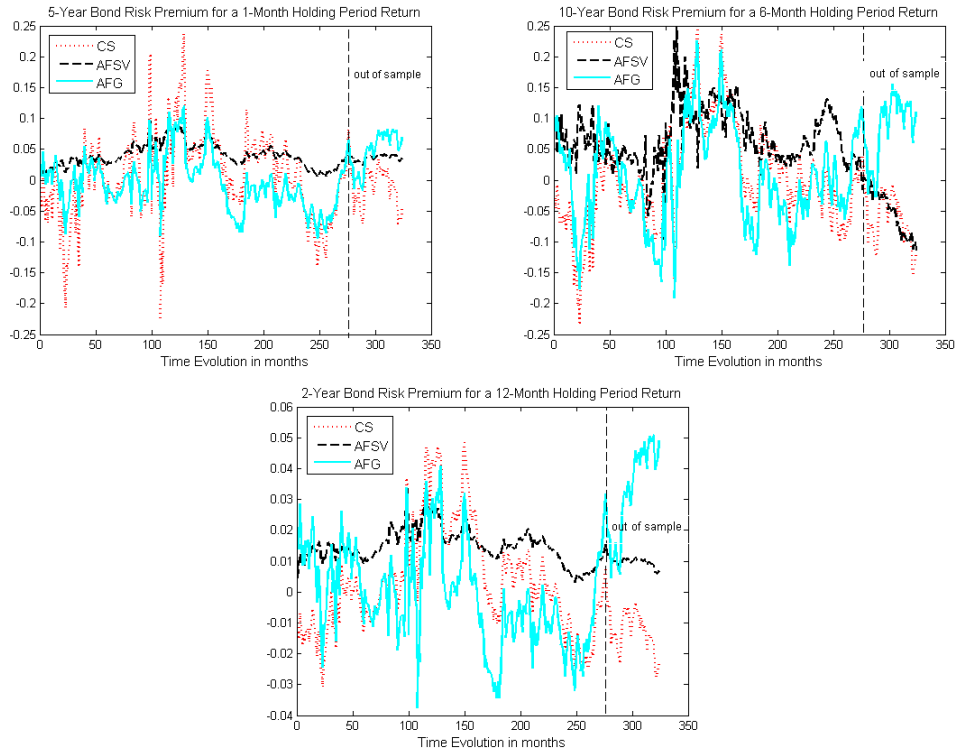
Figure 7: Bond Risk Premium for Different Maturities and Forecasting Horizons

This picture presents the time series of bond risk premium implied by each model version, for different maturities and forecasting horizons. Cyan solid line represents AFG, black dashed line represents AFSV, and red dotted line represents CS. In-sample period ranges from January 1972 to December 1994. Out-of-sample period ranges from January 1995 to December 1998.