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

Caio Almeida [a,*], José Vicente [b]

[a] *Graduate School of Economics, Getulio Vargas Foundation*

[b] *Research Department, Central Bank of Brazil and Faculdades Ibmec-RJ*

This version: June 2008

Abstract

Parametric term structure models have been successfully applied to numerous problems in fixed income markets, including pricing, hedging, managing risk, as well as to the study of monetary policy implications. In 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 whether no-arbitrage affects forecasting. We construct cross-sectional (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.

JEL Classification: G12;C51;C53
Keywords: Dynamic models; No-arbitrage; Forecasting; Bond risk premia

⋆ We thank Antonio Diez de los Rios, Darrell Duffie, Marcelo Fernandes, Jean-Sebastien Fontaine, René Garcia, and Lotfi Karoui for important comments. We also thank comments and suggestions from seminar participants at the Sofie 2008 Conference, the 26th Brazilian Colloquium of Mathematics, Getulio Vargas Foundation, HEC Montreal, and Catholic University in Rio de Janeiro. The views expressed are those of the authors and do not necessarily reflect those of the Central Bank of Brazil. The first author gratefully acknowledges financial support from CNPq-Brazil.

* Corresponding author. Tel.: 5521-2559-5828; fax: 5521-2553-8821.
Email addresses: calmeida@fgv.br (Caio Almeida), jose.valentim@bcb.gov.br (José Vicente).

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 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 return predictability and interest rate forecasting has emerged [1]. In particular, some models are not intertemporally consistent 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 for 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 a model's 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 premia, 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 objectively inform managers on whether 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 do not 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 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 dynamics of underlying factors. This invariant loading structure implies that bond risk premia relate to a common set of underlying factors (i.e. term structure movements) across different versions of the model. 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

[1] 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. 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), 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 predefined parametric functions, such as polynomials, exponentials, or trigonometric functions, among others. See, for instance, Chambers et al. (1984), Nelson and Siegel (1987), and Svensson (1994), among others.

polynomials. This framework supports flexible dynamics for term structure factors, including versions that allow for arbitrage opportunities and others that are arbitrage-free. By focusing the analysis on three-factor models [3], we compare a cross-sectional (CS) version, which allows for the existence of arbitrages, with 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 do not rule out arbitrages. In 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 ones. Interestingly, by affecting the dynamics of the three basic stochastic factors ("level", "slope" and "curvature"), the conditionally deterministic factors directly affect the bond risk premium structure.

More general arbitrage-free versions of the polynomial model exist and could also be analyzed [4]. However, in an attempt to achieve more 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 (2002), Duffee (2002), and Tang and Xia (2007). Duffee (2002) elects the three-factor affine Gaussian model as the best (within the affine family) 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, USA, 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's (1995) tests confirm the statistical significance of the obtained results.

[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. See, for example, Almeida (2005).

In order to analyze the effects of no-arbitrage on 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 the stylized facts of U.S. term structure, 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, which means including models that allow for arbitrages in our analysis.

Ang and Piazzesi (2003) show that imposing no-arbitrage restrictions to a VAR model 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 that allow for arbitrages. Those changes in factors and bond risk premia make it harder to isolate the pure effects of no-arbitrage on forecasting. In contrast, the parametric polynomial term structure 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. 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. Finally, 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 the Nelson and Siegel (1987) model, the Svensson (1994) model [5], and spline models with fixed knots, among others.

We provide evidence that no-arbitrage restrictions improve interest rate forecasting for a class of parametric models. However, what is the extent of this conclusion? Our results when coupled with those by Ang and Piazzesi (2003), Favero et al. (2007), and Christensen et al. (2007) indicate that the validity of no-arbitrage restrictions as a tool to improve a model's forecasting ability deserves some credibility [6]. In addition, as we show that no-arbitrage restrictions help to better econometrically identify risk premium parameters on parametric models, this identification improvement should be even more significant considering more complex dynamic term structure models. Models with time-varying conditional variances and nonlinear risk premia are becoming more common as tools to capture empirical stylized facts of the term structure of interest rates (see Dai and Singleton, 2003) [7]. Correspondingly, parametric models with more general dynamics should be tested both with the purpose of fitting and forecasting interest rates. This suggests that no-arbitrage restrictions as a tool to improve econometric identification of parameters (especially risk premium parameters) should be of fundamental importance for the implementation of such models.

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 the empirical results, including an interesting discussion relating bond risk premium to model forecasting ability. Section 4 offers the concluding remarks and presents possible topics for further research. An online Appendix at http://www.fgv.br/professor/calmeida/ presents details on the arbitrage-free versions of the polynomial model.

[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 (1987) and Svensson (1994) models, as shown for instance, by Sharef and Filipovic (2004), and Christensen et al. (2007).

[6] Nevertheless, we believe that a more detailed analysis of the contributions of no-arbitrage to interest rate forecasting should be accomplished with extensive tests including a variety of different dynamic models, complemented with robustness tests based on different sample periods of data. For instance, in a recent paper, Duffee (2008) suggests, for a class of *affine Gaussian models*, that no-arbitrage restrictions do not agregate additional forecasting ability for his proposed model, although they do not hinder its performance either.

[7] For instance, in a recent paper, Dai et al. (2006) propose a class of nonlinear discrete-time models whose market prices of risk are nonlinear functions of the state variables. They show, under a three-factor dynamic model, that the inclusion of a cubic term in the drift of the factor driving stochastic volatility improves out-of-sample forecasting ability when compared to a linear drift for the same factor.

2. Legendre Polynomial Model

Almeida et al. (1998) proposed modeling the term structure of interest rates $R(.)$ as a linear combination of Legendre polynomials [8]:

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

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 [9]. This model has been applied to problems involving scenario-based portfolio allocation, risk management, and hedging with non-parallel movements (see, for instance, Almeida et al., 2003).

In the estimation process, the number of Legendre polynomials is fixed according to some statistical criterion [10]. 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 previously selected Legendre polynomials ($P_n(\cdot)$'s). The cross-sectional 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 intertemporal restrictions** upon term structure movements, consequently allowing for the **existence of arbitrages** within the model [11].

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

---

[8] 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.

[9] 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.

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

[11] 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.

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

where W is an 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 on 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:

$$B(t,T) = \mathrm{e}^{-\tau G(\tau)' Y_t}, \tag{3}$$

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

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

(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 [12]) 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:*

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

where W^* is a Brownian 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[13] .

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

[13] 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.

$$
\begin{aligned}
&(6.1) \quad \sum_{j=2}^{N}(j-1)L_j Y_{t,j}\tau^{j-2} = \sum_{j=1}^{N} L_j \mu_j^{\mathbb{Q}}(Y_t)\tau^{j-1} - \sum_{j=1}^{[\frac{N}{2}]}\sum_{k=1}^{[\frac{N}{2}]} \Gamma_{jk}(Y_t)\frac{\tau^{j+k-1}}{k} \\
&(6.2) \quad \Gamma_{jk}(Y_t) = 0 \textit{ for } j > [\tfrac{N}{2}] \textit{ or } k > [\tfrac{N}{2}]
\end{aligned}
\qquad (6)
$$

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 is provided in the Appendix at http://www.fgv.br/professor/calmeida/.

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 drift will compensate for 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 belongs 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 [14]:

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

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 (Section 3), 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 and three conditionally deterministic). The first AF version is a Gaussian model ($\beta_i = 0, \ \forall i$) and the second is a stochastic volatility model with only one factor driving the volatility. In the online Appendix, we show in detail how to translate the AF restriction into the affine framework and how to further specialize the results to the Gaussian and stochastic volatility AF versions.

[14] 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.

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

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

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

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

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:

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

Equation 10 indicates that expected instantaneous zero-coupon bond excess return is a linear combination of model factors, with weights depending on matrices λ_Y, and Σ, and on a predefined 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 the online Appendix for details).

2.1. Forecasting with the Polynomial Model

Within the subclass 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 into two steps: first extract the time series Y_t^{CS} of term structure movements by running cross-sectional regressions and then fit a VAR(1) process to those series of term structure movements:

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

Given a fixed maturity τ and a fixed forecasting horizon (h-step horizon), forecasts are produced by calculating the conditional expectation of CS factors under the VAR(1) structure:

$$E_t\left(Y^{\mathtt{CS}}_{t+h}\right) = c\sum_{j=0}^{h-1}\phi^j + \phi^h Y^{\mathtt{CS}}_t. \tag{12}$$

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

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

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 subclass the drift of latent factors $Y^{\mathtt{arb.free}}$ can be written as $\mu(Y^{\mathtt{arb.free}}_t) = \kappa(\theta - Y^{\mathtt{arb.free}}_t)$, the time t conditional expectation of $Y^{\mathtt{arb.free}}_{t+h}$ is given by Duffee (2002):

$$E_t\left(Y^{\mathtt{arb.free}}_{t+h}\right) = (I_{2N} - e^{-\kappa h})\theta + e^{-\kappa h} Y^{\mathtt{arb.free}}_t \tag{14}$$

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:

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

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

The data consist of 324 monthly observations of bootstrapped smoothed Fama-Bliss U.S. Treasury zero-coupon yields (2-, 3-, 5-, 7-, and 10-year maturities) observed from January 1972 to December 1998 [15] . Based on a subsample 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 subsequent four years of monthly data (from 1995 to 1998), not included in the estimation process, are used to measure the forecasting ability of the models, and to study their risk premium structure.

[15] This dataset is an extended version of the same dataset used by Dai and Singleton (2002).

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 zero rates indicate that the AFSV version presents a better in-sample cross-sectional fitting than the AFG version (13.6 and 26.0 bps under the AFG versus 9.3 and 16.0 bps under the 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 line) achieves values up to 75 bps under the Gaussian version (in-sample), and up to 20 bps under the stochastic volatility version (in-sample). It does not 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 parametric term structure form, a linear combination of the first six Legendre polynomials, implying that any differences in the time series of the lower order movements should come from differences in 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 factor 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 in 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 clearly transferred to Figure 4 [16].

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 [17]. We argue that keeping estimated parameters fixed, as opposed to recursively re-estimating models out-of-sample (as performed in other studies), is an appropriate choice: With fixed estimated parameters, better out-of-sample forecasting suggests higher ability to capture the underlying 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 1995 to December 1998. For each maturity and forecasting horizon h, a total of 49-h forecasts is produced, with h-month ahead forecasts beginning in the h_{th} month of 1995, and ending in December 1998. Bias and RMSE are measured in basis points, and boldfaced 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,

[16] 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's (2006) model, finding that, across models, level factors are more homogenous, while slope and curvature factors present higher distances.
[17] 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 the 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.

no-arbitrage tremendously improves results, especially 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 line represents an average of the 48 curves that were observed between January 1995 and December 1998. Correspondingly, the dotted, the dash-dotted, and the 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, due to the conditionally deterministic factors, how arbitrage-free versions present much higher curvature than CS. This higher curvature produces two antagonistic effects: it makes arbitrage-free versions get much closer to observed yields for most maturities, but also generates strong bias for a few cases [18].

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. In turn, CS version is only better at 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 Diebold and Mariano's (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 [19]. 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 (two-tailed 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. In 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, vis-à-vis AFG, CS is strong on short-horizon forecasts and on forecasts for the 10-year yield. Compared to AFSV, CS is strong only on forecasts for the 10-year yield.

[18] The AFG presents high bias at the 7-, and 10-year maturities, and the AFSV, at the 10-year maturity.
[19] The significance of results was not affected when we tested forecasting ability with a quadratic loss function.

3.4. Discussion

3.4.1. The Effects of Bond Risk Premium on 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) [20] .

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

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

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:

$$E_t \left[R\left(t + \frac{h}{12}, \tau\right) \right] = f\left(t, \tau, \frac{h}{12}\right) - \left(\frac{1}{\tau}\right) BP(\tau, h) \tag{17}$$

Equation 17 shows 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 time-to-maturity).

This way, adopting Equation 17, conditional yields are decomposed into a forward rate and into 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 premia 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 for this difference, and to decrease bias. Interestingly, Table 6 indicates that both arbitrage-free versions indeed generate positive risk premia, while in contrast, the CS version generates negative premia. In other words, under a vector autoregressive structure of lag one, the

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

version that allows arbitrages does not capture risk premium correctly [21] . 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 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 premia. 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% [22] . 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 premia 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 premia 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 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 startling contrast, arbitrage-free versions present (for most table entries) values higher than 1, in addition to some entries with values much higher than

[21] It is important to say that the lack of CS ability to reproduce risk premia 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.
[22] The average of observed yields depends on the forecasting horizon because the horizon defines the beginning of the averaging window. See the description in Table 4 for further explanations.

1 [23], indicating that no-arbitrage tremendously increases the model's ability to correctly capture risk premium dynamics.

A dynamic picture of the risk premium effect described in 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 premia [24]. 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 forward rate corrected by inclusion of risk premium, that is, the yield forecast produced with 17. Once risk premium is included, it clearly improves forecasts in the two arbitrage-free versions: the solid line is much closer to the dotted line than the dashed line is. However, in 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 throughout 27 years, from 1972 to 1998, for different maturities and forecasting horizons. The goal of this graph is to show similarities and differences among risk premia 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 in Table 7. Risk premia coming from the other two versions (CS and AFG) have a more similar in-sample behavior, but clearly move away out-of-sample, with the AFG version generating positive premia, and the CS version generating negative ones. This out-of-sample separation of premia 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 graph 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, eventually 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 premia. This is a first indication that the polynomial family, at least under its affine subclass, might not be the best candidate to simulta-

[23] In the AFG version, 7 ratio values are higher than 3, and in 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 to less than one third of the bias value with zero premium.

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

neously describe the behavior of **the whole** cross-section of yields, and to guarantee intertemporal consistency of the underlying term structure factors.

The third graph 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 in 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 the future (-2.8 bps of bias) can be explained by its risk premium out-of-sample behavior. Graph 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 for those exaggerated forward rates, and bringing forecasts to values close to observed yields. On the other hand, the AFSV version presents a positive bias of 35.2 basis points, indicating that it should have produced higher risk premium values to decrease bias. The CS version clearly misses the premium as it should have been positive (see graph 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 intention 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 (the AFSV version), from the established random walk (RW) benchmark, and from the recently proposed Diebold and Li's (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, especially 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 such as 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 a 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-sectional 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, this 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 in the out-of-sample region [25]. 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, there appears to be room for further evaluation of important parametric families such as the classical polynomial-exponential family to which the models by Nelson and Siegel (1987), Diebold and Li (2006), and Svensson (1994) belong, and also for the analysis of more complex families such as "splines with fixed knots" (see Bowsher and Meeks, 2006) [26]. 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. We leave those topics for future research.

References

Almeida, C.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.

Almeida, C.R., Duarte, A.M., Fernandes, C.C., 1998. Decomposing and simulating the movements of term structures of interest rates in emerging eurobond markets. Journal of Fixed Income, 8 (1), 21-31.

Almeida C.R., Duarte, A.M., Fernandes, C.C., 2003. A generalization of principal component analysis for non-observable term structures in emerging markets. International Journal of Theoretical and Applied Finance, 6 (8) 885-903.

[25] 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 in both versions, at those "bad" maturities, while keeping the previous good results at other maturities. The results of this test are available upon request.

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

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics, 50, 745-787.

Bali T., Heidari, M., Wu, L., 2006. Predictability of interest rates and interest rate portfolios. Working Paper, Baruch College.

Bowsher, C., Meeks, R., 2006. High dimensional yield curves: Models and forecasting. Working Paper, Nuffield College, University of Oxford.

Campbell, J.Y., Shiller, R., 1991. Yield spreads and interest rate movements: A bird's eye view. Review of Economic Studies, 58, 495-514.

Chambers, D.R., Carleton, W.T., Waldman, D.W., 1984. A new approach to estimation of the term structure of interest rates. Journal of Financial and Quantitative Analysis, 19 (3), 233-251.

Cochrane, J., Piazzesi, M., 2005. Bond risk premia. American Economic Review, 95 (1), 138-160.

Chen, R.R., Scott, L., 1993. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. Journal of Fixed Income, 3, 14-31.

Christensen, J.H., Diebold, F., Rudebusch, G.D., 2007. The affine arbitrage-free class of the Nelson-Siegel term structure models. Working Paper, Federal Reserve Bank of San Francisco.

Dai, Q., Le, A., Singleton, K., 2006. Discrete-time dynamic term structure models with generalized market prices of risk. Working Paper, University of North Carolina at Chapel Hill.

Dai, Q., Singleton, K., 2002. Expectation puzzles, time-varying risk premia, and affine models of the term structure. Journal of Financial Economics, 63, 415-441.

Dai, Q., Singleton, K., 2003. Term structure modeling in theory and reality. Review of Financial Studies, 16, 631-678.

Diebold, F.X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics, 130, 337-364.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics, 13, 253-263.

Duffee, G.R., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance, 57, 405-443.

Duffee, G.R., 2008. Forecasting with the term structure: The role of no-arbitrage. Working Paper, University of California, Berkeley.

Duffie, D., 2001. Dynamic Asset Pricing Theory. Princeton University Press.

Duffie, D., Kan, R., 1996. A yield factor model of interest rates. Mathematical Finance, 6 (4), 379-406.

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

Favero, A.C., Niu, L., Sala, L., 2007. Term structure forecasting: No-arbitrage restrictions vs. Large Information Set. Working Paper, Bocconi University.

Filipovic, D., 1999. A note on the Nelson and Siegel family. Mathematical Finance, 9 (4), 349-359.

Filipovic, D., 2001. Consistency problems for Heath-Jarrow-Morton interest rate models. Lecture Notes in Mathematics, 1760, Springer-Verlag, Berlin.

Heath, D., Jarrow, R., Morton, A., 1992. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. Econometrica, 60 (1), 77-105.

Hordahl, P., Tristani, O., Vestin, D., 2006. A joint econometric model of macroeconomic and term structure dynamics. Journal of Econometrics, 131, 405-440.

Kim, D., Orphanides, A., 2007. The bond market term premium: What is it, and how can we measure it? BIS Quarterly Review, June.

Litterman, R., Scheinkman, J.A., 1991. Common factors affecting bond returns. Journal of Fixed Income, 1, 54-61.

Mönch, E., 2007. Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach. Working Paper, Humboldt Universität zu Berlin.

Nelson, C., Siegel, A., 1987. Parsimonious modeling of yield curves. Journal of Business, 60 (4), 473-489.

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

Svensson, L., 1994. Monetary policy with flexible exchange rates and forward interest rates as indicators. Institute for International Economic Studies, Stockholm University.

Tang, H., Xia, Y., 2007. An international examination of affine term structure models and the expectations hypothesis. Journal of Financial and Quantitative Analysis, 42 (1), 41-80.

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 h^{th} 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 was normalized by a factor $\frac{1}{\tau}$.

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.

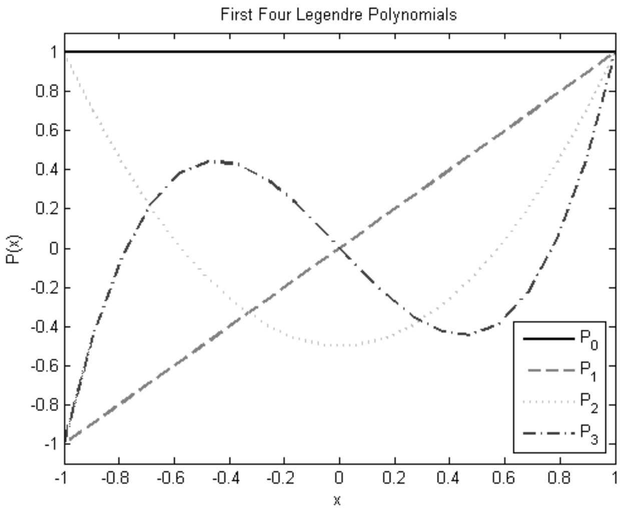

Fig. 1. The First Four Legendre Polynomials

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].

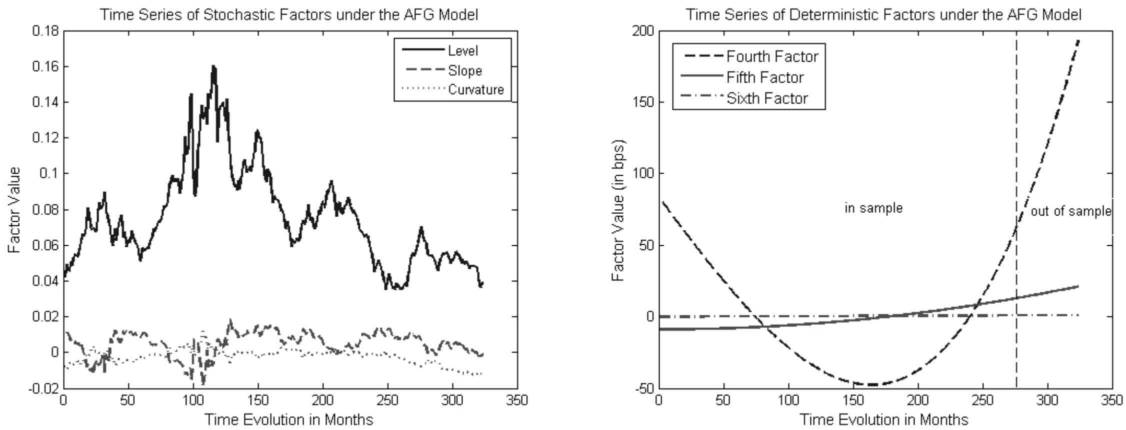

Fig. 2. Dynamic Factors in the AFG Polynomial Model

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.

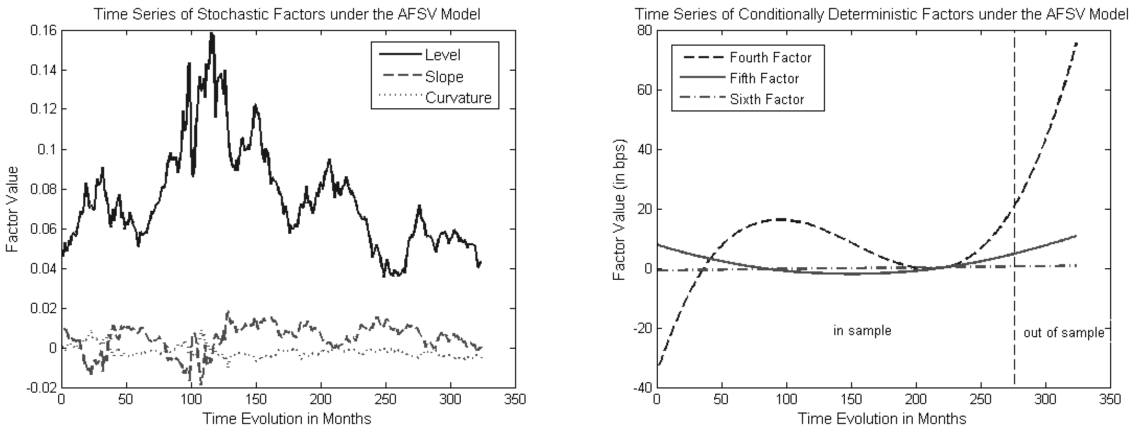

Fig. 3. Dynamic Factors in the AFSV Polynomial Model

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.

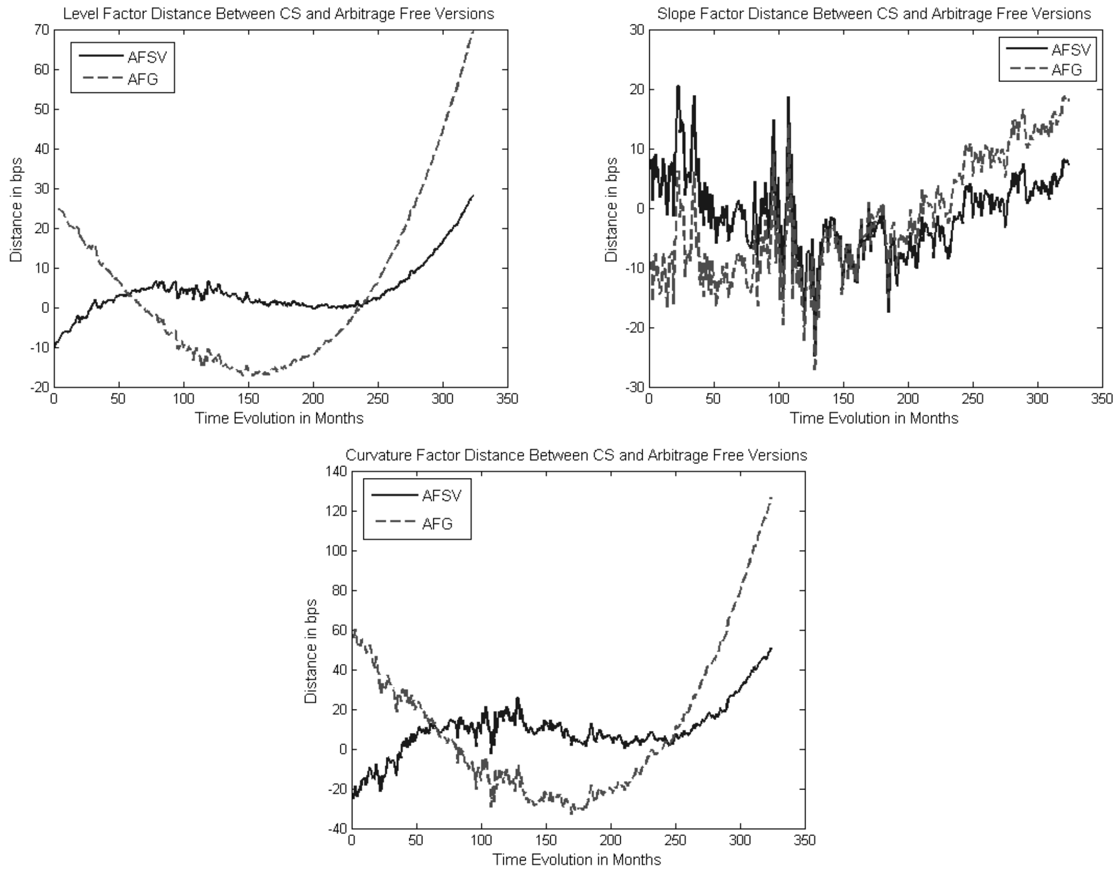

Fig. 4. Distance Between Factors from CS and Arbitrage-Free Versions

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. Full line captures the distance between a CS factor and the corresponding AFSV factor. Dashed line captures the distance between a CS factor and the corresponding AFG factor. In-sample period ranges from January 1972 to December 1994.

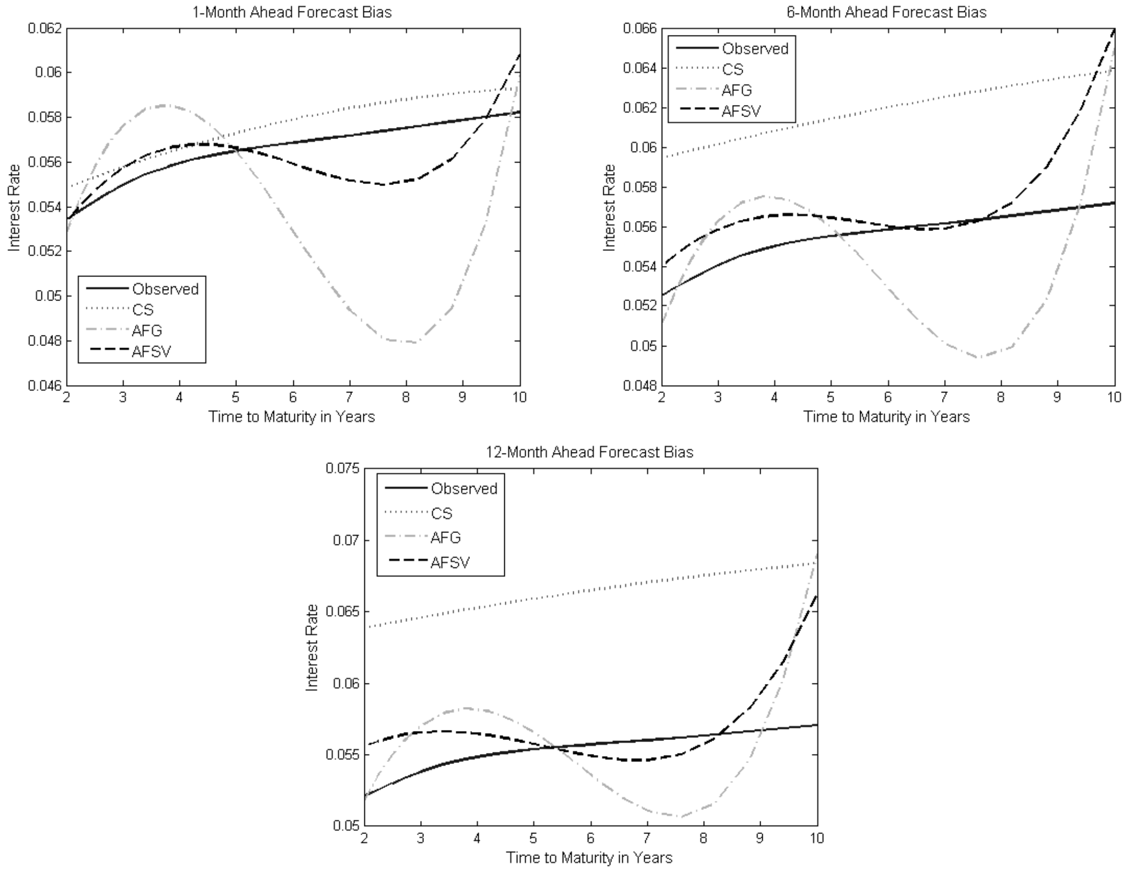

Fig. 5. Out-of-Sample Averaged Forecasts and Observed Yield Curves

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. Solid line represents the observed yield curve, dotted line represents the CS version, dash-dotted line represents the AFG version, and dashed line represents the AFSV version. In-sample period ranges from January 1972 to December 1994.

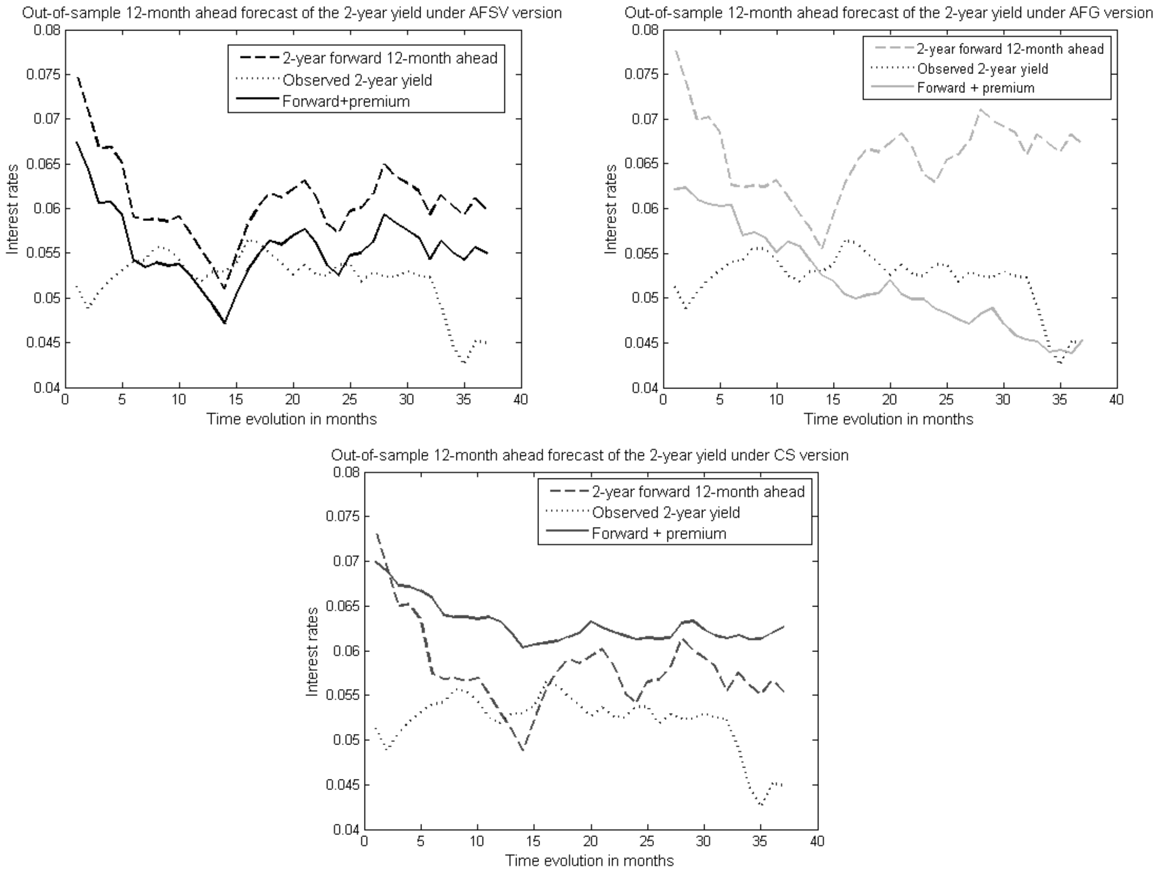

Fig. 6. 12-Month Ahead Out-of-Sample Forecasting of the 2-Year Yield

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.

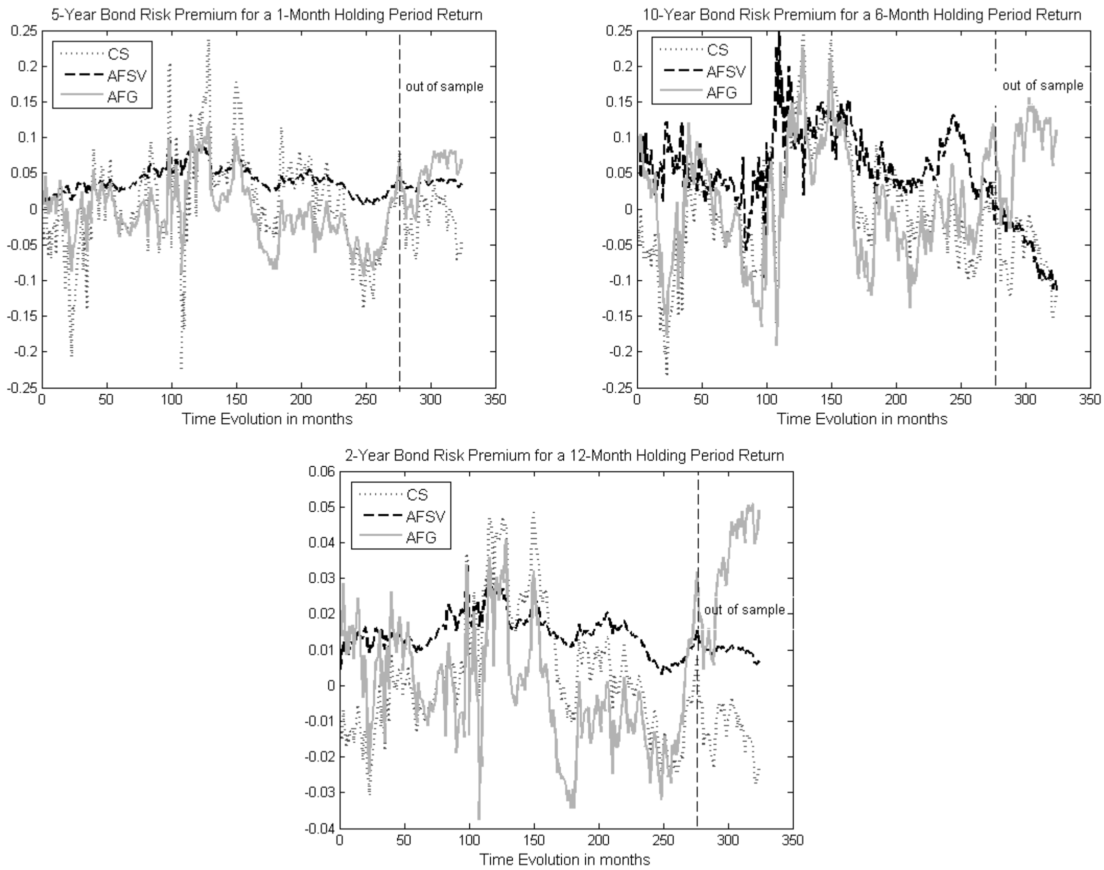

Fig. 7. Bond Risk Premium for Different Maturities and Forecasting Horizons

This picture presents the time series of bond risk premium implied by each model version, for different maturities and forecasting horizons. Solid line represents AFG, dashed line represents AFSV, and 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.