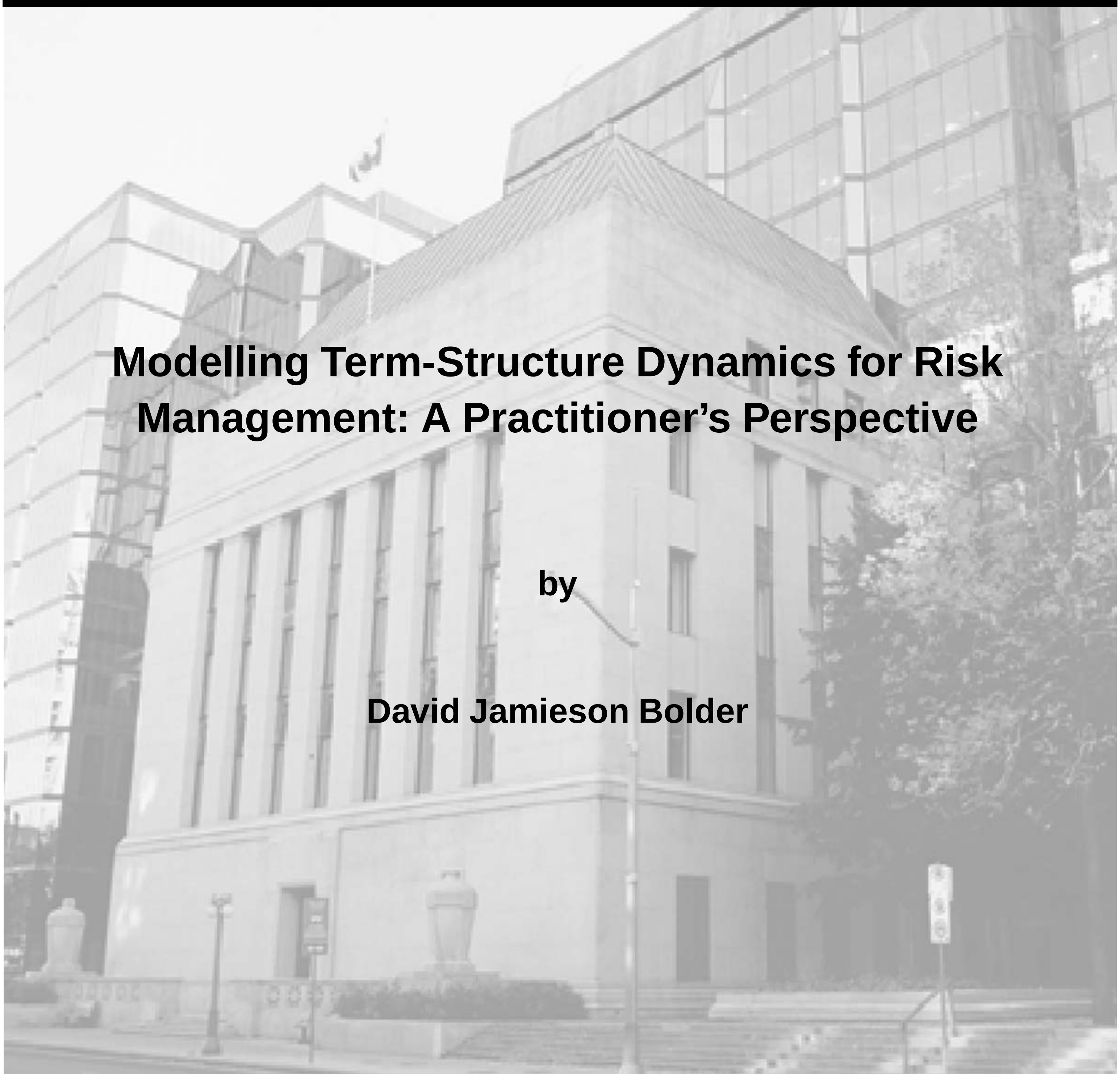
Bank of Canada Banque du Canada

Working Paper 2006-48 / Document de travail 2006-48



Modelling Term-Structure Dynamics for Risk Management: A Practitioner's Perspective

by

David Jamieson Bolder

ISSN 1192-5434

Printed in Canada on recycled paper

Bank of Canada Working Paper 2006-48

December 2006

Modelling Term-Structure Dynamics for Risk Management: A Practitioner's Perspective

by

David Jamieson Bolder

Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
dbolder@bankofcanada.ca

Contents

Acknowledgements

I would like to particularly thank Scott Gusba from the Bank of Canada for a number of useful mathematical discussions and technical assistance. I would also like to specifically thank Greg Bauer from the Bank of Canada for important and timely direction on the structure and content of the paper. Comments from Fousseni Chabi-Yo, Jun Yang, Antonio Diez de los Rios, and Scott Hendry from the Bank of Canada are also much appreciated. Finally, I would also like to thank Jeremy Graveline from the University of Minnesota, Mark Reesor from the University of Western Ontario, and Adam Kobor from the World Bank Treasury for their helpful input. All thanks are, of course, without implication. The responsibility for any mistakes, errors, or omissions is entirely my own.

Abstract

Modelling term-structure dynamics is an important component in measuring and managing the exposure of portfolios to adverse movements in interest rates. Model selection from the enormous term-structure literature is far from obvious and, to make matters worse, a number of recent papers have called into question the ability of some of the more popular models to adequately describe interest rate dynamics. The author, in attempting to find a relatively simple term-structure model that does a reasonable job of describing interest rate dynamics for risk-management purposes, examines two sets of models. The first set involves variations of the Gaussian affine term-structure model by modestly building on the recent work of Dai and Singleton (2000) and Duffee (2002). The second set includes and extends Diebold and Li (2003). After working through the mathematical derivation and estimation of these models, the author compares and contrasts their performance on a number of in- and out-of-sample forecasting metrics, their ability to capture deviations from the expectations hypothesis, and their predictions in a simple portfolio-optimization setting. He finds that the extended Nelson-Siegel model and an associated generalization, what he terms the “exponential-spline model,” provide the most appealing modelling alternatives when considering the various model criteria.

JEL classification: C0, C6, E4, G1
Bank classification: Interest rates; Econometric and statistical methods; Financial markets

Résumé

La modélisation de la dynamique de la structure des taux d’intérêt est un élément important de la mesure et de la gestion de l’exposition d’un portefeuille aux mouvements défavorables des taux d’intérêt. Il est toutefois difficile de choisir un modèle parmi ceux recensés dans la vaste littérature consacrée au sujet, tout particulièrement depuis la parution de récents articles qui remettent en question la capacité de certains des modèles les plus utilisés à décrire la dynamique des taux. L’auteur cherche à mettre au point un modèle simple qui parvienne relativement bien à rendre compte de cette dynamique aux fins de la gestion des risques. Pour ce faire, il examine deux catégories de modèles. Le premier modèle étudié consiste en une variante du modèle gaussien affine décrit par Dai et Singleton (2000) et Duffee (2002). Les modèles de la seconde catégorie s’inspirent, en les prolongeant, des travaux de Diebold et Li (2003). Après avoir présenté la dérivation mathématique de ces modèles et les avoir estimés, l’auteur compare, sur la base de différents critères, leur capacité à prévoir l’évolution des taux durant la période d’estimation et au-delà de celle-ci, leur capacité à rendre compte des écarts par rapport à l'hypothèse relative aux attentes, de même que leur pouvoir de prédiction dans un cadre simple d'optimisation des portefeuilles. Il constate que le modèle étendu de Nelson-Siegel et une variante généralisée de celui-ci, qu'il appelle « modèle spline exponentiel », constituent les modèles les plus prometteurs eu égard aux divers critères de sélection retenus.

Classification JEL : C0, C6, E4, G1
Classification de la Banque : Taux d'intérêt; Méthodes économétriques et statistiques; Marchés financiers

1 Introduction

Finance practitioners often use models of the term structure of interest rates to assist in addressing everyday problems. The type of model that they will employ depends, not surprisingly, on the specific problem at hand. One can, broadly speaking, decompose into two separate categories the problems related to interest rates that practitioners face: pricing and risk-management problems. A pricing problem involves the determination—conditioning on the set of current market data—of the value of some interest rate contingent claim. In this setting, through the fundamental theorem of finance, one works entirely with the dynamics of interest rates under an equivalent martingale measure, $\mathbb{Q}$.[1] This permits the pricing of interest rate derivative contracts without explicitly considering investors' risk preferences. Moreover, one can generally determine the model parameters by calibrating the model to a set of financial contracts at the current time.[2] Interest rate risk-management problems, however, involve measuring the exposure of one's portfolio to adverse movements in the term structure of interest rates. In this case, it is essential to model the dynamics of the term structure through time under the physical, or real-world, probability measure, denoted $\mathbb{P}$. This implies that one must explicitly consider investors' risk attitudes. The implication is that one needs to use a historical data set to estimate model parameters.

There is a catch. The financial literature related to describing dynamics of the term structure of interest rates under both $\mathbb{P}$ and $\mathbb{Q}$ is enormous. Many models, for example, can be used in either a pricing or risk-management setting. Not only is the number of alternative models bewildering, but their mathematical complexity and variety can be overwhelming. Even worse, the estimation algorithms used to determine the model parameters are also often quite involved. Finally, there is an enormous secondary literature dedicated, it appears, to entirely debunking the existing set of models. After a review of the literature, one could reasonably come to the conclusion that there *does not exist* a model that can adequately describe the dynamics of the term structure of interest rates. Yet, for a finance practitioner, there are many problems whose solution depends exactly on these interest rate dynamics. For the management of the Government of Canada's foreign and domestic debt portfolios, for example, an understanding of term-structure dynamics is essential for the selection of appropriate financing and investment strategies.

Thus, a practitioner is faced with a difficult situation. On the one hand, the finance literature provides simultaneously excessive and limited guidance on the best way to describe term-structure dynamics. Yet, on the other hand, a reliable, dynamic term-structure model is essential for solving the problems faced by practitioners on a daily basis. The practitioner's question is *not*, therefore, what is theoretically the most appealing model.

[1] The fundamental theorem of finance (briefly) states that an equivalent martingale measure, $\mathbb{Q}$, exists if and only if there is an absence of arbitrage; market completeness gives us a unique $\mathbb{Q}$. This is useful, because all contingent claims can be represented as the discounted expectation taken with respect to $\mathbb{Q}$.

[2] See Brigo and Mercurio (2001) for a detailed discussion of model calibration.

Instead, the question becomes "which of the existing models does the best job of describing the dynamics of the term structure of interest rates?" The next question is "under which probability measure?" We simplify things somewhat in this paper by focusing only on the risk-management problem, and thus restrict our attention to the physical measure. This still remains something of a daunting task, considering the wide range of risk-management models and the additional complication of measuring investors' risk attitudes. The objective of this paper, therefore, is to try to find a relatively simple term-structure model that does a reasonable job of describing interest rate dynamics under the physical measure. If one plans to compare a number of different models, however, it is necessary to predefine a set of criteria to help distinguish the best model or models. We hope to achieve this in a number of steps by:

- considering the intuition and details of the derivation of a number of different term-structure models;
- estimating the model parameters for each of these approaches using a common dataset;
- considering the ability of these models to forecast, both in- and out-of-sample, zero-coupon rates over different horizons;
- considering the ability of these models to forecast the out-of-sample excess holding-period returns over different horizons;
- using simulation and two well-known econometric tests to determine how well the alternative models capture deviations from the expectations hypothesis, and;
- performing a simplified portfolio-optimization exercise.

A fundamental criterion for model comparison, therefore, is a model's forecasting ability. Why? The central empirical fact about the term structure of interest rates is that the (pure) expectations hypothesis does not hold. There are many ways to express the expectations hypothesis, but in its simplest form it postulates that expected holding-period returns on bonds of different maturities should be equal. It is a well-established fact in the finance literature, however, that the existence of time-varying risk premia implies that holding-period returns for bonds across different maturities are *not* equal. Consequently, in recent years, there has been a move to use model forecasts to examine the ability of a term-structure model to describe the deviations from the expectations hypothesis. Duffee (2002), in particular, states that "if a model produces poor forecasts of future yields, and thus poor forecasts of future bond prices, it is unlikely that the model can shed light on the economics underlying the failure of the expectations hypothesis." In addition to the forecasting exercise, we also use simulation to compute and compare two direct measures of the expectations hypothesis, including the time-honoured LPY-regression test and a relatively new technique proposed by Backus et al. (2001). Finally, we compare how the various term-structure models perform in the context of a simplified portfolio-optimization exercise; this seems quite reasonable, since the principal practical application of these models, from the Bank of Canada's perspective, is found in strategic portfolio decisions.

To accomplish the objectives of this paper, we focus on two different classes of term-structure model. In the first class, we examine some variations on the Gaussian affine term-structure model by modestly building on the recent work of Dai and Singleton (2000), Duffee (2002), and Cheridito, Filipović, and Kimmel (2005). The second class of models, introduced into the literature by Diebold and Li (2003), works directly with interest rates under the physical measure. This is by no means an exhaustive examination of the models in this literature. We have, however, selected a subgroup of models that we believe have the potential to combine relative parsimony of implementation and the ability to describe deviations from the expectations hypothesis.

This paper is a piece of practitioner literature. Our target audience is other financial practitioners—with particular emphasis on central bankers and finance ministry staff—struggling through the enormous literature in an attempt to find a workable model to solve their underlying business problems. For this reason, this paper does not look much like a typical academic paper. For one, there is relatively little in this paper that is new. Second, the exposition is highly detailed and, by necessity, rather lengthy. Our objective is to provide a thorough derivation of each of the models, in the hope is that this will provide additional understanding and flexibility in model implementation and estimation.

The remainder of this paper includes four sections. Section 2 outlines the logic, intuition, and mathematical derivation of the term-structure models considered in this paper. Section 3 provides a detailed description of the estimation algorithms used to determine the model parameters. The results are described in section 4; this includes a review of the different criteria used to distinguish between the models. Section 5 provides some conclusions.

2 The Models

In this work, we will consider two alternative classes of models to describe the dynamics of the term structure of interest rates. The first class of models, which we will refer to as the *theoretical* class, is the collection of affine term-structure models. Affine term-structure models have been, and continue to be, the workhorse model in the finance literature. The advantage of these models is their analytical tractability as well as their appealing theoretical foundation. In recent years, however, there has been growing evidence that affine term-structure models have difficulty capturing deviations from the expectations hypothesis. Duffee (2002) and Cheridito, Filipović, and Kimmel (2005) provide some specific advice to help permit greater model flexibility. Questions about these models, however, remain. For this reason, we also consider a class of models recently introduced into the literature by Diebold and Li (2003). This collection of models, which we will refer to as the *empirical* class, are essentially a time-series description of the term structure of interest rates.[3] The disadvantage of this approach is the lack of a theoretical model foundation. This appears to be offset by better model forecasting performance.

There is one slight complication. The class of affine models is extensive, while the class of empirical models is rather small. We solve the first issue by restricting the set of affine models and deal with the second problem by introducing a number of alternative empirical models. In particular, we restrict our attention to the three-factor Gaussian model; this is the so-called $A_0(3)$ model in the vernacular of Dai and Singleton (2000). There are two reasons for this choice. First, the $A_0(3)$ model is in many respects the simplest of affine term-structure models. Second, there is some evidence that this simplicity, when combined with a sufficiently flexible specification of the market price of risk, leads to superior yield forecasts relative to other, more complex, affine term-structure models. We use the form of the market price of risk suggested by Leippold and Wu (2000) and Duffee (2002) with different restrictions on the form of the matrix pre-multiplying the state variables.

The empirical model suggested by Diebold and Li (2003) basically writes the term structure at a given point in time as a linear combination of a set of underlying factors. We generalize this idea somewhat and introduce—following from previous work in Bolder and Gusba (2002)—two alternative empirical models. The logic behind these models is predicated entirely on the logic proposed by Diebold and Li (2003), albeit with a different mathematical form. We do this, again, for two reasons. First, it seems natural to generalize the basic model and to examine the relative performance of the special case suggested by Diebold and Li (2003). Second, the Nelson-Siegel model, as a description of the term structure of zero-coupon rates at a given point in time, performs rather poorly—in terms of goodness of fit—relative to other alternatives. We wondered if a model that performs better in fitting bond prices at a given point in time would also do a better job of describing term-structure dynamics as well as deviations from the expectations hypothesis.

In the following sections, we perform a detailed derivation of each of the models examined in this paper. At first glance, these details may appear somewhat superfluous. We would argue, however, that from a practitioner's perspective, the derivation of the model is of paramount importance, because it permits a greater degree of flexibility in the model's implementation. Because of tight space limitations, academics often present their models in highly succinct, even terse, form. Many details are spread across different publications or skipped entirely. We operate under no such constraints and thus seek to provide a comprehensive view of the models. Model derivation also adds a higher degree of comfort with the model and helps build intuition about it and its parameters. This paper's comprehensiveness will also help us later with the estimation of the parameter set.

[3] There is, for example, no notion of risk premia in this model. The model works directly with interest rates under the physical measure ($\mathbb{P}$) and, as such, does not permit us to move into the risk-neutral setting.

2.1 A theoretical model

The theoretical affine term-structure model derived in this section is constructed from first principles. In particular, the model uses the fundamental theorem of finance to construct a description of the dynamics of the term structure of interest rates. In the following pages, we will work through the derivation of the affine term-structure model using what is termed the pricing-kernel approach.[4] The general idea is that pure-discount bond prices are a function of some (generally unobservable) set of state variables. Before we can start, we need to add a bit of structure, introduce some notation, and provide some technical conditions. We begin with a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$, a complete and right-continuous filtration, $\{\mathcal{F}_t, t \in [0, \mathcal{T}]\}$, where $\mathcal{T}$ represents some strictly positive future point in time. We further introduce $W \in \mathbb{R}^n$ as an n-dimensional Wiener process defined on $(\Omega, \mathcal{F}, \mathbb{P})$, where $\{\mathcal{F}_t, t \in [0, \mathcal{T}]\}$ has the usual $\mathbb{P}$-augmentation of the natural filtration of the Wiener processes.

In our derivation, we will have to specify the usual suspects, including the instantaneous short rate, the market price of risk, and the dynamics of our state variables. Let us begin with our state variables. Let $X_t \in \mathbb{R}$ be a Markov process taking values in an open set $\mathcal{D} \subset \mathbb{R}^n$ with dynamics described by the underlying stochastic differential equation,

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{1}$$

where $\mu(X_t) \in \mathbb{R}^{n\times 1}$ and $\sigma(X_t) \in \mathbb{R}^{n\times n}$ are deterministic functions of the state variables. As you can see, this is a fairly general specification for the dynamics of the state variable. It will turn out in practice to be a multidimensional Ornstein-Uhlenbeck process.

The pricing kernel approach is conceptually quite straightforward. The idea is to develop two alternative approaches to determining the price of a pure-discount bond of arbitrary tenor. One representation comes from the mathematical-finance approach of writing the price of a contingent claim as an expectation taken with respect to an equivalent martingale measure. The alternative representation comes from the economic idea of a pricing kernel. Essentially, the pricing kernel relates future cash flows, say $\{C_t, t \in [t, \mathcal{T}]\}$, to today's price. It has a number of different names, including the state-price density and stochastic discount factor. We denote the pricing kernel as $\{\xi_t, t \in [0, \infty\}$, and the price of a security, $\psi(t)$, paying cash flows $\{C_t, t \in [t, \mathcal{T}]\}$ can be written as

$$\psi(t) = \mathbb{E}^{\mathbb{P}}\left[\int_t^{\mathcal{T}} \frac{\xi_s C_s}{\xi_t} ds \,\middle|\, \mathcal{F}_t\right]. \tag{2}$$

There are a variety of results that discuss the relationship between the absence of arbitrage and the existence of the pricing kernel. In short, modulo a number of technical conditions, one can show that the absence of arbitrage implies the existence of a pricing kernel.

[4] For a much more detailed discussion of affine term-structure models, see Bolder (2001) and the many excellent academic references contained in that document.

The result in equation (2), while interesting, is not yet useful in its current form for modelling interest rates. We are in the business of modelling the term structure of interest rates and, as such, are not interested in an arbitrary security, $\psi(t)$. Instead, we are interested in the pure-discount bond price function, which we denote as

$$P(t,T) \equiv P(X_t, T-t) \equiv P(X_t, \tau), \tag{3}$$

to denote its dependence on the state variables, X_t, and the tenor of the security, $T-t$. For notation convenience, we are generally going to let $\tau = T - t$. This will help keep some lengthy expressions later in the text from becoming completely unwieldy. If we can describe, for a given point in time, the collection of pure-discount bond prices, then we can characterize the entire term structure of interest rates at that point in time. Recall that the zero-coupon rate function is given as

$$z(X_t, \tau) = -\frac{\ln P(X_t, \tau)}{\tau}. \tag{4}$$

The pure-discount bond has a unit cash flow occurring at time T. This implies that the cash flow is merely $C_T = 1$ for some $T \in [0, \mathcal{T}]$. More specifically,

$$P(X_t, 0) = P(X_t, T-T) = 1. \tag{5}$$

Thus, if we cast our pure-discount bond price into equation (2), we have

$$\begin{aligned} P(X_t, \tau) &= \mathbb{E}^{\mathbb{P}}\left[\int_t^{\mathcal{T}} \frac{\xi_T \overbrace{P(X_t, 0)}^{\text{Equation (5)}}}{\xi_t} ds \,\middle|\, \mathcal{F}_t \right], \\ &= \mathbb{E}^{\mathbb{P}}\left[\frac{\xi_T}{\xi_t} \,\middle|\, \mathcal{F}_t \right]. \end{aligned} \tag{6}$$

Equation (6) is quite a useful result, particularly considering that we can also state, from the fundamental theorem of finance, that the price of any contingent claim is its discounted expectation taken with respect to the equivalent martingale measure, $\mathbb{Q}$. Since the pure-discount bond price is a contingent claim on the interest rate, we can write

$$\begin{aligned} P(X_t, \tau) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r(X_s)ds} \overbrace{P(X_t, 0)}^{\text{Equation (5)}} \,\middle|\, \mathcal{F}_t \right], \\ &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r(X_s)ds} \,\middle|\, \mathcal{F}_t \right], \end{aligned} \tag{7}$$

where $r(X_t)$ is the instantaneous short rate. Using L'Hopital's rule, we define it as

$$\begin{aligned} r(X_t) &= \lim_{T \downarrow t} z(X_t, T-t), \\ &= \lim_{T \downarrow t} -\frac{\ln P(X_t, \tau)}{\tau}, \\ &= \lim_{T \downarrow t} -\frac{\frac{\partial \ln P(X_t,\tau)}{\partial T}}{\frac{\partial \tau}{\partial T}}, \\ &= -\frac{\partial \ln P(X_t, \tau)}{\partial T}, \end{aligned} \tag{8}$$

where this final derivative is the right-hand derivative evaluated at $\tau = 0$. The function $P(X_t, \tau)$ is assumed to be sufficiently differentiable and continuous for this limit to exist. At this point, observation of equations (6) and (7) describing $P(X_t, \tau)$ reveals that the expectations are taken with respect to different probability measures. We can, using the Cameron-Girsanov theorem, rewrite equation (7) as

$$P(X_t, \tau) = \mathbb{E}^{\mathbb{P}} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} e^{-\int_t^T r(X_s) ds} \bigg| \mathcal{F}_t \right], \tag{9}$$

where $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is the Radon-Nikodým derivative. It is an exponential martingale of the following form:

$$\begin{aligned} \frac{d\mathbb{Q}}{d\mathbb{P}} &= \exp\left(-\int_t^T \gamma(X_s)^T dW_s - \frac{1}{2} \left\langle -\int_t^T \gamma(X_s)^T dW_s, -\int_t^T \gamma(X_s)^T dW_s \right\rangle \right), \\ &= \exp\left(-\int_t^T \gamma(X_s)^T dW_s - \frac{1}{2} \int_t^T \left| \gamma(X_s)^T \gamma(X_s) \right| ds \right), \end{aligned} \tag{10}$$

where $\langle X_t, X_t \rangle$ denotes the quadratic-variation process of X_t and the $\mathcal{F}_t$-adapted process, $\gamma(X_t)$, represents the market price of risk. Note that under the physical measure, $\mathbb{P}$, risk-averse market participants require some compensation for the risk associated with holding fixed-income instruments.

Now, we have two expressions for $P(X_t, \tau)$—in equations (6) and (9)—written as expectations taken with respect to the same measure. Moreover, as of time t, these two expressions must be equal for all t. This is a consequence of the absence of arbitrage. This, however, can only be true if the integrands are equal:

$$\frac{\xi_T}{\xi_t} = \frac{d\mathbb{Q}}{d\mathbb{P}} e^{-\int_t^T r(X_s) ds}. \tag{11}$$

If we shift the time interval from $[t, T]$ to $[0, t]$ and observe that the value of the pricing kernel at time 0 is $\xi_0 = 1$, we have the following relationship:

$$\xi_t = \frac{d\mathbb{Q}}{d\mathbb{P}} e^{-\int_0^t r(X_s) ds}. \tag{12}$$

This expression represents the equilibrium relationship between the pricing kernel and the instantaneous short rate. It is exactly this relationship that we will exploit to solve for the bond pricing function, $P(X_t, \tau)$.

We want to understand the infinitesimal dynamics of both ξ_t and $P(X_t, \tau)$. This will require a bit of clever manipulation and a few applications of Itô's theorem. First, by plugging equation (10) into our pricing-kernel equation (12) and simplifying, we arrive at

$$
\begin{aligned}
\xi_t &= \frac{d\mathbb{Q}}{d\mathbb{P}} e^{-\int_0^t r(X_s)ds}, \qquad (13)\\
&= \underbrace{\exp\left(-\int_0^t \gamma(X_s)^T dW_s - \frac{1}{2}\int_0^t \left|\gamma(X_s)^T\gamma(X_s)\right| ds\right)}_{\text{Equation 10}} \exp\left(-\int_0^t r(X_s)ds\right),\\
&= \exp\left(-\int_0^t r(X_s)ds - \int_0^t \gamma(X_s)^T dW_s - \frac{1}{2}\int_0^t \left|\gamma(X_s)^T\gamma(X_s)\right| ds\right).
\end{aligned}
$$

To find $d\xi_t$, it is necessary to apply Itô's theorem. To use this result, let us first make a simple change of variables,

$$
\xi_t = \exp(Y_t), \tag{14}
$$

where,

$$
Y_t = -\int_0^t r(X_s)ds - \int_0^t \gamma(X_s)^T dW_s - \frac{1}{2}\int_0^t \left|\gamma(X_s)^T\gamma(X_s)\right| ds. \tag{15}
$$

If we represent equation (15) in differential notation, we have

$$
\begin{aligned}
dY_t &= \left(-r(X_t) - \frac{1}{2}\gamma(X_t)^T\gamma(X_t)\right) dt - \gamma(X_t)^T dW_t, \qquad (16)\\
&= a(X_t)dt + b(X_t)dW_t.
\end{aligned}
$$

Since $\xi_t = \exp(Y_t)$, we apply Itô's Lemma to the function $F(x,t) = \exp(x)$. This implies that $\frac{\partial F(x,t)}{\partial t}$ vanishes. By Itô, therefore, we have

$$
\begin{aligned}
d\xi_t &= \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial x}dY(t) + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}d\langle Y\rangle_t, \qquad (17)\\
&= \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial x}(a(X_t)dt + b(X_t)dW_t) + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}b(X_t)^2 dt,\\
&= \left(\frac{\partial F}{\partial x}\cdot a(X_t) + \underbrace{\frac{\partial F}{\partial t}}_{=0} + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\cdot b(X_t)^2\right) dt + \frac{\partial F}{\partial x}\cdot b(X_t)\, dW_t,\\
&= \left(e^{Y_t}a(X_t) + \frac{e^{Y_t}}{2}\cdot b(X_t)^2\right) dt + e^{Y_t}\cdot b(X_t)\, dW_t,\\
&= \exp(Y_t)\left[\left(-r(X_t) - \frac{1}{2}\gamma(X_t)^T\gamma(X_t) + \frac{1}{2}\gamma(X_t)^T\gamma(X_t)\right) dt - \gamma(X_t)^T dW_t\right],\\
&= \xi_t\left[-r(X_t)dt - \gamma(X_t)^T dW_t\right].
\end{aligned}
$$

Therefore, it follows from this development that

$$\begin{aligned}\frac{d\xi_t}{\xi_t} &= -r(X_t)dt - \gamma(X_t)dW_t, \\ &= a_1(X_t)dt + b_1(X_t)dW_t,\end{aligned} \tag{18}$$

where $a_1(X_t) = -r(X_t)$ and $b_1(X_t) = -\gamma(X_t)$. This expression will turn out to be quite useful.

We will leave $P(X_t, \tau)$ in its general form for the moment and make an observation. Recall that, from equation (6),

$$P(X_t, T-t) = \mathbb{E}^{\mathbb{P}}\left[\frac{\xi_T}{\xi_t}\middle|\mathcal{F}_t\right].$$

This expression, given the $\mathcal{F}_t$-measurability of ξ_t and the definition of a pure-discount bond, can be written as

$$P(X_t, T-t)\xi_t = \mathbb{E}^{\mathbb{P}}\left[\xi_T \underbrace{P(X_t, T-T)}_{=1}\middle|\mathcal{F}_t\right]. \tag{19}$$

From the martingale property (i.e., $\mathbb{E}(X_T|\mathcal{F}_t) = X_t$), we can see that $P(X_t, T-t)\xi_t$ is a martingale. Thus, it follows that

$$\begin{aligned}\mathbb{E}\left[d\left(P(X_t, T-t)\xi_t\right)\right] &= 0, \\ \mathbb{E}\left[\frac{d\left(P(X_t, T-t)\xi_t\right)}{P(X_t, T-t)\xi_t}\right] &= 0.\end{aligned} \tag{20}$$

What can we do now? The game plan is to expand equation (20) and to remark, from the fact that $P(X_t, T-t)\xi_t$ is a martingale, that the drift term must be equal to zero. By setting this drift term equal to zero, we eliminate the Brownian motion terms and arrive at a partial differential equation. An assumption is then made regarding the form of $P(X_t, \tau)$, which ensures the existence of the partial differential equation. Finally, we observe that the partial differential equation can, fortunately, be reduced to a system of ordinary differential equations that can be solved numerically with a finite-difference algorithm.

We can expand equation (20) with a stochastic version of the product rule. In particular, it is true that, for two continuous semi-martingales, X_t and Y_t,

$$d(X_tY_t) = X_tdY_t + Y_tdX_t + d\langle X, Y\rangle_t,$$

where $\langle X, Y\rangle_t$ denotes the co-quadratic variation process of X_t and Y_t. Applying this result to equation (20) yields

$$\begin{aligned}\mathbb{E}\left[\frac{d\left(P(X_t, \tau)\xi_t\right)}{P(X_t, \tau)\xi_t}\right] &= \frac{1}{P(X_t, \tau)\xi_t}\mathbb{E}\left[P(X_t, \tau)d\xi_t + \xi_t dP(X_t, \tau) + d\langle P(X_t, \tau), d\xi_t\rangle\right], \\ &= \mathbb{E}\left[\frac{d\xi_t}{\xi_t} + \frac{dP(X_t, \tau)}{P(X_t, \tau)} + \frac{d\langle P(X_t, \tau), d\xi_t\rangle}{P(X_t, \tau)\xi_t}\right].\end{aligned} \tag{21}$$

We have an expression for $\frac{d\xi_t}{\xi_t}$. To evaluate equation (21), therefore, we need to understand the infinitesimal dynamics of $P(X_t,\tau)$. This follows from another application of Itô's theorem. In particular,

$$
\begin{aligned}
dP(X_t,\tau) &= \frac{\partial P(X_t,\tau)}{\partial t}dt + \frac{\partial P(X_t,\tau)}{\partial X_t^T}dX_t + \sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 P(X_t,\tau)}{\partial X_{it}X_{jt}} d\langle X_i, X_j\rangle_t, \qquad (22)\\
&= \frac{\partial P(X_t,\tau)}{\partial t}dt + \frac{\partial P(X_t,\tau)}{\partial X_t^T}(\mu(X_t)dt + \sigma(X_t)dW_t) + \sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 P(X_t,\tau)}{\partial X_{it}X_{jt}}\sigma(X_{it})\sigma(X_{jt})dt,\\
&= \underbrace{\left(\frac{\partial P(X_t,\tau)}{\partial t} + \frac{P(X_t,\tau)}{\partial X_t^T}\mu(X_t) + \sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 P(X_t,\tau)}{\partial X_{it}X_{jt}}\sigma(X_{it})\sigma(X_{jt})\right)}_{a_2(X_t)} dt + \underbrace{\frac{\partial P(X_t,\tau)}{\partial X_t^T}\sigma(X_t)}_{b_2(X_t)} dW_t,\\
&= a_2(X_t)dt + b_2(X_t)dW_t.
\end{aligned}
$$

We now have all of the ingredients to evaluate equation (21). The evaluation proceeds as follows:

$$
\begin{aligned}
0 &= \mathbb{E}\left[\frac{d\xi_t}{\xi_t} + \frac{dP(X_t,\tau)}{P(X_t,\tau)} + \frac{d\langle P(X_t,\tau), d\xi_t\rangle}{P(X_t,\tau)\xi_t}\right], \qquad (23)\\
&= \mathbb{E}\left[P(X_t,\tau)\frac{d\xi_t}{\xi_t} + dP(X_t,\tau) + \frac{d\langle P(X_t,\tau), d\xi_t\rangle}{\xi_t}\right],\\
&= \mathbb{E}\left[P(X_t,\tau)\left(\underbrace{a_1(X_t)dt + b_1(X_t)dW_t}_{\text{Equation (18)}}\right) + \underbrace{a_2(X_t)dt + b_2(X_t)dW_t}_{\text{Equation (22)}} + \frac{b_1(X_t)\xi_t b_2(X_t)dt}{\xi_t}\right],\\
&= \mathbb{E}\left[\underbrace{\left(P(X_t,\tau)a_1(X_t) + a_2(X_t) + b_1(X_t)b_2(X_t)\right)}_{a_3(X_t)} dt + \underbrace{\left(P(X_t,\tau)b_1(X_t) + b_2(X_t)\right)}_{b_3(X_t)} dW_t\right],\\
&= \mathbb{E}\left[a_3(X_t)dt + b_3(X_t)dW_t\right].
\end{aligned}
$$

This work brings us, not surprisingly, to a stochastic process with a drift coefficient of $a_3(X_t)$ and a diffusion coefficient of $b_3(X_t)$. Recall that we derived this equation from the fact that $P(X_t,\tau)\xi_t$ is a martingale. As such, it must be that the drift of this process, $a_3(X_t)$, is identically zero. Let us work backwards to determine the implications of this conclusion:

$$
\begin{aligned}
0 &= a_3(X_t), \qquad (24)\\
0 &= P(X_t,\tau)a_1(X_t) + a_2(X_t) + b_1(X_t)b_2(X_t),\\
P(X_t,\tau)r(X_t) &= \frac{\partial P(X_t,\tau)}{\partial t} + \frac{P(X_t,\tau)}{\partial X_t^T}\mu(X_t) + \sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 P(X_t,\tau)}{\partial X_{it}X_{jt}}\sigma(X_{it})\sigma(X_{jt}) - \frac{\partial P(X_t,\tau)}{\partial X_t^T}\gamma(X_t)^T\sigma(X_t),\\
P(X_t,\tau)r(X_t) &= \frac{\partial P(X_t,\tau)}{\partial t} + \frac{P(X_t,\tau)}{\partial X_t^T}(\mu(X_t) - \gamma(X_t)^T\sigma(X_t)) + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 P(X_t,\tau)}{\partial X_{it}X_{jt}}\sigma(X_{it})\sigma(X_{jt}).
\end{aligned}
$$

Here we have a useful result. Equation (24) is a partial differential equation in $P(X_t, \tau)$. To actually make any progress in solving this equation, we need to make a few modelling choices. In particular, it is necessary to:

- postulate a form for the pure-discount bond price function, $P(X_t, \tau)$;
- derive an expression for the instantaneous short rate, $r(X_t)$;
- provide a form for the market price of risk function, $\gamma(X_t)$; and,
- make a choice regarding the form of the coefficients, $\mu(X_t)$ and $\sigma(X_t)$, in the stochastic differential equation.

Once we have made these choices, we can proceed to solve equation (24) and thereby arrive at an explicit expression for the pure-discount bond price. This will then be our characterization of the term-structure of interest rates.

Based on our list of required ingredients, the first step is to postulate a form for $P(X_t, \tau)$ that permits a solution of equation (24). As might be expected, $P(X_t, \tau)$ is assumed to have an exponential-linear form, as follows:

$$P(X_t, \tau) = e^{-c_\tau - b_\tau^T X_t}, \tag{25}$$

where $c_\tau \in \mathbb{R}$ and $b_\tau \in \mathbb{R}^{n\times 1}$ are deterministic functions of τ. Translating equation (25) into words, we have that the pure-discount bond price is an exponentially affine function in the state variables, X_t.[5] Given the form for $P(X_t, \tau)$, we proceed to compute the requisite derivatives in the partial differential equation in (24) with the hope it will lead to some simplification. We have, therefore, that

$$\frac{\partial P(X_t, \tau)}{\partial t} = \left(\frac{\partial c_\tau}{\partial t} + \frac{\partial b_\tau}{\partial t}^T X_t \right) P(X_t, \tau), \tag{27}$$

$$\frac{\partial P(X_t, \tau)}{\partial X_t} = -b_\tau^T P(X_t, \tau), \tag{28}$$

and,

$$\frac{\partial^2 P(X_t, \tau)}{\partial X_t \partial X_t^T} = b_\tau b_\tau^T P(X_t, \tau). \tag{29}$$

[5]This can be extended. Leippold and Wu (2000, 2001, 2003) describe a class of quadratic term structure models of the form,

$$P(X_t, \tau) = e^{-c_\tau - b_\tau^T X_t - X_t^T A_\tau X_t}, \tag{26}$$

where $c_\tau \in \mathbb{R}$, $b_\tau \in \mathbb{R}^{n\times 1}$, and $A_\tau \in \mathbb{R}^{n\times n}$ are deterministic functions of τ. This model nests the affine class of models (i.e., an affine model arises when $A_\tau \equiv 0$ for all τ). That is, the pure-discount bond price becomes an exponentially *quadratic* function in the state variables, X_t.

If we plug equations (27)—(29) into the partial differential equation in (24), we get some simplification:

$$P(X_t,\tau)r(X_t) = \underbrace{\left(\frac{\partial c_\tau}{\partial t} + \frac{\partial b_\tau}{\partial t}^T X_t\right) P(X_t,\tau)}_{\text{Equation (27)}} + \underbrace{-b_\tau^T P(X_t,\tau)}_{\text{Equation (28)}}(\mu(X_t) - \sigma(X_t)\gamma(X_t)) \tag{30}$$
$$+ \operatorname{tr}\left[\underbrace{b_\tau b_\tau^T P(X_t,\tau)}_{\text{Equation (29)}}\sigma(X_t)\sigma(X_t)^T\right],$$
$$r(X_t) = \frac{\partial c_\tau}{\partial t} + \frac{\partial b_\tau}{\partial t}^T X_t - b_\tau^T(\mu(X_t) - \sigma(X_t)\gamma(X_t)) + \frac{1}{2}\operatorname{tr}\left[b_\tau b_\tau^T \sigma(X_t)\sigma(X_t)^T\right],$$

where we see that the double summation in equation (24) is replaced with a trace operator, $\operatorname{tr}(\cdot)$. Observe that we have eliminated a $P(X_t,\tau)$ term from each element of our partial differential equation.

To make any more progress, it is necessary to address the second item on our list of ingredients; that is, we need to derive an expression for the instantaneous short rate, $r(X_t)$. The instantaneous short rate is, as we saw in equation (8), merely the limit of a continuously compounded zero-coupon interest rate as the tenor, $\tau = T - t$, tends to zero. Having specified the form of $P(X_t,\tau)$, we can proceed to evaluate this limit:

$$\begin{aligned} r(X_t) &= \lim_{T\downarrow t} -\frac{\partial \ln P(X_t, T-t)}{\partial T}, \\ &= \lim_{T\downarrow t} -\frac{\partial}{\partial T}\left(\ln\left(e^{-c(T-t)-b(T-t)^T X_t}\right)\right), \\ &= \lim_{T\downarrow t} \frac{\partial c(T-t)}{\partial T} + \frac{\partial b(T-t)}{\partial T}^T X_t, \\ &= c_r + b_r^T X_t, \end{aligned} \tag{31}$$

where $c_r \in \mathbb{R}$ and $b_r \in \mathbb{R}^{n\times 1}$ exist through the assumed continuity and differentiability of these functions. In the context of these models, however, we can think of the values of c_r and b_r as parameters that have to be determined from our estimation algorithm.[6] Ang and Piazzesi (2003) provide a useful motivation of equation (31) by comparing it with the Taylor rule. Essentially, equation (31) states that the short rate, targeted by most monetary authorities, is a linear function of a set of unobservable state variables. The Taylor rule, of course, describes the short-rate dynamics as a linear function of macroeconomic variables such as the output gap and inflation. Ang and Piazzesi (2003) consider both macreconomic and unobservable state variables in their approach.

Our penultimate ingredient involves the specification of the market price of risk, $\gamma(X_t)$. Leippold and Wu (2000) suggest the following choice, which we will adopt:

$$\gamma(X_t) = b_\gamma + A_\gamma X_t, \tag{32}$$

[6] We will see that the system of ordinary differential equations that arises from collecting the terms in equation (30) cannot be solved if we leave c_r and b_r in explicit derivative form.

where $b_\gamma \in \mathbb{R}^{n\times 1}$ and $A_\gamma \in \mathbb{R}^{n\times n}$. The idea is that the market price of risk is an affine function of the state variable, X_t. The advantage of this choice is that it permits the market price of risk, and thus risk premia, to vary through time. Duffee (2002) introduces a clever technique to allow more flexibility for the market price of risk; these are termed essentially affine models. The basic intuition is that it has been generally assumed that the market price of risk process is a fixed multiple to the instantaneous variance of the state-variable vector. This is reasonable insofar as risk and risk premia go to zero together, but the positivity of variance implies that risk premia can never change sign over time. The primary contribution of Duffee (2002), therefore, is to provide a general approach to relax this restriction on the market price of risk. Cheridito, Filipović, and Kimmel (2005) simplify the mathematical implementation of the approach suggested by Duffee (2002). In the context of the $A_0(3)$, however, the approach is relatively simple, given that the state variables follow Ornstein-Uhlenbeck processes. By providing an affine form of the market price of risk, as described in equation (32), the necessary flexibility is provided. Life becomes somewhat more complicated when one introduces square-root processes for the state variables.

Our final ingredient requires us to make specific choices for the general state-variable coefficients, $\mu(X_t)$ and $\sigma(X_t)$. Let us, therefore, define the instantaneous drift and volatility for X_t as

$$\mu(X_t) = \kappa(\theta - X_t), \tag{33}$$

with $\kappa \in \mathbb{R}^{n\times n}$ and $\theta \in \mathbb{R}^{n\times 1}$:

$$\sigma(X_t) = \Sigma S(X_t) \in \mathbb{R}^{n\times n}, \tag{34}$$

where $\Sigma \in \mathbb{R}^{n\times n}$ is the Cholesky decomposition of the instantaneous correlation matrix, and $S(X_t) \in \mathbb{R}^{n\times n}$ is a diagonal matrix where

$$[S(X_t)]_{ii} = \alpha_i + \beta_i^T X_t, \tag{35}$$

for $i = 1, ..., n$. Often, the collection of β vectors is denoted $\mathcal{B} = \{\beta_1, ..., \beta_n\}$. Dai and Singleton (2000) perform a thorough econometric specification analysis of the class of affine models that permits us to add some restrictions to the form of $\mu(X_t)$ and $\sigma(X_t)$. In particular, they introduce a classification scheme for different affine models. A given affine model is a member of the class $A_m(n)$, where n denotes the number of state variables and m represents the number of state variables that influence the conditional variance. In other words, m is the number of state variables whose dynamics follow a square-root process. We will be considering models of the form $A_0(3)$, which implies that the three state variables each follow an Ornstein-Uhlenbeck process. Dai and Singleton (2000) further demonstrate that, to econometrically identify these models, $\kappa\theta = 0$, $\mathcal{B} = 0$, $\{\alpha_i = 0, i = 1, ..., n\}$,

$\Sigma = I \in \mathbb{R}^{n\times n}$, and κ must be lower diagonal with positive eigenvalues.[7] This implies that

$$\mu(X_t) = -\kappa X_t, \tag{36}$$

and

$$\sigma(X_t) = I \in \mathbb{R}^{n\times n}. \tag{37}$$

The final step before attempting to solve equation (30) is to recall a few properties of the trace operator. In particular, for the constant $\alpha \in \mathbb{R}$ and two matrices A and B, and where the products AB and BA make sense, it is true that

$$\begin{aligned} \mathrm{tr}(\alpha A) &= \alpha \cdot \mathrm{tr}(A), \\ \mathrm{tr}(A) &= \mathrm{tr}(A^T), \\ \mathrm{tr}(\alpha) &= \alpha, \\ \mathrm{tr}(AB) &= \mathrm{tr}(BA). \end{aligned}$$

Now, plugging equations (31) to (34) into our partial differential equation in (30), we have

$$\begin{aligned} r(X_t) &= \frac{\partial c_\tau}{\partial t} + \frac{\partial b_\tau}{\partial t}^T X_t - b_\tau^T(\mu(X_t) - \underbrace{\sigma(X_t)}_{I\in\mathbb{R}^{n\times n}} \gamma(X_t)) + \frac{1}{2}\mathrm{tr}\left[b_\tau b_\tau^T \underbrace{\sigma(X_t)\sigma(X_t)^T}_{I\in\mathbb{R}^{n\times n}} \right], \\ &- \underbrace{(c_r + b_r^T X_t)}_{\text{Equation (31)}} + \frac{\partial c_\tau}{\partial t} + \frac{\partial b_\tau}{\partial t}^T X_t - b_\tau^T(-\kappa X_t - \underbrace{(b_\gamma + A_\gamma X_t)}_{\text{Equation (32)}}) + \frac{1}{2}\mathrm{tr}\left(b_\tau b_\tau^T\right) = 0. \end{aligned} \tag{38}$$

Inspection of equation (38) reveals that we will have two types of terms: constant terms and linear terms in X_t. If we collect these two sets of coefficents, they must be identically zero for all τ for equation (38) to hold. We will find that these two coefficients are, in fact, ordinary differential equations. The first step in getting to this system of ordinary differential equations is to expand each of the expressions in equation (38) and start collecting terms, as follows:

$$\begin{aligned} -c_r - b_r^T X_t + \frac{\partial c_\tau}{\partial t} + \frac{\partial b_\tau}{\partial t}^T X_t + b_\tau^T \kappa X_t + b_\tau^T\left(b_\gamma + A_\gamma X_t\right) + \frac{1}{2}\mathrm{tr}\left(b_\tau b_\tau^T\right) = 0. \\ \left(\underbrace{\frac{\partial c_\tau}{\partial t} - c_r + b_\tau^T b_\gamma \frac{1}{2}\mathrm{tr}\left(b_\tau^T b_\tau\right)}_{\text{Constant terms}} \right) + \left(\underbrace{\frac{\partial b_\tau}{\partial t}^T - b_r^T + b_\tau^T \kappa + b_\tau^T A_\gamma}_{\text{Linear (i.e., } X_t \text{) terms}} \right) X_t = 0. \end{aligned} \tag{39}$$

[7] This final condition ensures that the time evolution of the state variable vector is stable. Positive eigenvalues for a lower diagonal matrix imply that all of the diagonal elements—or the mean-reversion coefficient of the individual elements of the state-variable vector—are positive.

We can also simplify the expression inside the trace operator, by using the previously cited trace properties. Specifically, careful inspection of equation (39) reveals that we have rotated the order of the multiplication of the terms inside the trace operator (i.e., $\mathrm{tr}(AB) = \mathrm{tr}(BA)$) such that the dot product collapses to a scalar. In other words, $b_\tau^T b_\tau \in \mathbb{R}$ is a singleton and thus we can eliminate the trace operator:

$$\left(\frac{\partial c_\tau}{\partial t} - c_r + b_\tau^T b_\gamma + \frac{b_\tau^T b_\tau}{2} \right) + \left(\frac{\partial b_\tau}{\partial t}^T - b_r^T + b_\tau^T \kappa + b_\tau^T A_\gamma \right) X_t = 0. \tag{40}$$

Elimination of the trace operator permits us to write a system of ordinary differential equations. Moreover, if we solve the following system of ordinary differential equations, we will have the deterministic functions c_τ and b_τ, bond tenor (i.e., τ) that—along with the current value of the state variables, X_t—describe the price of a pure-discount bond at a given point in time. The first differential equation, for the scalar-valued function c_τ, is

$$\begin{aligned} 0 &= \frac{\partial c_\tau}{\partial t} - c_r + b_\tau^T b_\gamma + \frac{b_\tau^T b_\tau}{2}, \\ \frac{\partial c_\tau}{\partial t} &= c_r - b_\tau^T b_\gamma - \frac{b_\tau^T b_\tau}{2}. \end{aligned} \tag{41}$$

The second group of ordinary differential equations, for the vector-valued function b_τ, is

$$\begin{aligned} 0 &= \left(\frac{\partial b_\tau}{\partial t}^T - b_r^T + b_\tau^T \kappa + b_\tau^T A_\gamma \right)^T, \\ \frac{\partial b_\tau}{\partial t} &= b_r - \kappa^T b_\tau - A_\gamma^T b_\tau. \end{aligned} \tag{42}$$

We numerically solve the system of ordinary differential equations in (41) and (42) using a fourth-order Runge-Kutta method.[8] The initial conditions are $c(0) = 0$ and $b(0) = 0$.

There are, in the base version of this model, 22 parameters to be estimated. The instantaneous short-rate parameters, c_r and b_r, generate four parameters. The mean-reversion matrix, κ, with its lower-diagonal form, leads to six parameters. Finally, there are 12 market-price-of-risk parameters: three from b_γ and nine from A_γ. Since this large number of model parameters must be estimated in a non-linear optimization setting, we naturally become concerned about the curse of dimensionality. We therefore decide to consider how we might be able to reduce the number of parameters. Our focus is on the matrix $A_\gamma \in \mathbb{R}^{n \times n}$. In addition to the fully unrestricted form of A_γ (nine parameters), we also consider a symmetric version of A_γ (six parameters) and a diagonal version of A_γ (three parameters). This implies that we will consider three different implementations of the $A_0(3)$ model with varying restrictions on the A_γ matrix that pre-multiplies the state-variable vector (i.e., X_t) in the market price of risk introduced in equation (32). We note that this is not a terribly deep extension of the model, but it is introduced with a view towards providing some relief of the curse of dimensionality by reducing the number of parameters. As we will see in later sections, the non-linear optimization problem that arises in the estimation

[8]See Press et al. (1992, 710–712) for a detailed description of the Runge-Kutta technique.

of model parameters is computationally very difficult to solve. As such, any potential simplification is, in our view, worth exploring.

2.2 An empirical model

In the previous section, we considered a model for the dynamics of the term structure of interest rates that was constructed from first principles. That is, the point of departure was the fundamental expression for pricing contingent claims. Another strand of literature, however, has evolved in recent years from the following authors: Diebold and Li (2003); Diebold, Ji, and Li (2004); Diebold, Rudebusch, and Aruoba (2004); Diebold, Piazzesi, and Rudebusch (2005); and Bernadell, Coche, and Nyholm (2005). This work essentially marries the term-structure estimation problem—that is, the extraction of zero-coupon and forward interest rates from a collection of coupon-bond prices—with the description of the dynamics of the term structure of interest rates. In doing so, one essentially constructs a time-series model for the evolution of the interest rates. In other words, this is an empirical approach. For this reason, in this paper, we will refer to this approach as an *empirical* model. We will present three possible model choices. The first follows from the work of Diebold and Li (2003), while we develop the other two approaches in an effort to demonstrate the generality of the empirical approach.

2.2.1 Diebold and Li's extension of the Nelson-Siegel model

Diebold and Li (2003) use the work of Nelson and Siegel (1987) as the foundation for their description of term-structure dynamics. The work of Nelson and Siegel (1987)—as well as subsequent work by Svensson (1994) and Svensson and Söderlind (1997)—focuses entirely on the problem of extracting zero-coupon and forward interest rates from a collection of coupon-bond prices.[9] The initial idea behind these models is to create a parsimonious description of the term structure of interest rates by providing a continuous functional form of forward interest rates. To accomplish this, however, it is necessary to develop a notion of the forward rate that is mathematically continuous. As such, the departure point for these models is the definition of an *instantaneous* forward rate, defined as

$$f(t,\tau) = \lim_{T\downarrow\tau} f(t,\tau,T). \tag{43}$$

That is, the instantaneous forward rate is the limit as the tenor of the underlying zero-coupon contract tends (from above) towards the maturity of the forward contract. In practice, we can think of this as the forward *overnight* interest rate τ periods forward. We can, in fact, evaluate this limit by recalling the definition of the continuously compounded forward interest rate,

$$f(t,\tau,T) = \frac{1}{T-\tau}\ln\left(\frac{P(t,\tau)}{P(t,T)}\right), \tag{44}$$

[9] These models are considered in the Canadian context in Bolder and Streliski (1999) and in Bolder and Gusba (2002).

and an application of L'Hopital's rule. Note that this requires the necessary assumptions regarding the differentiability and continuity of P, which we assume to hold. The form of the instantaneous forward rate, therefore, is derived as

$$\begin{aligned} f(t,\tau) &= \underbrace{\lim_{T\downarrow\tau} \frac{1}{T-\tau} \ln\left(\frac{P(t,\tau)}{P(t,T)}\right)}_{\text{Equation (44)}}, \\ &= \lim_{T\downarrow\tau} \frac{\ln P(t,\tau) - \ln P(t,T)}{T-\tau}, \\ &= \lim_{T\downarrow\tau} \frac{\frac{\partial}{\partial T}\left(\ln P(t,\tau) - \ln P(t,T)\right)}{\frac{\partial}{\partial T}(T-\tau)}, \\ &= \lim_{T\downarrow\tau} \frac{\frac{P'(t,T)}{P(t,T)}}{1}, \\ &= \lim_{T\downarrow\tau} \frac{P_T(t,T)}{P(t,T)}, \\ &= -\frac{P_\tau(t,\tau)}{P(t,\tau)}, \end{aligned} \tag{45}$$

where $P_x(t,x)$ indicates partial differentiation with respect to the second argument, x. Observe that $f(t,\tau)$ represents, at time t, an entire curve, in τ, that characterizes the instantaneous forward term structure of interest rates. We can derive the instantaneous interest zero-coupon curve by a simple manipulation of equation (45) as follows:

$$\begin{aligned} -\frac{P_\tau(t,\tau)}{P(t,\tau)} &= f(t,\tau), \\ -\frac{\partial}{\partial\tau}\left(\ln P(t,\tau)\right) &= f(t,\tau), \\ -\frac{\partial}{\partial\tau}\left(\ln e^{-z(t,\tau)(\tau-t)}\right) &= f(t,\tau), \\ \int_t^\tau \frac{\partial}{\partial u}\left(z(t,u)(u-t)\right) du &= \int_t^\tau f(t,u)du, \\ z(t,\tau)(\tau-t) - z(t,t)(t-t) &= \int_t^\tau f(t,u)du, \\ z(t,\tau) &= \frac{1}{\tau-t}\int_t^\tau f(t,u)du. \end{aligned} \tag{46}$$

This expression permits the specification of both instantaneous forward and zero-coupon curves and a method for moving between them. Note from equations (45) and (46) that the forward rate is represented as a partial derivative of pure-discount bond prices, and that the zero-coupon rate is the sum (i.e., integral) of forward rates. Economically, therefore, we can interpret the zero-coupon curve as an average interest rate, while the forward rate is essentially the marginal interest rate concept.

The idea suggested by Nelson and Siegel (1987) is both simple and powerful. Since equation (45) is essentially a function of instantaneous forward rates in τ, they suggest a specific functional form. In particular, they suggest the following:

$$f(t,\tau) = x_0 + x_1 e^{-\lambda(\tau-t)} + x_2\lambda(\tau-t)e^{-\lambda(\tau-t)}, \tag{47}$$

with the parameters $x_i \in \mathbb{R}, i = 0,1,2$ and $\lambda \in \mathbb{R}$. One can, using equation (46) and a bit of basic calculus, derive the corresponding zero-coupon curve:

$$\begin{aligned}
z(t,\tau) &= \frac{1}{\tau-t}\int_t^\tau \left(x_0 + x_1 e^{-\lambda(s-t)} + x_2\lambda(s-t)e^{-\lambda(s-t)}\right) ds, \qquad (48)\\
&= \frac{1}{\tau-t}\left(x_0(\tau-t) + x_1\left[-\frac{e^{-\lambda(s-t)}}{\lambda}\right]_t^\tau + x_2\lambda\int_t^\tau (s-t)e^{-\lambda(s-t)}ds\right),\\
&= x_0 + x_1\left(\frac{1-e^{-\lambda(\tau-t)}}{\lambda(\tau-t)}\right) + \frac{x_2\lambda}{\tau-t}\left(\underbrace{\left[-\frac{(s-t)e^{-\lambda(s-t)}}{\lambda}\right]_t^\tau + \int_t^\tau e^{-\lambda(s-t)}ds}_{\text{Using integration by parts}}\right),\\
&= x_0 + x_1\left(\frac{1-e^{-\lambda(\tau-t)}}{\lambda(\tau-t)}\right) + \frac{x_2\lambda}{\tau-t}\left(-\frac{(\tau-t)e^{-\lambda(\tau-t)}}{\lambda} + \left[-\frac{e^{-\lambda(s-t)}}{\lambda^2}\right]_t^T\right),\\
&= x_0 + x_1\left(\frac{1-e^{-\lambda(\tau-t)}}{\lambda(\tau-t)}\right) + x_2\left(-e^{-\lambda(\tau-t)} + \frac{\lambda}{\tau-t}\left(\frac{1-e^{-\lambda(\tau-t)}}{\lambda^2}\right)\right),\\
&= x_0 + x_1\left(\frac{1-e^{-\lambda(\tau-t)}}{\lambda(\tau-t)}\right) + x_2\left(\frac{1-e^{-\lambda(\tau-t)}}{\lambda(\tau-t)} - e^{-\lambda(\tau-t)}\right).
\end{aligned}$$

The classical Nelson-Siegel approach suppresses the first argument in $f(t,\tau)$ and $z(t,\tau)$; instead, the functions have the form $f(\tau-t)$ and $z(\tau-t)$, because the curve, in their construction, depends only on the tenor (i.e., term to maturity) of the relevant interest rate, $\tau-t$. The time element, as measured by t, does not enter into the equations since this is a static model. This is where Diebold and Li (2003) enter into the picture. In their work, they make two insightful observations. First, they note that equation (48) is a linear combination of three functions with coefficients, x_0, x_1, and x_2. These functions are:

$$f_0(y) = 1, \tag{49}$$

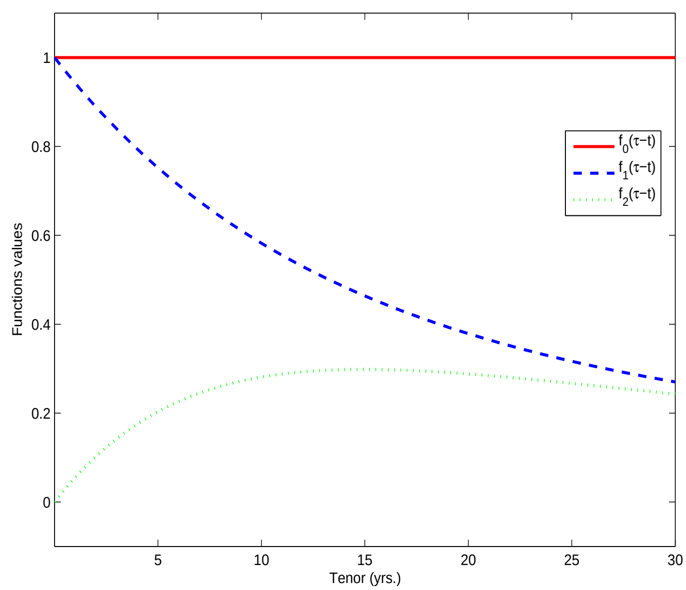
$$f_1(y) = \frac{1-e^{-\lambda y}}{\lambda y}, \tag{50}$$

$$f_2(y) = \frac{1-e^{-\lambda y}}{\lambda y} - e^{-\lambda y}. \tag{51}$$

Figure 1 plots these three functions for a value of $y \in [0,30]$ and a fixed value of $\lambda = 0.12$. Inspection of Figure 1 reveals that f_0 impacts all tenors equally, that f_1 has an unequal impact on the short and long ends of the curve, and that f_2 has a disproportionate impact on the middle part of the curve. Indeed, one can think of these functions as factor loadings that influence the term structure through the coefficient values, $x_i, i = 0,1,2$. Changes in x_0 will create parallel shifts up or down. A positive shock or change to x_2 will lead to a steepening of the zero-coupon curve, while a negative value will elicit a flattening, assuming that the initial curve is upward sloping. Different values of x_2 will change the overall curvature of the zero-coupon term structure; positive values increase curvature, while negative values decrease curvature. In summary—and this is perhaps the principal observation made by Diebold and Li (2003)—the functions f_0, f_1, and f_2 can be interpreted as the level, slope, and curvature of the term structure.

This was a very useful insight, because since the work of Litterman and Schenkman (1991), fixed-income practitioners and academics have used these three principal factors to characterize the dynamics of the term structure of interest rates. In other words, movements in the complex system of interest rates can be broken into changes in the level, slope, and curvature of the yield curve. Nelson and Siegel's model, therefore, could be represented as a linear combination of these three fundamental yield-curve factors.

Figure 1: **Nelson-Siegel Functions**: As demonstrated in equations (48—51), the zero-coupon function can be described as a linear combination of three functions, $\{f_i(\cdot), i = 0, 1, 2\}$. These three functions are graphed below for a fixed value of $\lambda = 0.12$. Observe that f_0 impacts all tenors equally, that f_1 has an unequal impact on the short and long ends of the curve, and that f_2 has a disproportionate impact on the middle part of the curve.



Diebold and Li (2003) then take further step. They propose a model whereby the coefficients—or rather the weights on the level, slope, and curvature of the term structure—vary through time. To see how this might work, let us define a few variables. We define our time-varying coefficients in matrix form as

$$X_t = \begin{bmatrix} x_{0,t} \\ x_{1,t} \\ x_{2,t} \end{bmatrix}. \tag{52}$$

We purposely use this suggestive notation to represent the Nelson-Siegel model coefficients, because, in the Diebold and Li (2003) approach, the model coefficients are essentially state variables. We can also write equations (49) to (51) as

$$F(\tau - t) = \begin{bmatrix} f_0(\tau - t) \\ f_1(\tau - t) \\ f_2(\tau - t) \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda y}}{\lambda y} \\ \frac{1-e^{-\lambda y}}{\lambda y} - e^{-\lambda y} \end{bmatrix}. \tag{53}$$

The attentive reader will have noted that we have assumed that λ remains constant. Strictly speaking, however, this is not true. Since the λ parameter is a non-linear parameter and it is well documented as numerically unstable, Diebold and Li (2003) suggest that it be treated as a constant value.[10] If one is willing to accept this assumption, then we can write the pure-discount bond price function at time t as the following linear function,

$$P(t,T) = e^{F(T-t)^T X_t}, \tag{54}$$

which is really quite similar in spirit to the formulations presented in the previous section used to decribe the $A_0(3)$ model.[11] In this case, $F(T-t)$ is a deterministic vector-valued function of the term to maturity of the pure-discount bond. X_t is a vector-valued stochastic process describing the evolution of the Nelson-Siegel coefficients. Implementing the Diebold and Li (2003) model is, therefore, merely a process of specifying and estimating the dynamics of $\{X_t, t \geq 0\}$. There are a variety of alternatives. We have, for the sake of consistency with previous models, made the following choice:

$$dX_t = \kappa(\theta - X_t)dt + C^T \Sigma dW_t, \tag{55}$$

where $\kappa, C, \Sigma \in \mathbb{R}^{3\times 3}$ and $X_t, \theta, dW_t \in \mathbb{R}^{3\times 1}$ where Σ is diagonal, C is the Cholesky decomposition of the instantaneous correlation matrix, and $\{W_t, t \geq 0\}$ is a standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. Observe that this model describes the dynamics of the term structure under the physical measure; indeed, there is no notion of an equivalent martingale measure or risk premia in this approach.

[10] See Cairns (1998), Bolder and Streliski (1999), and Cairns (2001) for a description of this phenomona in British, Canadian, and German fixed-income markets, respectively.

[11] Diebold, Rudebusch, and Aruoba (2004) relax this assumption by casting the problem in state-space form and using a Kalman filter for parameter estimation. In this paper, however, we use the traditional two-step procedure suggested by Diebold and Li (2003).

A straightforward Euler discretization of equation (55) yields

$$\begin{aligned} X_t - X_{t-1} &= \kappa(\theta - X_{t-1}) + \epsilon_t, \\ X_t &= \kappa\theta + (1-\kappa)X_{t-1} + \epsilon_t, \\ X_t &= \alpha + \beta X_{t-1} + \epsilon_t, \end{aligned} \tag{56}$$

where,

$$\epsilon_t \sim \mathcal{N}\left(0, \underbrace{\Sigma C^T C \Sigma}_{=\Omega}\right). \tag{57}$$

We can see, from a practical perspective, that we are proposing a vector-autoregressive model for the evolution of the factor coefficients.

2.2.2 Other possible models

As noted in the previous section, Diebold and Li (2003) take a model that describes the term structure of interest rates at an instant in time and transform it into a model of the movement of the term structure over time. This raises an interesting question: are there other models for which this transformation is possible? Bolder and Gusba (2002) review eight different term-structure estimation models and consider their relative performance at fitting the term structure. This collection of models, in fact, includes the extended Nelson-Siegel model proposed by Svensson (1994). Indeed, the relatively poor performance of the Svensson model in fitting bond prices is one of the reasons for considering alternative formulations. At first glance, however, there does not appear to be much hope. Many of these models are not appropriate for characterization of interest rate dynamics. This is particularly true for cubic-spline models that often involve an enormous number of parameters.

Two of these models appear to have some potential. The first example, proposed by Li et al. (2001), follows from the so-called exponential-spline methodology. This approach, inspired by the work of Vasicek and Fong (1981) and Shea (1985), describes the discount function as a linear combination of exponential basis functions. Recall that the discount function and the pure-discount bond price function are equivalent. This implies that we can write pure-discount bond prices as

$$P(t,T) = \sum_{k=1}^{n} \xi_k g_k(T-t), \tag{58}$$

where $\{g_k(T-t), t = 1, .., n\}$ is a collection of basis functions. Li et al. (2001) suggest

$$g_k(T-t) = e^{-k\alpha(T-t)}, \tag{59}$$

for $k = 1, ..., n$ and $\alpha \in \mathbb{R}$.[12] The parameter, α, can be interpreted as a long-term instantaneous forward rate.

[12] In actuality, we use an orthogonalized version of these basis functions computed using the Gram-Schmidt orthogonalization procedure; see Bolder and Gusba (2002) for more details.

As with the λ parameter in the Diebold-Li setting, it is fairly reasonable to assume that α is approximately constant.

The second example comes from Bolder and Gusba (2002), who suggest a Fourier-series basis of the following form:

$$g_k(T-t) = \begin{cases} 1 : k = 1 \\ \sin\left(\frac{\frac{k}{2}(T-t)}{10}\right) : \quad \text{mod } (k,2) = 0 \\ \cos\left(\frac{\frac{k-1}{2}(T-t)}{10}\right) : \quad \text{mod } (k,2) = 1 \end{cases} . \tag{60}$$

for $k = 1, ..., n$. Note that the horizontal stretch factor $\frac{1}{10}$ is arbitrarily selected to extend the wavelength of each basis function and avoid excessive oscillation. In their current form, described in equations (59) and (60), both of these choices for $g_k(\cdot)$ are used to estimate the term structure of interest rates at a given point in time; in other words, they are essentially curve-fitting techniques. While the Fourier-series basis performs relatively well at this task, it is typically dominated by the exponential basis in equation (59). This is due primarily to the fundamental form of the pure-discount bond price function, which essentially has a negative exponential form.

We can—borrowing liberally from the ideas of Diebold and Li (2003)—transform these models into a dynamic model for interest rates, by slightly adjusting equation (58):

$$P(t,T) = \sum_{k=1}^{n} \xi_{t,k} g_k(T-t). \tag{61}$$

Again, in a manner analogous to the Diebold-Li model, we can interpret the pure-discount bond function as a linear combination of n basis functions, where the relative weights vary through time according to the coefficients $\xi_{t,k}$ for $k = 1, ..., n$.[13] We will examine both the exponential and Fourier-series basis functions described in equations (59) and (60).

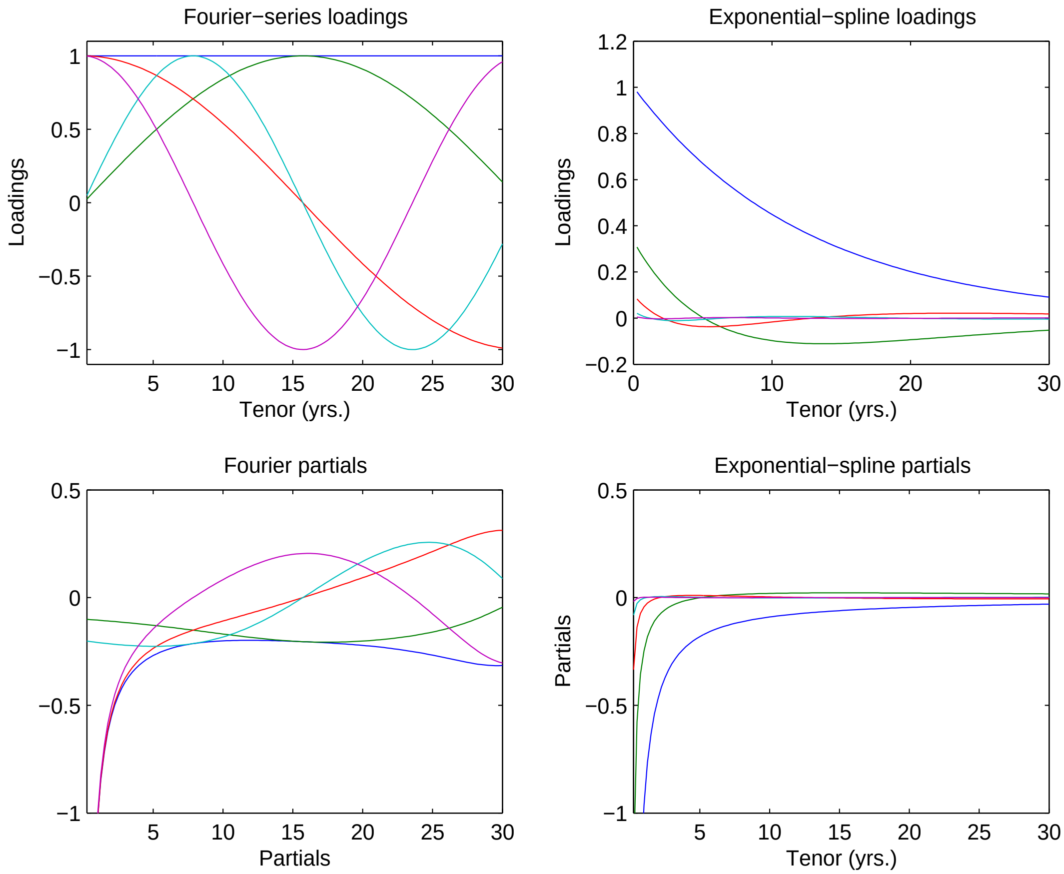
Clearly, neither set of basis functions has the intuitive interpretation of the three basis functions in the Diebold-Li methodology—$\{f_k, k = 0, 1, 2\}$ found in equation (53). Nevertheless, to the extent that we are interested in how well a model captures the term-structure dynamics over time, the relative superiority of a given model is ultimately an empirical question.

Let us introduce some notation to cast our model into the same form as the Diebold-Li approach. Set

$$\xi_t = \begin{bmatrix} \xi_{1,t} \\ \xi_{2,t} \\ \vdots \\ \xi_{n,t} \end{bmatrix}, \tag{62}$$

[13] Bolder and Gusba (2002) find that a choice of $n \approx 9$ is optimal in terms of describing the term structure at a given instant of time.

Figure 2: **Suggested Empirical Models**: For $n = 5$, the underlying figure outlines the exponential-spline and Fourier-series basis functions described (i.e., $g_n(\alpha, T-t)$ for n=1,...,5) in equations (59) and (60). This figure also includes the partial derivatives of each basis function with respect to the zero-coupon curve (see equation (66)).



and let,

$$G(T-t) = \begin{bmatrix} g_1(\alpha, T-t) \\ g_2(\alpha, T-t) \\ \vdots \\ g_n(\alpha, T-t) \end{bmatrix}. \tag{63}$$

This implies that we can write the pure-discount function as

$$P(t,T) = G(T-t)^T \xi_t, \tag{64}$$

and the associated zero-coupon rate function is

$$z(t,T) = -\frac{\ln\left(G(T-t)^T \xi_t\right)}{T-t}. \tag{65}$$

We will describe the dynamics of ξ_t in a manner exactly analogous to that used for the Diebold-Li model outlined in equations (55) to (57).

Moreover, we can describe the sensitivity of the zero-coupon curve to the model coefficients (or state variables), x_t, by the following partial derivatives:

$$\begin{aligned} \frac{\partial}{\partial \xi_k} z(t,T) &= \frac{\partial}{\partial \xi_k}\left(-\frac{\ln\left(G(T-t)^T \xi_t\right)}{T-t}\right), \\ &= -\frac{g_k(\alpha, T-t)}{\underbrace{G(T-t)^T \xi_t}_{\text{Equation (64)}}(T-t)}, \\ &= -\frac{g_k(\alpha, T-t)}{G(T-t)^T \xi_t (T-t)}. \end{aligned} \tag{66}$$

This computation is not necessary in the Diebold-Li approach, because the basis functions are also the partial derivatives.

The upper quadrants of Figure 2 illustrate the collection of Fourier-series and exponential-spline basis functions for $n = 5$. It is interesting to note how these basis functions cover the interval from $[0, 30]$ relative to the Diebold-Li basis functions in Figure 1. The partial derivatives, for the Fourier-series and exponential-spline basis, and again for $n = 5$, are also outlined in the lower quadrants of Figure 2. It is easy to see how linear combinations of the parameters could permit a wide range of term-structure outcomes; this appears to be particularly the case in the Fourier-series model. This incremental flexibility may or may not translate into a better description of term-structure dynamics. It should nonetheless be noted that this additional flexibility comes at the cost of less-intuitive basis functions.

The key point in this section, however, is that the Diebold-Li model involves a specific choice of basis functions for the description of the term structure. This could well turn out to be an optimal basis, but it can, of course, be generalized to consider different options. Indeed, the lack of theory behind this approach implies that no restrictions exist, and that the superiority of a given model rests solely upon empirical considerations.

3 Model Estimation

The various dynamic term-structure models discussed in section 2 require estimation. Indeed, the most critical aspect of implementing any of these models is finding a set of model parameters that ensures the model is consistent with observed interest rate outcomes. The importance of parameter estimation can hardly be overstated. From a practical perspective, it is observationally equivalent to say that a model does not fit the data, or to say that a model cannot be reliably estimated. In both scenarios, the result is identical: the model cannot adequately describe term-structure dynamics. As such, parameter estimation requires as much, if not more, attention as the mathematical specifics of the model.

In this section, therefore, we review in detail the important techniques used to estimate the model parameters. Not surprisingly, the algorithm employed for the theoretical models is rather more involved than for the empirical models. The primary reason is that, for the empirical models, the state variables can be estimated independently of the factor loadings. This permits a relatively straightforward two-step estimation algorithm. In the theoretical $A_0(3)$ model, it is necessary to jointly estimate the state variables and the factor loadings at the same time. As we will see in the subsequent development, this adds rather substantially to the complexity of the estimation algorithm.

3.1 The theoretical model

Three primary techniques are used for the estimation of multifactor affine term-structure models: maximum likelihood, the Kalman filter, and a number of variations on the simulated method of moments.[14] In this discussion, we will focus on maximum-likelihood estimation. This approach exploits *two* facts about the previously described theoretical model. First, the equation describing continuously compounded zero-coupon interest rates has a linear form. This implies that, given n zero-coupon rates, one can invert this equation to solve for the unobservable state variables embedded in an n-factor exponential-affine term-structure model. Second, the form of the stochastic differential equations that govern the dynamics of these models can be solved to determine the transition densities of the unobservable state variables. Knowledge of these transition densities, coupled with the ability to extract values of the state variables from zero-coupon rates, permits us to use the change-of-variables formula to determine the conditional joint density of the state variables. It is then straightforward to compute the log-likelihood function from this conditional joint density.

Let $Y(t, \tau_1)$ denote the yield of a zero-coupon bond with a maturity of τ_1, as at time t. Given that we are working with a three-factor model, we must select three separate zero-coupon yields. Using these yields, we can construct the following linear system using the affine pure-discount bond price function provided in equation

[14] The Kalman filter is conceptually very similar to the maximum-likelihood approach. The key difference is that a more complicated approach is used to extract the unobserved state variables from the set of observed bond prices. The most popular implementation of the simulated method of moments is termed the efficient method of moments (EMM). EMM involves the parameterization of an auxiliary function—which is selected by the analyst—that describes the model dynamics. One then simulates population moment conditions from this auxiliary function.

(25):

$$\begin{aligned} Y(t,\tau_1) &= -\frac{1}{\tau_1 - t}\ln e^{-c_\theta(\tau_1-t)-B_\theta(\tau_1-t)^T X(t)}, \\ Y(t,\tau_2) &= -\frac{1}{\tau_2 - t}\ln e^{-c_\theta(\tau_2-t)-B_\theta(\tau_2-t)^T X(t)}, \\ Y(t,\tau_3) &= -\frac{1}{\tau_3 - t}\ln e^{-c_\theta(\tau_3-t)-B_\theta(\tau_3-t)^T X(t)}. \end{aligned} \tag{67}$$

The θ subscript in each of the factor loadings (i.e., c and B) denotes the dependence on the parameter vector. Note that, as previously suggested, the linear form of this system makes it trivial to solve for the latent state variables. The solution is given as

$$\begin{aligned} \begin{bmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{bmatrix} &= \begin{bmatrix} \frac{B_\theta(\tau_1-t)^T}{\tau_1-t} \\ \frac{B_\theta(\tau_2-t)^T}{\tau_2-t} \\ \frac{B_\theta(\tau_3-t)^T}{\tau_3-t} \end{bmatrix}^{-1} \begin{bmatrix} Y(t,\tau_1) - \frac{c_\theta(\tau_1-t)}{\tau_1-t} \\ Y(t,\tau_2) - \frac{c_\theta(\tau_2-t)}{\tau_2-t} \\ Y(t,\tau_3) - \frac{c_\theta(\tau_3-t)}{\tau_3-t} \end{bmatrix} \\ x_t &= H^{-1} y_t, \end{aligned} \tag{68}$$

where we let the vector x_t denote the vector of state variables associated with the market zero-coupon yields, conditional on a choice of parameter vector $\theta \in \Theta$.

At this point, it is useful to briefly review the change-of-variables technique. The basic theorem is often used in mathematical statistics, but it finds its origins as an integration technique. Imagine that we want to solve the following integral:

$$\int\int_\Omega f(x,y)dxdy. \tag{69}$$

Unfortunately, however, it may turn out that finding the antiderivative of $f(x,y)$ might prove difficult. If there exists a transformation between the variables x and y to variables u and v of the form,

$$\begin{aligned} x &= x(u,v), \\ y &= y(u,v), \end{aligned} \tag{70}$$

the change-of-variables theorem then holds that,

$$\int\int_\Omega f(x,y)dxdy = \int\int_{\hat{\Omega}} f(x(u,v), y(u,v)) \det\left((J(u,v)\right) du\, dv, \tag{71}$$

where $\det\left(J(u,v)\right)$ is the determinant of the Jacobian matrix, defined as

$$J(u,v) = \begin{bmatrix} \frac{\partial x(u,v)}{\partial u} & \frac{\partial x(u,v)}{\partial v} \\ \frac{\partial y(u,v)}{\partial u} & \frac{\partial y(u,v)}{\partial v} \end{bmatrix}. \tag{72}$$

The statistical corollary follows almost directly from this deterministic example. Suppose that X and Y are vector-valued random variables. In particular, we have a situation where the density of X is known and $f_X(x)$

is known, while $f_Y(y)$ is not known. Moreover, assume that $g(X) = Y$ is a one-to-one (i.e., inverse is unique) once continuously differentiable function. The density of the vector-valued random variable, Y, is given as

$$\begin{aligned} f_{g(X)}(y) &= f_{g^{-1}(g(X))}(g^{-1}(y)) \det\left(J(g^{-1}(y))\right), \\ f_Y(y) &= f_X(x) \det\left(J(x)\right). \end{aligned} \tag{73}$$

This extremely useful result is exactly what is required to write out the joint conditional density of the unobserved state variables. In our setting, X_t and Y_t are our random vectors as defined in equation (68). Again, in our situation the joint transition density of the state variables, $f_{X_t}(x_t \mid x_s)$, is known for $s < t$.[15] The joint density of the zero-coupon yields $f_{Y_t}(y_t \mid y_s)$ is not. Using the theorem described in equation (73), we have that

$$f_{Y_t|Y_s}(y_t \mid y_s) = f_{X_t|X_s}(x_t \mid x_s) \det(J(x_t)). \tag{74}$$

In order to write out the maximum-likelihood function, we need to know the analytical form of each of the transition densities, as well as the Jacobian matrix. Let us begin with the Jacobian,

$$\begin{aligned} \det(J(x_t)) &= \det\left(\frac{\partial x_t}{\partial y_t}\right), \\ &= \det\left(\frac{\partial H^{-1} y_t}{\partial y_t}\right), \\ &= \det\left(H^{-1}\right), \\ &= \det\left(H\right)^{-1}, \\ &= \det\left(\begin{bmatrix} \frac{B_\theta(\tau_1 - t)^T}{(\tau_1 - t)} \\ \frac{B_\theta(\tau_2 - t)^T}{(\tau_2 - t)} \\ \frac{B_\theta(\tau_3 - t)^T}{(\tau_3 - t)} \end{bmatrix}\right)^{-1}. \end{aligned} \tag{75}$$

Next, we need to solve each of the stochastic differential equations to determine the form of the transition density, $f(X_t \mid X_s)$. Recall the stochastic dynamics of the instantaneous short rate,

$$dX_t = -\kappa X_t dt + dW_t, \tag{76}$$

where κ is a lower-diagonal matrix with distinct and positive eigenvalues (i.e., diagonizable) and $\{W_t, t \geq 0\}$ is an n-dimensional Brownian motion on $(\Omega, \mathcal{F}, \mathbb{Q})$ with the usual $\mathbb{Q}$-augmentation of its natural filtration $\{\mathcal{F}_t, t \geq 0\}$. Next, we let Y_t denote the instantaneous drift of our state variable,

$$Y_t = \kappa X_t. \tag{77}$$

We then apply Itô's rule to the function $e^{\kappa t} Y_t$. Note that $e^{\kappa t}$, as $\kappa \in \mathbb{R}^{n \times n}$, represents the matrix exponential. To solve for the transition density of X_t, we will need to exploit a few properties of the matrix exponential.

[15] We will derive this density later in the text.

Recall that the matrix exponential is defined as a power series of the form for a square matrix, $A \in \mathbb{R}^{n\times n}$:

$$e^A = \sum_{i=0}^{\infty} \frac{A^i}{i!} = I_n + A + \frac{A^2}{2} + \frac{A^3}{3} + ... \tag{78}$$

There is, of course, also a matrix logarithm defined as follows for a square invertible matrix $A \in \mathbb{R}^{n\times n}$:

$$\ln(A) = \sum_{i=1}^{\infty} \frac{(I_n - A)^i}{i} = I_n + \frac{(I_n - A)^2}{2} + \frac{A^2}{2} + \frac{(I_n - A)^3}{3} + ... \tag{79}$$

The matrix logarithm is not, in general, unique. For a matrix, such as κ with positive and distinct eiqenvalues, however, it is unique. Finally, there is an additional useful result regarding the matrix exponential that we need to consider: if two square matrices, A and C, commute (i.e., $AC = CA$), then it is true that

$$\begin{aligned} e^{A+C} &= e^A e^C, \\ &= e^C e^A. \end{aligned} \tag{80}$$

Armed with these results, we can proceed, as promised, to apply Itô's rule to the function $e^{\kappa t} Y_t$:

$$d\left(e^{\kappa t} Y_t\right) = \int_0^t \frac{\partial e^{\kappa u} Y_t}{\partial u} du + \int_0^t \frac{\partial e^{\kappa u} Y_t}{\partial X(u)} dX_u + \int_0^t \underbrace{\frac{\partial^2 e^{\kappa u} Y_t}{\partial X(u)^2}}_{=0} d\langle X\rangle_u, \tag{81}$$

$$\begin{aligned} d\left(e^{\kappa t} Y_t\right) &= \int_0^t \kappa e^{\kappa u} Y_u du - \int_0^t \kappa e^{\kappa u} dX_u, \\ &= \int_0^t \kappa e^{\kappa u} Y_u du - \int_0^t \kappa e^{\kappa u} \left(Y_u + dW_u\right), \\ &= \int_0^t \kappa e^{\kappa u} Y_u du - \int_0^t \kappa e^{\kappa u} Y_u du - \int_0^t \kappa e^{\kappa u} dW(u), \\ &= -\int_0^t \kappa e^{\kappa u} dW(u). \end{aligned}$$

We can then work with the right-hand side of equation (81) to solve for X_t, with liberal use of the commutative property of the matrix exponential:

$$
\begin{aligned}
e^{\kappa t}Y_t - \underbrace{e^{\kappa 0}}_{=I} Y_0 &= -\int_0^t \kappa e^{\kappa u} dW(u), \qquad (82)\\
-e^{\kappa t}\kappa X_t + \kappa X_0 &= -\int_0^t \kappa e^{\kappa u} dW(u),\\
e^{\kappa t}\kappa X_t &= \kappa X_0 + \int_0^t \kappa e^{\kappa u} dW(u),\\
X_t &= \kappa^{-1}e^{-\kappa t}\kappa X_0 + \int_0^t \kappa^{-1}e^{-\kappa t}\kappa e^{\kappa u} dW(u),\\
&= e^{\ln\left(\kappa^{-1}\right)}e^{-\kappa t}e^{\ln(\kappa)}X_0 + \int_0^t e^{\ln\left(\kappa^{-1}\right)}e^{-\kappa t}e^{\ln(\kappa)}e^{\kappa u} dW(u),\\
&= e^{\ln\left(\kappa^{-1}\right)-\kappa t+\ln(\kappa)}X_0 + \int_0^t e^{\ln\left(\kappa^{-1}\right)-\kappa t+\ln(\kappa)+\kappa u} dW(u),\\
&= \kappa^{-1}\kappa e^{-\kappa t}X_0 + \int_0^t \kappa^{-1}\kappa e^{-\kappa(t-u)} dW(u),\\
&= e^{-\kappa t}X_0 + \int_0^t e^{-\kappa(t-u)} dW(u).
\end{aligned}
$$

Thus, we see that X_t is a multivariate Gaussian random variable with the following transition density:

$$
f\left(X_t \mid X_s\right) \sim \mathcal{N}\left(e^{-\kappa(t-s)}X_s, \left(\int_s^t e^{-\kappa(t-u)} dW(u)\right)^2\right). \qquad (83)
$$

An application of Itô isometry and knowledge of the quadratic variation of the Brownian motion reveals that

$$
\left(\int_s^t e^{-\kappa(t-u)} dW(u)\right)^2 = \int_s^t \left(e^{-\kappa(t-u)}\right)\left(e^{-\kappa(t-u)}\right)^T du. \qquad (84)
$$

It is fairly difficult to solve this integral analytically, if it is even possible at all, but fortunately it is fast and simple to evaluate numerically. Let us therefore denote the conditional variance as

$$
V_{s,t} = \int_s^t \left(e^{-\kappa(t-u)}\right)\left(e^{-\kappa(t-u)}\right)^T du. \qquad (85)
$$

Observe, however, that for a fixed time interval $(t-s)$, the conditional variance is constant. We can therefore write $V \equiv V_{s,t}$. As such, the transition density of the state variables is given as

$$
f\left(X_t \mid X_s\right) \sim \mathcal{N}\left(e^{-\kappa(t-s)}X_s, V\right). \qquad (86)
$$

We now have all the necessary ingredients for a description of the estimation algorithm, which originated with Chen and Scott (1993). Ultimately, for a given data set $\{y_0, ..., y_T\}$, where each $y_t \in \mathbb{R}^n$ represents a vector of zero-coupon interest rates at time t, we want to maximize the following:

$$\begin{aligned} f_{Y_1,...,Y_T}(y_1,...,y_T) &= \prod_{t=1}^{T} f_{Y_t|Y_{t-1}}(y_t \mid y_{t-1}), \\ &= \prod_{t=1}^{T} f_{X_t|X_{t-1}}(x_t \mid x_{t-1}) \det(J(x_t)), \\ &= \prod_{t=1}^{T} \left(\frac{1}{\sqrt{2\pi}} \det(V)^{-\frac{1}{2}} e^{-\frac{1}{2}\left(e^{-\kappa\Delta t}X_{t-1}\right)^T V^{-1}\left(e^{-\kappa\Delta t}X_{t-1}\right)} \right) \frac{1}{\det(H^{-1})}, \end{aligned} \tag{87}$$

where, as demonstrated in equation (68), $X_{t-1} = H^{-1}Y_{t-1}$. This, of course, is equivalent to maximizing the log-likelihood function,

$$\max_{\kappa,c_r,b_r,b_\gamma,A_\gamma} -\frac{T}{2}\ln(2\pi) - T\ln\left(\det\left(H^{-1}\right)\right) - \frac{T}{2}\ln\left(\det(V)\right) - \frac{1}{2}\sum_{t=1}^{T}\left(e^{-\kappa\Delta t}X_{t-1}\right)^T V^{-1}\left(e^{-\kappa\Delta t}X_{t-1}\right). \tag{88}$$

If we have a three-factor term-structure model, as with the $A_0(3)$ models considered in this paper, the log-likelihood function in equation (88) permits us to price only three bonds without error. We would like, however, to use more than three zero-coupon bond prices in the estimation algorithm. These additional bonds will be assumed to be observed with error. Assume, for example, that we want to use n zero-coupon rates in our estimation algorithm. Further assume that we price m zero-coupon rates exactly; this implies that we will price $n-m$ zero-coupon rates with error. This changes our log-likelihood function. The idea is to once again use the change-of-variables trick to develop an extended log-likelihood function. The idea is to write out the zero-coupon bonds priced with error as

$$\begin{aligned} \begin{bmatrix} Y(t,\tau_{m+1}) \\ \vdots \\ Y(t,\tau_n) \end{bmatrix} &= \begin{bmatrix} \frac{c_\theta(\tau_{m+1}-t)}{\tau_{m+1}-t} \\ \vdots \\ \frac{c_\theta(\tau_n-t)}{\tau_n-t} \end{bmatrix} + \begin{bmatrix} \frac{B_\theta(\tau_{m+1}-t)^T}{\tau_{m+1}-t} \\ \vdots \\ \frac{B_\theta(\tau_n-t)^T}{\tau_n-t} \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{bmatrix} + \begin{bmatrix} \frac{D_\theta(\tau_{m+1}-t)^T}{\tau_{m+1}-t} \\ \vdots \\ \frac{D_\theta(\tau_n-t)^T}{\tau_n-t} \end{bmatrix} \begin{bmatrix} \xi(t,\tau_{m+1}) \\ \vdots \\ \xi(t,\tau_n) \end{bmatrix}, \\ y_t &= c + Bx_t + D\xi_t, \end{aligned} \tag{89}$$

where $c, y_t, e_t \in \mathbb{R}^{(n-m)\times 1}$, $B \in \mathbb{R}^{(n-m)\times m}$, and $D \in \mathbb{R}^{(n-m)\times(n-m)}$. If we solve for ξ_t, we have

$$\xi_t = D^{-1}\left(y_t - c - Bx_t\right). \tag{90}$$

Note, as before, that $x_t \in \mathbb{R}^{m\times 1}$ is known, since we use the values determined from the m zero-coupon rates assumed to be observed without error in equation (68). We therefore want to use the same change-of-variables idea as previously; this implies working with the joint conditional density of ξ_t and X_t. We assume, however, that X_t and ξ_t are independent, so that the joint conditional density is the product of the marginal conditional densities, as follows:

$$f_{Y_t|Y_s}(y_t \mid y_s) = \underbrace{f_{X_t|X_s}(x_t \mid x_s)\det J(x_t)}_{\text{Observations without error}}\ \underbrace{f_{\Xi_t|\Xi_s}(\xi_t \mid \xi_s)\det J(\xi_t)}_{\text{Observations with error}}, \tag{91}$$

where,

$$\det J(\xi_t) = \det\left(\frac{\partial \xi_t}{\partial y_t}\right), \tag{92}$$
$$= \det D^{-1},$$
$$= \frac{1}{\det D}.$$

In this analysis, we treat D with full generality. Having said that, for practical purposes, we merely assume that D is an $(n-m)\times(n-m)$ identity matrix. Let us next focus on developing the contribution of the second term on the right-hand side of equation (91), because, when we apply logarithms to our log-likelihood function, it will have an additive form. It is reasonable to assume that the zero-coupon pricing errors are multivariate normal distributed as

$$\xi_t \sim \mathcal{N}\left(\vec{0}, \Omega\right), \tag{93}$$

so that we have the following conditional density:

$$f_{\Xi_t|\Xi_{t-1}}(\xi_t \mid \xi_{t-1}) \cdot \det|J(\xi_t)| = (2\pi)^{-\frac{(n-m)}{2}} \cdot |\det \Omega|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}\xi_t^T \Omega^{-1}\xi_t} \cdot |\det D|^{-1}. \tag{94}$$

A consistent estimator for Ω is

$$\hat{\Omega} = \sum_{t=1}^{T} \xi_t \xi_t^T. \tag{95}$$

Let us next build the contribution of the zero-coupon rates observed with error to our log-likelihood function:

$$\sum_{t=1}^{T} \ln\left(f_{\Xi_t|\Xi_{t-1}}(\xi_t \mid \xi_{t-1}) \cdot \det|J(\xi_t)|\right) = \sum_{t=1}^{T} \ln\left((2\pi)^{-\frac{(n-m)}{2}} \cdot |\det \Omega|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}\xi_t^T \Omega^{-1}\xi_t} \cdot |\det D|^{-1}\right), \tag{96}$$
$$= T\left(\frac{-(n-m)}{2}\ln 2\pi - \frac{1}{2}\ln\left(|\det \hat{\Omega}|\right) - \ln\left(|\det D|\right)\right) - \frac{1}{2}\sum_{t=1}^{T}\xi_t^T \hat{\Omega}^{-1}\xi_t,$$
$$= -\frac{T}{2}\left((n-m)\ln 2\pi + \ln\left(|\det \hat{\Omega}|\right) + 2\ln\left(|\det D|\right)\right) - \frac{1}{2}\sum_{t=1}^{T}\mathrm{tr}\left(\hat{\Omega}^{-1}\xi_t\xi_t^T\right),$$
$$= -\frac{T}{2}\left((n-m)\ln 2\pi + \ln\left(|\det \hat{\Omega}|\right) + 2\ln\left(|\det D|\right)\right) - \frac{1}{2}\mathrm{tr}\left(\hat{\Omega}^{-1}\underbrace{\sum_{t=1}^{T}\xi_t\xi_t^T}_{\hat{\Omega}}\right),$$
$$= -\frac{T}{2}\left((n-m)\ln 2\pi + \ln\left(|\det \hat{\Omega}|\right) + 2\ln\left(|\det D|\right)\right) - \frac{1}{2}\mathrm{tr}\left(I_{n\times m}\right),$$
$$= -\frac{T}{2}\left((n-m)\ln 2\pi + \ln\left(|\det \hat{\Omega}|\right) + 2\ln\left(|\det D|\right)\right) - \frac{(n-m)}{2}.$$

3.2 The empirical model

The approach used to estimate all of the empirical models is essentially identical. As previously mentioned, we employ a two-step estimation algorithm. This is possible because the factor loadings are constructed independently of the model parameters.[16] In the first step, we use an optimization algorithm to extract on a day-by-day basis the state variables. Recall that, in this class of models, the state variables are the parameters of a term-structure model used to extract zero-coupon rates from coupon-bond prices. This process basically amounts to constructing a time series of state variables. The second step, which describes the term-structure model dynamics through time, involves fitting a time-series model to the state variables. We opt for a vector-autoregressive specification of the state-variable dynamics.

3.2.1 Extracting the state variables

The first step involves pricing coupon bonds. To price a coupon bond, of course, one merely needs to sum the product of the individual cash flows with their associated discount factors. The pure discount bond price is, in fact, the discount factor. Let us begin with a review of the pure-discount bond price function for each of the three empirical models under consideration. In each case, we write the pure-discount bond price as a function of some set of basis functions. In the Diebold and Li (2003) approach, we have that

$$P_x(t,\tau) = e^{F(\tau-t)^T x}, \tag{97}$$

with $F(\tau - t)$ defined in equation (53). Our extension of the exponential-spline approach suggested by Li et al. (2001) and the Fourier-series model has the form

$$P_x(t,\tau) = G(\tau-t)^T x. \tag{98}$$

The collection of basis functions (i.e., $G(\cdot)$) for the exponential-spline model is given as

$$G(\tau-t) = \begin{bmatrix} e^{-\alpha(\tau-t)} \\ e^{-2\alpha(\tau-t)} \\ \vdots \\ e^{-k\alpha(\tau-t)} \end{bmatrix}, \tag{99}$$

[16] This is not entirely true since, in each model, there is a non-linear parameter that arises in the factor loadings. Diebold and Li (2003) suggest that this non-linear parameter can be fixed, and we also adopt this approach.

for k basis functions and fixed $\alpha \in \mathbb{R}$, while the basis functions for the Fourier-series model are given as

$$G(\tau - t) = \begin{bmatrix} 1 \\ \sin\left(\frac{(\tau-t)}{\gamma}\right) \\ \cos\left(\frac{(\tau-t)}{\gamma}\right) \\ \vdots \\ \cos\left(\frac{\frac{(k-1)-1}{2}(\tau-t)}{\gamma}\right) \\ \sin\left(\frac{\frac{k}{2}(\tau-t)}{\gamma}\right) \end{bmatrix}, \tag{100}$$

again for k basis functions and a fixed wavelength stretch factor, $\gamma \in \mathbb{R}$.[17]

For the collection of dates $\{t_0, ..., t_T\}$, we have a set of coupon bond prices, coupon rates, and maturity dates. Moreover, at each time point, we have N_{t_q} coupon-bond observations. Let us introduce the following notation that is also used in Bolder and Gusba (2002):

$$\begin{aligned} B_i(t_q) &\triangleq \text{price of the } i\text{th coupon bond at time } t_q, \\ c_{ij} &\triangleq \text{the } j\text{th payment of the } i\text{th bond}, \\ \tau_{ij} &\triangleq \text{the time when the } j\text{th payment of the } i\text{th bond occurs}, \\ m(i, t_q) &\triangleq \text{the remaining number of payments for the } i\text{th bond at time } t_q. \end{aligned} \tag{101}$$

We can approximate, using the pure-discount bond price functions in equations (97) and (98)—or we can equivalently call this the discount function—the price of each of these coupon bonds. For an arbitrary set of state variables, $x \in \mathbb{R}^k$, we have that the estimated price is given:

$$\hat{B}_i(t_q, x) = \sum_{j=1}^{m(i,t_q)} c_{ij} P_x\left(t_q, \tau_{ij}\right), \tag{102}$$

for $z = 0, .., T$ and $i = 1, .., N_T$. In other words, the estimated price of the ith bond at time t_q is the sum of its discounted cash flows. This is hardly a surprising result. What we would like to do, however, is to find the x that provides the best possible price for each of the individual bonds observed at time t. This requires some optimization. Let

$$\hat{\mathcal{B}}(t_q, x) = \begin{bmatrix} \hat{B}_1(t_q, x) & \cdots & \hat{B}_{m(i,t_q)}(t_q, x) \end{bmatrix}^T \tag{103}$$

denote the vector of estimated coupon-bond prices at time t_q for an arbitrary set of state variables, x. The actual observed coupon-bond prices at time t_q are represented as

$$\mathcal{B}(t_q) = \begin{bmatrix} B_1(t_q) & \cdots & B_{m(i,t_q)}(t_q) \end{bmatrix}^T. \tag{104}$$

[17] Note, in this case, that k must be an even number, to ensure an equal number of sine and cosine members in the basis.

We then perform a sequence of $T+1$ optimizations of the form

$$\min_x \left(\mathcal{B}(t_q) - \hat{\mathcal{B}}(t_q, x)\right)^T W_{t_q} \left(\mathcal{B}(t_q) - \hat{\mathcal{B}}(t_q, x)\right), \tag{105}$$

for $q \in \{0, ..., T\}$ and with the diagonal weighting matrix W_{t_q}.[18] The sequence of optima representing the solution to the $T+1$ optimization problems in equation (105),

$$\{x^*_{t_0}, ..., x^*_{t_T}\}, \tag{107}$$

represent the time-series of state variables for our dynamic term-structure model. This completes the first step of the estimation algorithm for the empirical models.

3.2.2 Describing the state-variable dynamics

The second step in the empirical-model estimation algorithm involves applying a statistical time-series model to the results derived in the previous section. Figure 3 describes the evolution of each of the parameter estimates from the application of the optimization algorithm described in equation (105) to monthly data from January 1990 to August 2005 (187 months). Inspection of the second quadrant of Figure 3 suggests that there is relatively little interaction between the state variables in the Nelson-Siegel model. Indeed, the contemporaneous cross-correlation between x_0 and x_1 is approximately 0.08; it is -0.12 between x_1 and x_2, and it is about -0.38 for x_1 and x_3. Also note that the Nelson-Siegel state variables are highly correlated, with the principal components extracted from the variance-covariance matrix of zero-coupon yield differences over the same period.[19] This is very similar to the results found in Diebold and Li (2003) for American data.

Decomposing the relationship between the state variables and the principal components is rather more difficult in the exponential-spline and Fourier-series models. The contemporaneous correlation between the sum of the first and second Fourier-series state variables and the first principal component—computed on the pure-discount bond prices—is 0.97. This seems encouraging. The contemporaneous correlation between the fourth and fifth Fourier-series state variables and the first principal component, however, is 0.81. Similarly, the contemporaneous correlation between the second principal component and the first and fifth Fourier-series state variables *or* the third state variable is approximately 0.90. Indeed, the first plus the fifth Fourier-series state variable less the

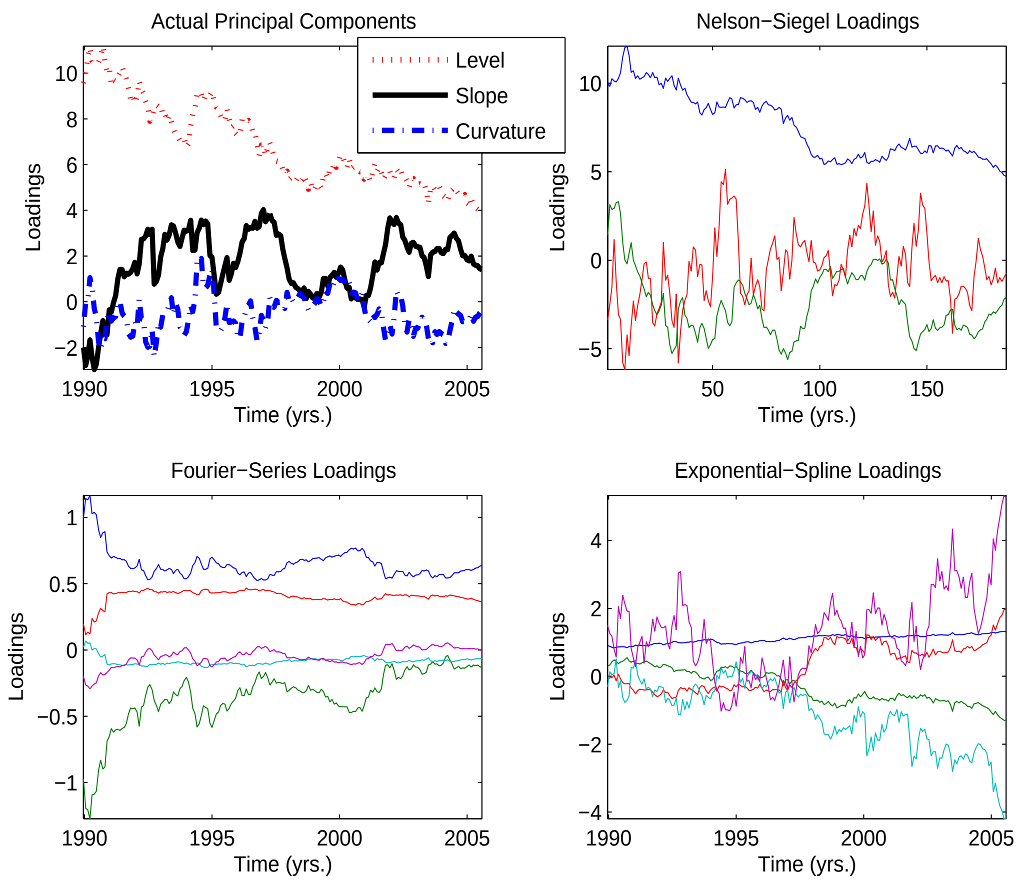
[18] The weighting matrix is required to correct for the fact that we are working in price space, rather than yield space. If we do not weight the price errors, we will actually overfit the long end of the curve at the expense of the short end. The weighting matrix, therefore, is a diagonal matrix with the entries

$$[W_{t_q}]_{ii} = \frac{1}{d(i, t_q)}, \tag{106}$$

where $d(i, t_q)$ denotes the modified duration of the i coupon bond at time t_q.

[19] The contemporaneous correlation with the first principal component and the Nelson-Siegel level factor is 0.95. It is -0.94 for the second principal component and the Nelson-Siegel slope factor, and 0.82 for the third principal component and the Nelson-Siegel curvature factor.

Figure 3: **The State Variables**: This figure shows the actual principal components for the 187 months of data beginning in January 1990 and ending in August 2005. It also shows the evolution of the state variables (i.e., parameter estimates) for the Nelson-Siegel, Fourier-series, and exponential-spline models.



third state variable yields a contemporaneous correlation of slightly more than 0.90. The situation is, if possible, even more complicated in the exponential-spline model. Clearly, this is a highly integrated system and there does not exist—as is the case with the Nelson-Siegel model—a simple mapping between the state variables and the principal components.[20]

How then do we model the dynamics of these state variables? The actual approach is really quite straightforward. For simplicity, we consider two related, indeed nested, statistical models for the dynamics of $\{x_{t_q}, q = 0, ..., T\}$. The first is to treat each individual state variable, $\{x_{i,t_q}, q = 0, ..., T\}$ for $i = 1, .., k$, as an independent AR(1) process. This yields the following well-known model:

$$x_{i,t_q} = \alpha_i + \phi_i x_{i,t_{q-1}} + \epsilon_{i,t_q}, \tag{108}$$

[20] In many ways, this is the principal advantage of the Diebold and Li (2003) approach predicated on the Nelson-Siegel model.

for $i = 1, .., k$ and $q = 1, ..., T$.

The second approach is to model the evolution of the state variables as a correlated system. To do this, we merely employ a VAR(1) model of the following form:

$$x_{t_q} = \alpha + \Phi x_{t_{q-1}} + \epsilon_{t_q}, \tag{109}$$

for $q = 1, ..., T$, $\alpha \in \mathbb{R}^{k \times 1}$, and $\Phi \in \mathbb{R}^{k \times k}$.

We expect that the VAR(1) model will be rather more useful for the exponential-spline and Fourier-series models, because there are five state variables in each setting and we observe rather large contemporaneous cross-correlations between the various state variables. Examination of the lower quadrants of Figure 3 should underscore this reasoning. Modelling the state variables as independent AR(1) processes, therefore, would ignore the interactions between the state variables. This would consequently reduce the ability of the model to describe the time-series dynamics of the term structure of interest rates. Consequently, we use the VAR(1) model to compute state variables dynamics.

Using our dataset, we proceed to use the estimation techniques, described in section 3, to determine the parameters for each of the models. For the empirical models, this is quite straightforward. The $A_0(3)$ models, however, require us to solve a high-dimensional non-linear optimization problem. In this setting, one can never—absent strong restrictions on the mathematical form of the objective functions—be certain of having found the global minimum. The best one can do is to perform sufficient computation to feel confident that a solution close to the global minimum is found. We employ, therefore, a rather extensive optimization algorithm to attempt to find the global minimum. First, we find a starting value by evaluating 500 randomly selected starting values. The actual starting value is the lowest objective value among this collection of objective function values. We then perform six alternations between 1,000 iterations of the Nelder-Meade (i.e., function-evaluation based method) and the sequential-quadratic programming (i.e., gradient-based method) implemented in Matlab. In each alternation, the best value from the previous step is used as the starting value for the subsequent step, to arrive at a final optimal parameter set. This sequence of steps is performed 500 times. We then look at the top 50 objective function values and select the set of parameters that provides the best fit to the data.

4 The Results

The primary objective of this paper is to examine the performance of the six different dynamic term-structure models, described in section 2, with respect to four different sets of criteria. The four criteria are: (i) in-sample zero-coupon rate forecasting ability; (ii) out-of-sample zero-coupon rate and expected excess holding-period return forecasting ability; (iii) the capacity of the simulated model to capture deviations from the expectations hypothesis; and, (iv) model performance in a simplified portfolio-optimization exercise. Having examined the mathematical details of the derivation and estimation of these models in sections 2 and 3, we can now address this important question. This section has five parts, corresponding to our different evaluation criteria. First, we briefly describe the approach used to constructing the necessary zero-coupon rate data. Second, we examine the in- and out-of-sample zero-coupon rate forecasting performance of the various models. Third, we look at how well the models produce out-of-sample forecasts of excess holding-period returns. Fourth, we investigate two specific econometric tests of the expectations hypothesis, and look at how our six models capture deviations from the expectations hypothesis. Finally, we perform a simplified portfolio-optimization exercise using the various term-structure models to forecast the mean and variance of the excess holding-period returns on a portfolio of pure-discount bonds.

4.1 The data

To estimate a model's parameters, of course, one requires data. Estimating a model of zero-coupon term-structure dynamics is complicated somewhat by the fact that we do not—for maturities beyond one year, at least—observe zero-coupon rates. Fortunately, substantial work has been done in this area for the Canadian government bond market. Bolder and Gusba (2002) and Bolder, Johnson, and Metzler (2004) examine a number of alternative zero-coupon curve estimation algorithms and proceed to construct a database of zero-coupon rates for the Canadian government bond market. We will make extensive use of this database.

The choice of a dataset raises an interesting question. The three empirical models that we consider stem from models used to extract these zero-coupon interest rates from bond prices: the Nelson-Siegel, the exponential-spline, and the Fourier-series models. Is it possible that, by choosing a specific model to extract zero-coupon rates, we will be giving an advantage to that model when performing the model comparison? In particular, we have selected an exponential-spline model as the representation of the true zero-coupon data. Does this provide a benefit to the exponential spline model? We think that, in principle, the answer to this question is yes.[21] Nevertheless, we have taken a number of measures to solve this potential problem. In the empirical exponential-spline model, we use a linear combination of seven orthogonal exponential basis functions with a fixed value of the parameter, α. The data used as the *true* representation of the zero-coupon term structure at each point in time are estimated using a linear combination of *nine* orthogonal exponential basis functions where the parameter α is determined on each date through a separate non-linear optimization routine.[22] The empirical model based on the use of the exponential-spline model is, therefore, a stylized version of the actual model to extract zero-coupon rates.

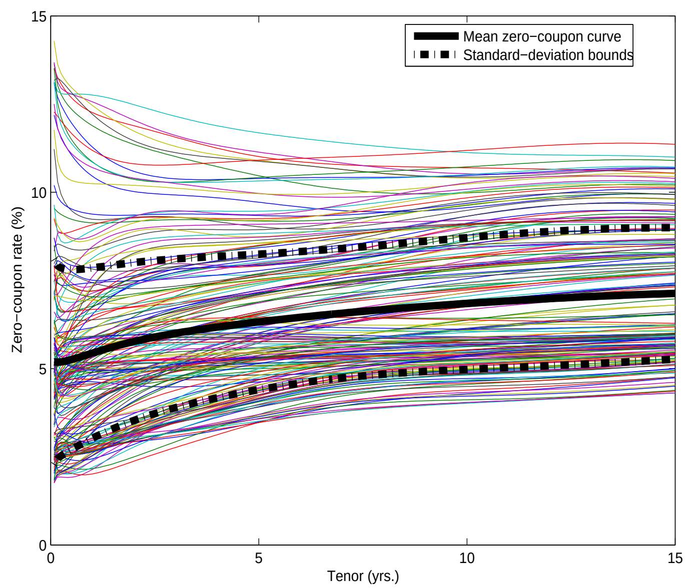
Figure 4 provides a summary of the 187 monthly zero-coupon curves beginning in January 1990 and running to August 2005; this is a period of slightly more than $15\frac{1}{2}$ years. Observe that the curve takes a number of shapes including flat, inverted, and quite steep forms. The solid black line represents the average curve of the period.

[21] Dai, Singleton, and Yang (2004) address this issue in their discussion of Cochrane and Piazzesi (2005).

[22] This model provides a very tight fit to observed zero-coupon bond prices. Bolder and Gusba (2002) find that this model has an average root-mean-squared error fit to a collection of Government of Canada bond prices of approximately four basis points.

The upward-sloping curve over the period corresponds to the general observation that long-term interest rates typically dominate short-term rates in Canada. The dotted lines bracketing the average yield curve indicate the standard deviation of the curves; we can observe that the short end of the zero-coupon curve appears to be slightly more volatile than the long end. This is also consistent with previous work on the Canadian term structure of interest rates.

Figure 4: **Another view of historical Canadian zero-coupon curves**: This figure outlines 187 monthly zero-coupon curves beginning in January 1990 until August 2005.



Tables 1 and 2 illustrate the traditional measure of model fit for each of the six different dynamic term-structure models considered in this paper. Table 1 provides summary statistics that attempt to describe how well the model fits the entire curve. Note that no forecasting is involved in these computations; indeed, one can think of these as the fitted versus actual values of an ordinary least-squares regression. Observe that the root-mean-squared errors (RMSE) are quite similar across all models. The exponential-spline model appears to have the closest fit, while the symmetric $A_0(3)$ demonstrates the worst fit. The other models appear, from the perspective of the entire curve, to be quite similar. Table 2 drills down somewhat and considers the fit at different specific zero-coupon tenors. Here we note that, while the $A_0(3)$ models generally appear to fit better at the short end of the curve, the empirical models seem to do better at the longer zero-coupon tenors.

We have, however, argued that the ability of a model to fit the data is *not* a good measure of the capacity of a model to capture interest rate dynamics. Recent work by Duffee (2002) and Diebold and Li (2003) suggests that one needs to consider a model's ability to forecast future zero-coupon rates. In the subsequent sections, we

Table 1: **Goodness of Fit**: In this table, we present the traditional *in-sample* goodness of fit of our empirical and theoretical models estimated with the 187 months of zero-coupon data. We present the overall fit of these models to the curve in terms of root-mean-squared and mean-absolute error. All values are in basis points.

Models	**RMSE**					**MAE**				
	Mean	**Median**	**Max**	**Min**	**STD**	**Mean**	**Median**	**Max**	**Min**	**STD**
Empirical Models										
Nelson-Siegel	9.27	8.35	27.45	1.98	4.50	7.41	6.49	23.79	1.52	3.98
Exponential spline	6.00	5.47	14.76	0.08	3.80	4.00	3.67	10.42	0.06	2.60
Fourier-series	8.45	7.53	27.40	0.51	5.46	4.58	3.85	11.78	0.34	2.76
Theoretical Models										
Diagonal A_γ	8.70	6.86	32.02	0.58	5.69	6.18	4.53	26.63	0.33	4.49
Symmetric A_γ	11.20	9.98	35.01	0.85	7.23	8.52	6.70	29.35	0.61	6.12
Unrestricted A_γ	9.30	7.79	31.67	0.96	5.80	6.72	5.25	25.83	0.80	4.49

will examine the forecasting ability of each of our models.

4.2 Forecasting zero-coupon rates

The mechanics of the forecasting exercise are straightforward. We provide a description of the out-of-sample approach, since the in-sample estimation is merely a special case of the more general out-of-sample forecasting technique. As previously discussed, we have 187 months of Canadian government zero-coupon interest rate data. Given this dataset, we start from the qth month and try to forecast the zero-coupon term structure in n periods from this point of time. A good way to think about this approach is to imagine that we are an analyst who is trying to use a model, at time t_q, to predict interest rates in n periods. In both the empirical and theoretical models this involves conditioning on the information available up until time t_q (i.e., the filtration $\mathcal{F}_{t_q}$) and projecting the t_{q+n} values of the state variables using their conditional expectation. This implies that, for each of the models, we estimate the parameters using the data up until time t_q to estimate the model and forecast the state variables.[23]

Using these forecasted state variables, we proceed to construct the projected zero-coupon curve associated with these forecasted state variables. For the theoretical models, we compute the expected value of the state variables using the conditional expectation implied by the model as

$$\mathbb{E}\left(X_{t_{q+n}} \mid \mathcal{F}_{t_q}\right) = e^{-\kappa(t_{q+n}-t_q)} X_{t_q}. \tag{110}$$

[23] For the empirical models, this is straightforward. The theoretical $A_0(3)$ models, however, are much more computationally demanding and require substantial expense for the estimation of model parameters.

Table 2: **Goodness of Fit by Specific Tenor**: In this table, we again present the goodness of fit of the empirical and theoretical models. The difference here is that we present the root-mean-squared and mean-absolute errors for specific zero-coupon maturities, rather than the entire zero-coupon curve, as presented in Table 1. All values are in basis points.

Tenor	**RMSE**			**MAE**		
	Nelson-Siegel	**Exponential spline**	**Fourier series**	**Nelson-Siegel**	**Exponential spline**	**Fourier series**
3 months	18.00	16.43	17.58	14.24	9.76	12.04
6 months	9.13	11.85	23.32	6.48	8.71	16.57
1 year	10.59	2.57	13.53	8.71	1.75	10.16
2 years	9.87	3.44	4.93	7.71	2.78	3.36
5 years	9.43	2.08	3.63	6.77	1.68	2.65
7.5 years	9.27	3.94	4.31	7.70	3.14	3.29
10 years	7.13	3.25	3.33	5.66	2.30	2.59
15 years	10.59	8.67	5.44	8.53	6.58	4.02
	Diagonal	**Symmetric**	**Unrestricted**	**Diagonal**	**Symmetric**	**Unrestricted**
3 months	12.21	13.07	13.50	9.09	10.09	10.49
6 months	0.00	0.00	0.00	0.00	0.00	0.00
1 year	6.07	6.85	7.35	4.31	5.14	5.63
2 years	0.00	0.00	0.00	0.00	0.00	0.00
5 years	14.86	17.34	16.30	10.41	12.45	11.57
7.5 years	11.67	14.91	11.92	8.00	10.98	8.57
10 years	0.00	0.00	0.00	0.00	0.00	0.00
15 years	10.32	18.67	10.79	8.30	14.77	8.59

Thus, the expected zero-coupon curve at time t_{q+n}, determined using equation (25), is

$$\mathbb{E}\left(z(t_{q+n},T) \mid \mathcal{F}_{t_q}\right) = \exp\left(-c(T-t_{q+n}) - b(T-t_{q+n})^T \underbrace{\mathbb{E}\left(X_{t_{q+n}} \mid \mathcal{F}_{t_q}\right)}_{\text{Equation (110)}}\right). \tag{111}$$

For the empirical models, the approach is similar. We forecast the t_{q+n} value of the VAR(1) form of the state variables using the following relation:

$$\mathbb{E}\left(x_{t_{q+n}} \mid \mathcal{F}_{t_q}\right) = \sum_{i=1}^{n-1} \Phi^i \alpha + \Phi^n x_{t_q}. \tag{112}$$

For an AR(1) model, there is a separate $\Phi_k \in \mathbb{R}$ for each of the state variables where k denotes the dimension of the state-variable vector. In a VAR(1) setting, however, the state variables are modelled as a system; thus, we have that $\Phi \in \mathbb{R}^{k\times k}$. The expected empirical zero-coupon curve at time t_{q+n} is

$$\mathbb{E}\left(z(t_{q+n},T) \mid \mathcal{F}_{t_q}\right) = F(T-t_{q+n})^T \underbrace{\mathbb{E}\left(x_{t_{q+n}} \mid \mathcal{F}_{t_q}\right)}_{\text{Equation (112)}}, \tag{113}$$

for the Nelson-Siegel approach suggested by Diebold and Li (2003). It is given as

$$\mathbb{E}\left(z(t_{q+n},T)\mid \mathcal{F}_{t_q}\right) = -\frac{\ln\left(F(T-t_{q+n})^T \overbrace{\mathbb{E}\left(x_{t_{q+n}}\mid \mathcal{F}_{t_q}\right)}^{\text{Equation (112)}}\right)}{T-t_{q+n}}, \tag{114}$$

for the exponential-spline and Fourier-series models.

Once the forecast is made, we can take a peek into the future and observe the actual zero-coupon curves at time t_{q+n}, and thereby measure the efficiency of the forecast. Having observed the actual value, we need some measure of distance, call it $\delta(\cdot)$, between the forecasted and observed values. Essentially, we are looking for a metric of the form

$$\delta\left(\mathbb{E}\left(z(t_{q+n},T)\mid \mathcal{F}_{t_q}\right) - z(t_{q+n},T)\right). \tag{115}$$

It is customary in the literature to use variations of the ℓ^1 and ℓ^2 norm to measure distance. The first measure, termed the mean-absolute error, is defined as

$$\begin{aligned}\text{MAE} &= \sum_{q=1}^{M} \frac{\left|\mathbb{E}\left(z(t_{q+n},T)\mid \mathcal{F}_{t_q}\right) - z(t_{q+n},T)\right|^T \mathbb{I}}{M}, \\ &= \frac{\left\|\mathbb{E}\left(z(t_{q+n},T)\mid \mathcal{F}_{t_q}\right) - z(t_{q+n},T)\right\|_1}{M},\end{aligned} \tag{116}$$

where $\mathbb{I}$ is a vector of ones, M is the number of zero-coupon bonds observed at time q_{t+n}, and $\|x\|_1$ denotes the ℓ^1 norm. The second measure of distance, called the root-mean-squared error, is defined similarly

$$\begin{aligned}\text{RMSE} &= \sum_{q=1}^{M} \sqrt{\frac{\left(\mathbb{E}\left(z(t_{q+n},T)\mid \mathcal{F}_{t_q}\right) - z(t_{q+n},T)\right)^T\left(\mathbb{E}\left(z(t_{q+n},T)\mid \mathcal{F}_{t_q}\right) - z(t_{q+n},T)\right)}{M}}, \\ &= \frac{\left\|\mathbb{E}\left(z(t_{q+n},T)\mid \mathcal{F}_{t_q}\right) - z(t_{q+n},T)\right\|_2}{\sqrt{M}},\end{aligned} \tag{117}$$

where $\|x\|_2$ denotes the ℓ^2 norm.

This process is repeated at time t_{q+1}. To summarize the various steps for an arbitrary time t_q:

- each model is estimated using the data up until time t_q;
- the model parameters and equations (110) and (112) are used to forecast the state variables at time t_{q+n};
- the forecasted state variables are used to forecast the zero-coupon term structure at time t_{q+n} using equations (111), (113), and (114);
- we then observe the actual zero-coupon term structure at time t_{q+n}, and proceed to compute the mean-absolute and root-mean-squared errors as described in equations (116) and (117).

The difference with in-sample forecasting is that the parameters are estimated only once using the entire sample period. This contrasts with the rolling parameter estimates performed in the out-of-sample forecasting exercise. As such, the filtration used to formulate the conditional expectations is $\mathcal{F}_{t_T}$. In-sample forecasting, therefore, is *not* really a natural experiment, because an analyst cannot actually condition on *future* information in constructing forecasts.[24] In-sample forecasting should thus be considered as a test of how well the model can actually fit the data. The out-of-sample exercise, conversely, represents a test of the consistency of the model with the actual data-generating process.

In both the in-sample and out-of-sample forecasting sections, we begin the exercise in February 2002 (i.e., the 145th monthly observation) and complete it 43 months later in August 2005. This allows us to perform 42 non-overlapping monthly forecasts, 40 overlapping three-month forecasts, and 37 overlapping six-month forecasts. To make this clear, for the monthly forecasts we compute the root-mean-squared and mean-absolute error, as described in equations (116) and (117), for each of the 42 forecasting periods. To perform this computation, we use 40 forecasted and observed zero-coupon rates. These rates are equally spaced from six months to twenty years. Using this information, we consider a number of summary statistics of the root-mean-squared and mean-absolute errors for the one-, three-, and six-month forecasting horizons. At each forecast horizon, for example, we compute the average, median, maximum, minimum, and the standard deviation of the corresponding root-mean-squared and mean-absolute forecast errors.

Note that, for each forecasting horizon, we compare the results with a random-walk model. The random-walk model assumes that zero-coupon interest rates are martingales and, as such, our best estimate of the future value is the current value. Thus, the random-walk forecast is also a no-change forecast. The random-walk model is a common comparator in the forecasting literature precisely because it is notoriously difficult to outperform.

4.2.1 In-sample forecasts

Table 3 illustrates the in-sample root-mean-squared and mean-absolute errors for the three empirical models where the dynamics of the state variables are assumed to follow a VAR(1) process. At the one-month forecasting horizon, the Fourier-series model exhibits the best in-sample forecasting performance and is the only empirical model that outperforms the random walk. At the three- and six-month forecasting horizons, all of the models outperform the random walk. The Fourier-series model exhibits the best in-sample forecasting performance at the three-month horizon, while the Nelson-Siegel model appears to do the best job at the six-month horizon. The exponential-spline models appear to be a close second.

The Fourier-series model is the strongest performer when we focus on the root-mean-squared forecast error; it beats the Nelson-Siegel and exponential-spline models at both the one- and three-month forecasting horizons. When we focus on the mean-absolute error, however, the exponential-spline model dominates the other models

[24] Although we are certain many analysts would be very happy to do so.

across all forecasting horizons. Yet, all three generate very comparable in-sample results. In most cases, the difference between models is no more than a basis point. As such, it seems that we can conclude that the in-sample performance of the empirical models is quite similar when considering the entire zero-coupon curve.

Table 3: **In-Sample Forecasts with VAR(1) Model**: In this table, we again present the root-mean-squared and mean-absolute error for a series of one-, three-, and six-month forecasts where the state variables are estimated using a standard VAR(1) model. All values are in basis points.

Models	**RMSE**					**MAE**				
	Mean	**Median**	**Max**	**Min**	**STD**	**Mean**	**Median**	**Max**	**Min**	**STD**
One-month forecast										
Random walk	18.24	13.82	57.47	3.35	12.61	15.61	12.72	54.28	0.28	13.29
Nelson-Siegel	19.22	14.96	58.84	4.70	12.16	15.71	11.96	56.99	0.53	13.16
Exponential spline	19.30	15.37	61.06	8.13	11.35	15.02	11.45	59.69	0.06	12.72
Fourier series	17.86	13.06	60.31	4.84	12.98	15.01	11.10	58.81	0.00	13.88
Three-month forecast										
Random walk	29.17	25.32	77.97	7.43	17.48	24.55	20.35	76.76	1.24	18.79
Nelson-Siegel	27.61	20.24	80.25	6.83	17.45	22.57	16.59	77.63	0.10	19.15
Exponential spline	26.42	19.75	81.05	7.11	16.94	20.79	13.03	79.38	0.04	18.92
Fourier series	25.27	18.08	81.51	4.02	19.22	21.18	13.88	79.60	0.14	20.24
Six-month forecast										
Random walk	35.69	30.09	79.34	7.99	17.03	30.18	26.91	74.77	3.31	18.67
Nelson-Siegel	30.87	30.23	58.02	18.02	13.05	27.35	24.22	54.66	16.20	13.36
Exponential spline	31.38	30.53	56.67	18.78	12.90	27.12	27.42	53.53	13.84	13.51
Fourier series	33.65	27.28	60.85	23.56	13.38	28.03	22.68	58.58	13.85	16.18

The approach that we have used so far—and which is described in equations (116) and (117)—is *not* the typical approach used in the forecasting literature. Typically, instead of examining the root-mean-squared or mean-absolute forecasting error of the entire zero-coupon yield curve, the focus is on specific zero-coupon rate tenors. By examining the full zero-coupon curve, we are essentially examining the average forecasting error across the entire yield curve. We believe, since we are interested in describing the dynamics of the entire term structure of interest rates, that this type of approach is preferable. We do admit, however, that with our approach there is some aggregration of information. In particular, we do not capture the ability of a given model to fit one area relative to other parts of the zero-coupon curve. For this reason, we supplement our analysis with the more traditional zero-coupon rate tenor-based measures. To this end, we examine the root-mean-squared and absolute forecast errors for a selection of zero-coupon tenors ranging from three months to 15 years. The results are provided in Table 4.

The key difference with an analysis of a specific tenor versus consideration of the entire curve is that the error computations occur across time, rather than at a given point in time. If, for example, we construct a

Table 4: **In-Sample Forecasts with VAR(1) Model by Specific Tenor**: In this table, we present the root-mean-squared and mean-absolute error for a series of one-month forecasts where the state variables are estimated using a standard VAR(1) models. The difference in this table, however, is that we focus on specific zero-coupon maturities, rather than the entire zero-coupon curve as presented in Table 3. All values are in basis points.

Tenor	**RMSE**				**MAE**			
	Random walk	**Nelson-Siegel**	**Exponential spline**	**Fourier series**	**Random walk**	**Nelson-Siegel**	**Exponential spline**	**Fourier series**
One-month forecast								
3 months	13.35	15.23	10.60	14.55	10.30	12.02	8.07	11.50
6 months	14.84	13.65	13.26	14.79	11.31	11.27	10.88	12.22
1 year	21.09	20.91	19.79	20.20	15.91	16.36	15.45	16.17
2 years	25.21	25.67	25.03	25.74	19.09	19.52	19.08	18.78
5 years	24.98	25.75	24.14	25.26	19.24	19.69	18.47	19.12
7.5 years	23.38	23.78	23.12	22.87	17.92	16.96	16.45	16.72
10 years	23.17	22.85	21.97	22.20	17.44	17.86	17.06	16.83
15 years	20.05	20.66	22.76	20.13	15.18	16.54	18.96	15.14
Six-month forecast								
3 months	47.36	37.88	37.59	43.44	41.49	30.52	30.70	34.63
6 months	44.98	37.46	38.70	40.28	37.38	30.93	32.35	32.86
1 year	43.13	41.12	41.15	42.16	35.49	34.85	34.86	35.88
2 years	44.21	43.23	40.05	40.42	38.05	35.64	34.84	34.68
5 years	44.71	39.99	34.73	35.08	36.37	33.84	29.07	29.30
7.5 years	41.11	32.38	27.03	29.30	34.51	26.30	21.49	23.63
10 years	38.93	32.96	27.16	28.63	33.19	26.19	21.19	22.59
15 years	34.94	28.90	30.26	25.96	29.17	22.60	23.80	21.07

sequence of one-month forecasts of the one-year rate, we are essentially computing a time series of forecast errors. We can then determine the root-mean-squared and mean-absolute forecast errors across the time series. When analyzing the entire curve, we compute a root-mean-squared and mean-absolute forecast error for each date in the forecasting interval; we then examine a variety of summary statistics of these forecast errors.

To make this idea more clear, let us work through the specific formulation of the tenor-based forecast errors. As before, we use equations (110) and (112) to forecast the state variables. In this case, however, we want to predict a specific zero-coupon rate, which we will denote as $z(t,\tau)$ where the tenor is $\tau - t$. We then modify equations (111), (113), and (114) as follows:

$$\mathbb{E}\left(z(t_{q+n}, t_{q+n}+\tau) \mid \mathcal{F}_{t_q}\right) = \exp\left(-c(\tau) - b(\tau)^T \underbrace{\mathbb{E}\left(X_{t_{q+n}} \mid \mathcal{F}_{t_q}\right)}_{\text{Equation (110)}}\right), \tag{118}$$

$$\mathbb{E}\left(z(t_{q+n}, t_{q+n}+\tau) \mid \mathcal{F}_{t_q}\right) = F(\tau)^T \underbrace{\mathbb{E}\left(x_{t_{q+n}} \mid \mathcal{F}_{t_q}\right)}_{\text{Equation (112)}}, \tag{119}$$

and,

$$\mathbb{E}\left(z(t_{q+n}, t_{q+n}+\tau) \mid \mathcal{F}_{t_q}\right) = -\frac{\ln\left(F(\tau)^T \overbrace{\mathbb{E}\left(x_{t_{q+n}} \mid \mathcal{F}_{t_q}\right)}^{\text{Equation (112)}}\right)}{\tau}. \tag{120}$$

The tenor-based forecast errors, in Table 4, provide a useful perspective. One clear conclusion from Table 4 is that the forecast errors are *not* constant across the spectrum of zero-coupon tenors. For the one-month forecast horizon, the exponential-spline model outperforms the Nelson-Siegel model at the three-month tenor, while it underperforms at the 10- and 15-year tenor. Also, at the six-month forecasting horizon, the Fourier-series model fairly dramatically underperforms at short zero-coupon tenors, but outperforms at longer zero-coupon tenors. Finally, there appears to be a general tendency for the forecast errors to increase with the zero-coupon rate tenor.

Table 5: **In-Sample $A_0(3)$ Model Forecasts with Different A_γ Matrix Restrictions**: In this table, we present the root-mean-squared and mean-absolute error for a series of one-, three-, and six- month forecasts where the A_γ matrix found in the market price of risk has a diagonal, symmetric, and unrestricted form. All values are in basis points.

Models	**RMSE**					**MAE**				
	Mean	**Median**	**Max**	**Min**	**STD**	**Mean**	**Median**	**Max**	**Min**	**STD**
One-month forecast										
Random walk	18.24	13.82	57.47	3.35	12.61	15.61	12.72	54.28	0.28	13.29
Diagonal A_γ	20.07	16.55	62.60	5.99	13.31	16.71	13.29	59.78	0.34	13.96
Symmetric A_γ	26.61	23.37	58.95	9.38	11.12	18.62	17.41	58.49	1.14	13.76
Unrestricted A_γ	21.03	17.49	61.17	4.89	13.43	17.56	13.70	57.48	1.13	14.05
Three-month forecast										
Random walk	29.17	25.32	77.97	7.43	17.48	24.55	20.35	76.76	1.24	18.79
Diagonal A_γ	32.29	29.02	80.80	8.34	18.50	27.55	25.49	78.19	1.34	20.14
Symmetric A_γ	37.54	35.50	83.40	14.88	15.70	29.34	27.12	82.76	0.70	20.14
Unrestricted A_γ	38.14	33.83	89.05	10.97	20.30	32.79	29.74	87.31	0.07	22.87
Six-month forecast										
Random walk	35.69	30.09	79.34	7.99	17.03	30.18	26.91	74.77	3.31	18.67
Diagonal A_γ	42.45	34.95	101.51	12.08	21.08	37.40	31.48	100.15	2.62	22.86
Symmetric A_γ	47.73	43.08	94.48	9.24	19.48	40.68	36.16	93.28	1.91	24.16
Unrestricted A_γ	54.13	51.18	139.80	13.61	28.67	49.09	44.86	135.70	2.27	30.85

We next examine the in-sample forecasting ability of the $A_0(3)$ models. Table 5 again illustrates a range of summary statistics for the root-mean-squared and mean-absolute forecast errors at the one-, three-, and six-month horizons. These figures are directly comparable with Table 3. The first thing to notice is that the model with the diagonal A_γ matrix substantially outperforms the symmetric and unrestricted forms of A_γ. This is the case using either the root-mean-squared or mean-absolute errors. Furthermore, at the three- and six-month forecasting horizons, the unrestricted form underperforms the other models.

Table 6: **In-Sample Forecasts for $A_0(3)$ Model by Specific Tenor**: In this table, we present the root-mean-squared and mean-absolute error for a series of one- and three-month forecasts for the different versions of the $A_0(3)$ model. The difference between this table and Table 5 is that in this table we focus on specific zero-coupon maturities, rather than the entire zero-coupon curve. All values are in basis points.

Tenor	RMSE				MAE			
	Random walk	Diagonal	Symmetric	Full	Random walk	Diagonal	Symmetric	Full
One-month forecast								
3 months	13.35	13.66	14.32	14.03	10.30	10.47	11.82	10.82
6 months	14.84	14.35	15.68	14.33	11.31	11.87	12.63	11.04
1 year	21.09	20.51	22.24	20.59	15.91	16.89	17.93	16.31
2 years	25.21	25.45	24.94	25.44	19.09	19.53	19.08	19.19
5 years	24.98	26.57	27.53	27.99	19.24	20.70	19.27	21.23
7.5 years	23.38	23.82	26.44	24.85	17.92	17.22	18.96	18.23
10 years	23.17	23.36	23.35	24.50	17.44	17.68	17.63	19.02
15 years	20.05	23.51	32.91	24.70	15.18	18.16	28.94	19.85
Six-month forecast								
3 months	47.36	49.88	60.30	46.18	41.49	40.19	45.86	38.86
6 months	44.98	48.50	60.43	44.37	37.38	38.83	48.20	36.52
1 year	43.13	52.03	60.77	46.69	35.49	42.16	50.22	39.38
2 years	44.21	56.81	56.17	54.77	38.05	45.97	45.70	45.18
5 years	44.71	54.41	46.41	65.65	36.37	46.83	37.65	54.68
7.5 years	41.11	43.83	39.29	59.30	34.51	37.20	31.85	49.32
10 years	38.93	43.93	43.38	61.34	33.19	37.55	37.23	51.72
15 years	34.94	44.28	56.24	63.82	29.17	37.86	50.54	54.78

Table 6, by focusing on forecast errors organized by specific zero-coupon rate tenors, provides a different perspective on the results in Table 5. Specifically, we can see that the symmetric form of A_γ has substantial in-sample forecasting difficulty at the short end of the curve for the six-month horizon. The root-mean-squared error of the symmetric model is approximately 10–15 basis points more than that observed in the diagonal and unrestricted models at the three-month tenor for both the six-month forecasting horizons. At the 10-year tenor, however, the forecast errors are quite comparable with the diagonal and restricted versions of the $A_0(3)$ model; indeed, they are superior to those posted by the unrestricted model. We also note that, while the diagonal and unrestricted models are quite comparable for the one-month forecasting horizon, the unrestricted model substantially underperforms at the six-month horizon for zero-coupon tenors beyond about five years.

It is also interesting, and useful, to compare Tables 3 and 5; this amounts to a fair in-sample forecasting comparison between the empirical and theoretical models described in this paper. The empirical models demonstrate superior in-sample forecasting ability relative to the $A_0(3)$ models. Indeed, we observe that the best theoretical model underperforms the worst empirical model for all forecasting horizons from one to nine basis points. The differential between the two model classes also appears to increase with the forecasting horizon.

At the one-month horizon, the results are actually quite close. At the three- and six-month horizons, however, the underperformance of the theoretical models widens. At the six-month horizon, for example, the in-sample average root-mean-squared error is often about ten basis points more than that observed among the empirical models. The results are similar, if somewhat less extreme, at the three-month horizon.

4.2.2 Out-of-sample forecasts

While interesting, the in-sample forecasting performance is of secondary relevance compared with the ability of these models to forecast out-of-sample. We next examine, therefore, the same summary statistics for the out-of-sample forecasting exercise. Table 7 illustrates the out-of-sample zero-coupon rate forecasting performance using a VAR(1) model for the state variables of the empirical models. The basic ordering is quite similar to the in-sample forecasting performance observed in Table 3. That is, by the root-mean-squared error metric, the Fourier-series model provides the best forecasts at the one- and three-month horizons, while the Nelson-Siegel model outperforms at the six-month horizon. The random-walk model is, as before, difficult to beat at both the one- and three-month horizons. When focusing on the mean-absolute error, the exponential-spline model provides the best out-of-sample forecasts across all horizons. Finally, we observe that the out-of-sample forecasting performance deteriorates relative to the in-sample results provided in Table 3. This is a natural outcome, because less information is available for estimating the model parameters in the out-of-sample exercise.

Table 8 examines the root-mean-squared and mean-absolute out-of-sample forecast errors by the specific zero-coupon tenor. This table is comparable with the in-sample version provided in Table 4. The results are also quite similar insofar as we observe that the models generally out-of-sample forecast the short end of the zero-coupon curve more accurately than the long end. The Fourier-series model again has trouble forecasting short-term zero-coupon tenors; this is a particular problem at the six-month horizon. We previously saw that the Nelson-Siegel and exponential-spline models look quite similar when we examine the entire curve. By investigating the individual zero-coupon tenors, however, we can see that the exponential-spline model does slightly better at the short end, while the Nelson-Siegel model outperforms at the longer tenors.

A final perspective on the out-of-sample forecasting errors of the empirical models is provided in Figure 5. This figure shows, for the three empirical models, the mean out-of-sample forecast errors by zero-coupon tenor for the one-, three-, and six-month forecast horizons. We would expect, if the model produced unbiased forecasts, that the average forecast error should be approximately zero. All three models, however, appear to produce *positive* out-of-sample forecast errors for the short-term tenors and *negative* out-of-sample forecast errors for the longer-term zero-coupon tenors. Since we define the errors as actual minus forecast, a positive bias in the forecast errors implies a persistent underestimation of future zero-coupon rates. This trend appears to reverse itself with negative forecast errors. Negative forecast errors suggest a persistent overestimation of future long-term zero-coupon rates over the forecasting horizon. We note that the random-walk model exhibits a rather

Table 7: **Out-Of-Sample Forecasts with VAR(1) Model**: In this table, we again present the root-mean-squared and mean-absolute error for a series of one-, three-, and six-month forecasts where the state variables are estimated using a standard VAR(1) model. All values are in basis points.

Models	**RMSE**					**MAE**				
	Mean	**Median**	**Max**	**Min**	**STD**	**Mean**	**Median**	**Max**	**Min**	**STD**
One-month forecast										
Random walk	18.24	13.82	57.47	3.35	12.61	15.61	12.72	54.28	0.28	13.29
Nelson-Siegel	19.77	16.22	61.94	4.71	12.12	16.35	12.91	59.90	1.63	13.00
Exponential spline	20.21	17.04	62.90	7.62	11.72	15.90	13.03	61.56	0.13	12.93
Fourier series	18.70	14.04	62.32	4.28	13.40	15.92	12.40	60.85	0.42	14.01
Three-month forecast										
Random walk	29.17	25.32	77.97	7.43	17.48	24.55	20.35	76.76	1.24	18.79
Nelson-Siegel	30.04	23.78	87.07	11.21	17.43	24.76	20.34	84.05	0.11	19.38
Exponential spline	29.34	21.51	85.70	10.73	17.64	23.48	15.72	84.03	1.16	19.53
Fourier series	28.42	19.61	86.92	8.81	20.26	23.79	17.49	85.05	1.64	21.19
Six-month forecast										
Random walk	35.69	30.09	79.34	7.99	17.03	30.18	26.91	74.77	3.31	18.67
Nelson-Siegel	31.65	28.72	59.55	20.67	13.37	27.96	21.85	56.13	16.05	13.93
Exponential spline	32.15	33.35	57.82	20.58	13.09	27.62	26.43	54.64	14.03	13.76
Fourier series	33.89	27.69	62.44	22.19	14.48	28.12	25.38	60.20	7.66	18.06

smooth pattern in average forecasts errors, while the three empirical models demonstrate an oscillatory pattern. The Fourier-series model, however, exhibits rather less variation relative to the other two models.

We next examine the out-of-sample performance of the theoretical models. Table 9 describes summary statistics for the one-, three-, and six-month out-of-sample forecasts of the $A_0(3)$ model with diagonal, symmetric, and unrestricted forms of the A_γ matrix that pre-multiplies the state variable vector in the market price of risk. Unlike the empirical models, the out-of-sample forecasting results are quite different relative to the in-sample exercise. The surprise is the performance of the unrestricted form of A_γ. At all forecasting horizons, the unrestricted model has the smallest root-mean-squared and mean-absolute forecast errors. It also has the lowest variation in forecast performance. In other words, therefore, the unrestricted model is the clear front-runner among the $A_0(3)$ models in terms of out-of-sample forecasting. The next best model is the diagonal model followed by the symmetric model, which performs quite poorly. None of the models, however, outperforms the random-walk model.

A view of the out-of-sample performance of the three versions of the $A_0(3)$ model by zero-coupon tenor is outlined in Table 11. What is striking is the quite poor forecasting performance of the diagonal and symmetric models for short-term zero-coupon tenors at the six-month horizon—either on the basis of root-mean-squared or mean-absolute error. The forecasting error of the unrestricted model at the three-month tenor is approximately

Table 8: **Out-of-Sample Forecasts with VAR(1) Model by Specific Tenor**: In this table, we present the root-mean-squared and mean-absolute error for a series of one- and six-month forecasts where the state-variable dynamics are estimated using a standard VAR(1) process. The difference in this table, however, is that we focus on specific zero-coupon maturities, rather than the entire zero-coupon curve as presented in Table 3. All values are in basis points.

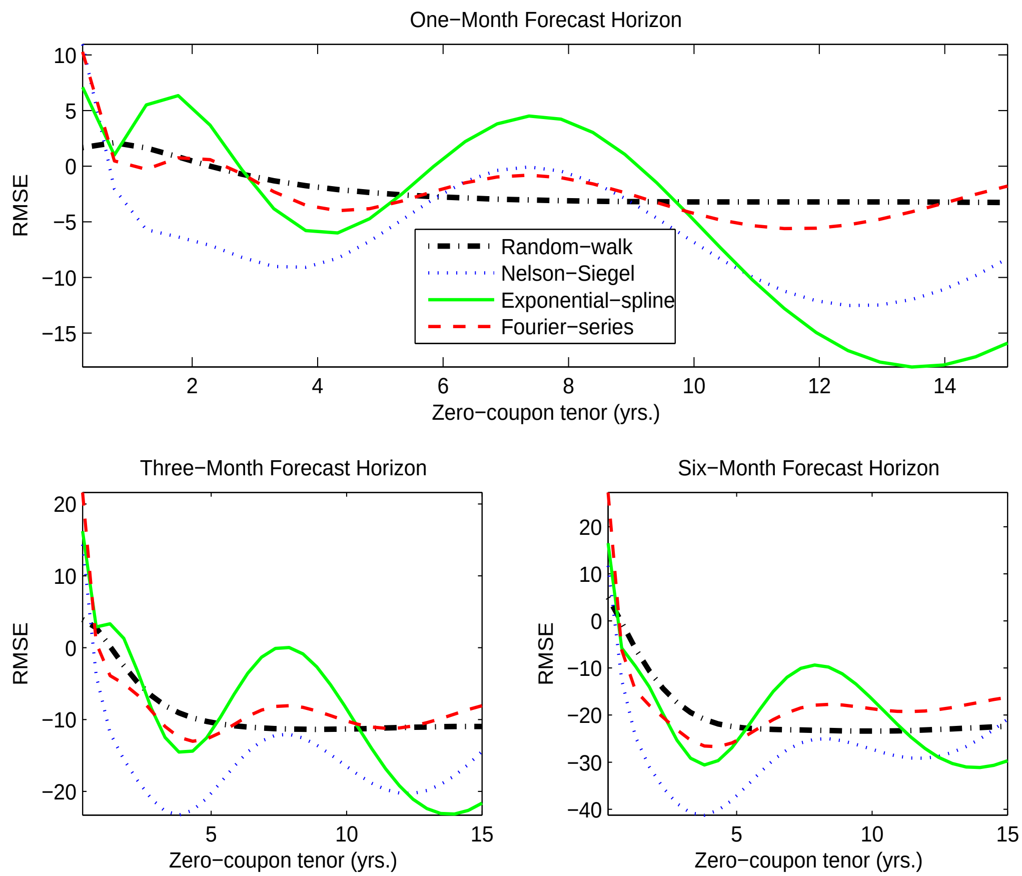
Tenor	**RMSE**				**MAE**			
	Random walk	**Nelson-Siegel**	**Exponential spline**	**Fourier series**	**Random walk**	**Nelson-Siegel**	**Exponential spline**	**Fourier series**
				One-month forecast				
3 months	13.35	16.04	13.37	19.87	10.30	12.78	10.71	14.60
6 months	14.84	13.99	14.10	16.92	11.31	11.55	11.56	12.99
1 year	21.09	21.32	21.38	22.79	15.91	16.67	16.27	17.32
2 years	25.21	26.40	26.62	28.25	19.09	20.08	19.59	19.94
5 years	24.98	26.39	25.14	26.36	19.24	20.61	19.51	20.21
7.5 years	23.38	23.70	23.68	23.51	17.92	17.42	16.76	17.61
10 years	23.17	23.37	22.72	22.90	17.44	18.45	17.72	17.46
15 years	20.05	21.26	24.35	20.47	15.18	17.15	20.58	15.59
				Six-month forecast				
3 months	47.36	40.60	45.22	62.53	41.49	32.99	36.47	46.14
6 months	44.98	39.71	43.97	52.93	37.38	32.77	36.45	39.81
1 year	43.13	44.54	46.40	52.12	35.49	37.76	38.94	42.78
2 years	44.21	49.86	46.38	49.45	38.05	40.70	39.44	40.99
5 years	44.71	48.64	41.41	42.60	36.37	42.05	35.03	36.64
7.5 years	41.11	38.33	30.12	34.87	34.51	32.58	24.52	29.42
10 years	38.93	38.49	31.41	33.96	33.19	31.04	25.16	27.52
15 years	34.94	32.83	37.76	30.34	29.17	25.70	30.77	24.33

one half the corresponding forecast errors observed for the diagonal models at the six-month horizon. The exact reason for this phenomenon is unclear. One probable explanation is that the reduced number of parameters in the A_γ matrix restricts the ability of the diagonal and symmetric models to accurately capture the dynamics of risk premia.

We also note that, for both the one- and six-month horizons at the short zero-coupon tenors, the unrestricted model is quite close to, or occasionally even outperforms, the random-walk model. As we lengthen the tenor, however, the out-of-sample forecasting performance of the unrestricted model deteriorates relative to the random-walk model. This contrasts with the empirical models that exhibited difficulty with the short-term zero-coupon tenors, but forecast better at longer zero-coupon tenors.

Figure 6 provides the final results for the forecasting exercise by outlining the mean $A_0(3)$ forecast error by zero-coupon tenor for the one-month forecast horizon. We also observe an oscillatory pattern in average forecast errors across the three models. Indeed, the oscillations are larger than those observed for the empirical models in Figure 5. Specifically, we observe negative errors at the short zero-coupon tenors, a trend towards positive (or

Figure 5: **Out-of-Sample Empirical VAR(1) Forecasts by Tenor**: This figure illustrates the mean out-of-sample forecast errors (in basis points) by specific zero-coupon rate tenors for the three empirical models across the one-, three-, and six-month forecasting horizons.



less negative) errors at the intermediate tenors, and a return to negative errors at the longer-term zero-coupon tenors. Across all forecasting horizons, however, the unrestricted model demonstrates the smallest oscillation and the closest relationship to the random-walk model. Note that, at the three- and six-month horizons, the diagonal and symmetric models post negative out-of-sample forecast errors across all zero-coupon tenors. This suggests that these models routinely overestimate the future value of zero-coupon rates over this forecasting interval.

We have examined the various empirical and theoretical models on the basis of their in- and out-of-sample forecasting ability on three primary dimensions: the ability to fit the entire zero-coupon curve, the capacity of these models to fit specific zero-coupon tenors, and the direction of the forecast errors. Overall, the results show that the empirical models forecast better than the theoretical models on basically every dimension on both the in- and out-of-sample forecasting exercises. The underperformance of the theoretical models seems to be most

Table 9: **Out-of-Sample $A_0(3)$ Model Forecasts with Different A_γ Matrix Restrictions**: In this table, we present the root-mean-squared and mean-absolute error for a series of one-, three-, and six-month forecasts where the A_γ matrix found in the market price of risk has a diagonal, symmetric, and unrestricted form. All values are in basis points.

Models	**RMSE**					**MAE**				
	Mean	**Median**	**Max**	**Min**	**STD**	**Mean**	**Median**	**Max**	**Min**	**STD**
One-month forecast										
Random walk	18.24	13.82	57.47	3.35	12.61	15.61	12.72	54.28	0.28	13.29
Diagonal A_γ	21.49	20.31	59.89	3.58	13.57	17.95	15.92	56.75	0.23	14.31
Symmetric A_γ	23.86	19.65	65.11	5.20	13.92	18.84	13.57	62.73	0.36	15.83
Unrestricted A_γ	21.34	17.55	58.46	5.44	12.82	17.46	13.13	57.49	0.95	13.67
Three-month forecast										
Random walk	29.17	25.32	77.97	7.43	17.48	24.55	20.35	76.76	1.24	18.79
Diagonal A_γ	37.97	32.80	90.57	8.62	19.80	33.09	29.54	88.56	0.41	22.21
Symmetric A_γ	44.00	37.37	124.68	11.91	28.74	38.15	35.61	119.74	1.22	31.30
Unrestricted A_γ	35.66	30.24	85.33	10.83	18.97	30.03	26.18	83.90	0.41	21.28
Six-month forecast										
Random walk	35.69	30.09	79.34	7.99	17.03	30.18	26.91	74.77	3.31	18.67
Diagonal A_γ	56.52	53.22	117.48	16.44	25.92	51.17	48.92	114.80	2.41	28.28
Symmetric A_γ	71.40	56.77	213.13	17.38	45.63	65.17	48.54	206.60	0.20	46.68
Unrestricted A_γ	46.08	42.30	119.39	8.56	24.16	40.36	37.74	115.19	3.68	26.30

evident as we increase the forecasting horizon and increase the zero-coupon tenor.

Within the class of empirical models, we conclude that the Nelson-Siegel model slightly outperforms the exponential-spline and Fourier-series models. It should be noted, however, that the Nelson-Siegel model does not dominate the other models. The exponential-spline model, for example, demonstrates the best forecasting performance using the mean-absolute error metric. Moreover, it also does relatively well using the root-mean-squared error. The Fourier-series model outperforms when we consider the overall fit to the yield curve at the one- and three-month forecasting horizons. When we look closer, however, we find that the Fourier-series approach has difficulty with short zero-coupon tenors and appears to generate more negative forecasts (i.e., overestimation) of zero-coupon rate forecasts. The Nelson-Siegel model, conversely, appears to be the most consistent of the empirical models, as evidenced by the relatively low volatility in the mean out-of-sample forecast errors. Specifically, it seems to handle the longer forecasting horizons better than the other two models.

Restricting our attention to the three variations on the $A_0(3)$ model, we can conclude rather comfortably that the unrestricted form of the $A_0(3)$ model exhibits the best forecasting behaviour. The unrestricted form dominates at all out-of-sample forecasting horizons and across almost all zero-coupon tenors. The diagonal and symmetric models appear to have difficulty in capturing the very short end of the zero-coupon curve in out-of-sample forecasts. Furthermore, these two models also generate relatively large negative (i.e., overestimation)

Table 10: **Out-of-Sample $A_0(3)$ Forecasts by Specific Tenor**: In this table, we present the root-mean-squared for three different versions of the $A_0(3)$ model. The difference, however, is that we focus on specific zero-coupon maturities, rather than the entire zero-coupon curve as presented in Table 9. All values are in basis points.

Tenor	RMSE				MAE			
	Random walk	Diagonal	Symmetric	Full	Random walk	Diagonal	Symmetric	Full
3 months	13.35	14.31	14.79	13.98	10.30	12.22	11.79	10.63
6 months	14.84	19.15	18.18	14.11	11.31	14.84	15.34	10.94
1 year	21.09	25.34	26.98	20.47	15.91	19.37	20.51	16.30
2 years	25.21	27.61	29.05	25.45	19.09	21.34	22.34	19.21
5 years	24.98	27.56	28.20	27.73	19.24	21.28	20.61	19.87
7.5 years	23.38	24.22	26.30	25.08	17.92	17.58	19.01	17.53
10 years	23.17	23.91	25.43	23.65	17.44	18.42	19.44	18.15
15 years	20.05	25.40	29.56	25.31	15.18	20.66	25.20	21.23
Six-month forecast								
3 months	47.36	90.80	74.50	46.04	41.49	79.91	63.49	38.17
6 months	44.98	94.03	92.54	43.84	37.38	82.36	75.53	36.04
1 year	43.13	95.54	114.26	45.45	35.49	83.10	90.77	38.41
2 years	44.21	88.15	121.41	52.69	38.05	76.11	94.95	43.74
5 years	44.71	68.77	99.76	59.10	36.37	60.29	77.16	47.13
7.5 years	41.11	53.26	81.01	50.46	34.51	46.18	63.05	39.73
10 years	38.93	52.81	74.79	49.84	33.19	45.40	62.75	41.66
15 years	34.94	52.32	71.47	52.77	29.17	46.47	61.23	44.78

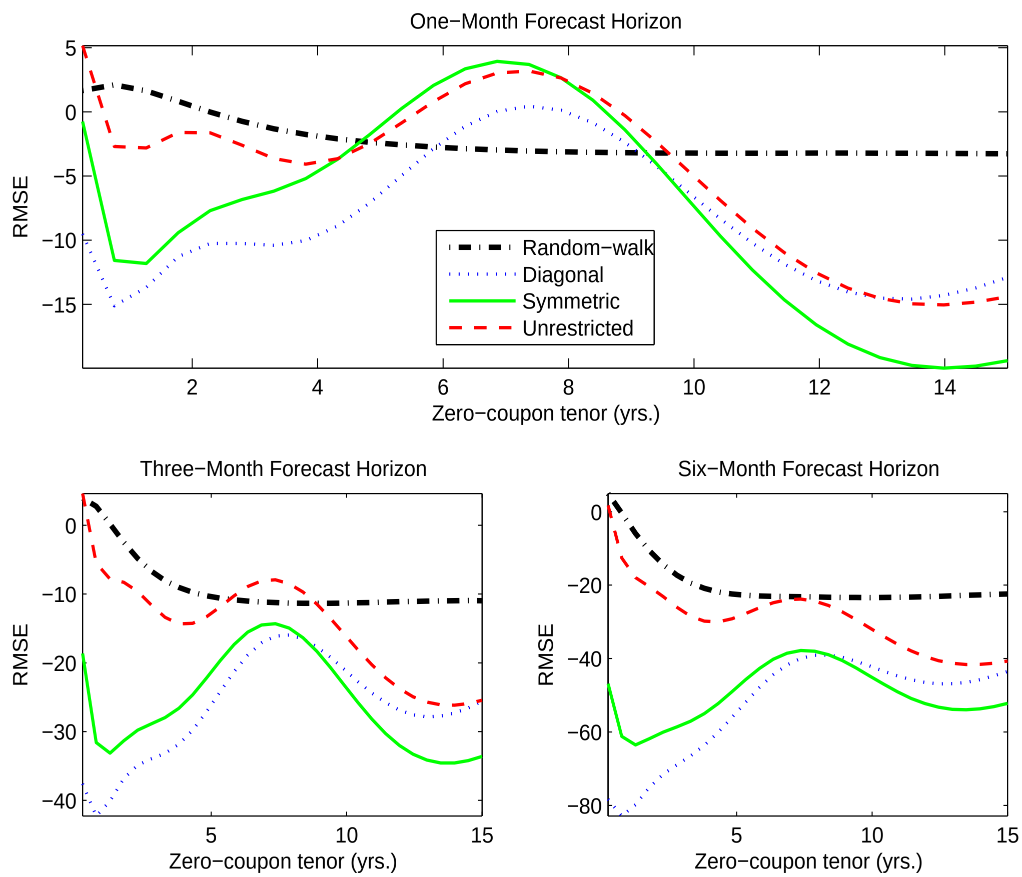
forecasts of future zero-coupon rates, as shown in Figure 6. This suggests that the extra parameters in the A_γ matrix found in the unrestricted model are quite important for forecasting future zero-coupon rates.

The Nelson-Siegel model, therefore, generally does the best job among both the empirical and theoretical models in the in- and out-of-sample exercise. We next examine an alternative out-of-sample forecasting exercise that examines how these models describe excess holding-period returns.

4.3 Forecasting holding-period returns

In the previous section, we focused on the ability of our six models to forecast future zero-coupon rates. This is similar to the analysis performed in both Duffee (2002) and Diebold and Li (2003). In this section, we change the focus somewhat by examining the ability of these models to forecast excess holding-period returns. A holding-period return is a rather simple quantity. It involves the return associated with purchasing an asset, holding it for some time interval, and then selling it. The *excess* holding-period return compares this return with what one would have earned by investing one's funds at the risk-free rate. We will focus on the excess holding-period returns associated with pure-discount bonds.

Figure 6: **Out-of-Sample $A_0(3)$ Forecasts by Tenor**: This figure illustrates the root-mean-squared out-of-sample forecast errors for the three theoretical $A_0(3)$ models across the one-, three-, and six-month forecasting horizons.



Let us make these ideas somewhat more precise. The n-period holding-period return, $h(t,T,n)$, is given as

$$h(t,T,n) = \frac{P(t+n,t+T) - P(t,T)}{P(t,T)}. \tag{121}$$

Simply put, assume that one buys a $(T-t)$-period pure-discount bond at time t. One holds this bond for n periods and then resells it at time $t+n$. The maturity of this pure-discount bond has become, over this time interval, $T-n$. Equation (121) represents the return on this investment strategy. We can also rewrite equation (121) in terms of zero-coupon rates as

$$h(t,T,n) = \frac{e^{-z(t+n,t+T)(T-n)} - e^{-z(t,T)(T-t)}}{e^{-z(t,T)(T-t)}}. \tag{122}$$

If one were to invest funds at the risk-free interest rate for n periods, one would purchase an n-period pure-discount bond and hold it to maturity. The return on this investment, which we denote as $g(t,n)$, is given as

$$g(t,n) = \frac{1 - P(t,t+n)}{P(t,t+n)}, \tag{123}$$
$$= \frac{1 - e^{-z(t,t+n)n}}{e^{-z(t,t+n)n}}.$$

The *excess* n-period holding-period return, therefore, is merely the n-period holding-period return less the risk-free investment; or, more specifically, it is equation (122) less equation (123). We denote it as $h_e(t,T,n)$, and it has the following form:

$$h_e(t,T,n) = \frac{e^{-z(t+n,t+T)(T-n)} - e^{-z(t,T)(T-t)}}{e^{-z(t,T)(T-t)}} - \frac{1 - e^{-z(t,t+n)n}}{e^{-z(t,t+n)n}}. \tag{124}$$

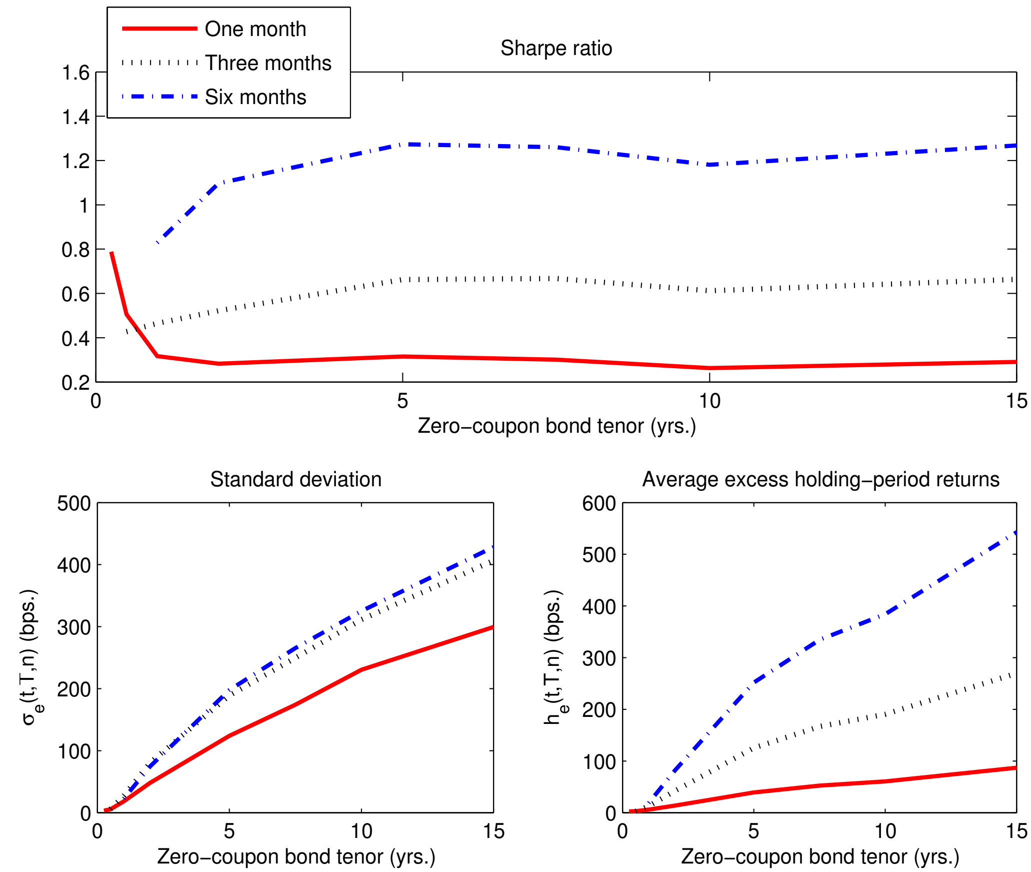
The plan is to use the zero-coupon rate forecasts from the previous section to construct excess holding-period return forecasts, since excess holding-period returns are intimately related to risk premia. Indeed, the expectations hypothesis—which essentially means that excess holding-period returns are constant across all maturities—holds in the absence of risk premia. The hope is that, by examining out-of-sample excess holding-period return forecasts, we can gain some additional insight into the relative performance of our models.

What do excess holding-period returns look like over the 42-month out-of-sample forecast horizon? Figure 7 illustrates the average excess-holding period returns for one-, three-, and six-month holding periods and a wide range of underlying pure-discount bond tenors.[25] The figure also includes the standard deviation of these excess holding-period returns and the associated Sharpe ratios. We note that excess holding-period returns are, in general, quite large and volatile. Moreover, the size and variability of excess holding-period returns increases as we extend the holding period and lengthen the tenor of the underlying pure-discount bond. This should not be surprising, since, by virtue of their lengthy duration, long-tenor pure-discount bonds are highly sensitive to interest rate movements.

Table 11 outlines the root-mean-squared and mean-absolute errors for the out-of-sample empirical-model forecasts of excess holding-period returns. These values are computed by using the zero-coupon rate forecasts in the previous sections and equation (124). It is immediately clear that all of the empirical models, including the random walk, have tremendous difficulty in forecasting excess holding-period returns. This is particularly evident at the longer pure-discount tenors. At tenors beyond five years, the root-mean-squared and mean-absolute errors exceed 100 basis points for the one-month forecast horizon. At the six-month forecast horizon, the errors are approximately twice as large, primarily because forecast errors are magnified by the strong interest rate sensitivity (i.e., long duration) of the long-tenor pure-discount bonds. This is combined with the fact, as evidenced in the previous section, that forecast errors tend to increase with both the forecast horizon and the pure-discount bond tenor.

[25] See Bolder, Johnson, and Metzler (2004) for a more detailed examination of excess holding-period returns on pure-discount bonds for the Canadian market. Note, however, that these values are broadly consistent with the values obtained in this study.

Figure 7: **Excess Holding-Period Returns**: This figure outlines the annualized average excess pure-discount bond holding-period returns (in basis points) for one-, three-, and six-month horizons. It also includes the standard deviations of these returns as well as their Sharpe ratios.



As with the previous forecasting exercise, the empirical models have difficulty in outperforming the random-walk model. The Nelson-Siegel model seems to do best in forecasting the shortest pure-discount tenor, the exponential-spline model does best at intermediate tenors (i.e., from about six months to ten years), and the Fourier-series model outperforms at the longest tenor. Figure 8 provides a bit more insight into the nature of the errors by outlining the time series of actual and forecast excess six-month holding-period returns (in basis points) for one-, five-, and ten-year pure-discount tenors. Observe that the actual excess holding-period returns are substantially more volatile than the model forecasts. We also see that the random-walk model exhibits an almost flat excess holding-period return profile. The three empirical models appear to do a slightly better job of capturing the variation in the excess holding-period returns; nevertheless, they fall quite short of the mark. We also observe that the forecast excess holding-period returns typically underestimate the actual excess holding-

Table 11: **Out-of-Sample Empirical Holding-Period Return Forecasts by Specific Tenor**: In this table, we present the annualized root-mean-squared holding-period forecast errors for our three empirical models, with the random-walk model included for comparison. We focus on specific zero-coupon maturities, rather than the entire zero-coupon curve. All values are in basis points.

Tenor	RMSE				MAE			
	Random walk	**Nelson-Siegel**	**Exponential spline**	**Fourier series**	**Random walk**	**Nelson-Siegel**	**Exponential spline**	**Fourier series**
One-month forecast								
3 months	2.20	2.77	3.43	5.58	1.68	2.16	3.00	3.90
6 months	5.86	5.77	5.45	6.85	4.56	4.64	4.48	5.16
1 year	18.57	18.61	18.65	20.12	14.07	14.62	14.40	15.35
2 years	48.27	50.42	51.07	53.90	36.58	38.41	37.89	38.19
5 years	123.56	131.11	124.81	130.96	95.18	102.38	96.94	100.26
7.5 years	174.17	176.71	176.19	175.26	133.86	130.18	125.13	131.72
10 years	230.22	232.21	225.73	227.98	173.79	182.94	175.60	173.63
15 years	301.53	321.83	369.65	309.33	228.82	259.63	311.05	235.65
Six-month forecast								
1 year	22.82	20.14	22.31	26.84	18.97	16.63	18.49	20.20
2 years	66.81	73.46	71.43	78.33	58.16	60.46	60.49	64.49
5 years	212.08	237.66	207.72	207.14	172.41	205.42	175.23	176.95
7.5 years	306.49	290.00	229.33	263.36	254.86	246.78	188.63	223.25
10 years	392.61	379.41	304.26	338.86	332.79	308.27	241.02	275.13
15 years	547.27	532.23	597.31	481.17	454.26	416.21	485.00	384.16

period returns. This is consistent with the general trend towards overestimating zero-coupon rates observed in the previous section.[26]

Table 12 provides the $A_0(3)$ forecast excess holding-period returns for a variety of zero-coupon tenors at the one- and six-month horizons. The performance of these models is also rather disappointing. With a few exceptions in the unrestricted model, none of the forecasts succeeds in outperforming the random-walk model. Indeed, as the forecast horizon increases, the model dramatically underperforms the random walk. As with the previous forecasting exercise, we observe that the unrestricted $A_0(3)$ model outperforms some of the empirical models at the shorter pure-discount tenors, but continues to underperform at longer tenors.

[26]To see this more clearly, consider the following representation of the difference in actual and forecast holding-period returns:

$$\underbrace{\left(\frac{P_1(z_1)-P_0(z_0)}{P_0(z_0)}\right)}_{\substack{\text{Actual}\\\text{return}}} - \underbrace{\left(\frac{\hat{P}_1(\hat{z}_1)-P_0(z_0)}{P_0(z_0)}\right)}_{\substack{\text{Forecast}\\\text{return}}} > 0, \tag{125}$$

$$\frac{P_1(z_1)-\hat{P}_1(\hat{z}_1)}{P_0(z_0)} > 0,$$

which implies that $P_1(z_1) > \hat{P}_1(\hat{z}_1)$ and, consequently, $z_1 < \hat{z}_1$, which is an overestimate of the zero-coupon rate.

Figure 8: **Empirical Excess Holding-Period Returns Time Series**: This figure outlines the time series of actual and forecast excess six-month holding-period returns (in basis points) for one-, five-, and ten-year pure-discount tenors.

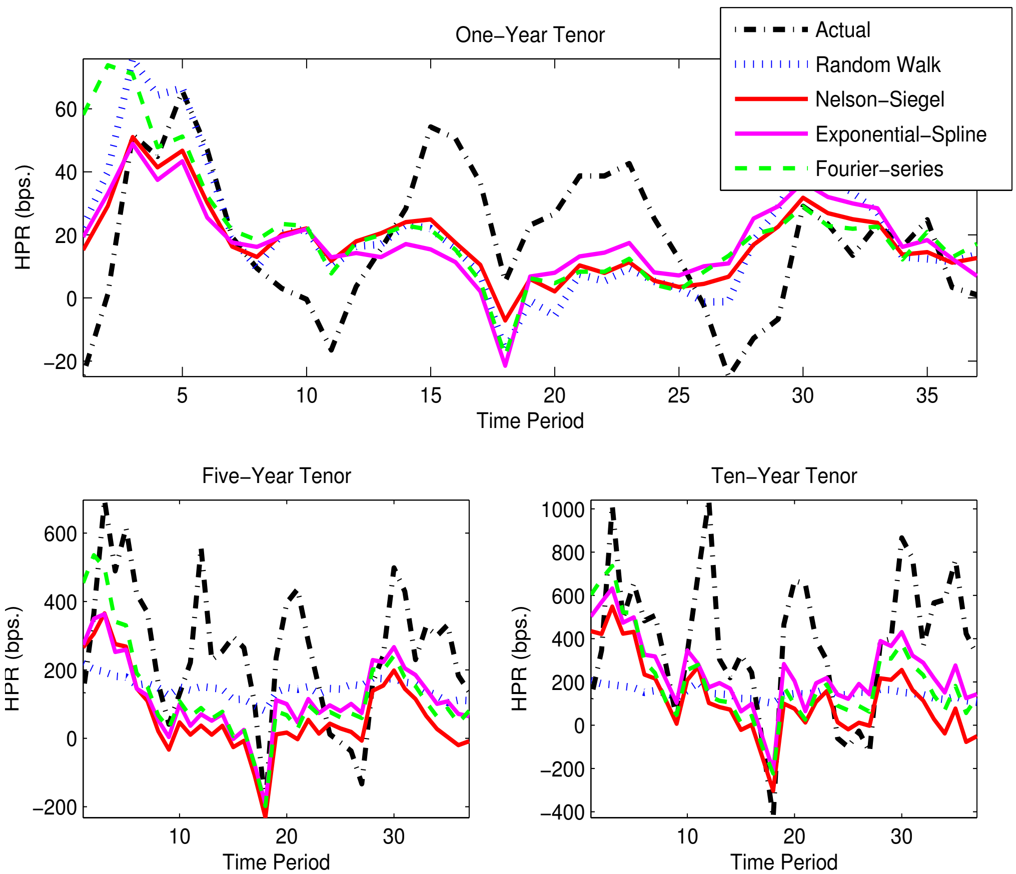


Figure 9 provides a more detailed view of the actual versus the forecast excess holding-period returns. Strikingly, the diagonal model exhibits even less volatility in its excess holding-period returns than the random-walk model. As the zero-coupon tenor increases, the diagonal model appears to suggest that the excess holding-period returns tend to zero. This is plainly contrasted by the actual data. The unrestricted model demonstrates somewhat more forecast variability, but it generates excess holding-period return forecasts that are negative correlated with the actual excess holding-period returns.[27] The underestimate of excess holding-period returns (i.e., overestimation of zero-coupon rates) is, if anything, even more pronounced among the $A_0(3)$ models.

[27]For the six-month forecasting horizon and the 10-year pure-discount bond tenor, the correlation between the actual excess holding-period returns and the unrestricted $A_0(3)$ model forecasts is -0.27. By contrast, this correlation is approximately 0.50 for the empirical models.

Table 12: **Out-of-Sample $A_0(3)$ Holding-Period Return Forecasts by Specific Tenor**: In this table, we present the root-mean-squared and mean-absolute holding-period forecast errors (in basis points) for the three variations on the $A_0(3)$ model, with the random-walk model included for comparison. We focus on specific zero-coupon maturities, rather than the entire zero-coupon curve. All values are in basis points.

Tenor	RMSE				MAE			
	Random walk	**Diagonal**	**Symmetric**	**Full**	**Random walk**	**Diagonal**	**Symmetric**	**Full**
One-month forecast								
3 months	2.20	2.59	3.15	2.74	1.68	2.20	2.65	2.08
6 months	5.86	7.25	6.68	5.54	4.56	5.69	5.87	4.22
1 year	18.57	22.79	24.05	18.09	14.07	17.41	18.33	14.41
2 years	48.27	52.84	55.81	48.37	36.58	40.90	42.91	36.68
5 years	123.56	137.00	139.56	137.41	95.18	105.92	102.22	98.81
7.5 years	174.17	180.53	196.10	186.98	133.86	131.32	142.35	131.25
10 years	230.22	237.25	251.82	234.79	173.79	182.68	192.04	180.00
15 years	301.53	381.88	442.49	380.78	228.82	310.58	376.46	318.62
Six-month forecast								
1 year	22.82	47.68	46.89	22.25	18.97	41.74	38.28	18.29
2 years	66.81	140.58	183.97	75.04	58.16	121.37	145.05	63.28
5 years	212.08	338.52	479.39	281.82	172.41	295.34	372.18	225.06
7.5 years	306.49	402.31	601.78	378.04	254.86	348.47	466.27	296.46
10 years	392.61	517.05	729.36	487.98	332.79	440.83	606.51	401.48
15 years	547.27	812.21	1074.39	809.04	454.26	718.16	923.82	684.51

4.4 Testing deviations from the expectations hypothesis

The so-called expectations hypothesis is the central theory of interest rates. To facilitate our discussion of this theory, let us introduce a bit of notation. We denote the zero-coupon interest rate for a claim at time t and maturing at time $t+n$ as

$$z(t, t+n) = z_t^n. \tag{126}$$

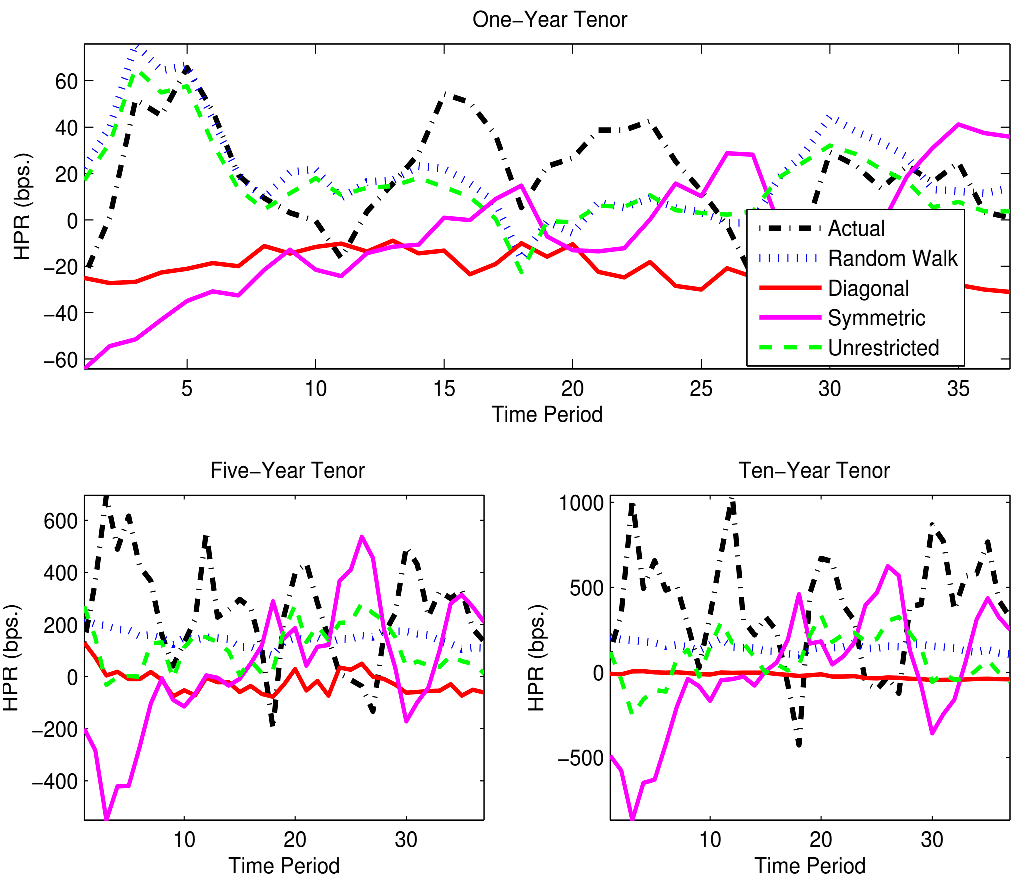
A forward interest rate beginning at time t, maturing at time $t+n$ for a tenor of τ, is denoted as

$$f(t, t+n, t+n+\tau) = f_{t,\tau}^n. \tag{127}$$

Note that, most of the time, we will be considering $\tau = 1$; that is, a one-year borrowing rate, n-periods forward. In this case, we will suppress the value of τ and denote the forward rate as f_t^n. Finally, we will denote the short rate as r_t and will use a short-term rate (i.e., the one-month rate) to approximate this value.

While there are a variety of flavours of the expectations hypothesis, in its most basic form it makes the

Figure 9: $A_0(3)$ **Excess Holding-Period Return Time Series**: This figure outlines the time series of actual and forecast excess six-month holding-period returns (in basis points) for one-, five-, and ten-year pure-discount tenors.



following statistical claim:

$$\underbrace{f(t, t+n, t+n+\tau)}_{f^n_{t,\tau}} = \mathbb{E}\left(\underbrace{z(t+n, t+n+\tau)}_{z^\tau_{t+n}} \mid \mathcal{F}_t\right). \tag{128}$$

What does equation (128) mean? It essentially states that, on average, the forward rate from time τ to T predicts the future spot rate over the same period. In other words, although we cannot know the future value of the zero-coupon rate from time τ to T (i.e., $z^{n+\tau}_t$), we can use the appropriate forward rate (i.e., $f^n_{t,\tau}$) as an unbiased predictor for this unknown rate. Were this to be true, an investor or borrower would be indifferent between different maturities. Essentially, the equation implies that borrowing for 10 years is equivalent to rolling over a one-year investment for 10 years.

An enormous amount of empirical work has been performed in the finance literature to test this theory. It turns out that the evidence overwhelmingly rejects the extreme version of the expectations hypothesis described in equation (128). In particular, the data suggest that there is a bias in the ability of the forward rate to predict future zero-coupon interest rates. This has led to a revision of the expectations hypothesis. The second version modifies equation (128) to make the following revised statistical claim:

$$f_{t,\tau}^{n} = \mathbb{E}\left(z_{t+n}^{\tau} \mid \mathcal{F}_t\right) + \xi^n. \tag{129}$$

In this case, the same interpretation applies, except for the presence of ξ, which is interpreted as a risk premium. The theory holds that fixed-income market participants demand a premium over and beyond the forward rate, to compensate for the various risks associated with holding fixed-income securities. The specific nature of these risks is not known with certainty, but there is a reasonable consensus that one of the primary risks relates to the risk of unexpected inflation eroding the real value of an investor's fixed-income claim.

Again, we can ask the question "does the empirical evidence support the theory summarized in equation (129)?" The answer is yes, but there is a twist. The financial literature finds overwhelming evidence of risk premia, but simultaneously rejects the hypothesis that ξ^n is a constant. Instead, empirical evidence strongly suggests the existence of a time-varying risk premium. This implies a new version of the expectations hypothesis to modify equation (129):

$$f_{t,\tau}^{n} = \mathbb{E}\left(z_{t+n}^{\tau} \mid \mathcal{F}_t\right) + \xi_t^n, \tag{130}$$

where ξ_t^n represents a time-varying risk premium.

There are a number of ways to econometrically test the expectations hypothesis. We will consider two alternatives. The first econometric test comes from Backus et al. (2001), who introduce both a useful test of the expectations hypothesis and indicate, quite cleverly, that deviations from the expectations hypothesis are, in fact, a statistical property of the term structure of interest rates. Dai and Singleton (2002) examine risk-premium-adjusted yield regressions. Our second econometric test will be the so-called LPY regression, although we do not adjust for risk premia. The important point is that any model purporting to describe the time-series dynamics of the term structure of interest rates should be able to replicate the observed deviations from the expectations hypothesis.

Let us begin with a description of the forward-rate regression suggested by Backus et al. (2001). We saw, in equation (128), a common form of the expectations hypothesis. Let us apply conditional expectations to both sides of equation (128), where we condition on the filtration, $\mathcal{F}_{t-1} \subset \mathcal{F}_t$. We can then proceed, as follows, to use the law of iterated expectations:

$$\begin{aligned}
\mathbb{E}(f(t,t+n,t+n+\tau) \mid \mathcal{F}_{t-1}) &= \mathbb{E}\left(\mathbb{E}\left(z(t+n,t+n+\tau) \mid \mathcal{F}_t\right) \mid \mathcal{F}_{t-1}\right), \qquad (131)\\
\mathbb{E}(f^n_{t,\tau} \mid \mathcal{F}_{t-1}) &= \underbrace{\mathbb{E}\left(z(t+n,t+n+\tau) \mid \mathcal{F}_{t-1}\right)}_{\text{By iterated expectations}},\\
&= \underbrace{f(t-1,(t-1)+(n+1),(t-1)+(n+1)+1)}_{\text{By definition: see equation (128)}},\\
&= f^{n+1}_{t-1,\tau}.
\end{aligned}$$

Adjusting the time scale forward by one period, we have that

$$f^n_{t,\tau} = \mathbb{E}(f^{n-1}_{t+1,\tau} \mid \mathcal{F}_t), \qquad (132)$$

which is equivalent to saying that f^n_t is a martingale.[28] As a consequence of this property, Backus et al. (2001) suggest the following *forward-rate regression* as a test of the expectations hypothesis:

$$f^{n-1}_{t+1} - r_t = \alpha_n + \beta_n(f^n_t - r_t) + \epsilon^n_t, \qquad (134)$$

for a wide range of forward-rate maturities, n. Were the expectations hypothesis to hold, we would expect to see an estimate of β_n approximately equal to unity. In actual fact, as we will see, this does not hold in the data.

The LPY regression, which is a mnemonic for linear projection coefficients in yield-based regressions, has the following form:

$$\begin{aligned}
z(t+1,t+n) - z(t,t+n) &= \gamma_n + \delta_n\left(\frac{z(t,t+n)-r_t}{n-1}\right) + \zeta^n_t, \qquad (135)\\
z^{n-1}_{t+1} - z^n_t &= \gamma_n + \delta_n\left(\frac{z^n_t - r_t}{n-1}\right) + \zeta^n_t.
\end{aligned}$$

The dependent variable is the difference between the $(n-1)$-period zero-coupon rate at time $t+1$ and the n-period zero-coupon rate at time t. One can think of this difference as related to the excess holding-period return from purchasing an n-period pure-discount bond at time t and selling it one period later when it becomes an $(n-1)$-period pure-discount bond. The independent variable is the excess yield, at time t, on the n-period zero-coupon rate relative to the risk-free rate, r_t, adjusted for n. This excess yield is what, under the expectations hypothesis, you would expect to earn on an n-period zero-coupon bond over the period. The idea is that the regression coefficient, δ_n, should be unity for all n. If this were to be true, it would have two implications. First, it would imply that, on average, the excess holding-period return on holding an n-period pure-discount bond for

[28] Recall that the martingale property for an arbitrary stochastic process, $\{X_t, t \geq 0\}$, holds that,

$$X_t = \mathbb{E}(X_s \mid \mathcal{F}_t), \qquad (133)$$

for all $s > t$.

one period should be approximately equal to the excess over the risk-free rate offered by the pure-discount bond. Second, it implies that this relationship will be the same across all pure-discount bond tenors. Again, we will see that, econometrically, this does not hold.

The goal of this section is to examine how simulated zero-coupon term-structure outcomes from the empirical and theoretical models capture this phenomenon. The basic approach to this simulation is quite simple. Since our dataset contains 187 months of data, we wish to compute the slope coefficients to the same number of simulated months of data. Our only concern is that, given the relatively short simulation horizon, there will be a strong dependence of the starting values of the state variables. We solve this problem by *burning-in* our simulations. That is, we start all of the state variables from their unconditional means and then simulate $M + 187$ months of data from each of the empirical and theoretical models; we then discard the first M months of simulated data and use the final 187 months for the estimation of the forward-rate regression in equation (134). For practical purposes, we set M=500 or about 40 years. This process is performed 2,500 times for each of the models. We then consider the distribution of regression coefficients.

The top graph in Figure 10 describes the actual forward-rate regression slope coefficients when we use equation (134) with our Canadian zero-coupon data. We consider a variety of forward-rate tenors out to 10 years (i.e., $n = 1, .., 120$). Again, were the expectations hypothesis to hold, we would expect to observe a straight line at unity for all forward-rate tenors. Instead, we note substantial deviations for short forward-rate tenors, with a gradual tendency towards unity as the forward-rate tenor increases; nevertheless, the slope coefficients are still less than one at the 10-year forward-rate tenor. Even when we take into account the standard errors—computed using the Newey-West algorithm with six lags—the probability of a slope coefficient equal to one at the 10-year tenor seems low.[29] These results are broadly consistent with the results of Backus et al. (2001) and Leippold and Wu (2001) using both American and European data.

The results of the simulation exercise for the forward-rate regression are presented in Table 13. The average forward rate coefficients and the average Newey-West errors are provided for each of the six empirical and theoretical models, as well as for the actual data. If we look down the column where the forward-rate tenor is two months (i.e., $n = 2$), we observe that the actual regression coefficient is 0.538. This is closely matched only by the average Nelson-Siegel model slope coefficient of 0.581. None of the $A_0(3)$ models demonstrates an average regression coefficient of less than approximately 0.95, while the exponential-spline and Fourier-series models are 0.782 and 0.885, respectively. Clearly, only the Nelson-Siegel model appears to be capable of capturing the observed deviations from the expectations hypothesis at the two-month forward-rate tenor.

If, however, we look down the column where the forward-rate tenor is 60 months (i.e., $n = 60$ or five years), we can see that the actual regression coefficient is 0.914. All of the models, both empirical and theoretical, are quite close. The smallest average regression coefficient is for the exponential-spline model, at 0.856, while the

[29]See Newey and West (1986) and Hamilton (1994) for detailed discussions of the Newey-West approach to the computation of standard errors.

largest is for the diagonal version of the $A_0(3)$ model, at 0.987.

Table 13: **Simulated forward-rate regression coefficients**: In this table, we present the average forward-rate regression coefficients estimated from 2,500 simulations for each of the six models. The values in parentheses are Newey-West errors computed with six lags.

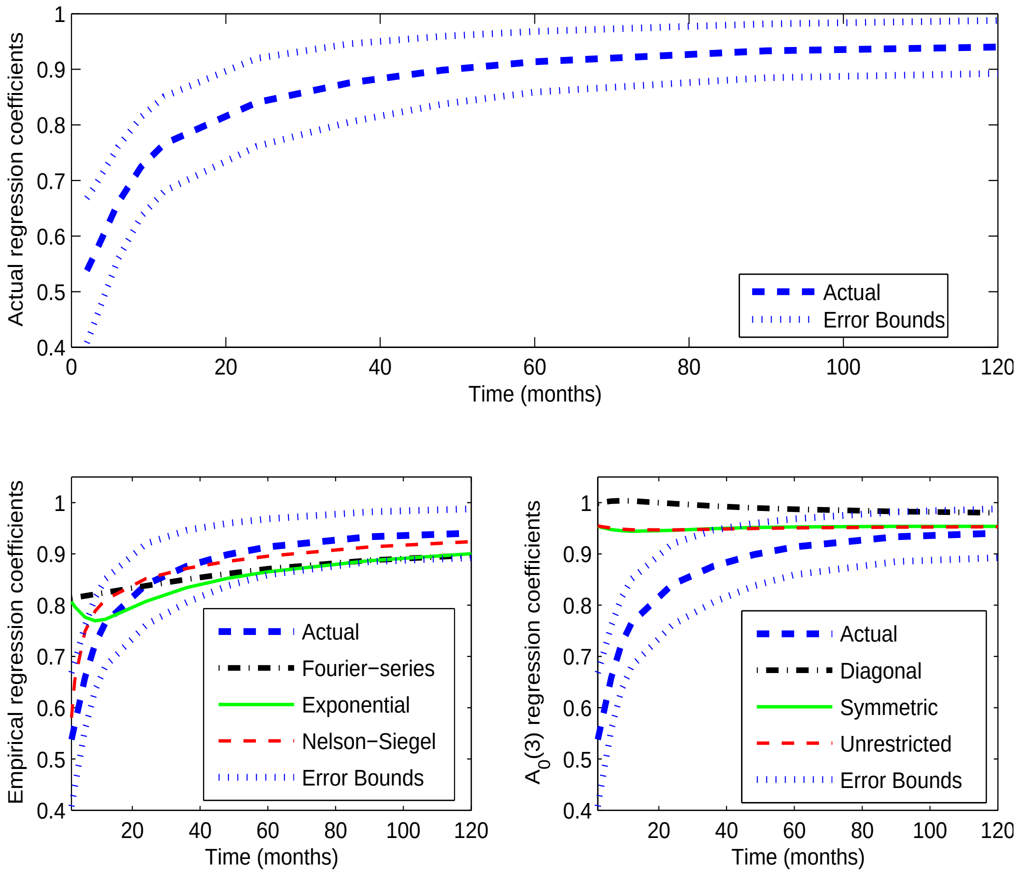
Models	**Forward-rate tenor in months: n**					
	2	**3**	**6**	**12**	**24**	**60**
No model						
Actual data	0.538	0.567	0.659	0.766	0.841	0.914
	(0.079)	(0.074)	(0.061)	(0.052)	(0.048)	(0.033)
Empirical models						
Nelson-Siegel	0.581	0.654	0.749	0.811	0.852	0.897
	(0.024)	(0.027)	(0.030)	(0.032)	(0.032)	(0.029)
Exponential spline	0.782	0.780	0.778	0.782	0.804	0.856
	(0.043)	(0.043)	(0.043)	(0.042)	(0.040)	(0.035)
Fourier series	0.815	0.816	0.819	0.825	0.838	0.871
	(0.039)	(0.039)	(0.039)	(0.038)	(0.036)	(0.033)
Theoretical models						
Diagonal A_γ	0.999	1.000	1.003	1.003	0.998	0.987
	(0.022)	(0.021)	(0.019)	(0.017)	(0.016)	(0.017)
Symmetric A_γ	0.954	0.952	0.947	0.944	0.946	0.953
	(0.028)	(0.028)	(0.028)	(0.026)	(0.024)	(0.022)
Unrestricted A_γ	0.955	0.953	0.950	0.947	0.947	0.951
	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.020)

The bottom two graphs in Figure 10 show the average regression coefficients for each of the empirical and theoretical models superimposed on the actual regression coefficients with the Newey-West standard error bounds. This provides a very clear view of the inability of the exponential-spline and Fourier-series models to capture deviations from the expectations hypothesis at the short end of the yield curve. Indeed, Figure 10 underscores that only the Nelson-Siegel model seems to capture this empirical relationship.

Figure 10 also indicates that none of the $A_0(3)$ models seems to be capable of capturing the observed deviations from the expectations hypothesis as described by the forward-rate regression. In fact, all of the models demonstrate the same general behaviour. The average regression coefficients are greater than 0.90 for the shorter forward-rate tenors, and they very gradually decline with the forward-rate tenor. The problem, however, is primarily restricted to the short end of the yield curve. Beyond about 30 months, or about 2.5 years, the average forward-regression coefficients do fall within the Newey-West standard error bounds. The notable exception is the diagonal model, which seems to stay at the upper range of the standard error bound on the actual regression coefficients.

The top graph in Figure 11 shows the actual LPY regression coefficients, along with the standard error

Figure 10: **Forward-rate regression coefficients**: This figure describes the forward-rate regression coefficients as described by Backus et al. (2001) in equation (134) for forward-rate tenors, $n = 2, ...120$ months. The error bounds are computed using Newey-West standard errors with six lags.



bounds, from the actual data. Indeed, the majority of the regression coefficients are negative. Furthermore, for all holding-period tenors, the point estimates of the regression coefficients are rather far from unity. Also note that the standard errors in this regression are larger than the forward-rate regression. The result is that unity lies with the standard error bounds for holding-period tenors between approximately one to five years. These results are also broadly consistent with the results of Dai and Singleton (2002) and Leippold and Wu (2001) using both American and European data.

The results of the simulation exercise for the LPY regressions are presented in Table 14. As before, the LPY regression coefficients and the average Newey-West errors are provided for each of the six empirical and theoretical models as well as for the actual data. Looking down the column, where the tenor of the underlying zero-coupon bond is two months (i.e., $n = 2$), we observe that the actual regression coefficient is -0.286 with

Table 14: **Simulated LPY regression coefficients**: In this table, we present the average LPY regression coefficients estimated from 2,500 simulations for each of the six models. The values in parentheses are Newey-West errors computed with six lags.

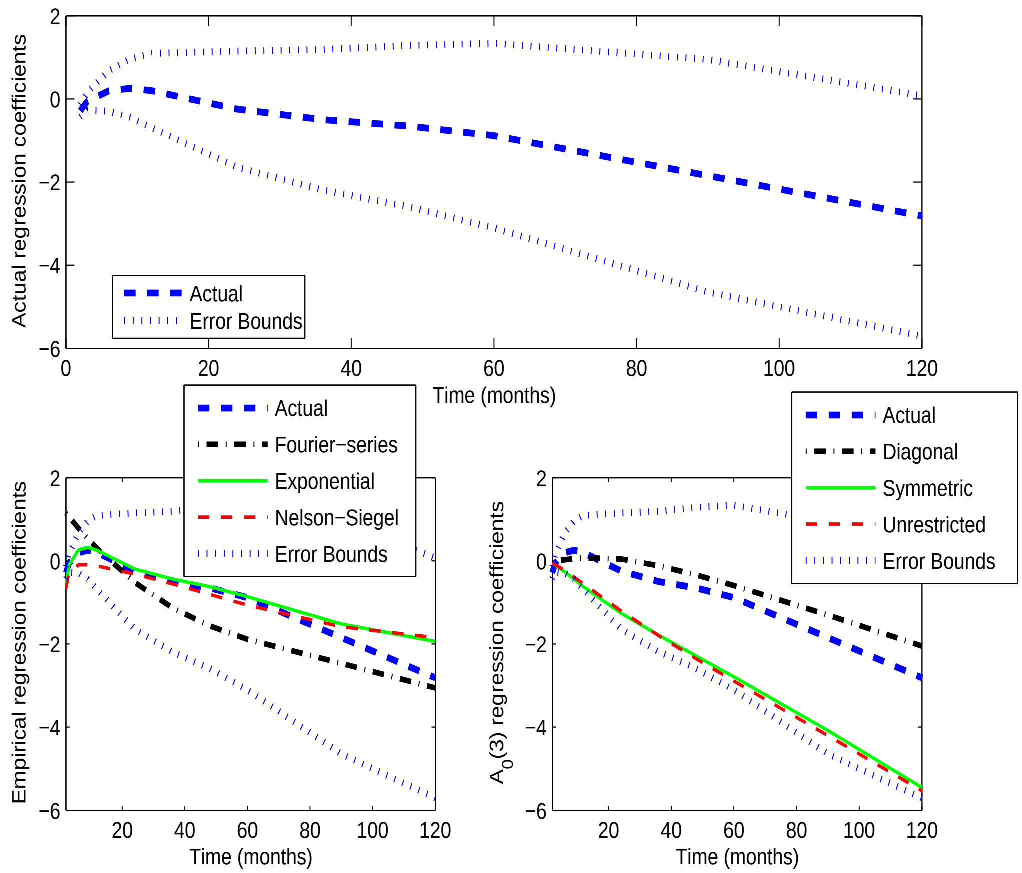
Models	**Forward-rate tenor in months:** n					
	2	**3**	**6**	**12**	**24**	**60**
			No model			
Actual data	-0.286	-0.056	0.192	0.200	-0.246	-0.883
	(0.089)	(0.132)	(0.295)	(0.543)	(0.849)	(1.351)
			Empirical models			
Nelson-Siegel	-0.668	-0.281	-0.096	-0.122	-0.316	-1.062
	(0.066)	(0.059)	(0.108)	(0.214)	(0.387)	(0.684)
Exponential spline	-0.570	-0.097	0.267	0.254	-0.195	-0.859
	(0.069)	(0.082)	(0.179)	(0.359)	(0.624)	(0.995)
Fourier series	1.148	1.055	0.786	0.293	-0.520	-1.889
	(0.794)	(0.794)	(0.800)	(0.841)	(0.990)	(1.478)
			Theoretical models			
Diagonal A_γ	-0.017	-0.007	0.024	0.076	0.042	-0.595
	(0.016)	(0.031)	(0.064)	(0.125)	(0.305)	(0.946)
Symmetric A_γ	-0.061	-0.104	-0.264	-0.624	-1.261	-2.786
	(0.029)	(0.058)	(0.143)	(0.299)	(0.571)	(1.311)
Unrestricted A_γ	-0.057	-0.094	-0.232	-0.553	-1.205	-2.890
	(0.016)	(0.033)	(0.084)	(0.192)	(0.425)	(1.153)

a standard error of 0.089. None of the point estimates of this regression coefficient for the other models falls within a 95 per cent confidence interval. The Fourier-series model is particularly far from the actual regression coefficient. As we extend the tenor, however, all of the models tend, fairly quickly, to the actual regression coefficient estimates.

The bottom two graphs in Figure 11 illustrate the data provided in Table 14. Among the empirical models, we can see that the exponential-spline and Nelson-Siegel models closely track the actual regression coefficients. Interestingly, the diagonal $A_0(3)$ model appears to provide the best fit to the actual LPY regression coefficients; the unrestricted and symmetric models generate substantially more negative regression coefficients. These point estimates do, however, lie within the lower boundary of the confidence interval for the actual regression coefficients.

The results of this section are not entirely conclusive. We can, however, carefully draw a few tentative conclusions. First, it appears that the empirical models, most notably the Nelson-Siegel model, capture the deviations from the expectations hypothesis fairly well across both econometric tests. The $A_0(3)$ models, conversely, seem to have trouble with shorter tenors in the forward-rate regression, but do fairly well in the LPY regressions. It is tempting to conclude that the superior forecasting performance of the empirical models stems from their ability to capture deviations from the expectations hypothesis. This may indeed be so, but the evidence does

Figure 11: **LPY regression coefficients**: This figure describes the LPY regression coefficients as described by Backus et al. (2001) in equation (134) for forward-rate tenors, $n = 2, ...120$ months. The error bounds are computed using Newey-West standard errors with six lags.



not appear to be sufficiently compelling to make a strong case. It is, nevertheless, suggestive that a model's forecasting ability is related to its capacity to describe deviations from the expectations hypothesis. Moreover, it is consistent with the view that forecasting future zero-coupon rate outcomes is intimately related to the description of risk premia and the associated excess returns inherent in the yield curve.

4.5 A simple portfolio exercise

We consider one final dimension for the comparison of these models. Our interest in these models, as previously noted, stems from our desire to incorporate them into a stochastic simulation model for taking debt-management portfolio decisions. It would seem logical, therefore, to consider how these models perform in the context of a portfolio-optimization exercise. The actual debt-strategy problem is, however, rather too involved for a clean comparison of our different models.[30] Instead, we will consider a portfolio of zero-coupon bonds in the context of a simplifed mean-variance setting.

The actual portfolio selection will depend upon the expected excess holding-period returns and their associated variance. Computing the expected excess holding-period returns comes directly from the model, while the determination of the variance requires a bit more effort. Recall from equation (124) that the *excess* n-period holding-period return, $h_e(t,T,n)$, has the following form:

$$\begin{aligned} h_e(t,T,n) &= \frac{e^{-z(t+n,t+T)(T-n)} - e^{-z(t,T)(T-t)}}{e^{-z(t,T)(T-t)}} - \frac{1-e^{-z(t,t+n)n}}{e^{-z(t,t+n)n}}, \\ &= \frac{P(t+n,t+T)-P(t,t+T)}{P(t,t+T)} - \frac{1-P(t,t+n)}{P(t,t+n)}. \end{aligned} \tag{136}$$

The expected n-period excess holding-period returns for a zero-coupon bond with a tenor of T years are

$$\mathbb{E}\left(h_e(t,T,n)|\,\mathcal{F}_t\right) = \frac{\mathbb{E}\left(P(t+n,t+T)|\,\mathcal{F}_t\right) - P(t,T)}{P(t,T)} - \frac{1-P(t,t+n)}{P(t,t+n)}, \tag{137}$$

where,

$$\mathbb{E}\left(P(t+n,t+T))|\,\mathcal{F}_t\right) = \exp\left(-\mathbb{E}\left(z(t+n,t+T)|\,\mathcal{F}_t\right)(T-n)\right), \tag{138}$$

and these expected future zero-coupon rates are formulated using our six different term-structure models, as described in equations (118–120). We can, for a given model, define the vector of expected n-period excess holding-period returns at time t_q for our portfolio of zero-coupon bonds as

$$\mu\left(t_q,\vec{T},n\right) = \begin{bmatrix} \mathbb{E}\left(h_e(t_q,T_1,n)|\,\mathcal{F}_{t_q}\right) \\ \vdots \\ \mathbb{E}\left(h_e(t_q,T_N,n)|\,\mathcal{F}_{t_q}\right) \end{bmatrix}, \tag{139}$$

for zero-coupon tenors, $\vec{T} = \begin{bmatrix} T_1 & \cdots & T_N \end{bmatrix}^T$. We compute $\mu\left(t_q,\vec{T},n\right)$ for each model at each point t_q in our out-of-sample period using the parameters obtained by estimating the model with data from the interval $[t_1,t_q]$.

Constructing the associated variance-covariance matrix for our n-period excess holding-period returns requires some numerical computation. In particular, we bootstrap from the set of historical n-period excess holding-period return forecasts. The j historical n-period excess holding-period return forecast error has the form

$$\xi_j = \mathbb{E}\left(z(t_j+n,t_j+T_i)|\,\mathcal{F}_{t_j}\right) - z(t_j+n,t_j+T_i), \tag{140}$$

for $j=1,...,q$ and $i=1,...,N$. Using these forecast errors, we proceed to construct a collection of randomly selected holding-period returns for each of the various tenors. The following expression describes the holding-

[30]See Bolder (2002, 2003) for comprehensive discussions of the debt-management problem.

period return including a randomly selected *shock* from the historical forecast errors described in equation (140):

$$h_e\left(t_q, T_i, n, \tilde{\xi}_k\right) = \frac{\exp\left(-\left(\mathbb{E}\left(z(t_q+n, t_q+T_i)\middle|\mathcal{F}_{t_q}\right) + \tilde{\xi}_k\right)(T_i - n)\right) - P(t_q, T_i)}{P(t_q, T_i)} - \frac{1 - P(t_q, t_q+n)}{P(t_q, t_q+n)}, \quad (141)$$

where $\tilde{\xi}_k$ denotes a uniform randomly selected forecast error from the set $\{\xi_1, ..., \xi_q\}$. This is repeated many times with different randomly selected error terms and pure-discount bond tenors. We can think of this as artificially generating a dataset from the historical forecast errors; as a consequence, it will, in the limit, have the same distributional characteristics as the true dataset. It also permits us to construct the following matrix of holding-period returns:

$$H\left(t_q, \vec{T}, n\right) = \begin{bmatrix} h_e(t_q, T_1, n, \tilde{\xi}_1) & \cdots & h_e(t_q, T_N, n, \tilde{\xi}_1) \\ \vdots & \ddots & \vdots \\ h_e(t_q, T_1, n, \tilde{\xi}_M) & \cdots & h_e(t_q, T_N, n, \tilde{\xi}_M) \end{bmatrix}, \quad (142)$$

where M is the number of randomly selected historical forecast errors used to construct the artificial dataset. This permits us to write the variance-covariance matrix of the n-period excess holding-period returns as

$$\begin{aligned} \Omega\left(t_q, \vec{T}, n\right) &= \text{var}\left(P(t_q, \vec{T}, n)\middle|\mathcal{F}_{t_q}\right), \\ &= \text{cov}\left(H\left(t_q, \vec{T}, n\right)\right). \end{aligned} \quad (143)$$

With the expected n-period excess holding-period returns and their associated variance-covariance matrix in hand, we can proceed to perform the portfolio optimization exercise. Let $\omega \in \mathbb{R}^{N\times 1}$ be a vector of portfolio weights on our N pure-discount bonds. We consider two alternative mean-variance portfolios. First, we compute the minimum-variance portfolio by solving the following optimization problem:

$$\min \ \omega^T \Omega\left(t_q, \vec{T}, n\right)\omega, \quad (144)$$

subject to:

$$0 \le \omega_i \le 1, \text{ for } i = 1, ..., N$$
$$\sum_{i=1}^{N} \omega_i = 1.$$

The second variation on the mean-variance portfolio is to determine the maximum Sharpe ratio. The associated optimization problem has the following form:

$$\max \frac{\mu\left(t_q, \vec{T}, n\right)^T \omega}{\sqrt{\omega^T \Omega\left(t_q, \vec{T}, n\right)\omega}}, \tag{145}$$

subject to:

$$0 \leq \omega_i \leq 1, \text{ for } i = 1, ..., N$$
$$\sum_{i=1}^{N} \omega_i = 1.$$

To be specific, we look at one-year excess holding-period returns and with a portfolio of zero-coupon bonds with tenors of $T = 13, 18, 24, 48, 60,$ and 120 months.[31] We perform this exercise in an out-of-sample manner, similar to the previous analysis, with 31 overlapping one-year periods. That is, at each period in the out-of-sample window, the model parameters are re-estimated, the expected one-year excess holding-period returns are computed, and the variance-covariance matrix is approximated via the bootstrap technique. Finally, the two optimization algorithms are run and the portfolio weights are estimated. The remainder of this section will compare the expected results with the actual results of using the portfolio weights to invest in the optimal portfolios.

Table 15: **Minimum-Variance Portfolio Exercise Results**: In this table, we present the summary statistics of the minimum-variance portfolio optimization performed using the three empirical models and the random-walk hypothesis. The expected and actual mean, standard deviation, maximum, and minimum excess holding periods (in basis points) are provided for a 12-month rolling optimization.

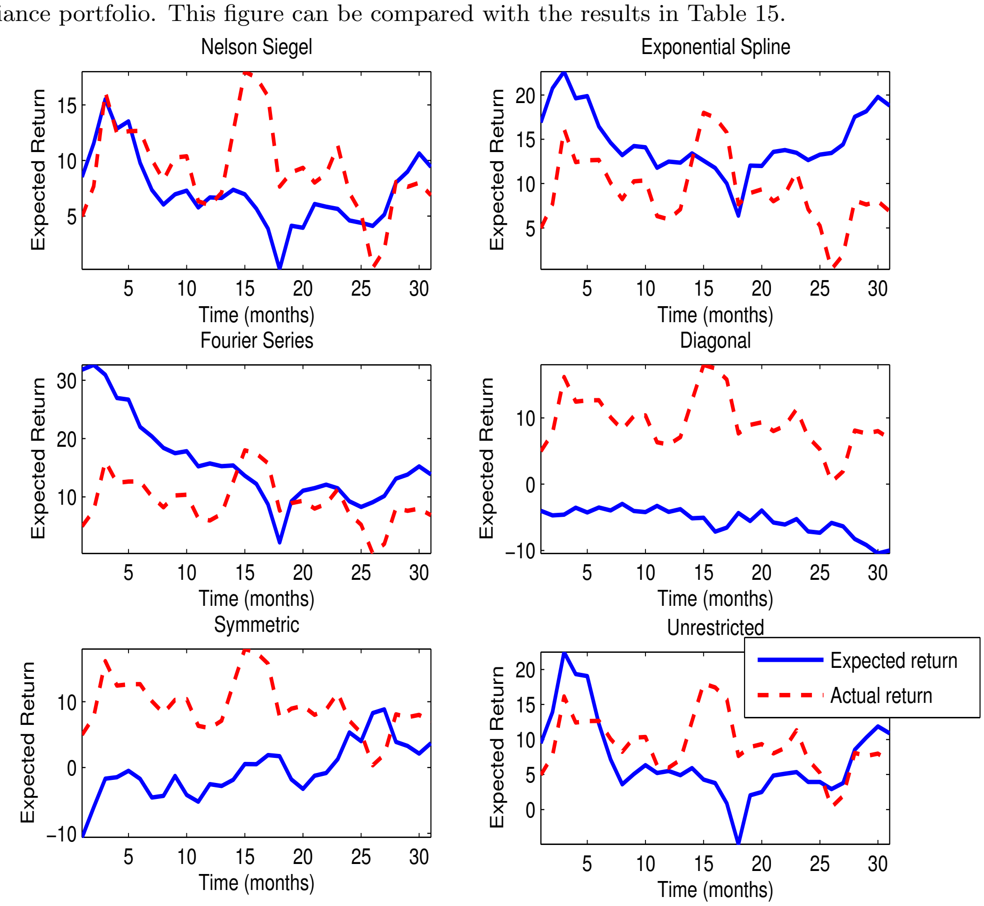
Models	**Expected**				**Actual**			
	Mean	σ	**Max**	**Min**	**Mean**	σ	**Max**	**Min**
Random walk								
Random walk	8.60	4.73	18.26	0.71	9.29	4.08	18.00	0.34
Empirical models								
Nelson-Siegel	7.21	3.21	15.51	0.22	9.29	4.08	18.00	0.34
Exponential spline	14.71	3.54	22.65	6.38	9.29	4.08	18.00	0.34
Fourier series	15.87	7.41	32.65	2.20	9.29	4.08	18.00	0.34
Theoretical models								
Diagonal A_γ	-5.49	1.97	-2.97	-10.42	9.29	4.08	18.00	0.34
Symmetric A_γ	-0.34	4.14	8.81	-10.59	9.29	4.08	18.00	0.34
Unrestricted A_γ	7.10	5.77	22.49	-4.96	9.29	4.08	18.00	0.34

Table 15 outlines the expected and actual results associated with the minimum-variance portfolio weights.

[31] Note that we cannot have $T_i \leq n$ for any $i = 1, ..., N$, since the bond will either mature during the holding period (i.e., $T_i < n$) or earn the risk-free rate (i.e., $T_i = n$).

The first four columns provide a number of summary statistics for the expected return associated with the optimal portfolio weights for the three empirical models, the three variations on the $A_0(3)$ affine model, and the random-walk model. The final four columns, conversely, demonstrate the same summary statistics for the actual results corresponding to an actual application of the ensuing optimal portfolio weights. The first point worth observing is that the actual results are the same for all portfolios; this is because the minimum-variance portfolio consists entirely of 13-month pure-discount bonds across all models. This should, however, be no surprise. Figure 7 clearly demonstrates that the standard deviation of the excess holding-period returns is an increasing function of the tenor of the underlying pure-discount bond. Thus, the optimization algorithm selects the shortest tenor pure-discount bond from among the potential portfolio candidates to minimize the portfolio's variance.[32]

Figure 12: **Minimum Variance Portfolios**: This figure summarizes the expected versus actual one-year excess holding-period returns of a portfolio of six zero-coupon bonds for each of our six term-structure models and the associated minimum-variance portfolio. This figure can be compared with the results in Table 15.



[32]Apparently, the covariance between the various elements of the portfolio is not sufficient to overcome this effect and lead to some level of diversification.

The second item of note is that the random-walk and empirical models do a relatively good job of forecasting the actual excess holding-period returns associated with the minimum-variance portfolio. The Nelson-Siegel and random-walk models appear to be the closest, while the exponential-spline and Fourier-series models tend to overestimate the expected excess holding-period return by around five basis points. A bit of reflection will reveal that this is essentially another out-of-sample forecasting exercise. At time t_q, one basically forecasts the value of a collection of zero-coupon bonds of the form $P(t_q + n, t_q + T_i)$ for $i = 1, ..., n$. In this case, we know that, since $n = 12$ and $T_i = 13$, one is essentially forecasting a single zero-coupon bond, $P(t_q + 12, t_q + 13)$. In other words, the key aspect to appropriately determining this optimal portfolio is to predict the one-month zero-coupon rate that will prevail in one year's time.

The diagonal and symmetric affine models, however, do not fare quite as well. Again, this is not much of a surprise, since the diagonal and symmetric $A_0(3)$ models—as previously demonstrated—have a substantial amount of difficulty in forecasting future excess holding-period returns as the holding period lengthens. Indeed, the expected portfolio returns for the diagonal and symmetric models are slightly negative. The exception is the unrestricted model that generally seems to forecast the actual portfolio returns quite well. These results are underscored in Figure 12, which compares the expected versus actual portfolio returns for each of the six term-structure models across the out-of-sample horizon. The diagonal and symmetric versions of the $A_0(3)$ model appear to fail quite dramatically at forecasting the actual portfolio returns of our pure-discount bond portfolio. The remaining models tend to have some difficulty over different regions, but generally follow the basic pattern of the actual portfolio returns.

Table 16: **Maximum Sharpe-Ratio Portfolio Exercise Results**: In this table, we present the summary statistics of the minimum-variance portfolio optimization performed using the three empirical models and the random-walk hypothesis. Again, the expected and actual mean, standard deviation, maximum, and minimum excess holding periods (in basis points) are provided for a 12-month rolling optimization.

Models	**Expected**					**Actual**				
	Mean	σ	**Sharpe**	**Max**	**Min**	**Mean**	σ	**Sharpe**	**Max**	**Min**
Random walk										
Random walk	85.16	78.07	0.5733	248.30	8.03	58.91	74.59	0.7607	231.23	-166.37
Empirical models										
Nelson-Siegel	59.74	114.94	0.6368	438.04	0.22	79.12	141.73	0.5368	496.12	1.96
Exponential	19.87	27.42	1.1652	116.10	6.38	14.90	28.65	0.2831	167.66	0.34
Fourier series	18.43	12.22	1.2196	66.84	2.20	12.96	13.58	1.2727	76.83	0.48
Theoretical models										
Diagonal A_γ	-0.06	36.84	0.0078	130.49	-47.50	645.77	400.70	1.6116	1509.44	-498.72
Symmetric A_γ	147.87	316.98	0.1880	1014.19	-477.05	299.71	225.57	1.3629	682.91	-12.10
Unrestricted A_γ	123.43	159.31	0.6557	453.44	4.27	105.28	153.26	0.6711	432.24	-198.92

We conclude, therefore, in this minimum-variance setting, that the three empirical models and the unrestricted version of the $A_0(3)$ model appear to be generally useful in formulating minimum-variance portfolios of pure-discount bonds. To be fair, this is not a terribly stringent test, given the uncomplicated nature of these portfolios.

Let us next examine the maximum Sharpe-ratio portfolios. In this case, differences among the portfolios recommended by the various term-structure models are evident. Reviewing these results will require us to examine Tables 16 and 17 simultaneously. For the random-walk model, we observe an average expected one-year excess holding-period return of about 85 basis points over the out-of-sample horizon. This compares with an actual portfolio return of around 58 basis points, and involves placing about half of the portfolio in the shortest tenor pure-discount bond, almost one-quarter of the portfolio in the two-year pure-discount bond, and the remainder spread among the remaining buckets (with the exception of the 10-year zero-coupon bond). This appears to be reasonable, given the fairly flat form of the Sharpe ratio for longer holding periods, as described in Figure 7; the Sharpe ratio does, however, ramp up somewhat for shorter tenors, and flattens out around two to five years. Moreover, we anticipate, and indeed observe, an increase in expected return associated with the maximum Sharpe-ratio portfolio relative to the minimum-variance portfolio.

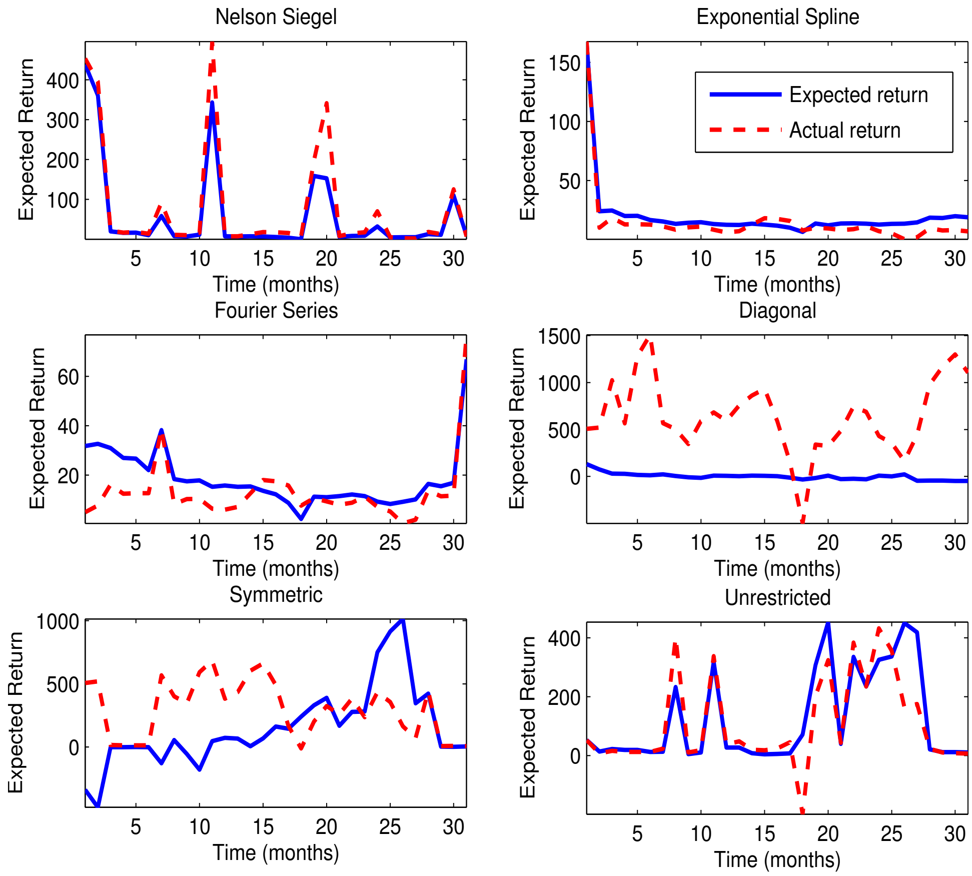
Table 17: **Maximum Sharpe-Ratio Portfolio Weights**: In this table, we present the average portfolio weights for our maximum Sharpe-ratio pure-discount portfolios across each of our term-structure models. These average portfolio weights are associated with the results in Table 16. All values are in basis points.

Tenor	**Random walk**	**Empirical models**			**Theoretical models**		
		Nelson-Siegel	**Exponential spline**	**Fourier series**	**Diagonal**	**Symmetric**	**Full**
13 months	0.51	0.79	0.99	0.97	0.00	0.25	0.62
1 1/2 years	0.07	0.06	0.00	0.02	0.00	0.00	0.00
Two years	0.23	0.03	0.00	0.01	0.00	0.05	0.00
Four years	0.10	0.00	0.00	0.00	0.00	0.03	0.06
Five years	0.09	0.05	0.01	0.00	0.26	0.58	0.31
Ten years	0.00	0.07	0.00	0.00	0.74	0.10	0.01

Among the empirical models, the Nelson-Siegel model appears to generate the most reasonable maximum Sharpe-ratio portfolios. It generates an average expected portfolio return of approximately 60 basis points, which compares favourably with the actual average return of about 80 basis points. The expected and actual realized Sharpe ratios are also actually quite similar. The exponential-spline and Fourier-series models, conversely, suggest maximum Sharpe-ratio portfolios that offer only about three to five basis points more than the minimum-variance portfolios. This is evidenced by Table 17, which indicates that the lion's share of the portfolio weights remain in the shortest tenor zero-coupon bond. One suspects that these models forecast generally quite flat Sharpe ratios as a function of all pure-discount bond tenors.

The unrestricted version of the $A_0(3)$ model again tends to outperform the two other affine alternatives.

Figure 13: **Maximum Sharpe-Ratio Portfolios**: This figure summarizes the expected versus actual one-year excess holding-period returns of a portfolio of six zero-coupon bonds for each of our six term-structure models and the associated maximum Sharpe-ratio portfolio. This figure can be compared with the results in Table 16.



The average portfolio weights for the unrestricted model involve about two-thirds in the shortest tenor and the remainder are concentrated in the four- to five-year sector. This yields an average expected portfolio return of 123 basis points compared with an actual portfolio return of 105 basis points. Moreover, Figure 13 illustrates that the expected portfolio returns track the actual returns quite closely. Indeed, only the unrestricted model has an expected Sharpe ratio (i.e., 0.6557) that is similar to the actual realized Sharpe ratio (i.e., 0.6711). Something rather odd, conversely, appears to occur with the diagonal implementation of the $A_0(3)$ model. It places all of the portfolio weight in the two longest tenor pure-discount bonds (i.e., five and ten years), to generate a zero expected excess holding-period return on the portfolio. The actual average returns on this portfolio, however, are in excess of 600 basis points, with a sizable variance. Clearly, the diagonal model has enormous difficulty in forecasting excess holding-period returns on longer-tenor pure-discount bonds. The symmetric model fares better than the diagonal model, although it still has an actual portfolio return of approximately twice the expected average portfolio return.

Figure 13 tells a very similar story to that told by Figure 12: the three empirical models and the unrestricted version of the $A_0(3)$ model tend to produce expected portfolio returns, rather than generally track the actual portfolio returns. The diagonal and symmetric implementations of the $A_0(3)$, conversely, do not closely track actual returns. We conclude, therefore, that there is substantial evidence that four of the six models are potentially useful in a portfolio exercise. Whether we can generalize these results from this simplified setting to the rather more complex debt-management setting remains, of course, an open question. This analysis, however, is nonetheless compelling.

5 Conclusion

In this paper, we examine six term-structure models that fall into two different classes. Given the size of the literature in this field, this is a very small subset of the available collection of models. With our focus on risk-management and the corresponding need to work under the physical measure, we think this is a reasonably good start. Moreover, we have restricted ourselves to models that, though not necessarily easy to derive and estimate, are substantially less complex than many competing models.[33]

The class of empirical models builds on the work of Diebold and Li (2003)). We examine their dynamic extension of the Nelson-Siegel model and create a dynamic extension of the exponential-spline model suggested by Li et al. (2001), as well as the Fourier-series model proposed by Bolder (2002). Each of these models proposes substantially different mathematical structures for the factor loadings of the term structure. Indeed, the Nelson-Siegel, exponential-spline, and Fourier-series models propose Laguerre, orthogonalized-exponential, and trigonometric basis functions, respectively.

The class of theoretical models considered in this paper is also quite small. We consider, within the context of an affine $A_0(3)$ model, different restrictions on the A_γ matrix pre-multiplying the state-variable vector in the market price of risk. Specifically, we consider a diagonal, symmetric, and unrestricted form for the A_γ matrix. We consider these alternatives principally through a general desire to reduce the number of model parameters; to be fair, this is a very modest change in the structure of the model. The differences in the performance of these versions of the $A_0(3)$ model in the forecasting exercise, and the deviations from the expectations hypothesis are, therefore, quite surprising.

Having worked through the technical details of these models, we compare the models based on four principal dimensions. First, we consider their in- and out-of-sample ability to forecast future zero-coupon rates on a variety of different forecasting horizons. This involves examination of how well the models forecast the entire yield curve, as well as a number of specific zero-coupon tenors. Second, we examine the ability of these models to forecast out-

[33] The quadratic models of Leippold and Wu (2000, 2001) and the positive models of Flesaker and Hughston (1996) are good examples.

of-sample excess holding-period returns associated with holding pure-discount bonds of different tenors. Third, we use simulation to examine the ability of these different models to capture two alternative econometric tests of the expectations hypothesis: these include the forward-rate regressions recently proposed by Backus et al. (2001) and the ubiquitous LPY regression. The idea behind these tests is that a good model should, in general, be able to reproduce the deviations from the expectations hypothesis observed in the actual zero-coupon data. Finally, we consider a practical application of these models to a portfolio-optimization problem. Specifically, we use each of the six term-structure models to select optimal weights for a pure-discount bond portfolio in a mean-variance setting.

There are several principal observations that we can draw from this work:

- the empirical models demonstrate superior in- and out-of-sample forecasting ability relative to the theoretical models on virtually every possible measure of forecasting performance;
- it is difficult for even the empirical models, which forecast well relative to the theoretical models, to outperform a random-walk model;
- the underperformance of the theoretical models is particularly evident at longer zero-coupon tenors and longer forecasting horizons;
- among the empirical models, the Nelson-Siegel model has the most consistent in- and out-of-sample forecasting performance, although the exponential-spline model is a fairly close second;
- among the theoretical models, the unrestricted model rather dramatically outperforms the diagonal and symmetric formulations of the A_γ matrix;
- over the forecasting period, all of the models demonstrate a tendency to overestimate zero-coupon rates and correspondingly underestimate excess holding-period returns;
- of the six models, only the Nelson-Siegel is capable of reasonably describing the observed deviations from the expectations hypothesis across both econometric tests;
- those models that exhibit the best performance in forecasting out-of-sample excess holding-period returns (i.e., the three empirical models and the unrestricted $A_0(3)$ model) provide the most reasonable results in our simplified portfolio-optimization exercise.

Why do the theoretical models underperform the empirical approach? One possible explanation relates to the fact that, over the forecasting period, the term structure of zero-coupon interest rates was very steep. Consequently, the $A_0(3)$ models may possibly be forecasting that the term structure will revert back to its normal shape more quickly than actually observed in the data. Duffee (2002) discusses the negative correlation between yield forecast errors and the slope of the yield curve. That is, when the yield curve is quite steep, the forecast errors are quite high; this would suggest that they are missing the generally high-risk premia evident in a steep yield-curve environment. This does not, however, appear to be the situation in this case. We can say this because we have somewhat more information. We see that all of the models, although the $A_0(3)$ models are the worst offenders, generally overestimate future zero-coupon rates and correspondingly underestimate excess holding-period returns. This would suggest that these models actually overestimate risk premia.

Note also that model performance in the forecasting exercise and the expectations hypothesis test does *not* appear to be independent. Those models—specifically, the Nelson-Siegel and the exponential-spline model—that best capture deviations from the expectations hypothesis also exhibit the best performance in the in- and out-of-sample forecasting exercises, as well as the portfolio-optimization analysis. This supports the view that forecasting future zero-coupon rate outcomes is closely related to the description of risk premia and the associated excess returns inherent in the yield curve. It also suggests that the $A_0(3)$ models' difficulty in forecasting the short zero-coupon tenors stems from the relative difficulty of these models to describe deviations from the expectations hypothesis.

A few caveats are in order. First, as is always the case, our results are sensitive to the data we use. We use approximately 15 years of data and allocate the final three-and-a-half years for the out-of-sample forecasting exercise. To the extent that this period is not representative, our results will be flawed or misleading. But, given that we have a limited amount of useful data, we have done the best that we can.[34]

A second caveat is that estimation of the $A_0(3)$ models requires the numerical solution of a high-dimensional and strongly non-linear optimization problem. One can never be certain of having attained a global minimum. In an effort to solve this problem, we perform a substantial amount of computation. It is nonetheless possible that parameters for the $A_0(3)$ models do exist that outperform the empirical models. While this is an important caveat, it does return us to the practitioner perspective of this paper. If a heroic computational effort, along with a healthy dose of good luck, is required to find a model's parameters, one is probably well advised to actively consider simpler alternatives.

We therefore conclude that, for a risk-management practitioner, the empirical models offer an appealing modelling alternative. The Nelson-Siegel and exponential-spline models offer a number of advantages relative to the $A_0(3)$ models. The derivation of these models involves relatively simple mathematics. Even better, the parameter estimation is fast and efficient, since one need only use basic linear econometrics. Perhaps most importantly, these models outperform the $A_0(3)$ models, and quite often the random walk, in their ability to out-of-sample forecast zero-coupon rates and excess holding-period returns. Moreover, the Nelson-Siegel and exponential-spline models also demonstrate an ability, based on two alternative econometric tests, to capture deviations from the expectations hypothesis. Finally, these models perform reasonably well in the portfolio-optimization exercise; this is important, since our ultimate practical application for these models involves portfolio

[34]It is difficult to use zero-coupon data in Canada prior to the 1990s, owing to the relatively fragmented and illiquid nature of the Government of Canada bond market.

selection.

These conclusions are not terribly surprising in the case of the Nelson-Siegel model, given the work of Diebold and Li (2003). What is perhaps slightly novel in this work is that an extension of the Diebold and Li (2003) approach to other basis functions also does quite well. In particular, the generalization to orthogonalized exponentials in the form of the exponential-spline model does as well, and occasionally better, than the Nelson-Siegel model. The parsimony of the Nelson-Siegel is a clear advantage of this model, but we conclude that the exponential-spline model merits more investigation.

Bibliography

Ahn, D.H., R.F. Dittmar, and A.R. Gallant. 2002. "Quadratic Term-Structure Models: Theory and Evidence." *Review of Financial Studies* 15: 243–88.

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: 745–87.

Backus, D., S. Foresi, A. Mozumdar, and L. Wu. 2001. "Predictable Changes in Yields and Forward Rates." *Journal of Financial Economics* 59: 281–311.

Bernadell, C., J. Coche, and K. Nyholm. 2005. "Yield Curve Prediction for Strategic Investors." European Central Bank Working Paper No. 472.

Billingsley, P. 1995. *Probability and Measure*. Third Avenue, New York, New York: Wiley, third edition.

Bolder, D.J. 2001. "Affine Term-Structure Models: Theory and Implementation." Bank of Canada Working Paper No. 2001-15.

———. 2002. "Towards a More Complete Debt Strategy Simulation Framework." Bank of Canada Working Paper No. 2002-13.

———. 2003. "A Proposed Stochastic Simulation Framework for the Government of Canada's Debt Strategy Problem." Bank of Canada Working Paper No. 2003-10.

Bolder, D.J. and S. Gusba. 2002. "Exponentials, Polynomials, and Fourier Series: More Yield Curve Modelling at the Bank of Canada." Bank of Canada Working Paper No. 2002-29.

Bolder, D.J., G. Johnson, and A. Metzler. 2004. "An Empirical Analysis of the Canadian Term Structure of Zero-Coupon Interest Rates." Bank of Canada Working Paper No. 2004-48.

Bolder, D.J. and D. Streliski. 1999. *Yield-Curve Modelling at the Bank of Canada*. Technical Report No. 84, Ottawa: Bank of Canada.

Brémaud, P. 1999. *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues*. 175 Fifth Avenue, New York, NY: Springer Verlag. Texts in Applied Mathematics.

Brigo, D. and F. Mercurio. 2001. *Interest Rate Models: Theory and Implementation*. Berlin, Germany: Springer Verlag.

Cairns, A.J.G. 1998. "Descriptive Bond-Yield and Forward-Rate Models for the British Government Securities Market." *British Actuarial Journal* 4: 265–321.

———. 2001. "Stability of Descriptive Models for the Term Structure of Interest Rates with Application to German Market Data." *British Actuarial Journal* 7: 467–507.

———. 2004. "A Family of Term-Structure Models for Long-Term Risk Management and Derivative Pricing." *Mathematical Finance* 14: 415–44.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton, New Jersey: Princeton University Press.

Casella, G. and R.L. Berger. 1990. *Statistical Inference*. Belmont, California: Duxbury Press.

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.

Cheridito, P., D. Filipović, and R.L. Kimmel. 2005. "Market Price of Risk Specifications for Affine Models: Theory and Evidence." Princeton University Working Paper.

Cochrane, J. and M. Piazzesi. 2005. "Bond Risk Premia." *American Economic Review* 95: 138–60.

Dai, Q. and K.J. Singleton. 2000. "Specification Analysis of Affine Term Structure Models." *Journal of Finance* 55: 1943–78.

———. 2002. "Expectation Puzzles, Time-Varying Risk Premia, and Dynamics Models of the Term Structure." *Journal of Financial Economics* 63: 415–41.

Dai, Q., K.J. Singleton, and W. Yang. 2004. "Predicitibility of Bond Risk Premia and Affine Term Structure Models." University of Stanford Working Paper.

Diebold, F.X., L. Ji, and C. Li. 2004. "A Three-Factor Yield Curve Model: Non-Affine Structure, Systemic Risk Sources, and Generalized Duration." University of Pennsylvania Working Paper.

Diebold, F.X. and C. Li. 2003. "Forecasting the Term Structure of Government Bond Yields." University of Pennsylvania Working Paper.

Diebold, F.X., M. Piazzesi, and G.D. Rudebusch. 2005. "Modeling Bond Yields in Finance and Macroeconomics." Federal Reserve Bank of San Fransisco Working Paper 2005-04.

Diebold, F.X., G.D. Rudebusch, and S.B. Aruoba. 2004. "The Macroeconomy and the Yield Curve: A Nonstructural Analysis." University of Pennsylvania Working Paper.

Duffee, G.D. 2002. "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance* 57: 405–43.

Duffie, D. 1996. *Dynamic Asset Pricing Theory.* Princeton, New Jersey: Princeton University Press, second edition.

Durrett, R. 1996. *Probability: Theory and Examples.* Belmont, California: Duxbury Press, second edition.

Filipović, D. 1999. "A Note on the Nelson-Siegel Family." *Mathematical Finance* 9: 349–59.

Fishman, G.S. 1995. *Monte Carlo: Concepts, Algorithms, and Applications.* 175 Fifth Avenue, New York, NY: Springer-Verlag. Springer Series in Operations Research.

Flesaker, B. and L. Hughston. 1996. "Positive Interest." *Risk* 9: 46–49.

Hamilton, J.D. 1994. *Time Series Analysis.* Princeton, New Jersey: Princeton University Press. Chapter 22.

Harvey, A.C. 1990. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press, first edition.

Johnson, N.L., S. Kotz, and N. Balakrishnan. 1997. *Continuous Univariate Distributions: Volume 2.* New York, NY: John Wiley and Sons. Wiley Series in Probability and Statistics.

Judge, G.G., W. Griffiths, R.C. Hill, H. Lütkepohl, and T.C. Lee. 1985. *The Theory and Practice of Econometrics.* New York, New York: John Wiley and Sons, second edition.

Karatzas, I. and S.E. Shreve. 1991. *Brownian Motion and Stochastic Calculus.* Berlin: Springer-Verlag, second edition.

Kim, C.J. and C.R. Nelson. 1999. *State-Space Models with Regime Switching.* Cambridge, Massachusetts: The MIT Press. Chapters 4-5.

Leippold, M. and L. Wu. 2000. "Quadratic Term-Structure Models." Swiss Institute of Banking and Finance Working Paper.

———. 2001. "Design and Estimation of Quadratic Term-Structure Models." Swiss Institute of Banking and Finance Working Paper.

———. 2003. "Design and Estimation of Multi-Currency Quadratic Models." Swiss Institute of Banking and Finance Working Paper.

Li, B., E. DeWetering, G. Lucas, R. Brenner, and A. Shapiro. 2001. "Merrill Lynch Exponential Spline Model." Merrill Lynch Working Paper.

Litterman, R. and P. Schenkman. 1991. "Common Factors Affecting Bond Returns." *Journal of Fixed Income* 1: 54–61.

Nelson, C.R. and A.F. Siegel. 1987. “Parsimonious Modeling of Yield Curves.” *Journal of Business* 60: 473–89.

Newey, W.K. and K.D. West. 1986. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” NBER Technical Working Paper No. 55.

Oksendal, B.K. 1995. *Stochastic Differential Equations*. Berlin: Springer Verlag, fourth edition.

Pesaran, M.H. and S.M. Potter. 1992. “Nonlinear Dynamics and Econometrics: An Introduction.” *Journal of Applied Econometrics* 7: S1–S7.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Trumpington Street, Cambridge: Cambridge University Press, second edition.

Rogers, L.C.G. 1997. “The Potential Approach to the Term Structure of Interest Rates and Foreign Exchange Rates.” *Mathematical Finance* 7: 157–76.

Rutkowski, M. 1997. “A Note on the Flesaker and Hughston Model of the Term Structure of Interest Rates and Foreign Exchange Rates.” *Applied Mathematical Finance* 4: 151–63.

Shea, G.S. 1985. “Interest Rate Term Structure Estimation with Exponential Splines: A Note.” *Journal of Finance* 40: 319–25.

Svensson, L.E. 1994. “Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994.” Centre for Economic Policy Research Discussion Paper No. 1051.

Svensson, L.E. and P. Söderlind. 1997. “New Techniques to Extract Market Expectations from Financial Instruments.” Centre for Economic Policy Research Discussion Paper No. 1556.

Vasicek, O.A. and H.G. Fong. 1981. “Term Structure Modeling Using Exponential Splines.” *Journal of Finance* 37: 339–48.

Bank of Canada Working Papers
Documents de travail de la Banque du Canada

Working papers are generally published in the language of the author, with an abstract in both official languages, and are available on the Bank's website (see bottom of page). ***Les documents de travail sont généralement publiés dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue. On peut les consulter dans le site Web de la Banque du Canada, dont l'adresse est indiquée au bas de la page.***

2006

2006-47	Stress Testing the Corporate Loans Portfolio of the Canadian Banking Sector	M. Misina, D. Tessier, and S. Dey
2006-46	Survey-Based Estimates of the Term Structure of Expected U.S. Inflation	S. Kozicki and P.A. Tinsley
2006-45	The Role of Debt and Equity Finance over the Business Cycle	F. Covas and W.J. den Haan
2006-44	The Long-Term Effects of Cross-Listing, Investor Recognition, and Ownership Structure on Valuation	M.R. King and D. Segal
2006-43	Efficient Hedging and Pricing of Equity-Linked Life Insurance Contracts on Several Risky Assets	A. Melnikov and Y. Romanyuk
2006-42	Linking Real Activity and Financial Markets: The Bonds, Equity, and Money (BEAM) Model	C. Gauthier and F. Li
2006-41	An Optimized Monetary Policy Rule for ToTEM	J.-P. Cayen, A. Corbett, and P. Perrier
2006-40	Education and Self-Employment: Changes in Earnings and Wealth Inequality	Y. Terajima
2006-39	Short-Run and Long-Run Causality between Monetary Policy Variables and Stock Prices	J.-M. Dufour and D. Tessier
2006-38	Conditioning Information and Variance Bounds on Pricing Kernels with Higher-Order Moments: Theory and Evidence	F. Chabi-Yo
2006-37	Endogenous Borrowing Constraints and Consumption Volatility in a Small Open Economy	C. de Resende
2006-36	Credit in a Tiered Payments System	A. Lai, N. Chande, and S. O'Connor
2006-35	Survey of Price-Setting Behaviour of Canadian Companies	D. Amirault, C. Kwan, and G. Wilkinson
2006-34	The Macroeconomic Effects of Non-Zero Trend Inflation	R. Amano, S. Ambler, and N. Rebei
2006-33	Are Canadian Banks Efficient? A Canada–U.S. Comparison	J. Allen, W. Engert, and Y. Liu
2006-32	Governance and the IMF: Does the Fund Follow Corporate Best Practice?	E. Santor

Copies and a complete list of working papers are available from:
Pour obtenir des exemplaires et une liste complète des documents de travail, prière de s'adresser à :

Publications Distribution, Bank of Canada
234 Wellington Street, Ottawa, Ontario K1A 0G9
Telephone: 1 877 782-8248
(toll free in North America)
Email: publications@bankofcanada.ca
Website: http://www.bankofcanada.ca

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9
Téléphone : 1 877 782-8248 (sans frais en Amérique du Nord)
Adresse électronique : publications@banqueducanada.ca
Site Web : http://www.banqueducanada.ca