BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2007-49

Examining Simple Joint Macroeconomic and Term-Structure Models: A Practitioner's Perspective

by David Jamieson Bolder and Shudan Liu

Bank of Canada Working Paper 2007-49

October 2007

Examining Simple Joint Macroeconomic and Term-Structure Models: A Practitioner's Perspective

by

David Jamieson Bolder and Shudan Liu*

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

* Shudan Liu worked on this paper during a six-month tenure as a visiting PhD student from the Applied Mathematics Department of the University of Western Ontario.

ISSN 1701-9397

Acknowledgements

We would like to thank Greg Bauer, Antonio Diez de los Rios, and Stephen Murchison from the Bank of Canada for their useful input into this paper. We would also like to thank Jeremy Rudin from the Department of Finance Canada, Mark Reesor from the Applied Mathematics Department of University of Western Ontario, and Jeremy Graveline from the Carlson School of Management at the University of Minnesota. As always, all acknowledgement is without implication. The authors are entirely responsible for any errors, omissions, or unclarities.

Abstract

The primary objective of this paper is to compare a variety of joint models of the term structure of interest rates and the macroeconomy. To this end, we consider six alternative approaches. Three of these models follow from the work of Diebold and Li (2003) with a generalization in Bolder (2006). The fourth model is a regression-based approach motivated entirely by empirical considerations. The fifth model follows from the seminal work of Ang and Piazzesi (2003), who suggest a joint macro-finance model in a discrete-time affine setting. The final model, which we term an observed-affine model, represents an adjustment to the Ang-Piazzesi model that essentially relaxes restrictions on the state-variable dynamics by making them observable. The observed-affine model is similar in spirit to work by Colin-Dufresne, Goldstein, and Jones (2005) and Cochrane and Piazzesi (2006). Using monthly Canadian data from 1973 to 2005, we compare each of these models in terms of their out-of-sample ability to forecast the transition density of zero-coupon rates. We also examine a simple approach aimed at permitting a subset of the parameters in the non-affine models to vary over time. We find, similar to Bolder (2006), that the Diebold and Li (2003) motivated approaches provide the most appealing modelling alternative across our different comparison criteria.

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

Résumé

Le but premier des auteurs est de comparer entre eux différents modèles qui formalisent à la fois la dynamique de la structure des taux d'intérêt et celle de l'économie. Des six modèles qu'ils considèrent, trois s'inspirent des recherches de Diebold et Li (2003) et d'une généralisation du modèle de ces derniers présentée par Bolder (2006). Le quatrième modèle fait appel à l'analyse de régression et est motivé exclusivement par des considérations empiriques. Le cinquième modèle s'inscrit dans la lignée des travaux novateurs d'Ang et Piazzesi (2003), qui suggèrent l'emploi d'un modèle macrofinancier de type affine formulé en temps discret. Le dernier, le modèle affine à facteurs observables, est une variante du modèle d'Ang et Piazzesi où l'on a recours à des variables d'état observables au lieu d'imposer des restrictions à la dynamique de ces variables. Le modèle affine à facteurs observables emprunte à la démarche de Colin-Dufresne, Goldstein et Jones (2005) et de Cochrane et Piazzesi (2006). À l'aide de données canadiennes mensuelles allant de 1973 à 2005, les auteurs comparent la capacité de chacun de ces modèles à prévoir la densité de transition des taux d'obligations coupon zéro au-delà de la période d'estimation. Ils analysent également une méthode simple devant permettre à un sous-ensemble des paramètres des modèles non affines de varier dans le temps. À l'instar de Bolder (2006), ils constatent que les modèles prenant appui sur les travaux de Diebold et Li (2003) sont les plus prometteurs eu égard aux critères de comparaison retenus.

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

1 Introduction

In a risk-management setting, one is generally interested in understanding the dynamics of the term-structure of interest rates under the physical probability measure. This differs from the dynamics of the term structure of interest rates under the risk-neutral measure in a number of ways, not least of which is the need to estimate one's parameters to historical data rather than to calibrate them to existing market conditions. This raises a number of practical econometric challenges and also naturally leads one to inquire about the in- and out-of-sample fit to the data. While there is a growing literature addressing term-structure dynamics under the physical measure, it is an unfortunate reality that the models offered by the literature generally perform poorly out-of-sample.[1] Indeed, most models are hard pressed to beat the random-walk assumption, which basically amounts to assuming that interest rates are martingales. In the words of Duffee (2002) "if a model produces poor forecasts of future rates [...] it is unlikely that the model can shed light on the economics underlying the failure of the expectations hypothesis." Moreover, if the model has little to say about the expectations hypothesis, the failure of which is the central empirical fact about the term structure of interest rates, then it is unlikely that it will be a suitable tool for describing interest-rate dynamics.

Recent work has addressed this deficiency in a variety of ways. Duffee (2002) and Cheridito, Filipović, and Kimmel (2005) introduce a more flexible market price of risk aimed at improving the class of affine term-structure models. Diebold and Li (2003) introduce an empirically motivated approach, which aims to improve the forecasting capacity of term-structure models. Ang and Piazzesi (2003) demonstrate how the incorporation of macroeconomic factors into a term-structure model can actually improve model performance. These are useful contributions and this paper examines all of these improvements, with a particular focus on the last two suggestions, from a practical risk-management perspective. In other words, this work investigates how we might use recent academic work for the joint description of Canadian term structure and macroeconomic dynamics.[2]

Throughout the course of this paper, we attempt to extend the current literature in a few unconventional directions. We do this out of a desire to identify relatively simple models that can be used for practical risk-management applications. It is, for example, absolutely essential that when we simulate from our model that the resulting outcomes are consistent with the financial and economic environment used to estimate the data. That is, it is critically important that our joint model be capable of describing the transition density of the term

[1] See, for example, Duffee (2002), Diebold and Li (2003), and Bolder (2006) among others for more detail.

[2] Bolder (2002, 2003) provide one possible approach for describing this joint distribution. The foundation of this methodology is the specification of a two-state Markov chain for Canadian output. The interest-rate dynamics are described by an affine term-structure model. The corresponding link between the macroeconomy and the term structure comes through the market price of risk, which is assumed to be a function of the probability of recession. It should be noted, however, that this model was very much an *ad hoc* construction that was built to address a specific risk-management question.

structure of interest rates. This is generally equivalent to having a model that generates good out-of-sample forecasts. As previously mentioned, the extant models in the finance literature demonstrate difficulty in this respect. We feel, therefore, that practical risk-management needs provide a certain degree of practitioner license to modify the existing models.

This paper is, in many ways, quite similar in spirit to Bolder (2006). Specifically, we compare a collection of empirically motivated models suggested by Diebold and Li (2003). We also examine the seminal joint macroeconomic term-structure affine model suggested by Ang and Piazzesi (2003). To this end, we compare a number of models in terms of their ability to forecast the first two moments of zero-coupon rates, and excess holding-period returns. There is, however, a major difference between this paper and Bolder (2006). In particular, we examine an unconventional adjustment to the Ang and Piazzesi (2003) model that essentially involves making the yield-curve state variables observable. The motivation for this adjustment is to relax the Ang and Piazzesi (2003) assumption of independence between macroeconomic and financial factors and simultaneously increase model flexibility without increasing the dimension of the parameter space and consequently making the optimization problem even more difficult. This suggested model is similar in spirit to work by Colin-Dufresne, Goldstein, and Jones (2005) and Cochrane and Piazzesi (2006).

Our adjustment essentially separates the estimation of the physical and pricing measure parameters. We term this the *observed-affine* model as our change implies that the state variables are no longer latent; instead, the yield-curve related state variables are extracted from a principal components analysis of the variance-covariance matrix of zero-coupon yield movements. We would encourage the reader *not* to immediately dismiss this unconventional adjustment. We believe it is worth investigation for two reasons. First, as will be seen, it does a reasonable job of generating out-of-sample estimates. As such, it appears to offer some usefulness from a practitioner's perspective. Second, we feel that it provides some insight into the nature of the out-of-sample underperformance of the affine class of term-structure models. In particular, it is our view that the estimation algorithm does not appear to place sufficient weight on the description of state-variable dynamics. This insight may be helpful to other researchers in constructing a more complete and theoretically appealing solution to the problem.

The paper is organized into two principal sections. Section 2 provides the motivation and description of the six alternative models considered in this study, with most of the technical details relegated to technical appendices. Section 3 compares these models on a number of dimensions including in-sample cross-sectional fit, out-of-sample fit to the first two moments of zero-coupon rate dynamics, and an out-of-sample description of excess holding-period returns. Section 4 provides concluding remarks.

2 Models

In this section, we describe the various models that we will consider in this study. Before we actually turn to the specific models, however, it is useful to first review the general structure of term-structure models. We provide this high-level overview to motivate the perhaps less than conventional modifications that we propose in this paper.

Most of the popular models of the term structure of interest rates, despite the broad literature in this area, have essentially the same form. Abstracting from the technical details, there are, in fact, really only *three* key assumptions involved in the modelling process. The first assumption is that the interest-rate system is a function of some set of state variables. The idea is that the *state* of the interest-rate system can be characterized by values of these variables, or factors, at each given point in time. These state variables may be latent (i.e., unobservable) or observable factors such as macroeconomic variables. Work by Litterman and Schenkman (1991) demonstrates that, generally speaking, the majority of the variance in interest-rate movements is well described by three or four factors. This useful paper provides confidence that working with a low-dimensional state-variable system is reasonable. Consequently, it is rare to see a term-structure model with more than three latent and/or—in the new stream of joint macro-finance term-structure models—two or three macroeconomic state variables.

The second assumption involves the dynamics of the state-variable vector. Generally, one can think of the state-variable vector as being any n-dimensional stochastic process. This is, however, a bit too general for practical work. There are a variety of possible assumptions. If one works in a continuous-time setting, it is common to use a stochastic differential equation—such as an Ornstein-Uhlenbeck or square-root process—to describe the state variable dynamics. In a discrete-time setting, one generally uses a vector autoregressive process—the discrete-time analogue of the Ornstein-Uhlenbeck process—to describe the evolution of the state-variable vector. This is not to say, however, that these are the only approaches to describing state-variable dynamics. In particular, there have been a number of papers considering discontinuous sample paths (i.e., Levy processes or jumps with Poisson arrival) that have been important in incorporating the skewness and kurtosis evident in interest-rate movements. We do not consider these complications in this paper.

The third, and final, assumption relates to the mapping between the state variables and the term structure of interest rates. This can, in principle, be quite *ad hoc* as any function that maps the state-vector into a collection of pure-discount bond prices (or correspondingly zero-coupon interest rates) would do the job. An arbitrary mapping can, however, be somewhat dangerous as it can lead to an interest-rate system that permits arbitrage. Building a system upon a model that frequently allows for arbitrage opportunities is probably a bad idea. Having said that, if a model has other positive properties, but permits a small probability of arbitrage opportunities, then it might *not* be a big problem.[3] One's opinion on this matter, of course, will depend importantly on one's ultimate application. If one is pricing exotic structured fixed-income securities, the presence of arbitrage opportunities in one's model could easily prove disastrous. If, however, one is trying—as we are in most risk-management settings—to describe interest-rate dynamics, then the permission of a small number of arbitrage opportunities may not be so problematic.

The majority of extant models, nevertheless, ensure that the mapping excludes arbitrage opportunities. The consequence is a number of restrictions on the nature of the state-variable dynamics and the corresponding form of the mapping. The most common model is the so-called affine model, where pure-discount bond prices are an exponential-affine function of the state variables. Dai and Singleton (2000) and Duffie, Filipovic, and Schachermayer (2003) are two excellent references for this very popular class of term-structure models. Leippold and Wu (2000) introduce the idea of an exponential-quadratic mapping between pure-discount bond prices and the state-variable vector leading to the quadratic term-structure approach. Flesaker and Hughston (1996) and Cairns (2004) consider a rather more complex integral expression for the price of a pure-discount bond as a function of the state variables; this has been termed the positive-interest rate model. There are, of course, other no-arbitrage mappings, but they find application in the pricing of contingent claims and are not particularly relevant given the practical risk-management focus of this paper.[4]

Diebold and Li (2003) introduce a mapping between a set of discrete-time, continuous-value state variables that is *not* motivated by no-arbitrage considerations. In this setting, zero-coupon interest rates are described as a linear combination of the state variables where the coefficients are relatively simple exponentially weighted combinations of Laguerre polynomials.[5] Bolder (2006) extends this idea to consider alternative mappings including orthogonalized-exponential and Fourier-series function coefficients. By virtue of the lack of no-arbitrage restrictions, these models operate under substantially fewer restrictions than the previously mentioned affine, quadratic, and positive interest-rate models. Moreover, there is a reasonable amount of empirical evidence suggesting that these models outperform the no-arbitrage models in their ability to generate out-of-sample forecasts and predict excess holding-period returns. These models do not, however, ensure a lack of arbitrage opportunities; exactly how many such opportunities may present themselves is an empirical question.

The natural question, therefore, is why do these models generate superior forecasts? Diebold and Li (2003) suggest that the reason is the relative parsimony of their model that leads to superior forecasting performance.

[3]The thinking here is analogous to the negative interest rates that occur with positive probability in Gaussian affine term-structure models. Everyone can agree that negative nominal interest rates are a fairly unreasonable model output, but a small probablilty of negative interest rates has not stopped these models from widespread application.

[4]One popular example is the well-known HJM-framework proposed by Heath, Jarrow, and Morton (1992), which has a wide range of practical implementations such as the approach suggested by Brace, Gatarek, and Musiela (1997).

[5]The details of this construction are found in Appendix C.

Bolder (2006) finds, however, that the relatively higher-dimensional exponential-spline and Fourier-series models also outperform affine term-structure models. This would suggest that there is perhaps a bit more going on than just increased parsimony. Duffee (2002) suggests that the underperformance of affine models stems from a lack of flexibility in the mathematical form of the market price of risk. His correction appears to improve the situation, but does not appear to entirely solve the problem. Dai, Le, and Singleton (2006) propose an even more general approach in the discrete-time setting that may be helpful. Given the range of mathematical restrictions on the market price of risk required to ensure the no-arbitrage condition, however, this is not a trivial exercise.

One hypothesis is that the difficulty associated with no-arbitrage models is related to the estimation procedure. There are essentially two sets of parameters in these models: those parameters related to the pricing of pure-discount bonds (i.e., the equivalent martingale measure, or $\mathbb{Q}$, parameters) and those parameters related to the dynamics of the state variables (i.e., the physical measure, or $\mathbb{P}$, parameters). It is precisely these $\mathbb{P}$ parameters that are critical for the prediction of future yield-curve outcomes. The estimation approach in these models uses panel data; that is, it is a time series of zero coupon rates, where for each date there is a cross section of rates across the maturity spectrum. The maximum-likelihood estimation approach proposed by Chen and Scott (1993) should, in principle, use the cross section to determine the $\mathbb{Q}$ parameters and the time series to identify the $\mathbb{P}$ parameters; we will refer to these as pricing and physical parameters, respectively. What appears to happen, however, is that the cross section dominates the estimation algorithm. The parameters generally fit the cross section quite well, but do a poor job of identifying the dynamics of the state variables. In other words, they appear to get the mapping between the state variables and pure-discount bonds correct, but miss the state-variable dynamics.[6] In short, the Chen and Scott (1993) estimation algorithm used for affine term-structure models is really quite ambitious insofar as it attempts to simultaneously extract the latent state variables, determine the dynamics of these state variables, and price the collection of zero-coupon bonds. Clearly, this is a tall order and it is not entirely surprising that it has difficulty.

To summarize, a term-structure model has three components: a collection of observable and/or unobservable state variables, a description of the dynamics of these state variables, and a mapping between these state variables and the term-structure of interest rates. The mapping can either be theoretically motivated and constructed so as to avoid arbitrage opportunities or constructed solely based upon empirical considerations. Estimation of these models occurs with panel data that simultaneously describes the time-series and cross section of the term

[6] It is important to note that this point is *not* entirely inconsistent with Duffee (2002), who argues that a more flexible market price of risk specification is required. It is possible, for example, that a sufficiently flexible mathematical form for the market price of risk—which is essentially the link between the physical and pricing worlds—could allow for the appropriate parametrization of the $\mathbb{P}$ and $\mathbb{Q}$ parameters. The mathematical restrictions on the form of the market price of risk, however, make this a daunting task. We leave this work for others.

structure of interest rates. Moreover, we can divide parameters describing the determination of the cross section and those describing the state-variable dynamics.

In this paper, we consider three fairly distinct approaches. Ang and Piazzesi (2003) represents the first approach and is the seminal paper in this nascent literature. This work begins by demonstrating that the instantaneous short-rate equation that arises in the affine-term structure literature can be extended—with the addition of macroeconomic state variables—to a Taylor rule. This permits the expansion of affine term-structure models to explicitly include observable macroeconomic outcomes. We address this model in Section 2.1. In the second approach, we transform the Ang-Piazzesi model in a manner that separates the estimation of the physical and pricing parameters and improves its out-of-sample forecast properties. This approach is termed the observed-affine model and is discussed in Section 2.2. The third approach, forwarded by Diebold, Rudebusch, and Aruoba (2004b), provides a natural extension of the Diebold and Li (2003) model by including macroeconomic variables as state variables. This model, along with an extension to include the modifications described in Bolder (2006) and finally an extremely simple, empirically motivated fifth approach, are described in Section 2.3.

2.1 An affine model

In this section, we discuss the model introduced by Ang and Piazzesi (2003). This paper is the seminal work in attempting to find a connection between macroeconomic and financial descriptions of the term structure of interest rates. Their approach is described in terms of our previously discussed three principal term-structure model assumptions. The entry point of the model, and the first term-structure assumption, is essentially a combination of the classical Taylor rule and the short rate equation used in the affine term-structure literature. The idea is that the short-term interest rate, targeted by the monetary authority, depends linearly on a set of *observable* macroeconomic state variables, represented by the vector X_t^o, and a set of *unobservable* latent yield-curve state variables denoted X_t^u. The short rate, r_t, has the form,

$$
\begin{aligned}
r_t &= \underbrace{\delta + \Delta_1^T X_t^o}_{\text{Taylor rule}} + \Delta_2^T X_t^u, \qquad (1)\\
&= \delta + \begin{bmatrix} \Delta_1^T & \Delta_2^T \end{bmatrix} \begin{bmatrix} X_t^0 \\ X_t^u \end{bmatrix}, \\
&= \delta + \Delta^T X_t,
\end{aligned}
$$

where δ, Δ_1, and Δ_2 have the necessary dimensions to ensure that r_t is a scalar, and $\Delta^T = \begin{bmatrix} \Delta_1 & \Delta_2 \end{bmatrix}^T$. In short, Ang and Piazzesi (2003) assume that the state variables influencing the term-structure of interest rates include unobservable financial variables (X_t^u) and observable macroeconomic variables (X_t^o). Ang and Piazzesi

(2003) recommend two macroeconomic state variables related to output and inflation. In our implementation of their model, we use the output gap (x_t) and consumer-price inflation (π_t). This implies that $X_t^o = \begin{bmatrix} x_t & \pi_t \end{bmatrix}^T$.

The second key term-structure modelling assumption relates to the state-variable dynamics. Ang and Piazzesi (2003) permit, in full generality, a VAR(p) specification for X_t^o and a VAR(1) approach for X_t^u. Essentially, this implies that the latent state variables are Markovian, while there is some memory in the macroeconomic variables. In our implementation of this model, we include only the contemporaneous macroeconomic state variables (i.e., VAR(1)) given that the evidence in Ang and Piazzesi (2003), and our own experimentation with the model, suggests that incorporating numerous lags for the economic data leads to overparametrization and poor out-of-sample fit. The actual state-variable dynamics, therefore, have the form,

$$\begin{bmatrix} X_t^o \\ X_t^u \end{bmatrix} = \begin{bmatrix} C^0 \\ C^u \end{bmatrix} + \begin{bmatrix} \Phi^0 & 0_{2\times 3} \\ 0_{3\times 2} & \Phi^u \end{bmatrix} \begin{bmatrix} X_{t-1}^o \\ X_{t-1}^u \end{bmatrix} + \begin{bmatrix} \epsilon_t^o \\ \epsilon_t^u \end{bmatrix}, \tag{2}$$

$$X_t = C + \Phi X_{t-1} + \epsilon_t,$$

where $C^u \equiv 0$ and,

$$\epsilon_t = \begin{bmatrix} \epsilon_t^o \\ \epsilon_t^u \end{bmatrix} \sim \mathcal{N}\left(\vec{0}, \underbrace{\begin{bmatrix} \Omega^0 & 0_{2\times 3} \\ 0_{3\times 2} & I_{3\times 3} \end{bmatrix}}_{\Omega}\right). \tag{3}$$

We also define Σ as the Cholesky decomposition of the variance-covariance matrix, Ω.[7]

Observation of equations (2) and (3) reveals that the dynamics of the latent state variables are determined independently of the macroeconomic variables.[8] In other words, the current model construction does not permit the macroeconomic variables to improve the description of the latent, yield-curve, state variables.[9] This is problematic as we might have different expectations, for example, associated with a flat term-structure with low inflation and high output than a flat term-structure with high inflation and low output. The influence of the macroeconomic factors on the term structure, therefore, arises elsewhere in the model.

The no-arbitrage mapping from the state variables to the term structure of interest rates, which is the third and final term-structure modelling assumption, has two components. The first is the specification of the market

[7]More specifically, $\Sigma = \text{chol}(\Omega)$, where Σ is a lower-triangular matrix such that $\Sigma\Sigma^T = \Omega$. See Press et al. (1992, 96–98) for more details on this decomposition.

[8]The orthogonality of the observable and unobservable state variables, while perhaps difficult to defend from an empirical perspective, makes the estimation substantially easier, and perhaps ensures that it is possible, using traditional estimation techniques. Ang and Piazzesi (2003), however, are quite clear that the imposition of independence between latent and macroeconomic variables is a drawback of their model.

[9]Equation (67) in appendix A also demonstrates that the Λ matrix pre-multiplying the state variables in the market price of risk is also block diagonal. This implies that state variables are also independent from a pricing perspective.

price of risk. The market price of risk is the link between the physical and pricing measures. Ang and Piazzesi (2003) suggest the following form stemming from the collection of essentially affine models proposed by Duffee (2002),

$$\lambda_t = \lambda + \Lambda X_t, \tag{4}$$

where $\lambda \in \mathbb{R}^{5\times 1}$ and $\Lambda \in \mathbb{R}^{5\times 5}$. If one works through the mathematics, one can construct an exponential affine mapping in the following recursive form,

$$P_{n+1,t} = \exp\left(\underbrace{A_n - \delta + \frac{1}{2}B_n^T\Sigma\Sigma^T B_n - B_n^T\Sigma\lambda}_{A_{n+1}} + \underbrace{\left(B_n^T\left(\Phi - \Sigma\Lambda\right) - \Delta^T\right)}_{B_{n+1}^T} X_t\right), \tag{5}$$

where $A_1 = -\delta$ and $B_1 = -\Delta$, and $P_{n+1,t}$ describes the price of an $(n+1)$-period pure-discount bond at time t. This equation allows us to use the current value of the state variables and our model parameters to recursively construct the term structure of interest rates. More detail on the mathematical derivation of this model, and the approach used to estimate the model parameters, is found in Appendix A.

Equation (5) illustrates how the macroeconomic variables influence the term structure of interest rates. Clearly, each macroeconomic variable is a factor that describes the cross section of the term structure at a specific point in time. Clearly, the influence of the macroeconomic variables on the yield curve occurs through this mapping. The zero-coupon yield curve, therefore, is represented as a linear combination of financial and macroeconomic variables. While this appears reasonable, it is a bit of a concern given that Litterman and Schenkman (1991) demonstrate that three financial factors are capable of describing approximately 95 per cent of the variance in zero-coupon curve movements. We keep this in mind as we examine the other approaches and compare their relative performance.

2.2 An observed-affine model

The motivation behind this model is quite straightforward. We attempt to relax the assumption of independence between the macroeconomic and latent yield-curve state variables and thereby permit the macroeconomic variables to provide some information on yield-curve dynamics. This is motivated, in part, by Colin-Dufresne, Goldstein, and Jones (2005) who attempt to find an invariant transformation of the latent variables in an affine model. It is also similar in spirit to the model suggested by Cochrane and Piazzesi (2006) in the context of building a factor-model for the term structure of interest rates.

The idea is to use the first three principal components from an eigenvalue decomposition of zero-coupon interest rates. This allows us to treat the yield-curve related variables, and thus all state variables, as observable.

This is the first term-structure modelling assumption relating to the choice of state variables. In one respect, this reduces the flexibilty of the model to adjust the latent state variables to tightly fit the term structure of interest rates. The consequence should, therefore, be a less tight fit to the cross section of interest rates. In another respect, this modification permits a substantial increase in the ease of estimation. Specifically, it permits us to split the parameter set into physical- and pricing-measure parameters; moreover, each set of parameters is estimated independently. We denote these measures as $\mathbb{P}$ and $\mathbb{Q}$, respectively. This helps to ensure that the physical parameters, associated with the $\mathbb{P}$-measure, are not distorted by the estimation algorithm in an effort to fit the zero-coupon cross section.

The first yield-curve assumption surrounds the choice of state variables. The state variable vector, therefore, has the form,

$$X_t = \begin{bmatrix} x_t & \pi_t & r_t & l_t & s_t & c_t \end{bmatrix}^T, \tag{6}$$

where l_t, s_t, and c_t are the level, slope, and curvature factors from the eigenvalue decomposition; we call these X_t^e. The x_t, π_t, and r_t are the macroeconomic variables representing the output gap, consumer-price inflation, and the overnight rate; we will denote these as X_t^o. The instantaneous short-rate, required in an affine setting, is thus merely a simple identity,

$$\begin{aligned} r_t &= \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} X_t, \\ &= \mathbb{I}_r X_t, \end{aligned} \tag{7}$$

which compares to equation (1) in the Ang-Piazzesi model.

For the second key yield-curve assumption, we provide a unrestricted VAR(2) specification for the state-variable dynamics. The actual state-variable dynamics are described as,

$$\begin{aligned} \begin{bmatrix} X_t^o \\ X_t^e \end{bmatrix} &= \begin{bmatrix} C^o \\ C^e \end{bmatrix} + \sum_{k=1}^{2} \begin{bmatrix} F_k^o & F_k^{eo} \\ F_k^{oe} & F_k^e \end{bmatrix} \begin{bmatrix} X_{t-k}^o \\ X_{t-k}^e \end{bmatrix} + \epsilon_t, \\ X_t &= C + \sum_{k=1}^{2} F_k X_{t-k} + \epsilon_t, \end{aligned} \tag{8}$$

where $C^e \neq 0$ and,

$$\epsilon_t \sim \mathcal{N}\left(0, \Omega\right), \tag{9}$$

and $\Sigma = \text{chol}(\Omega)$. The important point is that, in contrast to equations (2) and (3), interaction is permitted between the yield-curve and macroeconomic state variables. This implies that the future zero-coupon term-structure dynamics can now also vary depending on the state of the macroeconomy.

The corresponding price of an $(n+1)$-period zero-coupon bond—or rather the mapping between the state variables and the zero-coupon curve—has a similar form to equation (5) in the previous section. Namely,

$$P_{n+1,t} = \exp\left\{ \underbrace{A_n + B_n^T(C - \Sigma\lambda) + \frac{1}{2}B_n^T \Sigma\Sigma^T B_n}_{A_{n+1}} + \underbrace{\left(B_n^T\left(F - \Sigma\Lambda\right) - \mathbb{I}_r\right)}_{B_{n+1}^T} X_t \right\}. \tag{10}$$

where $A_1 = 0$ and $B_1 = -\mathbb{I}_r$.[10] Again, as with the Ang-Piazzesi model, the zero-coupon curve is a linear combination of yield-curve and macroeconomic state variables. Thus, with this approach, the macroeconomic variables can influence both the dynamics of the yield-curve state variables and the mapping between the state-variable vector and pure-discount bond prices.

This model is, in many respects, a reverse-engineering exercise. We found, after an enormous amount of optimizational effort, that the Ang-Piazzesi model did not appear to perform very well in out-of-sample forecasting analysis when applied to Canadian data.[11] We hypothesize that the principal issue was with the parametrization of state-variable dynamics. It appears that during the estimation algorithm, there is insufficient flexibility to simultaneously capture both the cross section and the dynamics of the term structure of interest rates. It appears to overweight the cross section and underweight the state-variable dynamics. By relaxing the existing restrictions on the state-variable dynamics and separating the estimation of the physical and pricing parameters, therefore, we hope to improve the model's out-of-sample forecasting performance. The observed-affine form, therefore, is really an *ad hoc* attempt to improve the description of these state-variable dynamics. We propose this model (i) from a practical perspective and (ii) as a direction for future research to improve this approach.

2.3 Some empirical models

The models covered in the previous two sections both involved describing the dynamics of the term structure of interest rates through an affine mapping of the state variables. Affine term-structure models have a number of theoretical and practical advantages. One of the principal advantages is the explicit description of market participants' aggregate attitude towards risk. This concept, captured by the Radon-Nikodým derivative in general and the market price of risk in particular, provides a clean and intuitive way for the analyst to understand deviations from the expectations hypothesis and simultaneously ensure the absence of arbitrage. In other words, one knows that the affine term-structure model is built upon a sound theoretical foundation.

In recent years, a new approach to describing term-structure dynamics has evolved. This methodology, introduced by Diebold and Li (2003), also assumes that the term structure of interest rates is a function of a set

[10] The details of the derivation of the observed-affine model are found in Appendix B.

[11] See Appendix A for a description of the specific optimization algorithm used to solve for the Ang-Piazzesi parameters.

of unobservable state variables. The key difference is that this approach does not have a theoretical foundation. It is essentially a time-series extension of a popular curve-fitting algorithm introduced by Nelson and Siegel (1987).

The lack of a solid theoretical foundation may appear, at first glance, to be a serious weakness. From a risk-management practitioner's perspective, however, the most important term-structure model criterion may be the ability of the model to adequately describe the empirical properties of the term structure.[12] To this end, Diebold and Li (2003) demonstrate that these models perform quite well relative to affine term-structure models. In recent work, Bolder (2006) also finds that these models compare quite favourably to the Gaussian class of affine term-structure models—the $A_0(3)$ models in the nomenclature of Dai and Singleton (2000)—in the context of Canadian data. There is, therefore, evidence that these models are a viable alternative, or supplement, to the traditional affine term-structure models. With the extension of term-structure models to include macroeconomic information, it is thus only natural that the Diebold and Li (2003) approach also be extended in this direction.

We consider the model suggested by Diebold, Rudebusch, and Aruoba (2004b). This paper has, as we understand it, two major contributions. First, it casts the so-called extended Nelson-Siegel model in a state-space form permitting estimation with a Kalman filter. Second, it extends the extended Nelson-Siegel model to include macroeconomic variables.[13] As this paper attempts to compare the range of models that combine macroeconomic and term-structure factors, it is important to examine this class of models in detail.

Let us begin with a brief description of the extended Nelson-Siegel model. The idea suggested by Nelson and Siegel (1987) is essentially a specific functional form of the instantaneous forward rate.[14] They suggested,

$$f(t,\tau) = x_0 + x_1 e^{-\lambda(\tau-t)} + x_2 \lambda(\tau-t) e^{-\lambda(\tau-t)}. \tag{12}$$

Exploiting the well-known fact that the continously compounded zero-coupon rate is related to the instantaneous forward rate as,[15]

$$z(t,\tau) = \frac{1}{\tau-t}\int_t^{\tau} f(t,u)du, \tag{13}$$

[12]To the extent that the model is being used to price contingent claims, of course, this will not be true.

[13]This is not the only paper that extends these models to include macroeconomic factors, but it is perhaps the most straightforward. Bernadell, Coche, and Nyholm (2005) also offer a competing model.

[14]The instantaneous forward rate is defined as,

$$f(t,\tau) = -\frac{P_{\tau}(t,\tau)}{P(t,\tau)}, \tag{11}$$

For more detailed discussion and derivation of the extended Nelson-Siegel model see Bolder (2006).

[15]One can understand this identity by recalling that the instantaneous forward rate is the marginal cost of borrowing over an infinitely short period of time, whereas the continuously compounded zero-coupon rate represents an average (hence the integral) of these marginal borrowing rates over a lengthier time period.

we can solve for the zero-coupon curve through some simple integration,

$$z(t,\tau) = \frac{1}{\tau - t}\int_t^{\tau} \left(x_0 + x_1 e^{-\lambda(s-t)} + x_2 \lambda(s-t)e^{-\lambda(s-t)}\right) ds, \tag{14}$$
$$= x_0 + x_1\left(\frac{1-e^{-\lambda(\tau-t)}}{\lambda(\tau-t)}\right) + x_2\left(\frac{1-e^{-\lambda(\tau-t)}}{\lambda(\tau-t)} - e^{-\lambda(\tau-t)}\right).$$

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 $z(\tau-t)$. The reason is that 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 as this is a static model. This is one of the main contributions of Diebold and Li (2003). Specifically, they noted that equation (14) is a linear combination of three functions with coefficients, x_0, x_1, and x_2. These functions are

$$f_0(y) = 1, \tag{15}$$
$$f_1(y) = \frac{1-e^{-\lambda y}}{\lambda y}, \tag{16}$$
$$f_2(y) = \frac{1-e^{-\lambda y}}{\lambda y} - e^{-\lambda y}. \tag{17}$$

Diebold and Li (2003) then proceeded to propose a model whereby the coefficients—which can be shown to resemble the level, slope, and curvature of the term structure—vary through time.[16] To see how this might work, we define our time-varying coefficients in matrix form as,

$$X_t = \begin{bmatrix} x_{0,t} \\ x_{1,t} \\ x_{2,t} \end{bmatrix}. \tag{18}$$

This suggestive notation is utilized because, in Diebold and Li (2003), the model coefficients are, in fact, the state variables. Thus, we have identified the first term-structure modelling assumption: the form of the state variables. We can now write equations (15) to (17) as,

$$F(\tau - t) = \begin{bmatrix} f_0(\tau-t) \\ f_1(\tau-t) \\ f_2(\tau-t) \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda y}}{\lambda y} \\ \frac{1-e^{-\lambda y}}{\lambda y} - e^{-\lambda y} \end{bmatrix}. \tag{19}$$

[16] Diebold and Li (2003) assumed that λ remains constant. This is because the λ parameter is a non-linear parameter and it is well documented as numerically unstable. While Diebold and Li (2003) suggest that λ be treated as a constant value, the representation of this model in state-space form and estimation with the Kalman filter in Diebold, Rudebusch, and Aruoba (2004b) solves this problem.

The model is easily extended to include macroeconomic variables by augmenting the state-variable vector in equation (18) as,

$$X_t = \begin{bmatrix} x_{0,t} & x_{1,t} & x_{2,t} & x_t & \pi_t & r_t \end{bmatrix}^T, \tag{20}$$

where, again, x_t, π_t, and r_t denote the output gap, consumer-price inflation, and the overnight rate at time t respectively. In other words, the state variables include, as in the Ang-Piazzesi and observed-affine approaches, both financial and macroeconomic elements. The factor loadings in equation (19) are also slightly modified as,

$$F(\tau - t) = \begin{bmatrix} f_0(\tau - t) & f_1(\tau - t) & f_2(\tau - t) & 0 & 0 & 0 \end{bmatrix}^T, \tag{21}$$

to indicate that the macroeconomic variables are not mapped into the term-structure of interest rates. Instead, the macroeconomic variables only assist in the description of the dynamics of the state variables. This is essentially the opposite of the Ang-Piazzesi approach. Indeed, one can now write the zero-coupon rate function at time t as the following linear function,

$$z(t, T) = F(T - t)^T X_t. \tag{22}$$

Here we have the third term-structure assumption, which describes the nature of the mapping between the state variables and zero-coupon rates.

We follow Diebold and Li (2003) and estimate this model in two steps. First, we use an optimization algorithm to fit equation (14) to a sequence of daily datasets of coupon bond data. For each date, we obtain three parameters (i.e., x_0, x_1, and x_2). At each time point, this vector is augmented with the values of x_t, π_t, and r_t. In the second step, we proceed to collect all of these parameter estimates together to generates a time-series of state variables: $\left\{ \begin{bmatrix} x_{0,t} & x_{1,t} & x_{2,t} & x_t & \pi_t & r_t \end{bmatrix}, t \in \{1, .., T\} \right\}$. The dynamics of these state variables, which is the second term-structure modelling assumption, are estimated with a VAR(2) specification.[17]

Following from Bolder (2006), we also consider two variations on the Diebold and Li (2003) approach. The first variation follows from the so-called exponential-spline methodology proposed by Li, DeWetering, Lucas, Brenner, and Shapiro (2001). 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,

$$P(t, T) = \sum_{k=1}^{n} \xi_k g_k(T - t), \tag{23}$$

[17] Diebold, Rudebusch, and Aruoba (2004b) suggest a Kalman-filter approach to the estimation of model parameters. We experimented with this technique and found that we obtained superior results with the two-step approach suggested in Diebold and Li (2003).

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

$$g_k(T-t) = e^{-k\alpha(T-t)}, \tag{24}$$

for $k = 1, ..., n$ and $\alpha \in \mathbb{R}$.[18] The parameter, α, can be interpreted as a long-term instantaneous forward rate. As with the λ parameter in the Diebold-Li setting, it is fairly reasonable to assume that α is approximately constant.

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

$$g_k(T-t) = \begin{cases} 1 : k = 1 \\ \sin\left(\frac{\frac{k}{2}(T-t)}{10}\right) : \quad \mathrm{mod}\ (k,2) = 0 \\ \cos\left(\frac{\frac{k-1}{2}(T-t)}{10}\right) : \quad \mathrm{mod}\ (k,2) = 1 \end{cases}, \tag{25}$$

for $k = 1, ..., n$. Note that the horizontal stretch factor $\frac{1}{10}$ was arbitrarily selected to extend the wave-length of each basis function and avoid excessive oscillation.

Equations (24) and (25) describe choices of $g_k(\cdot)$ for estimation of 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 (24). This is due primarily to the fundamental form of the pure-discount bond price function, which essentially has a negative exponential form.

We can now—liberally borrowing from the ideas of Diebold and Li (2003)—transform these models into a dynamic model for interest rates, by slightly adjusting equation (23),

$$P(t,T) = \sum_{k=1}^{n} \xi_{t,k} g_k(T-t). \tag{26}$$

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 $k = 1, ..., n$.[19] This idea was first introduced, in the absence of macroeconomic factors, by Bolder (2006). We also augment the state-variable vector and adjust the factor loadings in a fashion similar to that described in equations (20) and (21).

Clearly, neither the exponential-spline or Fourier-series 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 (97). Indeed, they

[18] 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.

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

require a substantially higher dimension to reasonably describe the yield curve. This dimensionality may prove problematic when we add the three additional macroeconomic state variables. Again, the work of Litterman and Schenkman (1991) reminds us that only three state variables are required to describe zero-coupon term-structure movements; we obviously use a substantially larger number of factors. The relative performance of these models, however, is ultimately an empirical question that will be addressed in Section 3.

The final model, which we term the OLS approach, is something of a strawman. While it is almost trivial, we feel that it is worth consideration by virtue of its extreme simplicity. Again, we can describe the model in the context of our three term-structure model assumptions. First, as with the observed-affine model, we use the first three principal components to describe the yield-curve related factors. All of the factors, therefore, are observable. The consequence is six state variables: three related to the yield curve and three macroeconomic variables (i.e., output, inflation, and monetary-policy rate). For the second assumption, the state-variable dynamics are estimated with a VAR(2) process as,

$$X_t = C + \sum_{i=1}^{2} F_i X_{t-i} + \nu_t, \tag{27}$$

where,

$$\nu_t \sim \mathcal{N}(0, \Omega). \tag{28}$$

The final assumption relates to the mapping between zero-coupon rates and the state variables. We assume the following linear form,

$$Z_t = A + HX_t + \epsilon_t, \tag{29}$$

where,

$$\epsilon_t \sim \mathcal{N}(0, \Xi). \tag{30}$$

Inspectation of equations (27) to (30) reveals that we have basically written out the term-structure system in state-space form. We do not, however, require a filtering approach to estimate the parameters as the state variables are observable. Consequently, we can estimate equation (27) and (29) using ordinary least-squares. The only restriction we have is that, given we have only six state variables, we can only use six zero-coupon rates to estimate the zero-coupon term structure.[20] We resort to linear interpolation for the intermediate points.

[20]In fact, we use three-month, two-, five-, seven-, ten-, and 15-year zero-coupon rates.

3 Results

This paper is, ultimately, a horserace between a number of competing macro-finance models. This is perhaps inevitable, given that the primary objective of this paper is to investigate, and hopefully identify, a macro-finance model for use in practical risk-management analysis. When holding a horserace, therefore, it is important that one is relatively explicit about the conditions required for good performance in the competition. In our case, fortunately, this is relatively straightforward. First, we desire models that are fairly simple to interpret, understand, and estimate. This is because well-designed sensitivity analysis is critical to the practical application and usage of risk-management models. As formulating sensitivity analysis with complex models is generally challenging, we have a preference for simple models. This condition, however, is hardly useful as a quantitative criterion for model comparison. We mention it because, when combined with other criteria, it is an important qualitative criterion that should not be ignored.

Second, a good macro-finance model should be able to provide reasonable out-of-sample forecasts of both the conditional expectation and variance of both interest-rate and macroeconomic outcomes across various forecasting horizons. One of the most common risk-management applications of term-structure modelling is stochastic simulation.[21] As stochastic simulation is essentially an out-of-sample forecasting exercise, it is essential that the model perform well out-of-sample.[22] Moreover, as pointed out by Duffee (2002) and Diebold and Li (2003), a model that generates poor out-of-sample forecasts is probably missing some important elements of term-structure dynamics.

Third, we require a certain degree of flexibility. In particular, as we will be estimating each model using a rather lengthy sample period with a variety of possible macroeconomic regimes, it is useful for a model to be able to, at least in a simple manner, handle non-time-homogeneous parameters. Indeed, as we will see shortly in the data, the Canadian economy over the last 30 years has been marked by a number of high-inflation episodes with attendant consequences for output and interest rates. Taking this into account will be of tremendous practical importance from a modelling perspective.

Armed with this understanding of our desired model criteria, the remainder of this section turns its attention to a comparison of our six term-structure models. Before the actual comparison can occur, however, we will briefly discuss the data used in this analysis and comment on a some of the specifics of the model implementation. The actual results can be divided into two parts. First, we examine each of these models based on a number of out-of-sample forecasting criteria with a constant-parameter assumption. This permits a clean comparison of the models in their simplest form. Second, we introduce a very simple approach to allow a subset of the model

[21] See Bolder (2003) for a detailed example of such a risk-management application.

[22] Bolder (2006) provides this type of analysis in the context of term-structure models without macroeconomic variables.

parameters to vary through time. We then proceed to examine the implications of this choice for our models.

3.1 The data

Let us now turn our attention to the data used for the analysis of our six models. We have 391 monthly datapoints from January 1973 to July 2005. The zero-coupon curves, outlined in Figure 1, are constructed using a nine-factor exponential-spline model with an independently estimated non-linear parameter.[23]

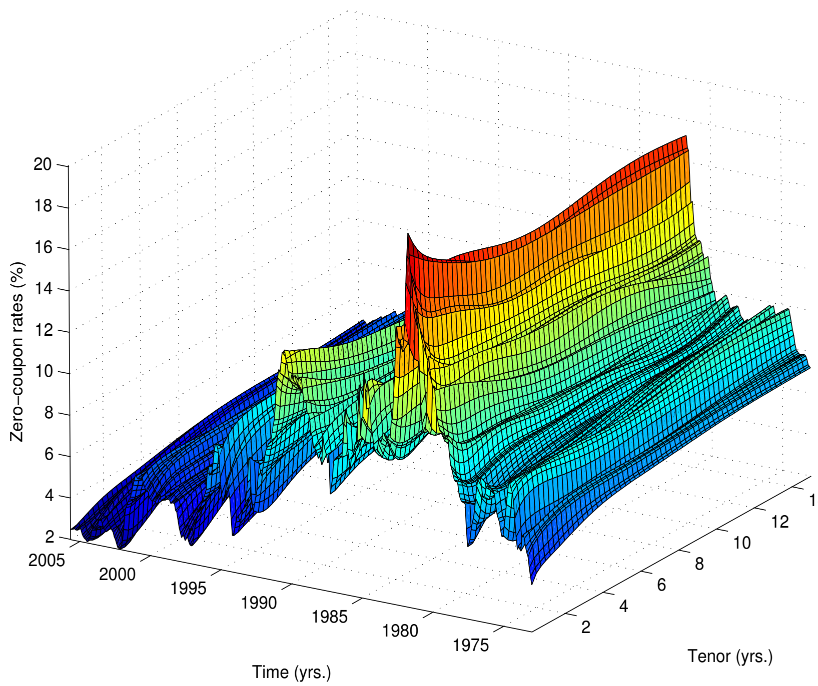

Figure 1: **The Zero-Coupon Data**: This figure illustrates the zero-coupon data ranging from January 1973 to July 2005. It was estimated from monthly Government of Canada treasury-bill and nominal bond prices with a nine-factor exponential-spline model described in Bolder and Gusba (2002).

We can identify periods of quite high interest rates during the early 1980s and early 1990s. Moreover, we observe periods of flat, inverted, and steep zero-coupon term structures over this 32-year period. Clearly, it will be something of a challenge for any model to handle what appear to be a number of different zero-coupon rate regimes. The macroeconomic data, as outlined in Figure 2, serve to underscore this difficulty. These macroeconomic time series appear to be constructed from two or more different regimes. There appears, for example, to be a high inflation regime associated with high short-term interest rates while there also appears to be a low-inflation regime associated with low short-term interest rates. Finally, one could characterize a third

[23]For more detail about this model, see Bolder and Gusba (2002).

regime describing the transition between the high- and low inflation regimes. Indeed, Demers (2003) constructs a Markov-switching model that identified three such regimes in the relationship between Canadian inflation and output.

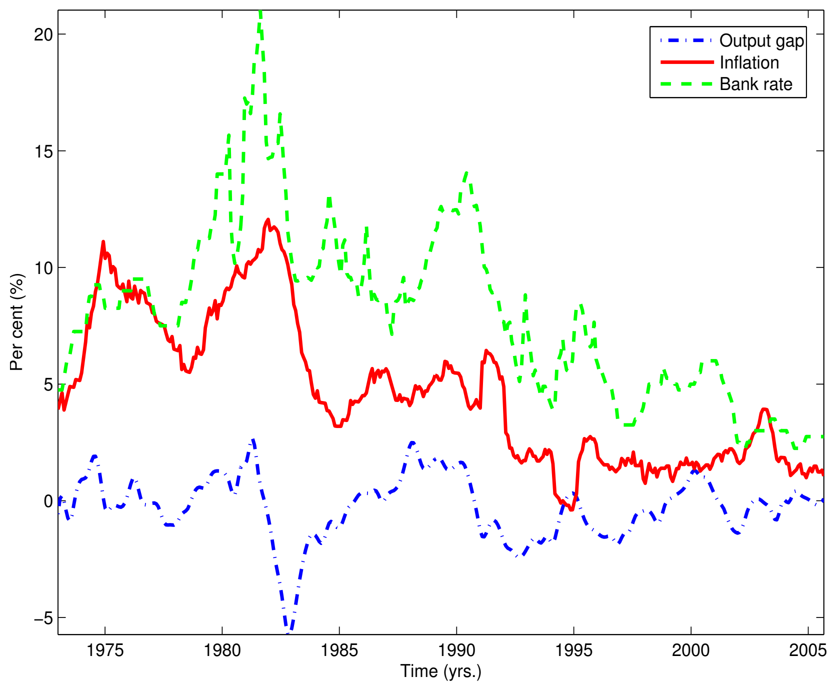

Figure 2: **The Macroeconomic Data**: This figure illustrates the macroeconomic data ranging from January 1973 to July 2005 including the output gap, the annual inflation rate, and the monetary-policy rate respectively. The output gap was computed by Bank of Canada staff following the methodology described by Butler (1996).

In the following analysis, we first restrict our attention to the constant-parameter version of these models. As most of the analysis is based on out-of-sample forecasting performance, with a rolling sample size, we felt this was a fair approach for comparison of the models. The constant-parameter assumption, however, becomes problematic when we actually attempt to use these models in a risk-management setting. There are, in fact, two problems. First, we can expect the parameter estimates, over the entire sample, to do a relatively poor job of characterizing future outcomes. This is because it will essentially represent an average across the various inflationary regimes.[24] This could be solved by restricting our parameter estimation to the most recent years that, many would argue, represent a period of stable inflation associated with well-anchored inflationary expectations. This is clearly an option, but it raises the second problem. We would like, in the context of our simulation model,

[24]It also leads to the Lucas critique as policy changes over our data interval will not be adequately represented in the model parameters.

to incorporate, at least, the possibility of a high-inflationary regime. While this, in the current period, is quite a low-probability event, it is nonetheless important to consider its inclusion from a risk-management perspective. We attempt to solve this problem by examining a fairly straightforward approach—borrowing from the work of Demers (2003)—to incorporating time-varying parameters.

3.2 Some model implementation specifics

As previously discussed, we consider six alternative term-structure models. These include the Ang-Piazzesi, OLS, observed-affine, Nelson-Siegel, exponential-spline, and Fourier series models. There are a number of practical details regarding the model implementation that must be mentioned. First, the Ang-Piazzesi already includes an explicit notion of a short-term interest rate. As a consequence, we use only the output gap and the inflation rate as observable macroeconomic state variables in the implementation of this model. Incorporation of the monetary-policy rate is not required and, indeed, may create problems.[25] The empirically motivated Nelson-Siegel, exponential-spline, Fourier series and OLS approaches as well as the observed-affine model include the three macroeconomic variables outlined in Figure 2.

Table 1: **In-Sample Model Fit**: In this table, we provide a view of the in-sample root-mean-squared error fit, in basis points, of each of our models to the actual zero-coupon term-structure data ranging from January 1973 to July 2005.

Tenor	**Nelson Siegel**	**Exponential Spline**	**Fourier Series**	**Ang-Piazzesi**	**OLS**	**Observed-affine**
3 months	37.65	30.85	38.77	21.72	26.17	46.40
6 months	29.05	26.21	36.20	0.00	19.39	58.85
1 year	27.32	21.86	24.24	10.98	27.67	47.78
2 years	22.46	21.56	21.46	0.00	10.60	32.61
5 years	23.40	18.40	18.48	19.06	7.45	24.27
7.5 years	21.00	18.19	18.31	17.13	9.50	36.34
10 years	20.46	19.92	19.29	0.00	8.77	37.77
15 years	30.24	24.75	19.18	26.72	13.98	34.24

Table 1 provides a summary of the in-sample cross-sectional fit of each of the various models to the zero-coupon data in our sample. The values presented represent the average root-mean-squared error differences between the model zero-coupon curves and the true zero-coupon curves over the 391-month sample period. The Ang-Piazzesi model, estimated with the maximum-likelihood approach, assumes that three zero-coupon rates are observed without error—these are, in our implementation, the six-month, two-year and ten-year zero-coupon

[25] Moreover, we already have substantial difficulty in estimating the parameters associated with this model. Addition of another state variable would add to the dimensionality and make the optimization algorithm even more cumbersome.

rates.[26] The in-sample fit of the Ang-Piazzesi model is, with the exception of the long end of the term structure, generally superior to the in-sample fit exhibited by the four empirical models. This appears to suggest that, even with the addition of the macroeconomic factors, the Ang-Piazzesi model does a good job of describing the cross section of zero-coupon rates. Note that the observed-affine model demonstrates a cross-sectional in-sample fit that is inferior, across the entire term structure, to each of the other five models. The reason is that, as the yield-curve factors are exogeneously provided, the model does not have the flexibility to adjust the state variables to improve the fit to the zero-coupon cross section. The model may only adjust the market price of risk parameters to fit the zero-coupon term structure, which leads to less accurate in-sample description. This is not a surprise, however, as this feature stems from the construction of the model.

The empirical term-structure models, with the exception of the OLS model at intermediate and long-term tenors, appear to demonstrate inferior in-sample fit relative to the affine models. This is particularly evident at the short end of the term structure, where the Nelson-Siegel and Fourier-series models underperform, on average, by as much as 11 basis points. This improves somewhat at the long end of the term structure where the exponential-spline and Fourier series models actually demonstrate a slightly better in-sample fit than the affine term-structure models.

Figure 3 shows the in-sample fit of our six macro-finance term-stucture models from another perspective. Here we show the root-mean-squared fit of each model to the entire zero-coupon yield curve for each individual datapoint in our sample.[27] The top graphic shows the Nelson-Siegel, exponential-spline, and Fourier-series models, while the lower graphic illustrates the OLS, Ang-Piazzesi, and observed-affine models. The Nelson-Siegel, exponential-spline, and Fourier-series models demonstrate substantial volatility in the in-sample fit, with the Nelson-Siegel model appearing to have more difficulty from the late 1970s to the late 1980s. The Ang-Piazzesi and OLS models track one another quite closely and, with the exception of the difficult period in the late 1970s and early 1980s, demonstrate an admirable fit to the zero-coupon data. The observed-affine model has trouble at discrete intervals across virtually the entire period. This is consistent with the results in Table 1 and supports the notion that the observed-affine model is less flexible in its ability to fit the cross section of zero-coupon rates.

Figure 4 outlines the state variables for each of our models. This graphic provides some insight into the time-series properties of the state variables associated with our various models. The Ang-Piazzesi and Nelson-Siegel models each have *three* latent term-structure state variables, while the exponential-spline and Fourier series were each constructed with *seven* state variables. It is admittedly difficult to extract much insight from the top two graphics in Figure 4; we do observe, however, that the high-inflation period appears to be characterized by

[26] The full estimation procedure for the affine term-structure models is described in Appendix A.

[27] We report the root-mean-squared error difference between 30 points on the true and model zero-coupon curves. One can think of this measure as essentially an average distance between the two curves across all zero-coupon tenors.

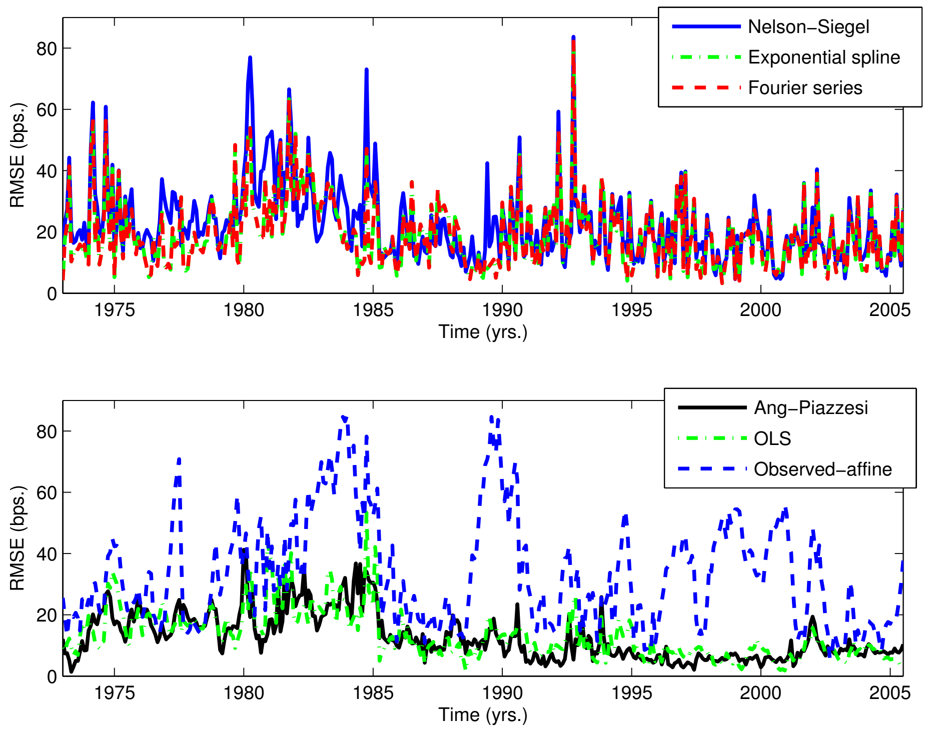

Figure 3: **The Evolution of In-Sample RMSE**: This figure displays in-sample root-mean-squared error fit of each of our six macro-finance term-structure models to the actual zero-coupon term structure of interest rates.

rather greater volatility relative to the latter periods. Clearly, the sheer dimensionality of the exponential-spline and Fourier-series models is unconventional. The yield-curve-related state variables for the OLS and observed-affine models are, as previously indicated, merely the first three principal components, which are found in the bottom graphic of Figure 4. As Diebold and Li (2003) point out, the Nelson-Siegel factors and the first three principal components of the zero-coupon term structure are highly correlated. For our data sample, this is verified insofar as we find a contemporaneous correlation between the Nelson-Siegel factors and the first three principal components of 0.918, 0.803, and -0.756, respectively.[28]

[28]The first three principal components are commonly, following from Litterman and Schenkman (1991), interpreted as level, slope, and curvature. We can proxy level, slope, and curvature as the ten-year zero-coupon rate, the difference between the ten-year and one-year zero-coupon rate, and twice the two-year zero-coupon rate less the sum of the ten-year and three-month rate. The contemporaneous correlation between these proxies and the first three principal components is 0.989, -0.851, and -0.790, respectively. This would suggest that we could presumably also use these simply constructed proxies in our analysis with similar results.

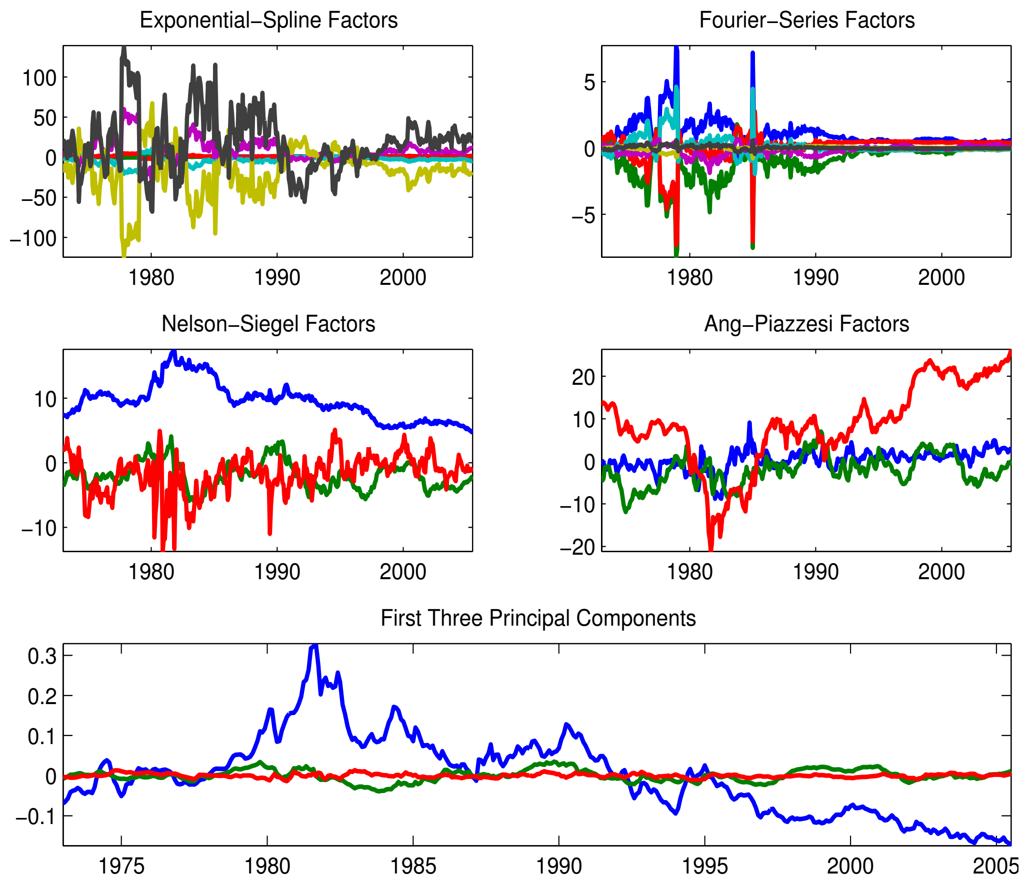

Figure 4: **The Model State Variables**: This figure displays state variables, or factors, associated with each of our six term-structure models. Note that the OLS and observed-affine model both make use of the first three principal components of zero-coupon term-structure movements.

3.3 Out-of-sample comparison

We are now ready to actually compare our six term-structure models. The comparison focuses on out-of-sample forecasting ability and occurs along the following four different dimensions:

Entire Curve Out-of-Sample Fit This measure describes the out-of-sample forecasting performance of the entire zero-coupon curve, where we select 30 evenly spaced tenors from three months to 15 years to represent the entire curve.

Individual Tenor Out-of-Sample Fit This is essentially the same as the previous measure, although the out-of-sample performance is organized by individual tenor to provide an alternative perspective on the results. A focus on the entire curve is useful, but it can hide under- or overperformance in different parts of the term structure.

Hit Ratio This measure denotes the proportion of zero-coupon forecasts that fall into a 95% confidence interval around the base forecast.[29] We would expect, for example, that for every 100 out-of-sample forecasts that only approximately five forecasts should fall outside of a 95% confidence interval. If none fall outside the confidence interval this might suggest an overestimate of the conditional variance, while if 20 forecasts fall outside the confidence interval we might suspect an underestimate of the conditional variance. This measure is a simple attempt to examine how well the models characterize conditional variance in addition to the conditional expectation generally examined in the literature.

Out-of-Sample Excess Holding-Period Return Forecasts Interpreting the performance of a zero-coupon rate forecast is not always straightforward. Is, for example, a 20 basis-point error large or small? For this reason, we use our six models to forecast excess holding period returns. Our hope is that this alternative perspective provides some additional insight into the absolute and relative performance of these models.

The out-of-sample forecasting begins in October 1987, which implies that the first forecasts are made using 178 months of data. After each set of forecasts, another month of data is added to the dataset, the model parameters are re-estimated, and a new set of forecasts are constructed. This sequence of operations is repeated until the end of the sample period. The consequence is 212 non-overlapping one-month forecasts, with 207 six-month, 201 12-month, 189 24-month, and 153 60-month overlapping forecasts respectively. Table 2 illustrates the out-of-sample root-mean-squared and mean-absolute forecast errors for each of our six macro-finance term-structure models at one-, six-, 12-, 24-, and 60-month forecasting horizons. Note that, in all cases, our six term-structure models are compared against the random-walk assumption. This is a useful benchmark because of its simplicity and also due to the fact that it is generally quite difficult to beat.[30] Table 2 compares the out-of-sample of the entire yield curve and includes a number of summary statistics regarding the forecasting performance such as mean, median, minimum, maximum, and standard deviation.

[29] The mechanics of constructing these conditional zero-coupon curve confidence intervals is described in Appendix D.

[30] The random-walk assumption essentially postulates that zero-coupon rates are martingales. That is, the conditional expectation of future zero-coupon rates, for all forecasting horizons, is the *current* zero-coupon curve.

Table 2: **Out-of-Sample Zero-Coupon Curve Forecasts**: In this table, we present the root-mean squared and mean-average error for a series of one-, six-, 12-, 24-, and 60-month zero-coupon curve forecasts for the entire term-structure for each of our six macro-finance term-structure models. Figures in bold represent an out-performance of the random-walk assumption.

Models	RMSE					MAE				
	Mean	Median	Max	Min	SD	Mean	Median	Max	Min	SD
One-month forecast										
Random walk	23.46	20.23	86.05	2.36	15.26	21.28	18.33	71.65	1.75	14.47
Nelson-Siegel	24.10	20.33	92.36	5.68	13.54	21.30	17.67	88.04	4.51	12.97
Exponential spline	**22.97**	**19.58**	94.32	3.86	14.46	**20.45**	**17.20**	87.85	2.57	13.62
Fourier-series	25.04	21.86	102.53	3.58	15.27	21.83	18.68	93.90	2.93	13.85
Ang-Piazzesi	25.65	21.83	85.58	3.94	15.57	22.92	18.81	66.47	2.89	14.72
Observed affine	30.89	29.85	80.10	9.32	13.87	26.37	25.30	71.87	6.19	12.37
OLS	**23.20**	20.29	88.96	3.36	13.87	**20.48**	**17.20**	76.72	2.80	13.12
Six-month forecast										
Random walk	74.98	64.60	307.09	9.29	46.84	67.03	56.09	303.34	8.23	46.10
Nelson-Siegel	**72.45**	65.34	296.50	11.86	41.85	**66.28**	57.72	294.71	9.46	41.32
Exponential spline	**69.19**	**63.18**	301.74	8.00	42.62	**62.37**	**55.35**	298.83	5.86	41.62
Fourier-series	**70.76**	**61.21**	309.73	11.71	43.96	**63.63**	**54.31**	306.56	9.47	42.85
Ang-Piazzesi	80.40	71.73	262.21	5.99	48.19	74.34	62.61	259.78	4.57	48.47
Observed affine	75.84	66.33	326.05	13.56	43.54	68.63	60.12	323.72	10.86	42.81
OLS	**73.03**	**63.65**	328.66	13.66	44.99	**66.68**	56.92	326.24	10.78	44.25
12-month forecast										
Random walk	110.96	95.61	346.24	13.39	68.16	99.57	81.70	337.50	10.00	64.75
Nelson-Siegel	111.46	101.54	310.60	10.59	62.24	101.82	93.53	308.10	8.76	60.63
Exponential spline	**102.52**	**94.84**	306.38	15.18	55.83	**93.18**	84.90	301.45	11.31	54.26
Fourier-series	**102.77**	**91.39**	334.74	8.09	58.27	**93.34**	83.23	328.14	5.81	56.80
Ang-Piazzesi	129.08	125.72	335.02	13.91	61.91	119.84	117.34	324.48	10.22	60.38
Observed affine	117.38	112.27	354.87	15.04	62.83	107.45	101.57	343.01	11.90	61.85
OLS	115.58	110.82	360.49	8.93	65.16	105.62	100.48	346.83	7.91	63.47
24-month forecast										
Random walk	156.87	140.23	487.90	21.21	95.28	142.54	123.84	451.76	15.36	90.06
Nelson-Siegel	186.66	181.05	377.68	14.19	86.24	174.48	172.77	357.58	12.75	84.35
Exponential spline	163.16	156.73	403.08	12.27	89.56	151.21	145.36	381.22	9.89	87.81
Fourier-series	161.39	155.15	366.91	18.29	85.66	149.47	137.64	348.78	14.88	83.78
Ang-Piazzesi	193.72	197.64	458.93	8.94	110.78	179.74	186.17	443.36	7.89	106.84
Observed affine	193.99	193.82	406.94	37.20	85.95	182.05	184.71	386.41	32.26	84.90
OLS	193.28	191.69	416.96	17.26	87.68	181.10	181.71	393.30	14.45	85.54
60-month forecast										
Random walk	223.55	206.75	521.97	65.52	98.52	207.93	189.42	492.87	57.13	96.86
Nelson-Siegel	349.48	351.35	515.68	171.61	77.72	342.98	337.86	509.84	170.03	76.03
Exponential spline	319.30	320.00	521.81	35.50	79.18	312.10	315.68	514.54	24.42	78.28
Fourier-series	310.85	303.74	525.24	45.21	74.45	303.47	292.47	518.32	40.42	72.86
Ang-Piazzesi	367.56	383.10	522.44	90.72	93.65	346.27	361.17	512.92	51.26	98.27
Observed affine	353.99	357.30	512.26	177.21	78.10	347.56	349.63	507.12	172.05	76.66
OLS	356.19	358.59	514.87	176.01	78.16	349.25	349.73	508.32	174.51	76.35

A variety of observations can be made from Table 2. First, we note that the exponential-spline and Fourier-series models exhibit the best performance across all forecasting horizons. In particular, we observe that the exponential-spline model beats the random-walk assumption at one-, six-, and 12-month forecasting horizons; this result applies equally to the root-mean-squared and mean-absolute error measures. The Fourier-series model exhibits more difficulty at the one-month horizon—as it does not succeed in beating the random walk—but exhibits the strongest performance across all models at the 24- and 60-month forecasting horizons. These results for the exponential-spline and Fourier-series models are generally robust to the choice of error measure (i.e., root-mean-squared mean-absolute error) or forecasting statistic (i.e., mean or median). The OLS model also outperforms the random-walk hypothesis at the one- and six-month forecasting horizons, although its relative performance deteriorates as we lengthen the forecasting horizon. We should note, however, that none of the empirically motivated models succeed in beating the random-walk approach at these longer forecasting horizons. Indeed, all of the models underperform the random-walk assumption by a range of approximately 90 to 145 basis points. Moreover, the forecasting performance deteriorates as the forecasting horizon increases.

The next observation from Table 2 relates to the Ang-Piazzesi model. At the one-month forecasting horizon, the Ang-Piazzesi model posts results that are within a few basis points of the other models. As we increase the forecasting horizon, however, we note that the Ang-Piazzesi model is the worst performer in all cases, with the exception of the 24-month horizon. In this case, it demonstrates slightly better results than the observed-affine model. The Ang-Piazzesi model does not, at any forecasting horizon, succeed in beating the random-walk assumption. Moreover, the Ang-Piazzesi model seems to demonstrate the least consistency in out-of-sample forecasts as evidenced by generally exhibiting the largest standard deviation of both root-mean-squared and mean-absolute forecasts.

The results for the observed-affine model are somewhat different. By relaxing the dynamics of the Ang-Piazzesi model, we find that the forecasting performance is improved from five to 14 basis points; the only exceptions are the one- and 24-month forecasting horizons where the observed-affine model underperforms the Ang-Piazzesi model by about 5 and 0.5 basis points, respectively. This is quite an accomplishment considering the relatively poor in-sample fit of the observed-affine model to the zero-coupon term structure. Nevertheless, this is hardly an indictment of the Ang-Piazzesi model. It merely suggests that permitting interaction between the macroeconomic and latent yield-curve state variables allows for better out-of-sample forecasting performance. It would, in fact, be rather a surprise where this not to be the case. This improved out-of-sample forecasting performance, we should recall, comes at the price of losing some flexibility in the ability of the model to fit the cross section of zero-coupon rates.

Figure 5 illustrates the evolution of the root-mean-squared out-of-sample forecasting fit to the entire curve at a 12-month forecasting horizon across the entire 16.75-year forecasting period. Each separate graphic in Figure 5

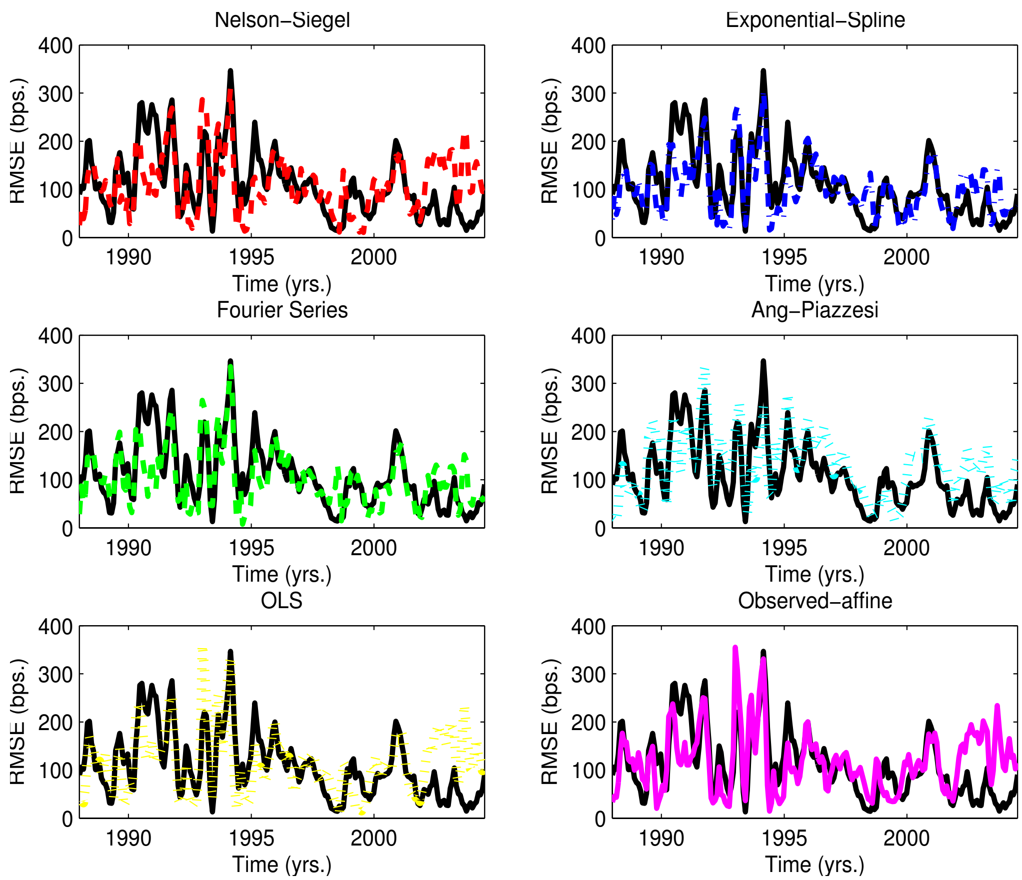

Figure 5: **The Evolution of RMSE for Each Model for a Short Forecasting Horizon**: This figure displays the evolution of the entire-curve RMSE across each of our models, with the incorporation of macroeconomic factors, for a 12-month forecasting horizon. The solid black line in each graphic represents the random-walk forecasting assumption.

represents one of our six term-structure models, while the black line in each graphic represents the random-walk forecast error. It appears that all of the models—although slightly less for the exponential-spline and Fourier-series approaches—have difficulty with the period from late 2001 to mid-2005. Moreover, all models appear to outperform the random-walk assumption during the early 1990s. With these two exceptions, the models tend to track the random-walk forecast errors quite closely, which again underscores the difficulty of outperforming this simple assumption.

Table 2 reveals quite clearly that the out-of-sample forecasting performance of *all* the models deteriorates quite substantially with the forecast horizon. This is somewhat worrisome, since we would expect that the mean-reversion characteristics in all of these models would overwhelm the random-walk hypothesis over longer time horizons. The problem, in our opinion, relates to the changes in interest rates over our out-of-sample period.

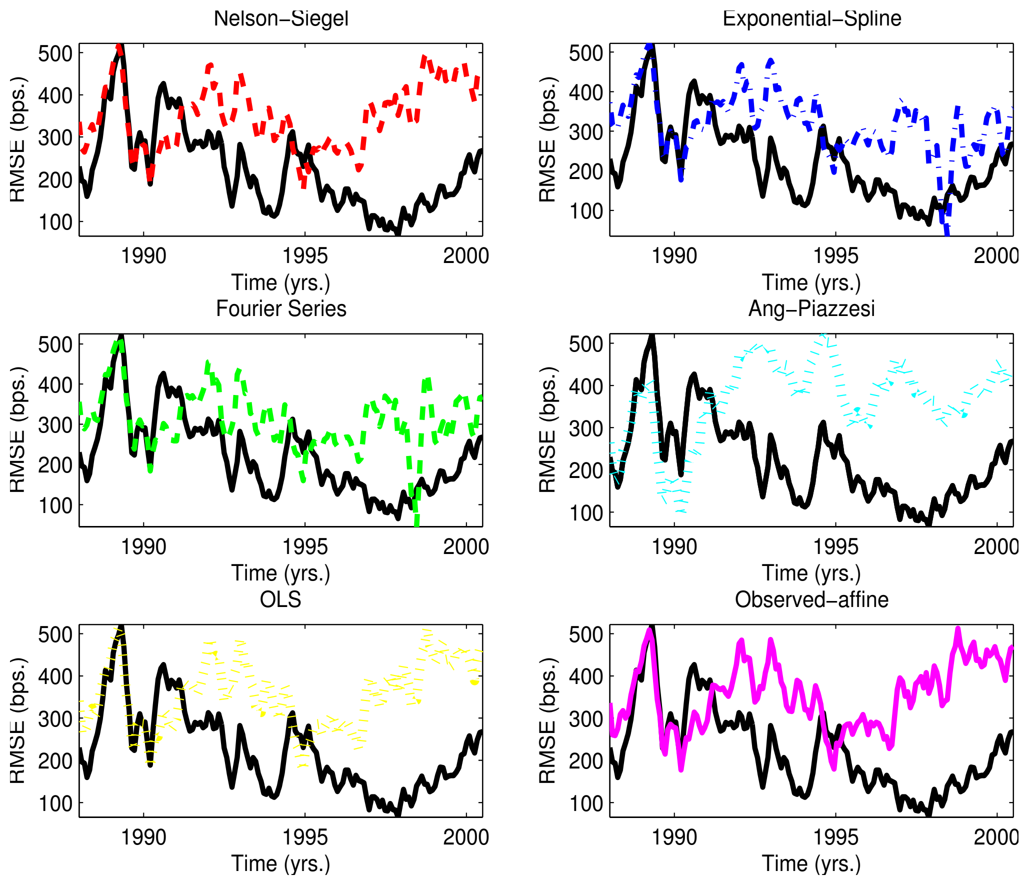

Figure 6: **The Evolution of RMSE for Each Model for a Long Forecasting Horizon**: This figure displays the evolution of the entire-curve RMSE across each of our models, with the incorporation of macroeconomic factors, for a 60-month forecasting horizon. The solid black line in each graphic represents the random-walk forecasting assumption.

Over the period of 1990 to 2003 there was an extended period of decreases in interest rates. In the early part of this period, there was a fairly unexpected decrease in interest rates. Our models, estimated with historical data on a rolling basis, did not expect such decreases in interest rates and correspondingly overestimated interest rates over a longer time horizon. The random-walk model was faster to predict interest-rate decreases to the extent that actual interest rates declined; this is faster than what was predicted by the model using historical data with relatively high interest-rate levels. Figure 6 supports this explanation insofar as all of the models appear to track the random-walk, in terms of root-mean-squared forecast error, until about 1990 when they begin to dramatically underperform the random-walk forecasts.

Figure 7 graphically illustrates the out-of-sample root-mean-squared forecast error by individual zero-coupon tenor, ranging from one month to 15 years, for each model across 12-, 24-, and 60-month forecasting horizons.

As before, the solid black line represents the random-walk model. At the 12-month horizon, we can see that the exponential-spline and Fourier-series models outperform the random-walk assumption across all tenors albeit by a greater margin at the shorter zero-coupon tenors. The Nelson-Siegel model appears to track the random-walk quite closely across all tenors while the OLS model, more or less, evenly underperforms the random-walk hypothesis across all tenors. The Ang-Piazzesi model, conversely, does a reasonable job at the shortest tenor, but has difficulty at intermediate and long tenors. In short, the Ang-Piazzesi model appears to consistently post larger errors than all of the other models across virtually all zero-coupon tenors. Interestingly, the observed-affine model tracks the random-walk quite closely out to about the five-year tenor, then deteriorates until the 12-year tenor, and subsequently improves to about the same level as the random-walk at the longest tenor.

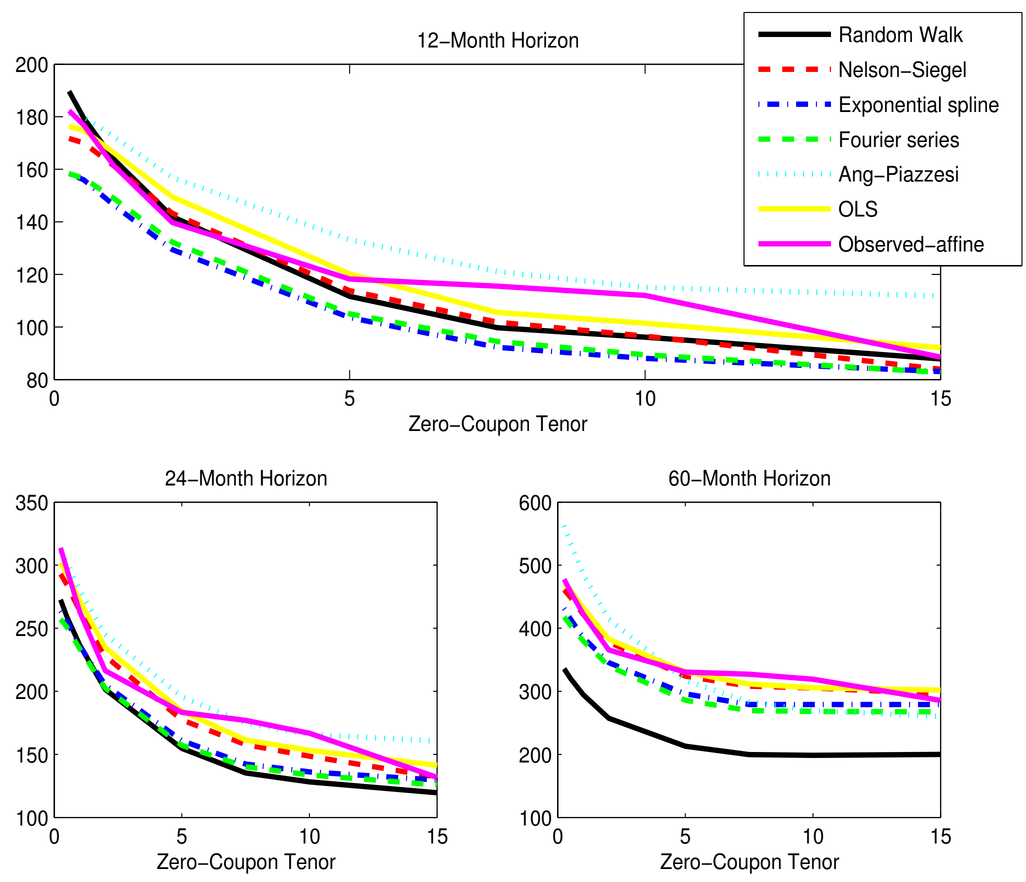

Figure 7: **Zero-Coupon Forecasting Performance by Tenor**: This figure displays the out-of-sample forecasting performance of our macro-finance term-structure models by individual zero-coupon tenor from one month to 15 years. We consider 12-, 24-, and 60-month forecasting horizons.

At the 24-month forecasting horizon, the ordering of the models is quite similar, although they all appear shifted to the right. That is, none of the models appear consistently capable of beating the random-walk approach. Again, the exponential-spline and Fourier-series models show the best performance followed by the Nelson-Siegel, OLS, and Ang-Piazzesi models, respectively. Once again, the observed-affine model demonstrates difficulty in forecasting the intermediate zero-coupon tenors; this behaviour may be related to its relative difficulty in describing the cross section of zero-coupon rates. While the Ang-Piazzesi model performs similarly to the observed-affine model in aggregate, the observed-affine model outperforms the Ang-Piazzesi model at moderate and long-term tenors. Finally, all of the models perform quite poorly at the 60-month forecasting horizon. Again, the ordering of the models is, more or less, maintained. We do note, however, that the Ang-Piazzesi model appears to have substantially more difficulty than the other models at the shorter tenors and that, at the longer tenors, the Nelson-Siegel, OLS, Ang-Piazzesi, and observed-affine models are generally indistinguishable. Again, the strongest performers are the exponential-spline and Fourier-series models.

To this point, we have focused on how well the various models perform in an out-of-sample forecasting exercise. Such an exercise essentially provides information on how well the models describe the first moment of the state-variable transition density; that is, models that do a good job of describing the conditional expectation of the state-variable vector will typically do a good job in an out-of-sample forecasting exercise. In a risk-management setting, we are also quite interested in the higher moments of the transition density. As all of these models are Gaussian, it does not make much sense to examine skewness or kurtosis, but we are nonetheless quite interested in understanding how well our six term-structure models describe the conditional variance of the term structure of interest rates.

To examine the accuracy of the conditional variance forecasts, we borrow an idea that is similar in spirit to Diebold and Mariano (1995). The idea of Diebold and Mariano (1995) is to compare the forecasting accuracy of two competing forecasts by taking into account the conditional variance of the forecasts. Our variation on this idea is somewhat different. We construct a 95% confidence interval for each out-of-sample zero-coupon forecast; a detailed description of how these confidence intervals are constructed is found in Appendix D. For each forecast, we then proceed to determine whether or not the actual outcome lies inside this confidence interval. If it does, then we consider this a "hit." We term the number of hits relative to the total number of forecasts as the *hit ratio*. One would expect, therefore, that if we repeated this 100 times, and the conditional variance was well specified, then we would observe approximately 95 hits or, equivalently, a hit ratio of 0.95. A hit ratio of, say, 0.60 would presumably indicate an underestimate of the zero-coupon conditional variance, while a hit ratio of 1.00 could be interpreted as a conditional-variance overestimate.

We freely admit that this is something of a primitive statistic as it will not help us distinguish between hit ratio of, say, 0.94 and 0.96. The idea, however, is not to make a definitive statement, but to better understand the distinctions between the models with respect to their characterization of the conditional variance of term-structure movements. In this respect, the hit ratio is more akin to the type of tests used in assessing different Value-at-Risk measures. Often different portfolio variance specifications are tested on their ability to generate historical Value-at-Risk values that are close to their supposed cut-off values (i.e., 99%).

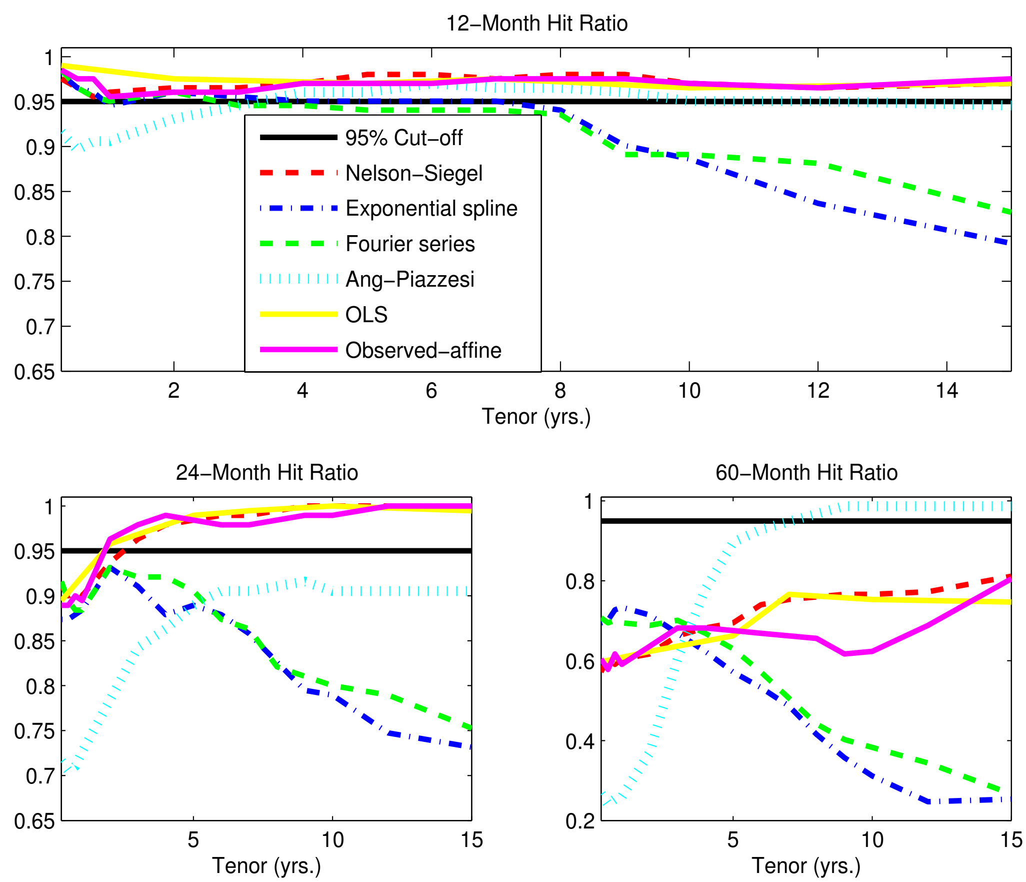

Figure 8: **<u>The Hit Ratio</u>**: This figure displays the hit ratio for a variety of tenors ranging from one month to 15 years across one-, two-, and five-year forecasting horizons for our six alternative macro-finance term-structure models. Note that the solid black line represents the 95% cut-off level for the confidence interval constructed by the hit-ratio measure.

Figure 8 illustrates the hit ratio for each of our six models at the 12-, 24-, and 60-month forecasting horizon. The solid horizontal black line denotes the 0.95 hit ratio that we would expect to observe from an ideal model. The results are quite interesting. At the 12-month forecasting horizon, we observe that all the models demonstrate a hit ratio in the range from 0.90 to about 0.98 until the zero-coupon tenor attains approximately nine years. At this point, the exponential-spline and Fourier-series models hit ratios fall almost linearly to about 0.80 at the 15-year tenor. This suggests that we can conclude that, at the 12-month forecasting horizon, these two models underestimate the conditional variance of zero-coupon rates at the long-end of the term structure. The remaining models, conversely, appear to do a relatively good job of describing the conditional variance for all tenors at this forecasting horizon.

At the 24-month forecasting horizon, the exponential-spline and Fourier-series models' hit ratio deteriorates from a reasonable level at the two-year tenor to approximately 0.75 at the longest zero-coupon tenor. The deterioration, however, begins somewhat earlier at around the five-year tenor. Once again, it appears that these two models underestimate the conditional variance of the longer part of the yield curve. The Ang-Piazzesi model also appears to have difficulty. In particular, the hit ratio increases from 0.70 at the short end to approximately 0.90 at approximately the five-year tenor. This suggests that, at the 24-month forecasting horizon, the Ang-Piazzesi model substantially underestimates the conditional variance of short-term tenors, but only slightly underestimates the conditional variance of medium- to long-term zero-coupon rates. The Nelson-Siegel, OLS, and observed-affine models follow one another closely; at very short tenors, they appear to slightly underestimate the conditional variance, while at longer-term tenors, they would appear to overestimate the conditional variance with a hit ratio of approximately unity.

Table 3: **Excess Holding-Period Return Forecast Errors**: In this table, we present the annualized root-mean squared holding-period forecast errors for our six term-structure 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. Note that RW, NS, ES, FS, AP, OLS, and OA denote the random-walk, Nelson-Siegel, exponential-spline, Fourier-series, Ang-Piazzesi, OLS, and observed-affine models, respectively. Bold values denote an out-performance of the random-walk assumption.

Tenor	**RMSE**						
	RW	**NS**	**ES**	**FS**	**AP**	**OLS**	**OA**
12-month forecast							
2 years	1.79	**1.73**	**1.58**	**1.61**	1.87	1.79	**1.71**
5 years	5.22	5.22	**4.78**	**4.83**	6.09	5.45	**5.20**
7.5 years	7.43	**7.42**	**6.80**	**6.93**	8.96	7.73	7.90
10 years	9.62	**9.58**	**8.78**	**8.94**	11.62	9.95	11.35
15 years	14.22	**13.41**	**13.29**	**13.18**	17.59	14.57	16.62
24-month forecast							
5 years	6.69	7.37	6.77	**6.65**	8.03	7.59	7.65
7.5 years	10.45	11.53	10.68	**10.44**	12.81	11.88	12.54
10 years	13.79	15.34	14.23	14.03	17.16	15.60	17.19
15 years	22.05	23.24	22.85	22.35	27.37	24.10	24.96
60-month forecast							
7.5 years	10.14	13.66	12.90	12.62	14.66	13.87	15.17
10 years	18.15	24.81	23.66	22.76	24.76	25.27	28.42
15 years	36.78	49.55	47.67	45.78	45.89	49.38	56.68

All of the models exhibit substantial difficulty with the conditional variance of the zero-coupon curve at the 60-month forecasting horizon. The exponential-spline and Fourier-series models demonstrate a hit ratio that ranges from approximately 0.70 at the short end to about 0.30 at the long end of the term structure. Clearly, this is evidence of a fairly dramatic underestimate of the conditional variance across the entire zero-coupon curve. The Nelson-Siegel, OLS, and observed-affine models exhibit a hit ratio increasing from 0.60 at the shortest tenor to roughly 0.80 at the longest zero-coupon tenor. Again, this appears to suggest a significant underestimate of the conditional variance of the zero-coupon curve, particularly at the short end. Finally, the Ang-Piazzesi hit ratio exhibits rather erratic behaviour. It roughly follows a logistic curve ranging from 0.25 to 1 over the 15-year range of tenors. This implies a dramatic underestimate at the short-end and an overestimate at the long end. Moreover, it begs a comparison with the comparatively more stable observable-affine model. We suspect that the variance-covariance matrix for the state-variable vector in the Ang-Piazzesi model is misspecified given, by virtue of the estimation algorithm, the values in this matrix have adjusted to better fit the cross section of zero-coupon rates. This contrasts to the observed-affine model where the state-variable variance-covariance matrix is estimated only with the state-variable time series.

The final element of consideration is how well our different models forecast excess holding-period returns for different zero-coupon bond tenors. The idea behind a holding-period return is quite simple. At time t, one purchases a T-period zero-coupon bond and holds it for τ periods, where $\tau < T$. One then sells the $(T-\tau)$-period bond, at time $t+\tau$. The return on this investment is termed a holding-period return. If we subtract the risk-free rate one could have earned from investing in a τ-period zero-coupon bond at time t, one has the excess holding-period return. We use our six separate term-structure models to forecast a variety of excess holding-period returns at 12-, 24-, and 60-month forecasting horizons. The results are summarized in Table 3.

The first thing to note is that the out-of-sample forecast errors increase monotonically with the zero-coupon tenor. Second, the magnitude of the errors is quite substantial across all models. The figures in Table 3 are in annualized percentage terms. This implies that, at the two-year forecasting horizon, the root-mean-squared-error excess holding-period forecast error of the random-walk model is 14.2%.[31] Part of the problem is that excess holding-period returns are notoriously difficult to forecast for any financial asset. Complicating matters, long-term pure-discount bond prices are extremely sensitive to interest-rate movements, which makes them even more difficult to forecast.[32] Nevertheless, excess holding-period returns are critical in portfolio optimization as they partially describe the risk premia inherent in the zero-coupon term structure. For this reason, it is very interesting to understand how well our term-structure models forecast these returns out-of-sample.

[31] This does not imply that there is a 14.2% difference between the two forecasts, but rather that the units of the error are percentage points. That is, there is a root-mean-squared difference of 14.2 percentage points between the two forecasts.

[32] See Bolder, Johnson, and Metzler (2004) for a review of historical excess holding-period returns on Canadian zero-coupon bonds.

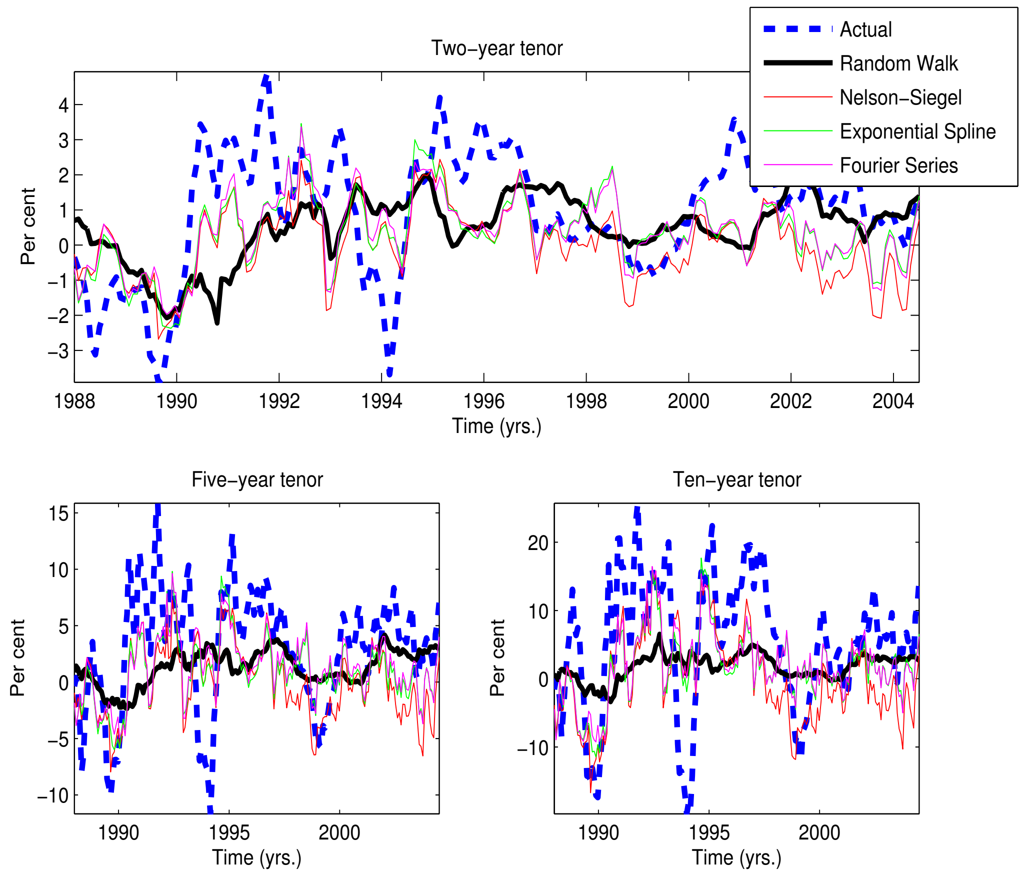

Figure 9: **Empirical Excess Holding-Period Return Forecasts**: This figure displays the 12-month excess holding period return forecasts for two-, five-, and ten-year zero-coupon bond tenors across our sample period for each of our empirical term-structure models including macroeconomic factors.

We can see, from Table 3, that the Nelson-Siegel, exponential-spline, and Fourier-series models outperform the random-walk approach at almost all of the 12-month forecasting horizons and for the shorter zero-coupon tenors for the 24-month horizon. The observed-affine model only slightly outperforms the random-walk assumption at the shortest zero-coupon tenors for the 12-month forecasting horizon, while the Ang-Piazzesi approach never succeeds in outperforming the random walk. At the 60-month forecasting horizon, none of the models outpeform the random-walk assumption. Moreover, none of the models, including the random-walk approach, provide a very convincing description of excess holding-period returns.

Another perspective on the accuracy of the excess holding-period return forecasts is provided in Figures 9 and 10. The figures provide the actual and forecast 12-month excess holding period returns for the two-, five-, and ten-year tenors across the out-of-sample forecast period; Figure 9 describes the Nelson-Siegel, exponential-spline,

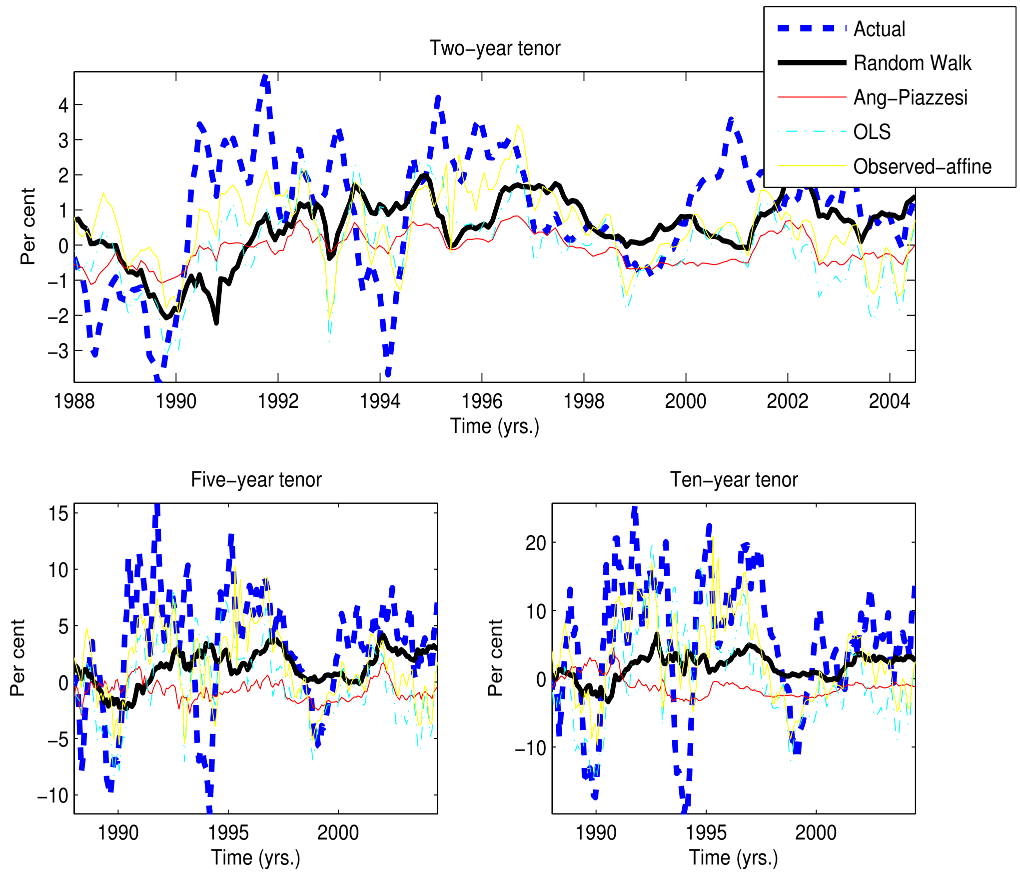

Figure 10: **Affine Excess Holding-Period Return Forecasts**: This figure displays the 12-month excess holding period return forecasts for two-, five-, and ten-year zero-coupon bond tenors across our sample period for each of the Ang-Piazzesi and observed-affine term-structure models.

and Fourier-series models while Figure 10 describes the Ang-Piazzesi, OLS, and observed-affine models. We can see rather clearly that, at the two-year tenor, all of the models, with the possible exception of the Ang-Piazzesi model, appear to track the actual excess holding-period returns fairly closely. For the longer zero-coupon tenors, however, the models are incapable of matching the dramatic variation in actual excess holding-period returns.

We can draw a number of interesting lessons from the previous analysis. It is clear that the exponential-spline and Fourier-series models demonstrate the best performance in terms of out-of-sample zero-coupon and excess holding-period return forecasting. This result is robust across virtually all zero-coupon tenors and forecasting horizons. In terms of out-of-sample forecasing performance, these two models are followed by the Nelson-Siegel, OLS, and observed-affine approaches. Finally, the Ang-Piazzesi model demonstrates the poorest, and generally least consistent, out-of-sample forecasts. When we turn our attention to a measure of conditional variance, however, we find rather different results. In particular, we find that the exponential-spline and Fourier-series models tend to underestimate the conditional variance of the zero-coupon curve at longer tenors; moreover, this tendency deteriorates as we extend the forecasting horizon. The Nelson-Siegel, OLS, and observed-affine models appear to provide a better characterization of the conditional variance, albeit with a tendency to both over- and underestimate at longer zero-coupon tenors. Finally, the Ang-Piazzesi model demonstrates rather erratic behaviour with underestimates at the short end of the zero-coupon curve and simultaneous overestimates of conditional variance at longer tenors. At this point, therefore, we would likely conclude that the most reasonable model is the Nelson-Siegel approach suggested by Diebold and Li (2003). It is also worth mentioning that, despite the fact that some of the models outperform the random-walk horizon at short forecasting horizons, we need to be cautious with respect to what can be accomplished with these models. It would be virtually impossible, for example, to successfully use these models to generate highly accurate forecasts of future financial and macroeconomic conditions. In the next section, we extend our analysis in a very simple manner to include a time-varying dimension to our parametrization and see how this might change our results and conclusions.

3.4 Incorporating a time-varying perspective

An examination of our data in Figures 1 and 2 reveals quite clearly that the constant-parameter assumption is not ideal. More specifically, the period from about 1995 to the present seems to be characterized by low and stable inflation, while the late-1970s to mid-1980s exhibit both high inflation and high interest rates. If we were to estimate any of our six term-structure models with the entire data sample and use the corresponding parameter set to simulate outcomes for a given risk-management analysis, it is hard to believe that the results would be reasonable. The results, most probably, would not be a reasonable forward-looking description of the current low and stable inflation and interest-rate regime, but rather a weighted average of low-, intermediate-, and high-inflationary regimes. We could, of course, estimate the model only using the data from 1995 to the present. This would avoid the constant-parameter problem, but it would imply that we would have a rather short period over which to characterize the long-term interaction between the term structure of interest rates and the macroeconomy.

The alternative to restricting the size of the dataset is to attempt to permit, to some degree, the model parameters to vary through time. One could, for example, apply a regime-switching model to the VAR dynamics of the state variables in each of the models. This is, however, easier said than done. All of the models—particularly the exponential-spline and Fourier-series approaches—have a large number of state variables. Given the VAR(2) specification for the state-variable dynamics, this implies a large number of model parameters. This is not a problem in itself, but implementing a full regime-switching structure for such a large number of parameters is unworkable.[33] We could attempt to reduce the size of the parameter space, but the VAR dynamics for the state-variables appears to represent a reasonable forecasting approach. We experimented, in this vein, with lower-dimensional specifications for the state-variable dynamics and found a general decrease in out-of-sample forecasting performance.

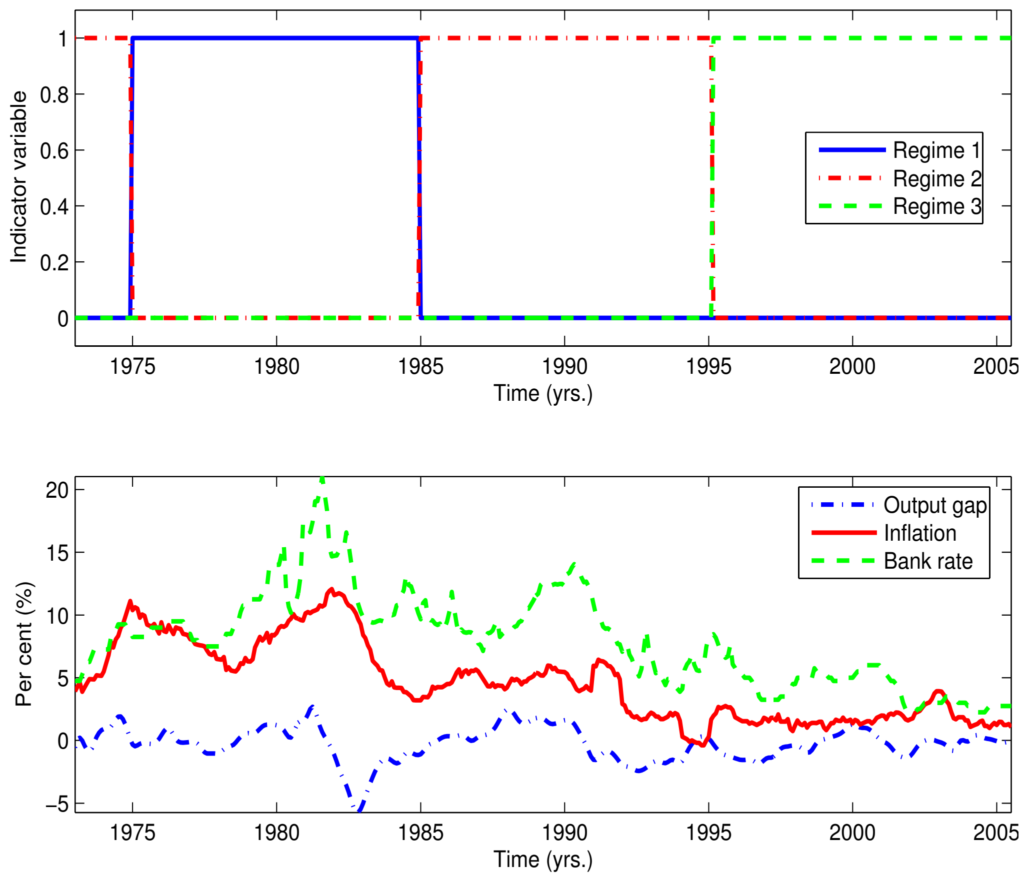

Figure 11: **An Exogeneous Indicator Variable**: This figure displays the exogeneous indicator variable that stems from the work of Demers (2003). The first regime denotes a high-inflation regime, the third regime describes a low-inflation regime with anchored inflationary expectations, while the second regime is a transition regime between the high- and low-inflationary regimes.

We consequently opted for a very simple approach. We decided to exogeneously impose the three regimes identified in Demers (2003) as indicator variables to the state-variable dynamics. We apply the indicator variables,

[33]Consider, for example, a VAR(1) specification for the exponential-spline model. This model, without regime switching, has 132 parameters where the intercept is $C \in \mathbb{R}^{11\times 1}$, while the slope matrix is $F \in \mathbb{R}^{11\times 11}$. A full Markov-chain switching model, following from Hamilton (1989), with three separate regimes would include more than 300 parameters to be determined by a non-linear optimization algorithm.

however, only to the intercept terms. This is something of a compromise solution, as there is compelling evidence that entirely separate equations are required to properly characterize inflation and output dynamics under these different regimes. In the context of our model, such a degree of complexity cannot easily be implemented.

Figure 11 illustrates the three separate regimes in conjunction with the evolution of our three primary macroeconomic variables. The mid-1970s to the mid-1980s are a high-inflation regime, the period from 1995 to the present is a low- and stable-inflation regime, while the remaining periods represent a third, transitory regime. Our model incorporates the indicator function values as dummy variables as applied to the intercept vector in the VAR(2) specifications for the state-variable dynamics. To be clear, there is neither a notion of a transition matrix in our approach, nor are the slope matrices and variance-covariance matrices permitted to vary through time. The absence of a transition matrix is not really an issue given our application of these models. As we intend to use our model to generate forward-looking simulations of future macroeconomic and interest-rate outcomes, it is not clear that a historical transition matrix will provide a strong description of the probability of a future inflationary episode. It is our view that it would be preferrable to examine how the results of a risk-management analysis vary for different *calibrated* probabilities of a high-inflation regime.

What happens to our results, therefore, when we add an exogenously determined indicator variable to the intercept term in each of the VAR models? It is not exactly fair to consider the same out-of-sample forecasting statistics as in the previous section. This is primarily because we are providing information about the current state of the world that would presumably be unknown at the time of the out-of-sample forecast. In other words, the different regimes where estimated by Demers (2003) using the entire dataset and, strictly speaking, should not be used in formulating out-of-sample forecasts constructed using only a subset of the dataset; this is often termed a *look-ahead* bias. Keeping this caveat in mind, it is still interesting to see how the various models perform. The experiment, however, is slightly different. Imagine that we provide an analyst with data up to time t and asked her to forecasting various zero-coupon rates, the variance of these zero-coupon rates, and excess holding-period returns for different horizons. We also tell the analyst the current and past regimes as described in Figure 11. We can think of this as a partial information forecast. Moreover, we can interpret the results as helping us understand the efficiency of adding this simple time-varying element into our model.[34]

Table 4 illustrates the results of the partial-information out-of-sample forecasts of the entire zero-coupon curve. First, note that there are no results for the Ang-Piazzesi and observed-affine models. We had to exclude these models as there is no straightforward way to incorporate regime-switching into the affine setting. There is, of course, a literature that incorporates regime-switching into affine term-structure models, but it is already

[34] We did experiment with a variety of different breakpoints for the three inflationary regimes and found that the results were essentially unchanged.

Table 4: **Out-of-Sample Zero-Coupon Curve Forecasts with Intercept Indicator**: In this table, we present the root-mean squared and mean-average error for a series of one-, six-, 12-, 24-, and 60-month zero-coupon curve forecasts for four of our macro-finance term-structure models with an exogeneously specified indicator variable from Demers (2003). Figures in bold denote an out-performance of the random-walk hypothesis.

Models	**RMSE**					**MAE**				
	Mean	**Median**	**Max**	**Min**	**SD**	**Mean**	**Median**	**Max**	**Min**	**SD**
One-month forecast										
Random walk	23.46	20.23	86.05	2.36	15.26	21.28	18.33	71.65	1.75	14.47
Nelson-Siegel	23.91	**19.95**	97.66	5.03	13.86	**21.16**	**17.59**	93.50	3.78	13.24
Exponential spline	**22.93**	**18.99**	95.63	3.42	14.92	**20.44**	**16.50**	89.27	2.53	14.12
Fourier-series	24.93	21.58	102.87	3.36	15.36	21.78	18.60	94.33	2.90	13.95
OLS	**23.18**	**20.20**	93.90	3.35	13.95	**20.42**	**17.99**	81.74	2.69	13.18
Six-month forecast										
Random walk	74.98	64.60	307.09	9.29	46.84	67.03	56.09	303.34	8.23	46.10
Nelson-Siegel	**67.81**	**60.35**	285.47	8.46	40.01	**61.53**	**55.06**	283.66	7.12	39.42
Exponential spline	**66.50**	**58.58**	281.29	9.20	40.97	**59.48**	**50.52**	278.61	7.62	40.38
Fourier-series	**67.92**	**56.83**	294.83	6.89	42.29	**60.88**	**49.27**	291.77	5.15	41.41
OLS	**68.34**	**58.96**	317.02	5.84	41.59	**61.71**	**53.12**	314.59	4.92	40.67
12-month forecast										
Random walk	110.96	95.61	346.24	13.39	68.16	99.57	81.70	337.50	10.00	64.75
Nelson-Siegel	**96.17**	**83.20**	295.62	9.97	56.34	**86.25**	**73.92**	283.49	8.25	53.78
Exponential spline	**90.70**	**82.44**	365.16	11.35	56.46	**81.00**	**68.49**	353.30	8.51	54.71
Fourier-series	**91.72**	**76.92**	315.42	13.24	58.17	**82.48**	**69.30**	307.66	11.50	56.39
OLS	**96.71**	**80.48**	350.41	14.51	60.56	**86.11**	**73.98**	336.83	11.88	57.60
24-month forecast										
Random walk	156.87	140.23	487.90	21.21	95.28	142.54	123.84	451.76	15.36	90.06
Nelson-Siegel	**146.38**	**123.40**	429.73	22.22	92.05	**131.90**	**106.25**	406.33	20.00	86.81
Exponential spline	**143.77**	**121.28**	394.45	24.05	84.72	**129.52**	**108.35**	376.94	19.13	80.77
Fourier-series	**138.79**	**117.05**	391.70	26.91	85.61	**124.76**	**100.30**	368.69	23.50	81.72
OLS	**148.93**	**127.73**	452.76	20.87	93.76	**133.25**	**107.63**	429.15	17.28	87.05
60-month forecast										
Random walk	223.55	206.75	521.97	65.52	98.52	207.93	189.42	492.87	57.13	96.86
Nelson-Siegel	**184.64**	**170.23**	407.46	15.68	103.45	**177.07**	**160.31**	402.00	11.67	103.88
Exponential spline	**167.18**	**124.53**	422.28	20.76	113.99	**158.30**	**114.35**	404.86	14.19	114.41
Fourier-series	**181.07**	**156.16**	389.92	15.22	98.12	**170.45**	**150.21**	386.53	12.64	99.72
OLS	**182.25**	**158.52**	405.79	25.82	104.99	**174.31**	**146.98**	401.49	16.95	105.17

a challenge to estimate the parameters of the Ang-Piazzesi model.[35] Observe that with a few exceptions at the one-month forecasting horizon, all of the models outperform the random-walk approach. Indeed, permitting the model to allow the intercept term in the VAR(2) specification to vary over time leads to substantial improvements

[35] In particular, we found it quite challenging to obtain consistent results for the non-linear optimization problem stemming from the estimate of these models. Addition of time-varying parameters makes matters even worse.

in the out-of-sample forecasting ability of the models. This trend is particularly evident at longer forecasting horizons. Moreover, it is consistent with our suspected reason for the underperformance of our models, relative to the random-walk hypothesis, at longer time horizons. Specically, it would appear that the world changed rather importantly in the early 1990s—as it became apparent that the Bank of Canada would be successful in controlling inflation—and this led to a secular decrease in interest rates. By providing this information to our models, they proceed to outperform the random walk.

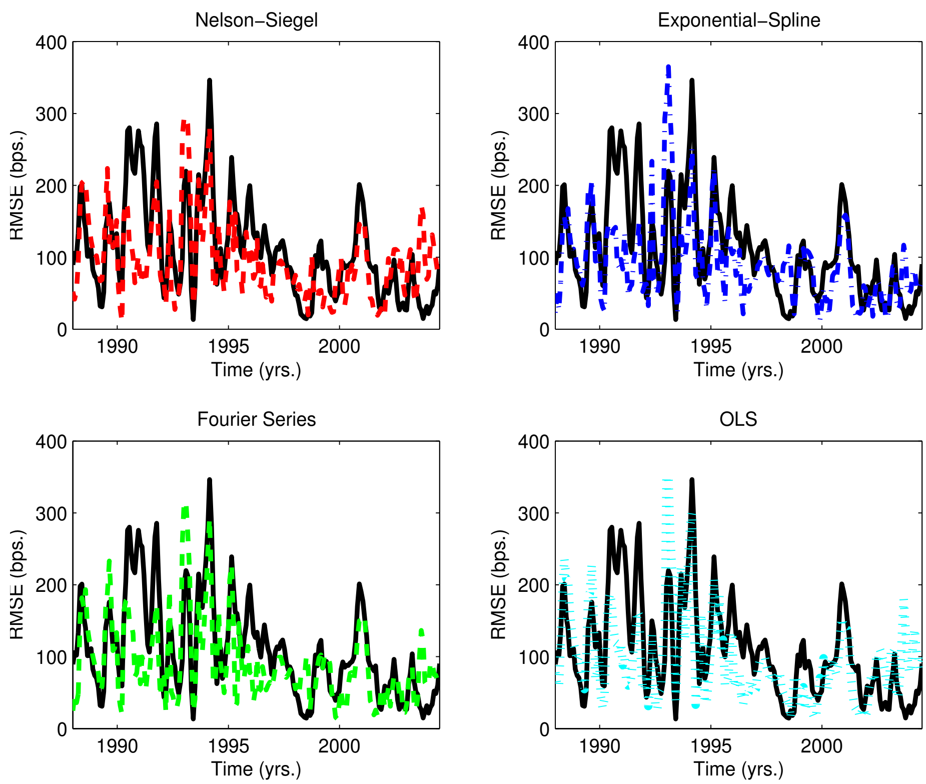

Figure 12: **The Evolution of RMSE with Intercept Indicator**: This figure displays the evolution of the entire-curve RMSE across four macro-finance term-structure models with an exogeneously specified indicator function, for a 12-month forecasting horizon. The solid black line in each graphic represents the random-walk forecasting assumption.

Figure 12 summarizes the evolution of the root-mean-squared forecast error for each of our six macro-finance term-structure models at the 12-month forecasting horizon. Again, the solid black line represents the random-walk forecasts. This figure, which compares to Figure 5 computed without the intercept indicator, demonstrates better out-of-sample forecasts of the zero-coupon curve. The improvement appears to be particularly evident in the period from approximately 2000 to the present. Again, this demonstrates the value of partial information and underscores the fact that the data is composed of distinct macroeconomic regimes.

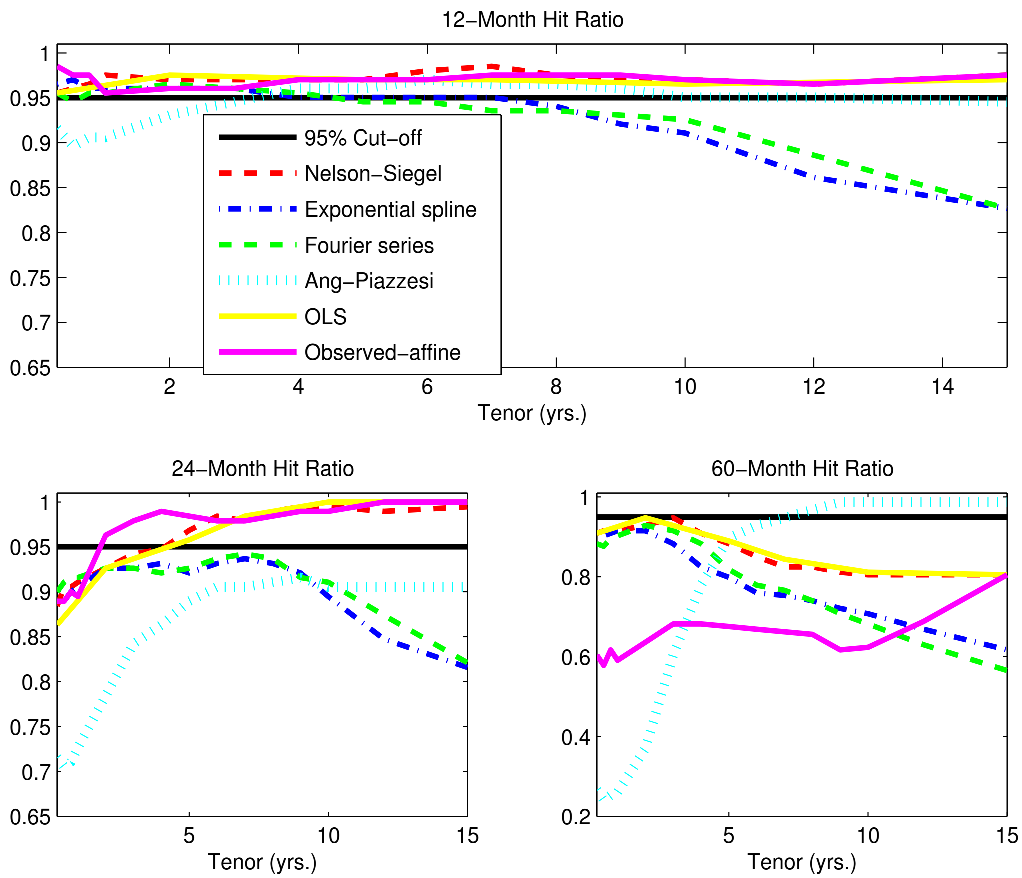

Figure 13: **Hit Ratio with Indicator Variable**: This figure displays the hit ratio for a variety of tenors ranging from three months to 15 years across one-, two-, and five-year forecasting horizons for our six alternative macro-finance term-structure models. An indicator variable is added to the intercept vector for the Nelson-Siegel, exponential-spline, Fourier-series, and OLS models. The Ang-Piazzesi and observed-affine models, neither or which has time-varying parameters, are added for comparison purposes. Note that the solid black line represents the 95% cut-off level for the confidence interval constructed by the hit-ratio measure.

While permitting the intercept vector in the state-variable dynamics to vary through time does not imply time-varying slope and variance-covariance matrices, it does allow them to take different values. Our hope is that allowing for a time-varying intercept vector should permit more realistic estimates of these matrices. In principle, for example, the intercept should capture some portion of the time-varying relationship between the variables. Not considering the time-varying aspect of the parameters could lead to distorted estimates of the slope and variance-covariance matrices. How true this is in practice, of course, depends on the reasonableness of imposing only a time-varying intercept vector. We are particularly interested in the variance-covariance matrix given its importance in the conditional variance of the zero-coupon curve.[36]

[36]For the mathematical details of this relationship, see Appendix D.

Table 5: **Excess Holding-Period Return Forecast Errors with Intercept Indicator**: In this table, we present the annualized root-mean-squared holding-period forecast errors for our six term-structure models along 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**				
	RW	**NS**	**ES**	**FS**	**OLS**
12-month forecast					
2 years	1.79	**1.55**	**1.46**	**1.48**	**1.58**
5 years	5.22	**4.40**	**4.23**	**4.40**	**4.50**
7.5 years	7.43	**6.18**	**6.02**	**6.23**	**6.34**
10 years	9.62	**8.04**	**7.73**	**8.00**	**8.13**
15 years	14.22	**11.63**	**11.64**	**11.79**	**12.24**
24-month forecast					
5 years	6.69	**6.23**	**6.01**	**5.99**	**6.31**
7.5 years	10.45	**9.27**	**9.19**	**8.96**	**9.34**
10 years	13.79	**11.82**	**11.45**	**11.25**	**11.26**
15 years	22.05	**17.67**	**17.49**	**17.51**	**16.75**
60-month forecast					
7.5 years	10.14	**8.07**	**7.66**	**8.07**	**8.07**
10 years	18.15	**15.20**	**14.37**	**14.32**	**15.11**
15 years	36.78	**32.63**	**32.92**	**32.65**	**31.44**

To help answer this question, Figure 13 illustrates the hit ratio for each of our six term-structure models—the Ang-Piazzesi is included for comparison purposes—at 12-, 24-, and 60-month forecasting horizons. It generally appears that the addition of the indicator variable to the intercept vector permits an improved characterization of the state-variable variance-covariance matrix. The exponential-spline and Fourier-series models, for example, continue to demonstrate a tendency to underestimate the longer-tenor zero-coupon conditional variance across all forecasting horizons. The magnitude of the conditional-variance underestimates is, however, rather less than without the intercept indicator variable. Moreover, the tendency to underestimate does not generally begin, at the 12- and 24-month forecasting horizons, until approximately the ten-year zero-coupon tenor. This is clearly an improvement. The Nelson-Siegel and OLS models demonstrate generally quite acceptable hit ratios across all tenors and forecasting horizons, albeit with a relatively slight tendency to overestimate longer-tenor conditional variance at the 24-month horizon and simultaneously to slightly underestimate longer-tenor conditional variance at the 60-month horizon. Despite a few criticisms, therefore, the Nelson-Siegel and OLS models seem to produce the most stable and reasonable characterization of the conditional variance of the zero-coupon curve.

We can now finally turn to see if the incorporation of an indicator variable in the intercept vector improves the out-of-sample forecasting performance of the excess holding-period returns. Table 5 summarizes the results. The Nelson-Siegel, exponential-spline, Fourier-series, and OLS models outperform the random-walk assumption across all zero-coupon tenors and forecasting horizons. Moreover, the differences between the various models are quite small. What is clear, however, is that the provision of partial information about the state of the economy leads to a dramatic improvement in the out-of-sample forecasting performance of these models. We also note that, as compared to Figure 8, that the empirical models appear to better describe excess holding-period returns.

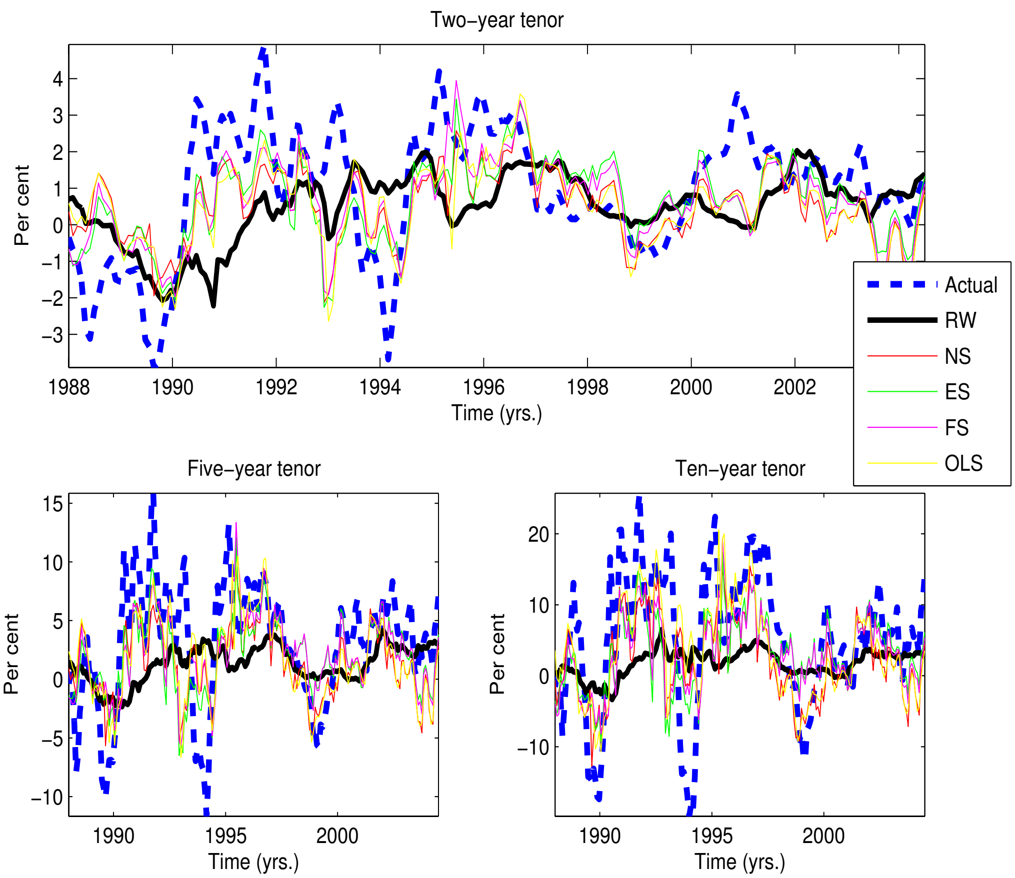

Figure 14: **Empirical Excess Holding-Period Return Forecasts with Indicator Variable**: This figure displays the 12-month excess holding period return forecasts for two-, five-, and ten-year zero-coupon bond tenors across our sample period for each of our empirical term-structure models including macroeconomic factors and an exogenous indicator variable.

What can we conclude from this discussion? First, it is hardly a surprise that permitting the intercept vector in our state-variable dynamics to vary through time improves the out-of-sample forecasting performance of all models. We have, after all, provided incremental information about the state of the world to the forecast model. For this reason, the improvement was particularly evident at the longer zero-coupon tenors. We can, however, make two principal conclusions. First, the addition of this partial information permits all four empirical models to outperform the random-walk hypothesis across virtually all tenors and forecasting horizons. Second, the conditional variance estimates of the zero-coupon curve stemming Nelson-Siegel, exponential-spline, and Fourier-series models are substantially improved. This suggests that the variance-covariance and slope matrices describing the state-variable dynamics were distorted, and presumably still are to some extent, by the various inflationary regimes present in the data. These two conclusions are, practically speaking, quite useful. It implies that the three states of the world that we have borrowed from Demers (2003) are a reasonable specfication of the various inflationary regimes over the last 30 years. Moreover, it also suggests that incorporating these regimes can improve how the empirical models characterize the second moment of the transition density of zero-coupon rates. We do wish to underscore that adding the time-varying intercept vector is not a suggested change to the model, but rather a technique to permit us to reasonably estimate the model parameters.

4 Conclusion

This is really quite a simple paper. Our primary objective is to compare a variety of simple joint macro-finance models with a particular view towards assessing their appropriateness as an input for a risk-management analysis. To this end, we consider six alternative joint models of the macroeconomy and the term structure of interest rates. The first model implements the seminal work of Ang and Piazzesi (2003), who implement a joint macro-finance model in a discrete-time affine setting. The next three of these models follow from the work of Diebold and Li (2003) and include the dynamic implementations of the Nelson-Siegel, exponential-spline, and the Fourier-series models. The fifth model is a regression-based model motivated entirely by empirical considerations.

The sixth, and final, model follows Ang and Piazzesi (2003) and is also similar in spirit to work by Colin-Dufresne, Goldstein, and Jones (2005) and Cochrane and Piazzesi (2006). This approach, which we term the observed-affine model, relaxes restrictions on the state-variable dynamics of the Ang-Piazzesi model by making them observable. Two important points should be noted about the observed-affine model. First, by forcing the state variables to be observable, we reduce the flexibility of the model to adjust the state variables to fit the cross section of zero-coupon interest rates. Second, we introduced this model from a practical perspective—indeed, by reverse-engineering—in an attempt to improve the out-of-sample forecasting properties of the affine model. In this respect, therefore, we are essentially trying to identify, from a practitioner's perspective, how the affine model might fruitfully be altered.

Using monthly data from 1973 to 2005, we compare each of the joint macro-finance models along four dimensions with a strong out-of-sample focus. First, we consider out-of-sample forecasts of the entire zero-coupon curve for forecasting horizons of one month to five years. Second, we examine the out-of-sample forecasts of specific zero-coupon tenors across the same range of forecasting horizons. Third, we look to see how the various models perform in terms of their ability to produce out-of-sample forecasts, again over one-month to five-year forecasting horizons, of excess holding-period returns. Each of these measures provides, from different perspectives, the ability of these models to characterize the first moment of the transition density of zero-coupon rates. The final measure, that we term the hit ratio, describes the proportion of out-of-sample forecasts that fall within a conditional 95% confidence interval around the forecast value. This measure, while imperfect, provides some insight into the ability of these models to capture the second moment of the transition density of zero-coupon rates.

What, therefore, did we learn from examination of these measures? First, the exponential-spline and Fourier-series models clearly demonstrate the best performance in terms of out-of-sample zero-coupon and excess holding-period return forecasting. This result is robust across virtually all zero-coupon tenors and forecasting horizons. In terms of out-of-sample forecasing performance, these two models are followed respectively by the Nelson-Siegel, OLS, and observed-affine approaches, while the Ang-Piazzesi model produces the weakest out-of-sample forecasts. When we turn our attention to a measure of conditional variance, however, we find rather different results. In particular, we find that the exponential-spline and Fourier-series models tend to underestimate the conditional variance of the zero-coupon curve at longer tenors; moreover, this tendency increases as we extend the forecasting horizon. The Nelson-Siegel, OLS, and observed-affine models appear to provide a better characterization of the conditional variance, albeit also with a tendency to both over- and underestimate at longer zero-coupon tenors. Finally, the Ang-Piazzesi model demonstrates rather erratic behaviour with sizable underestimates at the short end of the zero-coupon curve with simultaneous overestimates of conditional variance at longer tenors. Based on these results, we conclude that the most reasonable model for capturing Canadian term-structure dynamics is the Nelson-Siegel approach suggested by Diebold and Li (2003).

Examination of our data, however, reveals that the constant-parameter assumption is not ideal. That is, the period from about 1995 to the present is to be characterized by low and stable inflation, while the late-1970s to mid-1980s exhibit both high inflation and high interest rates. Unadjusted parametrization of our models—using the entire data sample and application corresponding constant-parameter set—would likely yield unreasonable results. Specifically, we fear that models would not provide a reasonable forward-looking description of the current low and stable inflation and interest-rate regime, but rather a weighted average of low-, intermediate-, and high-inflationary regimes. How do we propose to solve this problem? We opt for a very simple approach. Specifically, we exogeneously impose the three regimes identified in Demers (2003) as indicator variables to the state-variable dynamics. We apply these indicator variables, however, only to the intercept terms. This is something of a compromise solution, as there is compelling evidence that entirely separate equations are required to properly characterize inflation and output dynamics under these different regimes.

We then turn to examine the same four dimensions using these indicator variables. Note, it is not exactly fair to consider the raw out-of-sample forecasting statistics, because the technique provides more information about the current state of the world than would presumably be known at the time of the out-of-sample forecast. We can think of this as a partial-information forecast exercise (i.e., there is a *look-ahead* bias). Nevertheless, we can interpret the results as helping us understand the efficiency of adding this simple time-varying element into our model for the purposes of estimating model parameters that can be used in a forward-looking simulation exercise.

What, therefore, did we learn? First, it is hardly a surprise that permitting the intercept vector in our state-variable dynamics to vary through time improves the out-of-sample forecasting performance of all models. The improvement was particularly evident at forecasting horizons. Interestingly, it appears to almost uniformly improve all models. Second, the conditional variance estimates of the zero-coupon curve—particularly, the exponential-spline and Fourier-series models—are all substantially improved. This suggests that the variance-covariance and slope matrices describing the state-variable dynamics were distorted by the various regimes present in the data. These two conclusions are, practically speaking, quite useful. It implies that the three states of the world that we have borrowed from Demers (2003) are a reasonable specfication of the various inflationary regimes over the last 30 years. It also suggests that incorporating these regimes can improve how the empirical models characterize the second moment of the transity density of zero-coupon rates.

Overall, therefore, we can conclude from this analysis that if we wish to select a single model, our first choice would be the Nelson-Siegel approach suggested by Diebold and Li (2003). We should note also that the exponential-spline, Fourier-series, and OLS models should not be overlooked given their strong out-of-sample forecasting performance. Moreover, with the addition of the indicator variable, their characterization of the conditional variance of the zero-coupon term structure improves quite substantially. We can also conclude that our experiment with the observed-affine model, while interesting, was *not* a resounding success. The observed-affine model improves the out-of-sample forecasting performance of the Ang-Piazzesi model. It does not, however, improve the out-of-sample forecasts by as much as we might have hoped. We suspect that the principal reason is the observed-affine model's difficulty in fitting the cross section of zero-coupon rates. This stems from the fact that many of the model parameters are fixed before we attempt to econometrically fit the cross section. This would be a useful direction for future research.

We still believe, however, that the Ang-Piazzesi model is, in principle, useful. Nevertheless, it requires some additional work for its application to the Canadian setting. We suggest, therefore, two directions for future research. First, it would be interesting to examine other alternative adjustments to the affine approach to improve its practical usefulness. Second, we believe it would be useful to examine alternative estimation approaches for the Ang-Piazzesi model to improve its performance.

A Deriving the Ang-Piazzesi Model

This technical appendix works through the technical details of the construction and estimation of the Ang-Piazzesi model. The observable macroeconomic variables in the Ang and Piazzesi (2003) setting involve the extraction of the first principal component from a collection of inflation and output data. That is, the first observable state variable is the first principal component from an eigenvalue decomposition of a number of inflation variables, while the second observable state variable stems from an identical technique applied to a collection of variables related to output. This is a clever and useful method for reducing the dimensionality of the macroeconomic data. We will, however, restrict our attention to the output gap and the inflation rate. As such, we can define the vector of contemporaneous macroeconomic variables X_t^o as,

$$X_t^0 = \begin{bmatrix} x_t \\ \pi_t \end{bmatrix}, \tag{31}$$

where $n = \dim\left(X_t^0\right)$. The dynamics of X_t^0 are assumed to follow a VAR(p) form as,

$$X_t^0 = \sum_{k=1}^{p} \Phi_k X_{t-k}^o + \Omega u_t^0, \tag{32}$$

where $\Phi_1, .., \Phi_p \in \mathbb{R}^{n\times n}$.[37] If we stack the observable state variables as follows,

$$F_t^o = \begin{bmatrix} X_t^o \\ X_{t-1}^o \\ \vdots \\ X_{t-(p-1)}^o \end{bmatrix}, \tag{33}$$

with $F_t^0 \in \mathbb{R}^{(np)\times 1}$ then we can re-write equation (32) in companion form as,

$$F_t^o = \Phi^o F_{t-1}^o + \Sigma^o \epsilon_t^o, \tag{34}$$

where $\epsilon_t^0 \sim \mathcal{N}(0, I)$. The coefficients $\Phi^o, \Sigma^o \in \mathbb{R}^{(np)\times(np)}$ are defined as,

$$\Phi^o = \begin{bmatrix} \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p \\ I_{n\times n} & \cdots & 0_{n\times n} & 0_{n\times n} \\ \vdots & \ddots & \vdots & \vdots \\ 0_{n\times n} & \cdots & I_{n\times n} & 0_{n\times n} \end{bmatrix}, \tag{35}$$

$$\Sigma^0 = \begin{bmatrix} \Omega & \cdots & 0_{n\times n} \\ \vdots & \ddots & \vdots \\ 0_{n\times n} & \cdots & 0_{n\times n} \end{bmatrix}, \tag{36}$$

[37] Note that the intercept term $\Phi_0 = 0$ as the state variables are normalized to mean zero.

and,

$$\epsilon_t^0 = \begin{bmatrix} u_t^0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{37}$$

The unobservable state variables, F_t^u, are assumed to follow a first-order autoregression as,

$$F_t^u = \Phi^u F_{t-1}^u + \epsilon_t^u, \tag{38}$$

where, to ensure the identification of the model parameters in the associated affine model, the matrix $\Phi^u \in \mathbb{R}^{3\times 3}$ is lower triangular.[38] The dimension of F_t^u was determined by the substantial empirical evidence suggesting that the term structure of interest rates is well explained by three latent state variables–see Litterman and Schenkman (1991). Stacking the observable and unobservable state variables, F_t^o and F_t^u, as follows

$$X_t = \begin{bmatrix} F_t^o \\ F_t^u \end{bmatrix} \in \mathbb{R}^{(np+3)\times 1}, \tag{39}$$

permits us to write the dynamics of the entire state-variable vector as,

$$X_t = \Phi X_{t-1} + \Sigma \epsilon_t \tag{40}$$

where,

$$\Phi = \begin{bmatrix} \Phi^o & 0_{(np)\times 3} \\ 0_{3\times(np)} & \Phi^u \end{bmatrix}, \tag{41}$$

$$\Sigma = \begin{bmatrix} \Sigma^o & 0_{(np)\times 3} \\ 0_{3\times(np)} & I_{3\times 3} \end{bmatrix}, \tag{42}$$

and,

$$\epsilon_t = \begin{bmatrix} \epsilon_t^u \\ \epsilon_t^o \end{bmatrix}. \tag{43}$$

This brings us to the derivation of the affine term-structure model, which involves using the state-variable vector to construct a discrete-time affine term-structure model. Other descriptions of these models can be found in Backus et al. (1998), Backus et al. (1999), and Campbell et al. (1997). The foundation of the discrete-time

[38] It is also necessary that the conditional variance of F_t^u be an appropriately dimensioned identity matrix.

approach comes from the first-order condition of the following utility maximization problem,

$$\max \mathbb{E}\left(\sum_{j=0}^{\infty} \delta^j U(C_{t+j} \mid \mathcal{F}_t\right) \tag{44}$$

where δ is a discount factor, $U(\cdot)$ is a time-separable utility function, and C_t denotes the investor's consumption. One of the first-order conditions from this optimal-consumption problem, which describes the investor's optimal mix of consumption and investment, can be manipulated into a useful identity,

$$\begin{aligned} U'(C_t) &= \delta\mathbb{E}\left((1+R_{n,t+1})U'(C_{t+1})\middle| \mathcal{F}_t\right), \qquad (45) \\ 1 &= \mathbb{E}\left((1+R_{n,t+1})\underbrace{\frac{\delta U'(C_{t+1})}{U'(C_t)}}_{\substack{\text{Pricing}\\ \text{kernel}}}\middle| \mathcal{F}_t\right), \\ 1 &= \mathbb{E}\left((1+R_{n,t+1})M_{t+1}\middle| \mathcal{F}_t\right), \end{aligned}$$

where $R_{n,t+1}$ is the real rate of return on an arbitrary asset with n periods left to maturity of the underlying asset from time t to $t+1$ and M_{t+1} is termed the stochastic discount factor or the pricing kernel.[39] The pricing kernel can be interpreted as the intertemporal marginal rate of substitution—or, more simply, the investor's preferences regarding substitution of consumption and investment through time. The meaning of this expression, therefore, is that the rate of return on any asset will depend on its covariance with the intertemporal marginal rate of substitution. In a fixed-income setting, however, the cash-flows are deterministic and, as such, the covariance of the cash-flows with the pricing kernel depends only on time.

We can now specialize the general relation in equation (45) to the fixed-income setting. The holding-period return on a bond is given as,

$$1+R_{n,t+1} = \frac{P_{n-1,t+1}}{P_{n,t}}, \tag{47}$$

where P represents the value of a pure discount bond and n is the number of periods left to maturity of the pure discount bond at t. Now, substituting equation (47) into (45),

$$\begin{aligned} 1 &= \mathbb{E}\left(\frac{P_{n-1,t+1}}{P_{n,t}}M_{t+1}\middle| \mathcal{F}_t\right), \qquad (48) \\ P_{n,t} &= \mathbb{E}\left(P_{n-1,t+1}M_{t+1}\middle| \mathcal{F}_t\right) \end{aligned}$$

[39] Note that, as usual, we are defining these processes on a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, we can consider $\mathcal{F}_t$ to be the natural filtration generated by M_{t+1}. This filtration is defined in the usual manner, as follows:

$$\mathcal{F}_t \triangleq \sigma\{M_s, s=0,..,t\}. \tag{46}$$

We are left with, therefore, the current price of a zero-coupon bond as the conditional expectation of the product of its future price and the stochastic discount factor. For a one-period bond, this has an appealing form. We know, for example, that $P_{0,t} = 1$ for all t, because the price of a dollar today is one dollar. Thus, we have that,

$$P_{1,t} = \mathbb{E}\left(\underbrace{P_{0,t+1}}_{P_{0,t+1}=1} M_{t+1} \middle| \mathcal{F}_t \right), \tag{49}$$
$$P_{1,t} = \mathbb{E}\left(M_{t+1} \middle| \mathcal{F}_t \right).$$

Modelling pure-discount bond prices, and thus the term structure of interest rates, amounts to describing the dynamics of the pricing kernel. Ang and Piazzesi (2003)) assume the following form for the pricing kernel,

$$M_{t+1} = e^{-r_t} \left(\frac{\varphi_{t+1}}{\varphi_t} \right), \tag{50}$$

where the dynamics of the Radon-Nikodým derivative, $\frac{\varphi_{t+1}}{\varphi_t}$, follow a log-normally distribution,

$$\frac{\varphi_{t+1}}{\varphi_t} = e^{-\frac{1}{2}\lambda_t^T \lambda_t - \lambda_t^T \epsilon_{t+1}}, \tag{51}$$

implying that,

$$M_{t+1} = e^{-r_t - \frac{1}{2}\lambda_t^T \lambda_t - \lambda_t^T \epsilon_{t+1}}, \tag{52}$$

where,

$$\epsilon_{t+1} \sim \mathcal{N}(0, I). \tag{53}$$

The market price of risk, λ_t, has the essentially-affine form, suggested by Duffee (2002), as follows

$$\lambda_t = \lambda + \Lambda X_t. \tag{54}$$

The market price of risk specification describes the aggregrate market risk preferences for holding fixed-income securities.

It is important to note that the same source of randomness, ϵ_{t+1}, arises in the dynamics of the Radon-Nikodým derivative and our state-variable system. We will make use of this in the derivation of the recursion relation used to represent the price of an arbitrary pure-discount bond, $P_{n,t}$. The idea is that the pure-discount bond price is an exponential-affine function of the state variables. Using the pricing kernel—which embeds the

Radon-Nikodým derivative and the market price of risk—the pure-discount bond price can be determined as,

$$
\begin{aligned}
P_{n+1,t} &= \exp\left\{A_{n+1} + B_{n+1}^T X_t\right\}, && (55)\\
&= \mathbb{E}\left(P_{n,t+1} M_{t+1} \middle| \mathcal{F}_t\right),\\
&= \mathbb{E}\left(\exp\left\{A_n + B_n^T X_{t+1}\right\} \underbrace{\exp\left\{-r_t - \frac{1}{2}\lambda_t^T \lambda_t - \lambda_t^T \epsilon_{t+1}\right\}}_{\text{Equation (52)}} \middle| \mathcal{F}_t\right),\\
&= \mathbb{E}\left(\left(e^{A_n + B_n^T X_{t+1}}\right) \exp\left\{\underbrace{-\delta - \Delta^T X_t}_{\text{Equation (1)}} - \frac{1}{2}\lambda_t^T \lambda_t - \lambda_t^T \epsilon_{t+1}\right\} \middle| \mathcal{F}_t\right),\\
&= \mathbb{E}\left(\exp\left(A_n + B_n^T X_{t+1} - \delta - \Delta^T X_t - \frac{1}{2}\lambda_t^T \lambda_t - \lambda_t^T \epsilon_{t+1}\right) \middle| \mathcal{F}_t\right),\\
&= \mathbb{E}\left(\exp\left(A_n + B_n^T\left(\Phi X_t + \Sigma \epsilon_{t+1}\right) - \delta - \Delta^T X_t - \frac{1}{2}\lambda_t^T \lambda_t - \lambda_t^T \epsilon_{t+1}\right) \middle| \mathcal{F}_t\right),\\
&= \mathbb{E}\left(\exp\left(A_n - \delta + (B_n^T \Phi - \Delta^T) X_t - \frac{1}{2}\lambda_t^T \lambda_t + \left(B_n^T \Sigma - \lambda_t^T\right)\epsilon_{t+1}\right) \middle| \mathcal{F}_t\right),\\
&= \exp\left(A_n - \delta + (B_n^T \Phi - \Delta^T) X_t - \frac{1}{2}\lambda_t^T \lambda_t\right) \mathbb{E}\left(\exp\left(\left(B_n^T \Sigma - \lambda_t^T\right)\epsilon_{t+1}\right) \middle| \mathcal{F}_t\right).
\end{aligned}
$$

We use the properties of a lognormal random variable, $e^{\epsilon_{t+1}}$, to resolve the conditional expectation. In particular, if $X \in \mathbb{R}^{n \times 1}$ is normally distributed and $b \in \mathbb{R}^{n \times 1}$, then

$$
\mathbb{E}(e^{b^T X}) = \exp\left(b^T \mathbb{E}(X) + \frac{b^T \mathrm{var}(X) b}{2}\right), \tag{56}
$$

and this applies to both conditional and unconditional expectations. As ϵ_{t+1} has a zero mean, the corresponding form of equation (56) is,

$$
\mathbb{E}(e^{b^T \epsilon_{t+1}}) = \exp\left(\frac{b^T \mathrm{var}(\epsilon_{t+1}) b}{2}\right). \tag{57}
$$

This useful result implies that the expectation term in the final line of equation (55) can now be written as,

$$
\begin{aligned}
\mathbb{E}\left(\exp\left(\left(B_n^T \Sigma - \lambda_t^T\right)\epsilon_{t+1}\right) \middle| \mathcal{F}_t\right) &= \exp\left(\frac{1}{2}\left(B_n^T \Sigma - \lambda_t^T\right) \underbrace{\mathrm{var}\left(\epsilon_{t+1} \mid \mathcal{F}_t\right)}_{I \in \mathbb{R}^{(np+3)\times(np+3)}} \left(B_n^T \Sigma - \lambda_t^T\right)^T\right), && (58)\\
&= \exp\left(\frac{1}{2}\left(B_n^T \Sigma \Sigma^T B_n - 2 B_n^T \Sigma \lambda_t + \lambda_t^T \lambda_t\right)\right).
\end{aligned}
$$

If we now substitute the results of equation (58) into equation (55) to solve for the recursive form of the pure-discount bond price,

$$
\begin{aligned}
P_{n+1,t} &= \exp\left(A_n - \delta + (B_n^T\Phi - \Delta^T)X_t - \frac{1}{2}\lambda_t^T\lambda_t\right)\mathbb{E}\left(\exp\left(\left(B_n^T\Sigma - \lambda_t^T\right)\epsilon_{t+1}\right)\middle|\,\mathcal{F}_t\right), \qquad (59)\\
&= \exp\left(A_n - \delta + (B_n^T\Phi - \Delta^T)X_t - \frac{1}{2}\lambda_t^T\lambda_t\right)\underbrace{\exp\left(\frac{1}{2}\left(B_n^T\Sigma\Sigma^T B_n - 2B_n^T\Sigma\lambda_t + \lambda_t^T\lambda_t\right)\right)}_{\text{Equation (58)}},\\
&= \exp\left(A_n - \delta + \frac{1}{2}B_n^T\Sigma\Sigma^T B_n + (B_n^T\Phi - \Delta^T)X_t - B_n^T\Sigma\left(\underbrace{\lambda + \Lambda X_t}_{\text{Equation (54)}}\right)\right),\\
&= \exp\left(\underbrace{A_n - \delta + \frac{1}{2}B_n^T\Sigma\Sigma^T B_n - B_n^T\Sigma\lambda}_{A_{n+1}} + \underbrace{\left(B_n^T\left(\Phi - \Sigma\Lambda\right) - \Delta^T\right)}_{B_{n+1}^T}X_t\right).
\end{aligned}
$$

To complete the recursion relation, we require the values of A_1 and B_1. These values, in turn, rely on the boundary condition that must come from the economics of the problem. In particular, it is clear that the value of a zero coupon bond at maturity is unity. That is, $P_{0,t} = 1$. This implies that A_0 and B_0 are identically zero.[40] Thus, to find A_1, we merely evaluate equation (60) using our boundary condition as follows,

$$
\begin{aligned}
P_{1,t} &= \exp\left(A_1 + B_1 X_t\right), \qquad (60)\\
&= \exp\left(A_0 - \delta + \frac{1}{2}B_0^T\Sigma\Sigma^T B_0 - B_0^T\Sigma\lambda + (B_0^T\left(\Phi - \Sigma\Lambda\right) - \Delta^T)X_t\right).\\
&= \exp\left(-\delta - \Delta^T X_t\right),\\
&= \exp\left(A_1 + B_1^T X_t\right),
\end{aligned}
$$

implying that $A_1 = -\delta$ and $B_1 = -\Delta$.

We can see, therefore, that the model has a fairly large number of potential parameters. The parameter vector has the elements,

$$
\theta = \{\delta, \Delta, \Phi, \Sigma, \lambda, \Lambda\}. \qquad (61)
$$

We saw in equations (41) and (42) that Φ and Σ are sparse block matrices. Given their importance, it is worth mentioning a few points about the market price of risk.

$$
\lambda_t = \lambda + \Lambda X_t, \qquad (62)
$$

where $\lambda \in \mathbb{R}^{(np+3)\times 1}$ and $\Lambda \in \mathbb{R}^{(np+3)\times(np+3)}$. These are rather sizable matrices and without some restrictions would lead to enormous difficulty in solving the estimation problem. As a consequence, in Ang and Piazzesi

[40]More specifically, $P_{0,t} = \exp(A_0 + B_0^T X_t)$. Clearly, this can only equate to unity if $A_0, B_0 = 0$.

(2003), λ and Λ are assumed to be sparse matrices. The basic idea is that only the contemporaneous macroeconomic and latent term-structure variables can influence the market price. Moreover, the market price of risk specification for the macroeconomic factors are assumed to be independent of the market price of risk specification for the latent term-structure factors. This implies that only 5 (i.e., $p+3$) elements of λ are non-zero,

$$\lambda = \begin{bmatrix} \lambda(1) & \lambda(2) & 0 & \cdots & 0 & \lambda(np+1) & \lambda(np+2) & \lambda(np+3) \end{bmatrix}^T. \tag{63}$$

The Λ matrix that pre-multiplies the state-variable vector, X_t, is a sparse block matrix of the form,

$$\Lambda = \begin{bmatrix} \Lambda^o_{n\times n} & & 0_{n\times(n(p-1)+3)} \\ & 0_{n(p-1)\times(np+3)} & \\ 0_{3\times np} & & \Lambda^u_{3\times 3} \end{bmatrix}^T, \tag{64}$$

implying that only 13 (i.e., n^2+3^2) elements of Λ are non-zero. These dimension-reducing assumptions greatly ease the estimation of the model parameters.

How, therefore, do we estimate this model? In principle, the estimation algorithm follows from Chen and Scott (1993), but it has a few twists. The biggest difference comes from the fact that we need to incorporate the macroeconomic factors into the likelihood, but we need to ensure that the model can be reasonably estimated. The assumption of the independence of the observable and unobservable term-structure factors implies that the model can be estimated in two steps. The first step involves estimation of VAR(p) describing equation (32).[41] This yields the upper left-hand block in equations (41) and (42): the matrices Φ^o and Σ^o. We then proceed to estimate the parameters δ and Δ_1 in equation (1) using OLS. This implies that only the parameters relating to the latent state variables and the affine term-structure model remain to be estimated. This involves 27 parameters from three sources. There remain six parameters in the lower-triangular matrix, Φ^u, there are three parameters in Δ_2, and a final 18 parameters in λ and Λ.

Another way to see this is that we have actually partitioned the parameter space according to the two steps of the estimation algorithm,

$$\theta_1 = \{\delta, \Delta_1, \Phi^o, \Sigma^o\}, \tag{65}$$

and,

$$\theta_2 = \{\Delta_2, \Phi^u, \lambda, \Lambda\}, \tag{66}$$

[41] We should note that the general Ang-Piazzesi formulation permits a VAR(p) model, whereas in our application we specify on VAR(1) dynamics. Our results, which are consistent with Ang and Piazzesi (2003), suggested that a VAR(p) could not produce useful out-of-sample forecasts.

where, of course, $\theta = \{\theta_1, \theta_2\}$. The parameters θ_1 are estimated in the first stage, while θ_2 are estimated in the second stage with the assumption that $\bar{\theta}_2$.

The Chen and Scott (1993) approach assumes that k zero-coupon rates are observed without error—and used to infer the values of the latent state variables—and m zero-coupon rates are observed with error. The slight twist is that, in this setting, it is necessary to include the macroeconomic variables in the likelihood function. The consequence is a slightly different mathematical form. We begin with writing the system of k zero-coupon rates observed without error. It is necessary to partition the matrix of factor loadings on the state-variable vector. To ease the notation, let us set $K = \dim(X_t) = np + 3$, and solve for the implied unobserved state-variable vector as follows,

$$\begin{aligned}
\begin{bmatrix} y_{n_1,t} \\ \vdots \\ y_{n_k,t} \end{bmatrix} &= \begin{bmatrix} A_{n_1} \\ \vdots \\ A_{n_k} \end{bmatrix} + \begin{bmatrix} B_{n_1}^T \\ \vdots \\ B_{n_k}^T \end{bmatrix} X_t, \qquad (67) \\
&= \begin{bmatrix} A_{n_1} \\ \vdots \\ A_{n_k} \end{bmatrix} + \left[\underbrace{\begin{bmatrix} B_{n_1}^T(1) & \cdots & B_{n_1}^T(K-k) \\ \vdots & \ddots & \vdots \\ B_{n_k}^T(1) & \cdots & B_{n_k}^T(K-k) \end{bmatrix}}_{B_k^o} \overbrace{\begin{bmatrix} B_{n_1}^T(K-k+1) & \cdots & B_{n_1}^T(K) \\ \vdots & \ddots & \vdots \\ B_{n_k}^T(K-k+1) & \cdots & B_{n_k}^T(K) \end{bmatrix}}^{B_k^u} \right] \begin{bmatrix} X_t^o \\ X_t^u \end{bmatrix}, \\
Y_t^k &= A_k + B_k^o X_t^o + B_k^u X_t^u, \\
\hat{X}_t^u &= (B_k^u)^{-1} \left(Y_t^k - A_k - B_k^0 X_t^0 \right).
\end{aligned}$$

Armed with the implied unobservable state variables, in $\hat{X}_t^u$, we can now construct a larger system of zero-coupon rates. There are two points in this step. First, it is important for the estimation of the model to use more than k zero-coupon yields. Let us assume that we will use $N > k$ zero-coupon rates in the estimation algorithm, where $(N-k)$ zero-coupon rates are assumed to be observed without error. Second, we need to write the likelihood function in terms of both the observable and unobservable state variables to ensure that the information in the observable macroeconomic variables is used to estimate the term-structure model parameters.[42] The full model, therefore, has N zero-coupon rates that are a function of $K-k$ observable macroeconomic state variables, k unobservable term-structure state variables, and a vector $u_t^m \in \mathbb{R}^{m \times 1}$ of observation errors. It is represented as

[42]Note that in the structural macro-finance model proposed by Hördahl, Tristani, and Vestin (2004) the observable macroeconomic state variables are used, along with two zero-coupon rates, to infer the value of the unobservable state variables. The likelihood in this setting is smaller, but the structural macroeconomic model enters indirectly through the matrix, M.

follows,

$$Y_t^N = A_N + B_N^o X_t^o + B_N^u \hat{X}_t^u + B_N^m u_t^m \tag{68}$$

where $Y_t^N, A_N \in \mathbb{R}^{N\times 1}$, $B_N^o \in \mathbb{R}^{N\times(K-k)}$, $B_N^u \in \mathbb{R}^{N\times k}$, and $B_N^m \in \mathbb{R}^{N\times m}$ with non-zero elements corresponding to the zero-coupon rates in Y_t^n that are observed with error. The observation errors are assumed to be independent of the state-variable vector, X_t. Moreover, they are assumed to follow an independent, identically distributed Gaussian distribution as,

$$B_N^m u_t^m \sim \mathcal{N}\left(0, B_N^m \left(B_N^m\right)^T\right). \tag{69}$$

This is potentially problematic. If, for example, $N = 4$, $k = 2$, and the first and third zero-coupon rates are assumed to be observed with error (i.e., $m = 2$) then B_4^2 would have the form,

$$B_4^2 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & 0 \\ 0 & \sigma_3 \\ 0 & 0 \end{bmatrix}. \tag{70}$$

The associated variance-covariance matrix of the observation errors would have the form,

$$B_4^2 = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{71}$$

which is, of course, singular. We will need to be accordingly cautious when constructing the contribution of the observation errors to the likelihood function; more specifically, it will be necessary to avoid direction determinant and inverse computations of equation (71).

How do we, therefore, construct the likelihood function? The idea is that we would like to write down the joint conditional density of the zero-coupon rates, the observable macroeconomic variables, and the observation errors. That is, we would like to be able to maximize

$$\prod_{t=2}^{T} f\left(Y_t^N, X_t^o \mid \mathcal{F}_t\right), \tag{72}$$

where $\mathcal{F}_t \triangleq \sigma\{Y_{t-1}^N, X_{t-1}^o\}$ denotes the sigma-algebra generated by Y_{t-1}^N and X_{t-1}^o. The problem is that we do not observe this density, nor do we know its form. Indeed, we only observe X_t^o. We can, however, infer the value of the unobservable state variables using equation (67) and k zero-coupon rates observed without error.

Finally, there is information in the observation errors that can also assist in the estimation algorithm. We use, as suggested by Chen and Scott (1993), the change of variables formula to map the conditional density of the things that we know something about, $(X_t^o, \hat{X}_t^u, u_t^m)$ into the conditional density of what we need, (X_t^o, Y_t^N). This mapping has the usual form,

$$f\left(X_t^o, \hat{X}_t^u, u_t^m \,\middle|\, \underbrace{\sigma\{X_t^o, \hat{X}_t^u\}}_{\mathcal{G}_t}\right) = f\left(X_t^o, Y_t^N \mid \mathcal{F}_t\right) \det|J|, \tag{73}$$

$$\begin{aligned}
f(X_t^o, Y_t^N \mid \mathcal{F}_t) &= \frac{1}{\det|J|} f(X_t^o, \hat{X}_t^u, u_t^m \mid \mathcal{G}_t), \\
&= \frac{1}{\det|J|} f(X_t^o, \hat{X}_t^u \mid \mathcal{G}_t) f(u_t^m \mid \mathcal{G}_t), \\
&= \frac{1}{\det|J|} f(X_t^o, \hat{X}_t^u \mid \mathcal{G}_t) f(u_t^m),
\end{aligned}$$

where the final two steps follow from the independence of the observation errors from the state vector (i.e., X_t and u_t^m are independent). The Jacobian matrix, J, has the form

$$\begin{aligned}
J &= \begin{bmatrix} \frac{\partial X_t^o}{\partial X_t^o} & \frac{\partial X_t^o}{\partial \hat{X}_t^u} & \frac{\partial X_t^o}{\partial u_t^m} \\ \frac{\partial Y_t^N}{\partial X_t^o} & \frac{\partial Y_t^N}{\partial \hat{X}_t^u} & \frac{\partial Y_t^N}{\partial u_t^m} \end{bmatrix}, \\
&= \begin{bmatrix} I_{(K-k)\times(K-k)} & 0_{(K-k)\times k} & 0_{(K-k)\times m} \\ B_N^o & B_N^u & B_N^m \end{bmatrix},
\end{aligned} \tag{74}$$

so that $J \in \mathbb{R}^{(N+K-k)\times(N+K-k)}$ is a square matrix whose determinant exists should it be non-singular.

All that remains is to derive the associated log-likelihood function,

$$\begin{aligned}
\prod_{t=2}^{T} f\left(Y_t^N, X_t^o \mid \mathcal{F}_t\right) &= \prod_{t=2}^{T} \frac{1}{\det|J|} f(X_t^o, \hat{X}_t^u \mid \mathcal{G}_t) f(u_t^m), \\
&= \prod_{t=2}^{T} \Bigg(\det|J|^{-1} (2\pi)^{-\frac{N}{2}} \det\left(\Sigma\Sigma^T\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\phi_t^T \left(\Sigma\Sigma^T\right)^{-1} \phi_t\right) \\
&\qquad (2\pi)^{-\frac{m}{2}} \det\left(B_N^m \left(B_N^m\right)^T\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \left(u_t^m\right)^T \left(B_N^m \left(B_N^m\right)^T\right)^{-1} u_t^m\right)\Bigg), \\
\sum_{t=2}^{T} \ln f\left(Y_t^N, X_t^o \mid \mathcal{F}_t\right) &= -(T-1)\ln\det|J| - \frac{(N+m)(T-1)}{2}\ln(2\pi) - \frac{(T-1)}{2}\ln\det\left(\Sigma\Sigma^T\right) \\
&\quad - \frac{(T-1)}{2}\ln\det\left(B_N^m \left(B_N^m\right)^T\right) - \frac{1}{2}\sum_{t=2}^{T}\left(\phi_t^T \left(\Sigma\Sigma^T\right)^{-1}\phi_t + \left(u_t^m\right)^T \left(B_N^m \left(B_N^m\right)^T\right)^{-1} u_t^m\right), \\
\ln\mathcal{L}(\bar{\theta}_1, \theta_2) &= -(T-1)\ln\det|J| - \frac{(N+m)(T-1)}{2}\ln(2\pi) - \frac{(T-1)}{2}\ln\det\left(\Sigma\Sigma^T\right) \\
&\quad - \frac{(T-1)}{2}\ln\left(\sum_{j=1}^{m}\sigma_j^2\right) - \frac{1}{2}\sum_{t=2}^{T}\left(\phi_t^T \left(\Sigma\Sigma^T\right)^{-1}\phi_t + \sum_{j=1}^{m}\frac{\left(u_t^m(j)\right)^2}{\sigma_j^2}\right),
\end{aligned} \tag{75}$$

where $\phi_t = X_t - \Phi X_{t-1}$ and the final manipulations are required to avoid the problems with the application of determinant and inverse operators to the $B_N^m (B_N^m)^T$ matrix as described in equation (71).

Using our dataset, we then proceed to use a non-linear optimization algorithm to determine the parameters for each of the Ang-Piazzesi models. This is, however, a high-dimensional non-linear optimization problem. In this setting, one can never—absent strong restrictions on the mathematical form of the objective functions, which cannot be made in this case—be certain to have 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 similar to that used in Bolder (2006). First, we find a starting value by evaluating 1,000 uniform randomly selected starting parameter vectors. 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., a gradient-based method that is a generalization of the well-known Gauss-Newton algorithm) 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 then 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.

B Deriving the Observed-Affine Model

This technical appendix walks through the calculations required for the construction of the observed-affine model. We will examine two cases. The first case describes a VAR(1) process for the state-variable dynamics while the second cases considers the VAR(2) process used in our analysis.

The first case is a relatively straightforward application of the approach used by Ang and Piazzesi (2003). We begin with the state-variable vector,

$$X_t = \begin{bmatrix} X_t^{\mathrm{pc}} \\ X_t^o \end{bmatrix}, \tag{76}$$

where X_t^{pc} denotes the first three principal components associated with zero-coupon curve movements and X_t^o are output (i.e., x_t), inflation (i.e., π_t), and the monetary policy rate (i.e., r_t) respectively. In this way, all of the state-variables are observable. The state-variable dynamics have the form,

$$X_{t+1} = C + FX_{t-1} + \Sigma\epsilon_{t+1}, \tag{77}$$

the market price of risk has the usual form,

$$\lambda_t = \lambda + \Lambda X_t, \tag{78}$$

the pricing kernel is,

$$M_{t+1} = e^{-r_t}\frac{\varphi_{t+1}}{\varphi_t}, \tag{79}$$

and the Radon Nikodỳm derivative is,

$$\frac{\varphi_{t+1}}{\varphi_t} = e^{-\frac{1}{2}\lambda_t^T\lambda_t - \lambda_t^T\epsilon_{t+1}}. \tag{80}$$

If we note that the instantaneous short rate, as proxied by the monetary policy rate, is observed. It is the final element in our state-variable vector, thus we define,

$$\mathbb{I}_r = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{81}$$

so that we have,

$$\mathbb{I}_r X_t = r_t. \tag{82}$$

This, therefore, implies that the pricing kernel has the following straightforward form,

$$M_{t+1} = \exp\left\{ -\overbrace{\mathbb{I}_r X_t}^{\substack{\text{Equation}\\(82)}} - \frac{1}{2}\lambda_t^T\lambda_t - \lambda_t^T\epsilon_{t+1} \right\}. \tag{83}$$

We can now proceed to determine the form of the affine mapping between the state variables and pure-discount bond prices. Consider the following,

$$
\begin{aligned}
P_{n+1,t} &= e^{A_{n+1}+B_{n+1}^T X_t}, \qquad (84)\\
&= \mathbb{E}\left(P_{n,t+1}M_{t+1}\middle|\,\mathcal{F}_t\right),\\
&= \mathbb{E}\left(\exp\left\{A_n + B_n^T X_{t+1}\right\}\underbrace{\exp\left\{-\mathbb{I}_r X_t - \frac{1}{2}\lambda_t^T\lambda_t - \lambda_t^T\epsilon_{t+1}\right\}}_{\text{Equation (83)}}\middle|\,\mathcal{F}_t\right),\\
&= \mathbb{E}\left(\exp\left\{A_n + B_n^T\underbrace{\left(C+FX_t+\Sigma\epsilon_{t+1}\right)}_{\text{Equation (77)}} - \mathbb{I}_r X_t - \frac{1}{2}\lambda_t^T\lambda_t - \lambda_t^T\epsilon_{t+1}\right\}\middle|\,\mathcal{F}_t\right),\\
&= \mathbb{E}\left(\exp\left\{A_n + B_n^T C + \left(B_n^T F - \mathbb{I}_r\right)X_t - \frac{1}{2}\lambda_t^T\lambda_t + \left(B_n^T\Sigma - \lambda_t^T\right)\epsilon_{t+1}\right\}\middle|\,\mathcal{F}_t\right),\\
&= \exp\left\{A_n + B_n^T C + \left(B_n^T F - \mathbb{I}_r\right)X_t - \frac{1}{2}\lambda_t^T\lambda_t\right\}\mathbb{E}\left(\exp\left\{\left(B_n^T\Sigma - \lambda_t^T\right)\epsilon_{t+1}\right\}\middle|\,\mathcal{F}_t\right),\\
&= \exp\left\{A_n + B_n^T C + \left(B_n^T F - \mathbb{I}_r\right)X_t - \frac{1}{2}\lambda_t^T\lambda_t\right\}\exp\left\{\frac{1}{2}B_n^T\Sigma\Sigma^T B_n - B_n^T\Sigma\lambda_t + \frac{1}{2}\lambda_t^T\lambda_t\right\},\\
&= \exp\left\{A_n + B_n^T C + \left(B_n^T F - \mathbb{I}_r\right)X_t + \frac{1}{2}B_n^T\Sigma\Sigma^T B_n - B_n^T\Sigma\underbrace{\left(\lambda+\Lambda X_t\right)}_{\text{Equation (78)}}\right\},\\
&= \exp\left\{\underbrace{A_n + B_n^T(C-\Sigma\lambda) + \frac{1}{2}B_n^T\Sigma\Sigma^T B_n}_{A_{n+1}} + \underbrace{\left(B_n^T\left(F-\Sigma\Lambda\right)-\mathbb{I}_r\right)}_{B_{n+1}^T}X_t\right\}.
\end{aligned}
$$

Given that A_0 and B_0 must be identically zero, it follows that the boundary conditions for starting the recursion are $A_1 = 0$ and $B_1 = -\mathbb{I}_r$.

The second case requires a small trick in order to permit a reasonable solution. In particular, use of a VAR(2) process to describe the state-variable dynamics implies the following specification,

$$
X_{t+1} = C + F_1 X_t + F_2 X_{t-1} + \Sigma\epsilon_{t+1}. \qquad (85)
$$

If we attempt to solve for A_{n+1} and B_{n+1} directly, we will lose the affine form given that an X_{t-1} term will arise. We solve this problem—and this is the trick—by making the following adjustment to the market price of risk,

$$
\lambda_t = \lambda + \Lambda X_t + \Sigma^{-1}F_2 X_{t-1}. \qquad (86)
$$

This may seem somewhat odd, but it is perfectly legitimate. Moreover, it permits us to maintain the convenient exponential affine pure-discount price form, while simultaneously allowing the use of a VAR(2) specification for the state-variable dynamics.

If we proceed to follow the same process as described in equation (84) we arrive at an identical form for the affine mapping between the state variables and pure-discount bond prices as,

$$
\begin{aligned}
P_{n+1,t} &= e^{A_{n+1}+B_{n+1}^T X_t}, \\
&= \mathbb{E}\left(P_{n,t+1} M_{t+1} \middle| \mathcal{F}_t\right), \\
&= \mathbb{E}\left(\exp\left\{A_n + B_n^T X_{t+1}\right\} \underbrace{\exp\left\{-\mathbb{I}_r X_t - \frac{1}{2}\lambda_t^T \lambda_t - \lambda_t^T \epsilon_{t+1}\right\}}_{\text{Equation (83)}} \middle| \mathcal{F}_t\right), \\
&= \mathbb{E}\left(\exp\left\{A_n + B_n^T \underbrace{(C + F_1 X_t + F_2 X_{t-1} + \Sigma\epsilon_{t+1})}_{\text{Equation (85)}} - \mathbb{I}_r X_t - \frac{1}{2}\lambda_t^T \lambda_t - \lambda_t^T \epsilon_{t+1}\right\} \middle| \mathcal{F}_t\right), \\
&= \mathbb{E}\left(\exp\left\{A_n + B_n^T C + \left(B_n^T F_1 - \mathbb{I}_r\right) X_t + B_n^T F_2 X_{t-1} - \frac{1}{2}\lambda_t^T \lambda_t + \left(B_n^T \Sigma - \lambda_t^T\right)\epsilon_{t+1}\right\} \middle| \mathcal{F}_t\right), \\
&= \exp\left\{A_n + B_n^T C + \left(B_n^T F - \mathbb{I}_r\right) X_t + B_n^T F_2 X_{t-1} - \frac{1}{2}\lambda_t^T \lambda_t\right\} \mathbb{E}\left(\exp\left\{\left(B_n^T \Sigma - \lambda_t^T\right)\epsilon_{t+1}\right\} \middle| \mathcal{F}_t\right), \\
&= \exp\left\{A_n + B_n^T C + \left(B_n^T F - \mathbb{I}_r\right) X_t + B_n^T F_2 X_{t-1} - \frac{1}{2}\lambda_t^T \lambda_t\right\} \exp\left\{\frac{1}{2} B_n^T \Sigma\Sigma^T B_n - B_n^T \Sigma\lambda_t + \frac{1}{2}\lambda_t^T \lambda_t\right\}, \\
&= \exp\left\{A_n + B_n^T C + \left(B_n^T F - \mathbb{I}_r\right) X_t + B_n^T F_2 X_{t-1} + \frac{1}{2} B_n^T \Sigma\Sigma^T B_n - B_n^T \Sigma\left(\underbrace{\lambda + \Lambda X_t + \Sigma^{-1} F_2 X_{t-1}}_{\text{Equation (86)}}\right)\right\}, \\
&= \exp\left\{A_n + B_n^T C + \left(B_n^T F - \mathbb{I}_r\right) X_t + B_n^T F_2 X_{t-1} + \frac{1}{2} B_n^T \Sigma\Sigma^T B_n - B_n^T \Sigma\lambda - B_n^T \Sigma\Lambda X_t - B_n^T F_2 X_{t-1}\right\}, \\
&= \exp\left\{\underbrace{A_n + B_n^T (C - \Sigma\lambda) + \frac{1}{2} B_n^T \Sigma\Sigma^T B_n}_{A_{n+1}} + \underbrace{\left(B_n^T \left(F - \Sigma\Lambda\right) - \mathbb{I}_r\right)}_{B_{n+1}^T} X_t\right\}.
\end{aligned}
\tag{87}
$$

As we have the same form as in equation (84), it follows that the boundary conditions for starting the recursion remain $A_1 = 0$ and $B_1 = -\mathbb{I}_r$.

The estimation of this model is now straightforward. It is performed in two steps. In the first step, the physical-measure parameters (i.e., C, F_1, F_2, and Σ), are estimated from the observed dataset using an ordinary least-squares regression. In the second step, the remaining pricing-measure parameters (i.e., λ and Λ), are determined using a non-linear least-squares regression with the zero-coupon panel data and equation (87).

C Deriving the Nelson-Siegel Model

The intent of this section is to establish the basic logic behind the construction of the Nelson and Siegel (1987) model. They suggested the following form for the instantaneous forward rate,

$$f(v) = x_0 + x_1 e^{-\lambda v} + x_2 \lambda v e^{-\lambda v}. \tag{88}$$

By exploiting the fact that $z(\tau) = \frac{1}{\tau}\int_0^\tau f(v)dv$, one can solve for the zero-coupon curve through a straightforward integration by parts,[43]

$$\begin{aligned} z(\tau) &= \frac{1}{\tau}\int_0^\tau \left(x_0 + x_1 e^{-\lambda v} + x_2 \lambda v e^{-\lambda v}\right) dv, \\ &= x_0 + x_1\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + x_2\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right). \end{aligned} \tag{89}$$

Diebold and Li (2003) transformed this into a dynamic term-structure model by letting the coefficients x_0, x_1, and x_2 become state variables with the following factor loadings,

$$f_0(y) = 1, \tag{90}$$

$$f_1(y) = \frac{1-e^{-\lambda y}}{\lambda y}, \tag{91}$$

$$f_2(y) = \frac{1-e^{-\lambda y}}{\lambda y} - e^{-\lambda y}, \tag{92}$$

for $y \in \mathbb{R}$.

This model has attained a fairly comfortable spot in the term-structure literature. An interesting question, however, is what is the intuition behind these factor loadings. Hurn, Lindsay, and Pavlov (2005) provide an extension of the Nelson-Siegel model that provides some insight into the origins of the factor loadings and, more importantly, a hint as to how their approach might be extended. The key result comes from the following representation theorem used in spectral analysis.

Theorem C.1 *If $\gamma(x)$ is continuous on $(0,\infty)$ and $\beta_1 \in \mathbb{R}$ such that,*

$$\int_0^\infty e^x \left(\gamma(x) - \beta_1\right)^2 dx < \infty, \tag{93}$$

then $\gamma(x)$ has a pointwise Fourier-Laguerre series representation of the following form,

$$\gamma(x) = \beta_1 + e^{-\lambda x}\sum_{k=0}^{\infty} c_k L_k(\lambda x), \tag{94}$$

where $L_k(\lambda x)$ is a kth degree Laguerre polynomial and $\beta_1, \lambda, c_0, c_1, \ldots \in \mathbb{R}$ are coefficients to be determined.

[43] See Bolder (2006) for the details.

This result essentially applies that a positive, bounded, continuous function can be described as a relatively simple function of an infinite sum of Laguerre polynomials. Reconsidering equation (88), we see that Theorem C.1 can be applied to the instantaneous forward-rate equation. That is, if we let $x = v$, then the instantaneous forward-rate has the following representation,

$$f(v) = \beta_1 + e^{-\lambda v} \sum_{k=0}^{N} c_k L_k(\lambda v), \tag{95}$$

where the sum of Laguerre polynomials is truncated to N terms. What is required is a way to represent the Laguerre polynomial, $L_k(x)$.[44] There are a variety of ways to describe Laguerre polynomials, but probably the most straightforward arises from the Rodrigues formula, which is a more general result.[45] The specific result for the Laguerre polynomial has the form,

$$L_k(x) = \frac{e^x}{k!} \frac{\partial^k}{\partial x^k} \left(x^k e^{-x} \right). \tag{97}$$

We can now easily evaluate the first few Laguerre polynomials for $k = 0$,

$$\begin{aligned} L_0(x) &= \frac{1}{0!} e^x \frac{\partial^0}{\partial (x)^0} \left((x)^0 e^{-x} \right), \\ &= 1, \end{aligned} \tag{98}$$

for $k = 1$,

$$\begin{aligned} L_1(x) &= \frac{1}{1!} e^x \frac{\partial^1}{\partial (x)^1} \left((x)^1 e^{-x} \right), \\ &= e^x \left(e^{-x} - x e^{-x} \right), \\ &= 1 - x, \end{aligned} \tag{99}$$

and for $k = 2$

$$\begin{aligned} L_2(x) &= \frac{1}{2!} e^x \frac{\partial^2}{\partial x^2} \left(x^2 e^{-x} \right), \\ &= \frac{1}{2} e^x \frac{\partial}{\partial x} \left(2x e^{-x} - x^2 e^{-x} \right), \\ &= \frac{1}{2} e^x \left(2e^{-x} - 4x e^{-x} + x^2 e^{-x} \right), \\ &= \frac{1}{2} \left(x^2 - 4x + 2 \right). \end{aligned} \tag{100}$$

[44]Laguerre polynomials arise as the solution to the following differential equation,

$$x \frac{\partial^2}{\partial x^2} f(x) + (1 - x) \frac{\partial}{\partial x} f(x) + n f(x) = 0, \tag{96}$$

for $n \in \{0, \mathbb{Z}^+\}$.

[45]See Abramovitz and Stegun (1965, 774-776) for more detail on Rodrigues' formula.

This is all that is required to derive the functional form of the Nelson and Siegel (1987) zero-coupon curve. We take the corresponding zero-coupon curve, which follows from equation (89), and use the representation theorem as described in equation (95) with only the first two terms (i.e., $N = 1$). This yields the following expression that can be simplified as,

$$
\begin{aligned}
z(\tau) &= \frac{1}{\tau}\int_0^{\tau}\left(\beta_1 + e^{-\lambda v}\sum_{k=0}^{1} c_k L_k(\lambda v)\right) dv, \qquad (101)\\
&= \beta_1 + \frac{1}{\tau}\sum_{k=0}^{1} c_k \int_0^{\tau} e^{-\lambda v} L_k(\lambda v) dv,\\
&= \beta_1 + \frac{1}{\lambda\tau}\sum_{k=0}^{1} c_k \underbrace{\int_0^{\lambda\tau} e^{-x} L_k(x) dx,}_{\substack{\text{Change of variables}\\ \text{with } x=\lambda v}}\\
&= \beta_1 + \frac{c_0}{\lambda\tau}\int_0^{\lambda\tau} e^{-x}\underbrace{L_0(x)}_{=1} dx + \frac{c_1}{\lambda\tau}\int_0^{\lambda\tau} e^{-x}\underbrace{L_1(x)}_{=1-x} dx,\\
&= \beta_1 + \frac{c_0}{\lambda\tau}\int_0^{\lambda\tau} e^{-x} dx + \frac{c_1}{\lambda\tau}\int_0^{\lambda\tau} e^{-x}(1-x) dx,\\
&= \beta_1 + \frac{c_0}{\lambda\tau}\left(1-e^{-\lambda\tau}\right) + \frac{c_1}{\lambda\tau}\left(\int_0^{\lambda\tau} e^{-x} dx - \int_0^{\lambda\tau} x e^{-x} dx\right),\\
&= \beta_1 + \frac{c_0}{\lambda\tau}\left(1-e^{-\lambda\tau}\right) + \frac{c_1}{\lambda\tau}\left(\left(1-e^{-\lambda\tau}\right) + \underbrace{\left(x e^{-x}\Big|_0^{\lambda\tau} - \int_0^{\lambda\tau} e^{-x} dx\right)}_{\text{Using integration by parts}}\right),\\
&= \beta_1 + \frac{c_0}{\lambda\tau}\left(1-e^{-\lambda\tau}\right) + \frac{c_1}{\lambda\tau}\left(\left(1-e^{-\lambda\tau}\right) - \lambda\tau e^{-\lambda\tau} - \left(1-e^{-\lambda\tau}\right)\right),\\
&= \beta_1 + c_0\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + c_1 e^{-\lambda\tau}.
\end{aligned}
$$

If we now set $\beta_1 = x_0$, $c_0 = x_1 + x_2$, and $c_1 = -x_2$, we arrive at

$$
\begin{aligned}
z(\tau) &= \underbrace{x_0}_{\beta_1} + \left(\underbrace{x_1 + x_2}_{c_0}\right)\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)\underbrace{-x_2}_{c_1} e^{-\lambda\tau}, \qquad (102)\\
&= x_0 + x_1\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + x_2\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right),
\end{aligned}
$$

which is exactly the form suggested by Nelson and Siegel (1987) as shown in equation (89). We see, therefore, that the Nelson-Siegel model is, in fact, a consequence of the represention result summarized in Theorem C.1.

Hurn, Lindsay, and Pavlov (2005) proceed to extend this result. In particular, they demonstrate that one can easily create additional terms to the Nelson-Siegel model merely by changing the truncation of the infinite sum of

Laguerre polynomials. More specifically, they set $N = 2$ and $N = 3$ to create four- and five-factor Nelson-Siegel models respectively. To obtain the fourth and fifth terms, we can proceed with direct integration as we did in equation (101). This could get a bit tedious. Conversely, we may—as suggested by Hurn, Lindsay, and Pavlov (2005)—use a small trick. In particular, we manipulate Rodrigues' formula, as shown in equation (97), to get the following very useful result,

$$
\begin{aligned}
L_k(x) &= \frac{e^x}{k!}\frac{\partial^k}{\partial x^k}\left(x^k e^{-x}\right), \\
e^{-x}L_k(x) &= e^{-x}\left(\frac{e^x}{k!}\frac{\partial^k}{\partial x^k}\left(x^k e^{-x}\right)\right), \\
\int_0^{\lambda\tau} e^{-x}L_k(x)dx &= \frac{1}{k!}\int_0^{\lambda\tau}\left(\frac{\partial^k}{\partial x^k}\left(x^k e^{-x}\right)\right)dx, \\
&= \frac{1}{k!}\left(\frac{\partial^{k-1}}{\partial x^{k-1}}\left(x^k e^{-x}\right)\right)\Bigg|_0^{\lambda\tau}, \\
&= \frac{1}{k!}\frac{\partial^{k-1}}{\partial x^{k-1}}\left(x^k e^{-x}\right)\Big|_{x=\lambda\tau}.
\end{aligned}
\tag{103}
$$

This result permits us to compute the fourth and fifth terms quite easily. In particular, for $k = 2$ we have that,

$$
\begin{aligned}
\int_0^{\lambda\tau} e^{-x}L_2(x)dx &= \frac{1}{2!}\frac{\partial^{2-1}}{\partial x^{2-1}}\left(x^2 e^{-x}\right)\Big|_{x=\lambda\tau}, \\
&= \frac{1}{2}\frac{\partial}{\partial x}\left(x^2 e^{-x}\right)\Big|_{x=\lambda\tau}, \\
&= \frac{1}{2}\left(2xe^{-x} - x^2 e^{-x}\right)\Big|_{x=\lambda\tau}, \\
&= \frac{1}{2}\left(2\lambda\tau e^{-\lambda\tau} - (\lambda\tau)^2 e^{-\lambda\tau}\right), \\
&= \lambda\tau e^{-\lambda\tau}\left(1 - \frac{\lambda\tau}{2}\right),
\end{aligned}
\tag{104}
$$

and $k = 3$ has the following form,

$$
\begin{aligned}
\int_0^{\lambda\tau} e^{-x}L_3(x)dx &= \frac{1}{3!}\frac{\partial^{3-1}}{\partial x^{3-1}}\left(x^3 e^{-x}\right)\Big|_{x=\lambda\tau}, \\
&= \frac{1}{6}\frac{\partial^2}{\partial x^2}\left(x^3 e^{-x}\right)\Big|_{x=\lambda\tau}, \\
&= \frac{1}{6}\left(x^3 e^{-x} - 6x^2 e^{-x} + 6xe^{-x}\right)\Big|_{x=\lambda\tau}, \\
&= \frac{1}{6}\left((\lambda\tau)^3 e^{-\lambda\tau} - 6(\lambda\tau)^2 e^{-\lambda\tau} + 6\lambda\tau e^{-\lambda\tau}\right), \\
&= \lambda\tau e^{-\lambda\tau}\left(\frac{(\lambda\tau)^2}{6} - \lambda\tau + 1\right).
\end{aligned}
\tag{105}
$$

Now that we have all the pieces, we can write down the five-factor Nelson-Siegel model as follows,

$$
\begin{aligned}
z(\tau) &= x_0 + x_1\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + x_2\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \frac{c_2}{\lambda\tau}\underbrace{\int_0^{\lambda\tau} e^{-x}L_2(x)dx}_{\text{Equation (104)}} + \frac{c_3}{\lambda\tau}\underbrace{\int_0^{\lambda\tau} e^{-x}L_3(x)dx}_{\text{Equation (105)}}, \qquad (106)\\
&= x_0 + x_1\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + x_2\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \frac{c_2}{\lambda\tau}\left(\lambda\tau e^{-\lambda\tau}\left(1-\frac{\lambda\tau}{2}\right)\right) + \frac{c_3}{\lambda\tau}\left(\lambda\tau e^{-\lambda\tau}\left(\frac{(\lambda\tau)^2}{6} - \lambda\tau + 1\right)\right),\\
&= \underbrace{x_0 + x_1\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + x_2\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)}_{\text{Classical Nelson-Siegel model}} + \underbrace{x_3 e^{-\lambda\tau}\left(1-\frac{\lambda\tau}{2}\right) + x_4 e^{-\lambda\tau}\left(\frac{(\lambda\tau)^2}{6} - \lambda\tau + 1\right)}_{\text{Additional factors from Fourier-Laguerre expansion}},
\end{aligned}
$$

where we've set $x_3 = c_2$ and $x_4 = c_3$. With this specific construction, Hurn, Lindsay, and Pavlov (2005) demonstrate how one can cleverly maintain the basic structure of the Nelson-Siegel model while simultaneously adding more flexibility through two new state variables, x_3 and x_4.

D The Conditional Out-of-Sample Forecast Distribution

In order to construct a conditional confidence interval for our out-of-sample forecasts, which are actually the conditional expectation of the zero-coupon forecast, one needs to begin by deriving the conditional variance of the zero-coupon forecast. To obtain the conditional variance of the zero-coupon forecast, we must first compute the conditional variance of the state-variable vector associated with the specific term-structure model. Fortunately, in the collection of models examined in this paper, all of the state-variable dynamics are described by a VAR(p) process. Consider, therefore, the following VAR(1) process,

$$X_t = FX_{t-1} + \epsilon_t, \tag{107}$$

where $X_t \in \mathbb{R}^{n \times 1}$, $F \in \mathbb{R}^{n \times n}$, and

$$\epsilon_t \sim \mathcal{N}\left(\vec{0}, \Omega\right). \tag{108}$$

We can proceed to perform all of the subsequent analysis, without loss of generality, by assuming that the intercept vector, $C \equiv 0$; it turns out to have no impact on the final result and merely complicates the subsequent expressions. The one-period forward conditional expectation, therefore, has the following form,

$$\mathbb{E}\left(X_t \middle| \mathcal{F}_{t-1}\right) = FX_{t-1}, \tag{109}$$

while the one-period forward conditional variance can easily be determined by recalling that X_{t-1} is $\mathcal{F}_{t-1}$-measurable and, therefore, a constant as well as the fact that expecations involving only ϵ_t vanish. Here is the result,

$$\begin{aligned} \text{var}\left(X_t \middle| \mathcal{F}_{t-1}\right) &= \mathbb{E}\left(X_t X_t^T \middle| \mathcal{F}_{t-1}\right) - \mathbb{E}\left(X_t \middle| \mathcal{F}_{t-1}\right)\mathbb{E}\left(X_t \middle| \mathcal{F}_{t-1}\right)^T, \\ &= \mathbb{E}\left((FX_{t-1}+\epsilon_t)(FX_{t-1}+\epsilon_t)^T \middle| \mathcal{F}_{t-1}\right) - \mathbb{E}\left(FX_{t-1}+\epsilon_t \middle| \mathcal{F}_{t-1}\right)\mathbb{E}\left(FX_{t-1}+\epsilon_t \middle| \mathcal{F}_{t-1}\right)^T, \\ &= (FX_{t-1})(FX_{t-1})^T + \mathbb{E}\left(\epsilon_t \epsilon_t^T \middle| \mathcal{F}_{t-1}\right) - FX_{t-1}(FX_{t-1})^T, \\ &= \Omega. \end{aligned} \tag{110}$$

The consequence is that the one-period forward conditional distribution of our VAR(1) state-variables are,

$$X_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}\left(FX_{t-1}, \Omega\right). \tag{111}$$

While this is a useful result, what we require is the n-period forward conditional variance so that we can construct an n-period forward confidence interval. Let's do this in steps by computing the two-period forward confidence interval. First, we require an expression for X_{t+1} in terms of X_{t-1},

$$\begin{aligned} X_{t+1} &= FX_t + \epsilon_{t+1}, \\ &= F(FX_{t-1} + \epsilon_t) + \epsilon_{t+1}, \\ &= F^2 X_{t-1} + F\epsilon_t + \epsilon_{t+1}. \end{aligned} \tag{112}$$

implying that the expectation of X_{t+1} conditioning on the filtration $\mathcal{F}_{t-1}$ is described as,

$$\mathbb{E}\left(X_{t+1} \middle| \mathcal{F}_{t-1}\right) = F^2 X_{t-1}. \tag{113}$$

We now perform the same calculation as in equation (110), by working from the definition of the conditional variance as,

$$\begin{aligned} \operatorname{var}\left(X_{t+1} \middle| \mathcal{F}_{t-1}\right) &= \mathbb{E}\left(X_{t+1}X_{t+1}^T \middle| \mathcal{F}_{t-1}\right) - \mathbb{E}\left(X_{t+1} \middle| \mathcal{F}_{t-1}\right)\mathbb{E}\left(X_{t+1} \middle| \mathcal{F}_{t-1}\right)^T, \qquad (114) \\ &= \mathbb{E}\left((F^2X_{t-1} + F\epsilon_t + \epsilon_{t+1})(F^2X_{t-1} + F\epsilon_t + \epsilon_{t+1})^T \middle| \mathcal{F}_{t-1}\right) \\ &\quad - \mathbb{E}\left(F^2X_{t-1} + F\epsilon_t + \epsilon_{t+1} \middle| \mathcal{F}_{t-1}\right)\mathbb{E}\left(F^2X_{t-1} + F\epsilon_t + \epsilon_{t+1} \middle| \mathcal{F}_{t-1}\right)^T, \\ &= F^2X_{t-1}X_{t-1}^T\left(F^2\right)^T + F\,\mathbb{E}\left(\epsilon_t\epsilon_t^T \middle| \mathcal{F}_{t-1}\right)F^T + \mathbb{E}\left(\epsilon_{t+1}\epsilon_{t+1}^T \middle| \mathcal{F}_{t-1}\right) - F^2X_{t-1}X_{t-1}^T\left(F^2\right)^T, \\ &= F\Omega F^T + \Omega, \\ &= \sum_{j=0}^{1} F^j\Omega\left(F^j\right)^T. \qquad (115) \end{aligned}$$

This implies that the two-period forward conditional distribution for a VAR(1) is described as,

$$X_{t+1} \mid \mathcal{F}_{t-1} \sim \mathcal{N}\left(F^2X_{t-1}, F\Omega F^T + \Omega\right). \tag{116}$$

At this point, we have two choices. One may either, depending on his or her level of patience, continue to solve this expression for an arbitrary n or, logically work out the general n-period term. We prefer the latter approach. Consequently, we can see from equation (114) that if we construct the three-period forward conditional variance, that we will add an additional term of the form $F^2\mathbb{E}\left(\epsilon_t\epsilon_t^T \middle| \mathcal{F}_{t-1}\right)\left(F^2\right)^T$, which leads to the additional term $F^2\Omega\left(F^2\right)^T$.[46] The implication, therefore, is that the general term for the n-period forward conditional variance of a VAR(1) state-variable vector has the form,

$$\operatorname{var}\left(X_{t+n} \middle| \mathcal{F}_{t-1}\right) = \sum_{j=0}^{n} F^j\Omega\left(F^j\right)^T. \tag{118}$$

[46]This stems from equation (114) and the fact that,

$$X_{t+2} = F^3X_{t-1} + F^2\epsilon_t + F\epsilon_{t+1} + \epsilon_{t+2}. \tag{117}$$

There is a slight catch. The state variables are, in our analysis, assumed to have VAR(2) dynamics. In principle, however, this is not a problem as we can place our VAR(2) specification into companion form and use the results from the VAR(1) specification. In particular, let's start with our VAR(2) model,

$$X_t = F_1 X_{t-1} + F_2 X_{t-2} + \epsilon_t, \tag{119}$$

with $X_t \in \mathbb{R}^{n\times 1}$, $F_1, F_2 \in \mathbb{R}^{n\times n}$, and where ϵ_t is distributed as in equation (108). We then proceed to define,

$$Y_t = \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix}, \tag{120}$$

which allows us to write the companion form of Y_t as,

$$\begin{aligned} \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} &= \begin{bmatrix} F_1 & F_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}, \\ Y_t &= GY_{t-1} + u_t, \end{aligned} \tag{121}$$

where the variance-covariance matrix of u_t has dimensionality $2n\times 2n$ and the following distributional properties,

$$u_t \sim \mathcal{N}\left(\vec{0}, \underbrace{\begin{bmatrix} \Omega & 0 \\ 0 & 0 \end{bmatrix}}_{Q}\right). \tag{122}$$

The consequence, therefore, is that the conditional distribution of a one-period forward VAR(2) state-variable follows from equation (111) as,

$$Y_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}\left(GY_{t-1}, Q\right). \tag{123}$$

The conditional variance of the n-period forward VAR(2) state-variable vector is, therefore,

$$\text{var}\left(Y_{t+n} \middle| \mathcal{F}_{t-1}\right) = \sum_{j=0}^{n} G^j Q \left(G^j\right)^T. \tag{124}$$

We are interested in the $n\times n$ matrix in the top right-hand corner of the matrix, var $(Y_{t+n}|\mathcal{F}_{t-1})$, which is var $(X_{t+n}|\mathcal{F}_{t-1})$. This is because we are not interested in the conditional variance of Y_{t+n}, but rather that of X_{t+n}.

Now that we have expressions for the conditional variance of X_{t+n}, we may now proceed to consider the conditional variance of the n-period forward zero-coupon rate forecast. For the discrete-time affine term-structure models used in this paper, the n-period forward zero-coupon rate forecasts, with a tenor of τ periods, has the form,

$$\begin{aligned}\mathbb{E}\left(Z^{\tau}_{t+n}\middle|\mathcal{F}_{t-1}\right) &= \mathbb{E}\left(A_\tau + B_\tau^T X_{t+n}\middle|\mathcal{F}_{t-1}\right), \\ &= A_\tau + B_\tau^T\,\mathbb{E}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right).\end{aligned} \tag{125}$$

while the conditional variance is described as,

$$\begin{aligned}\mathrm{var}\left(Z^{\tau}_{t+n}\middle|\mathcal{F}_{t-1}\right) &= \mathrm{var}\left(A_\tau + B_\tau^T X_{t+n}\middle|\mathcal{F}_{t-1}\right), \\ &= \mathrm{var}\left(B_\tau^T X_{t+n}\middle|\mathcal{F}_{t-1}\right), \\ &= B_\tau^T \underbrace{\mathrm{var}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right)}_{\substack{\text{See equations} \\ \text{(118) and (124)}}} B_\tau.\end{aligned} \tag{126}$$

The implication is that a 95% confidence interval for the n-period forward, τ-tenor zero-coupon forecast from a discrete-time affine term-structure model is,

$$\underbrace{A_\tau + B_\tau^T\,\mathbb{E}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right)}_{\text{Equation (125)}} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\underbrace{B_\tau^T\,\mathrm{var}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right)B_\tau}_{\text{Equation (126)}}}, \tag{127}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the normal cumulative distribution function and α for our purposes is equal to 0.05.

For the Nelson-Siegel model, which represents zero-coupon rates as a linear combination of Laguerre polynomials, the n-period forecast of a τ-period zero-coupon rate,

$$\begin{aligned}\mathbb{E}\left(Z^{\tau}_{t+n}\middle|\mathcal{F}_{t-1}\right) &= \mathbb{E}\left(F(\tau)^T X_{t+n}\middle|\mathcal{F}_{t-1}\right), \\ &= F(\tau)^T\,\mathbb{E}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right),\end{aligned} \tag{128}$$

where $F(\tau)$ denotes the factor-loading vector. As a consequence, the conditional variance of this forecast is,

$$\begin{aligned}\mathrm{var}\left(Z^{\tau}_{t+n}\middle|\mathcal{F}_{t-1}\right) &= \mathrm{var}\left(F(\tau)^T X_{t+n}\middle|\mathcal{F}_{t-1}\right), \\ &= F(\tau)^T \underbrace{\mathrm{var}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right)}_{\substack{\text{See equations} \\ \text{(118) and (124)}}} F(\tau).\end{aligned} \tag{129}$$

This implies that a 95% confidence interval for the n-period forward, τ-tenor zero-coupon forecast from Nelson-Siegel term-structure model is,

$$\underbrace{F(\tau)^T\,\mathbb{E}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right)}_{\text{Equation (128)}} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\underbrace{F(\tau)^T\,\mathrm{var}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right)F(\tau)}_{\text{Equation (129)}}}. \tag{130}$$

Finally, we turn our attention to the exponential-spline and Fourier-series models, which are linear combinations of basis functions operating on the discount function. The forecast and forecast variance will, consequently, have a slightly different form. Indeed, we merely operate in pure-discount bond price space and perform the necessary transformation at the end of the process.[47] In particular, the n-period forecast of a τ-period pure-discount bond price is,

$$\begin{aligned}\mathbb{E}\left(P_{t+n}^{\tau}\middle|\mathcal{F}_{t-1}\right) &= \mathbb{E}\left(G(\tau)^T X_{t+n}\middle|\mathcal{F}_{t-1}\right), \\ &= G(\tau)^T\,\mathbb{E}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right).\end{aligned} \tag{132}$$

The associated conditional variance of the n-period forward τ-period pure-discount bond price, therefore, is

$$\begin{aligned}\operatorname{var}\left(P_{t+n}^{\tau}\middle|\mathcal{F}_{t-1}\right) &= \operatorname{var}\left(G(\tau)^T X_{t+n}\middle|\mathcal{F}_{t-1}\right), \\ &= G(\tau)^T \underbrace{\operatorname{var}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right)}_{\substack{\text{See equations}\\ \text{(118) and (124)}}} G(\tau).\end{aligned} \tag{133}$$

The 95% confidence interval for the n-period forward, τ-tenor zero-coupon bond price is accordingly a slight variation on equation (130) involving the transformation in equation (131) as follows,

$$-\frac{1}{\tau}\ln\left(\underbrace{G(\tau)^T\,\mathbb{E}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right)}_{\text{Equation (132)}} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\underbrace{G(\tau)^T\,\operatorname{var}\left(X_{t+n}\middle|\mathcal{F}_{t-1}\right)G(\tau)}_{\text{Equation (133)}}}\right). \tag{134}$$

[47] Recall that the aforementioned transformation is,

$$z_{t+n}^{\tau} = -\frac{\ln P_{t+n}^{\tau}}{\tau}. \tag{131}$$

Bibliography

Abramovitz, M. and I.A. Stegun. 1965. *Handbook of Mathematical Functions*. New York: Dover Publications.

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

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

Backus, D., S. Foresi, and C. Telmer. 1998. "Discrete-Time Models of Bond Pricing." Graduate School of Industrial Administration, Carnegie Mellon University.

Backus, D., C. Telmer, and L. Wu. 1999. "Design and Estimation of Affine Yield Models." Graduate School of Industrial Administration, Carnegie Mellon University.

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 2001-15.

———. 2002. "Towards a More Complete Debt Strategy Simulation Framework." Bank of Canada: Working Paper 2002-13.

———. 2003. "A Proposed Stochastic Simulation Framework for the Government of Canada's Debt Strategy Problem." Bank of Canada: Working Paper 2003-10.

———. 2006. "Modelling Term-Structure Dynamics for Risk Management: A Practitioner's Perspective." Bank of Canada: Working Paper 2006-48.

———. 2007. "Optimization in a Simulation Setting: Use of Function Approximation in Debt-Strategy Analysis." Bank of Canada: Working Paper (forthcoming).

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 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 2004-48.

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

Brace, A., D. Gatarek, and M. Musiela. 1997. "The Market Model of Interest Rate Dynamics." *Mathematical Finance* 7: 127–155.

Brigo, D. and F. Mercurio. 2001. *Interest-Rate Models: Theory and Practice*. Berlin, Germany: Springer Verlag, first edition.

Butler, L. 1996. "A Semi-Structural Method to Estimate Potential Output: Combining Economic Theory with a Time-Series Filter." Bank of Canada: Technical Report No. 77.

Cairns, A.J.G. 2004. "A Family of Term-Structure Models for Long-Term Risk Management and Derivative Pricing." *Mathematical Finance* 14: 415–444.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton, New Jersey: Princeton University 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.H. and M. Piazzesi. 2006. "Decomposing the Yield Curve." Unpublished Working Paper.

Colin-Dufresne, P., R.S. Goldstein, and C.S. Jones. 2005. "Can Interest Rate Volatility be Extracted from the Cross Section of Bond Yields? An Investigation of Unspanned Volatility." Unpublished Working Paper.

Dai, Q., A. Le, and K.J. Singleton. 2006. "Discrete-time Dynamic Term Structure Models with Generalized Market Prices of Risk." University of Stanford Working Paper.

Dai, Q. and K.J. Singleton. 2000. "Specification Analysis of Affine Term Structure Models." *The Journal of Finance* 55: 1943–1978.

———. 2002. "Expectation Puzzles, Time-Varying Risk Premia, and Dynamics Models of the Term Structure." *Journal of Financial Economics* 63: 415–441.

Deacon, M., A. Derry, and D. Mirfendereski. 2004. *Inflation-indexed securities: bonds, swaps and other derivatives*. West Sussex, England: Wiley Finance Series, second edition.

Demers, F. 2003. "The Canadian Phillips Curve and Regime Switching." Bank of Canada: Working Paper 2003-32.

Diebold, F.X., L. Ji, and C. Li. 2004a. "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. and R.S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 20: 134–144.

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. 2004b. "The Macroeconomy and the Yield Curve: A Nonstructural Analysis." University of Pennsylvania Working Paper.

Duffee, G.R. 2002. "Term Premia and Interest Rate Forecasts in Affine Models." *The Journal of Finance* 57: 405–443.

Duffie, D., D. Filipovic, and W. Schachermayer. 2003. "Affine Processes and Applications in Finance." *Annals of Applied Probability* 13: 984–1053.

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

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. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57: 357–384.

———. 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.

Heath, D., R. Jarrow, and A. Morton. 1992. "Bond Pricing and the Term Structure of Interest Rates: A New Methodology." *Econometrica* 60: 77–105.

Hördahl, P., O. Tristani, and D. Vestin. 2004. "A Joint Econometric Model of Macroeconomic and Term Structure Dynamics." European Central Bank Working Paper: No. 405.

Hurn, A.S., K.A. Lindsay, and V. Pavlov. 2005. "Smooth Estimation of Yield Curves by Laguerre Functions." Queensland University of Technology Working Paper.

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.

Kim, C.J. 1994. "Dynamic Linear Models with Markov Switching." *Journal of Econometrics* 60: 1–122.

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.

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: 473–489.

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.

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: 157–176.

Shea, G.S. 1985. "Interest Rate Term Structure Estimation with Exponential Splines: A Note." *The Journal of Finance* 40: 319–325.

Söderlind, P. 1999. "Solution and Estimation of RE macromodels with optimal policy." *European Economic Review* 43: 813–823.

Vasicek, O.A. and H.G. Fong. 1981. "Term Structure Modeling Using Exponential Splines." *The Journal of Finance* 37: 339–348.