Comparing Multifactor Models of the Term Structure

Michael W. Brandt*
The Wharton School
University of Pennsylvania
and NBER

David A. Chapman*
McCombs School
University of Texas at Austin

May 07, 2002

Abstract

There are a large number of reduced-form, multifactor term structure models available in the literature. However, the basic question of which model does the best job at explaining the key stylized facts of U.S. Treasury yields remains unanswered. We propose a set of economic moment conditions as the key facts to be explained, and we use these moments to estimate and directly compare three-factor quadratic and affine term structure models. Using 45 years of monthly data, we conclude that essentially affine models cannot reconcile both the dynamics of conditional expected bond returns and their conditional volatility, but the three-factor Gaussian-quadratic model is generally consistent with the full set of moments.

Preliminary and incomplete.

*Brandt: Philadelphia, PA 19104-6367. Phone: (215) 898-3609. E-mail: brandtm@wharton-.upenn.edu. **Chapman:** Austin, Texas 78712-1179. Phone: (512) 471-6621. E-mail: chapman@eco-.utexas.edu. We would like to thank Rob Engle for helpful discussions.

1 Introduction

The recent literature on the dynamics of default-risk free bond yields includes dramatic and important developments in both the specification and estimation of multifactor reduced-form, arbitrage free, dynamic term structure models. At this point, however, we still lack satisfactory answers to some of the most fundamental questions concerning relative model performance. In particular, which multifactor term structure model provides the best description of the data on U.S. Treasury yields, and what do we even mean by "the best description"? These questions are difficult to answer, in part, precisely because of the rapid proliferation in the classes of alternative models. The models are often nonlinear in important ways, and they are not encompassed by a general likelihood function that can be used for classical model comparison. The fact that, because of the nonlinearities, most models do not even possess closed-form likelihood functions has also led researchers to employ a variety of moments-based and approximate likelihood estimators, which introduces the possibility that some of the observed differences in model performance across studies reflect differences in the econometric approaches. Finally, the existing evidence is based on data measured at different frequencies, over different time spans, and from different markets.

We address these questions by first asking: What features of the U.S. Treasury data have received the most attention in the vast empirical term structure literature? Not surprisingly, these features include the unconditional means and volatilities of yields for different maturities, the cross-sectional and auto-correlations of these yields, and the conditional mean and volatility of holding period returns as functions of the current yield curve. We use these *economic moments* to estimate and directly compare a variety of three-factor quadratic and affine term structure models using a simulated moments estimator (SMM) of the type described in Lee and Ingram (1991) and Duffie and Singleton (1993).

There are at least two important advantages to using an SMM estimator with economic moments, as opposed to more direct attempts to achieve an approximate maximum likelihood (ML) estimator. First, and foremost, the successes and failures of alternative models are much more transparent using economic moments. For example, it is easy to see that a particular model can match the observed cross- and autocorrelations but not the conditional volatility structure. In contrast, when models are estimated with moment conditions generated by scores from an actual or approximate likelihood function, it is much more difficult to trace a model rejection to a particular feature of the data. In fact, the feature of the data responsible for the rejection may be in some obscure higher-order dimension that is of little interest to an economic researcher. The second advantage of an economic moments-based approach is that by setting out a common benchmark across all models, it allows for direct model comparison (along the important dimensions of the data).

These benefits, of course, come at a price. The SMM estimator with economic moments is consistent and asymptotically normal, but it is asymptotically inefficient relative to estimators that can accurately approximate the true likelihood estimator. However, as in other econometric applications, even if tests based on economic moments are not the most powerful, there is information in the ability to reject some alternative models. In spirit, we are advocating a trade-off similar to the "efficiency versus robustness" argument raised in Cochrane (1996) in the context of cross-sectional models of stock returns. Finally, the finite-sample performance of many of the estimators of multifactor term structure models (including our version of the SMM estimator) remains an open and important question.

Using a common data set of monthly observations on synthetic zero-coupon U.S.

Treasury yields from 1953 to 1998, we find that the economic moments above allow for powerful distinctions between affine and quadratic models. In particular, simultaneously matching the conditional mean of holding period returns and the conditional return volatilities is beyond the capabilities of *any* of the three-factor essentially affine models that we estimate using the SMM. Clearly, the Gaussian affine model never had a chance at matching conditional volatilities, but the mixed affine-square root models and the pure correlated square root models introduce conditional volatility in a way that appears to be wholly incompatible with the dependence of the conditional return volatility on the yield curve.

In contrast, the unconstrained three-factor Gaussian-quadratic model emerges as the clear winner in this "horse race." The overall specification test for this model is rejected, when the moments are weighted using the statistically optimal weighting matrix. However, this overall rejection is inconsistent with an entire battery of other tests. All of the tests performed on smaller groups of similar moments, for example all of the slope coefficients in holding period regressions or all cross- and autocorrelations, fail to reject the equality of model and sample moments. Furthermore, the individual point estimates from the quadratic model reproduce the magnitude and patterns found in the sample economic moments.

The remainder of the paper is organized as follows: Section 2 reviews the basic features of the affine and quadratic classes of dynamic term structure model (DTSM). The important facts of U.S. Treasury yields are summarized in Section 3, and Section 4 briefly reviews the key recent developments in the empirical term structure literature. Section 5 discusses model estimation in detail, and the data are described in Section 6. Our primary results for the SMM estimator are in Section 7, and Section 8 provides results for a variant of the efficient method of moments (EMM), for comparison purposes. We conclude in Section 9.

2 Multifactor Models

The recent empirical literature on term structure modelling has concentrated on two general classes of models: those in which the instantaneous interest rate is an affine function of the states and those in which the instantaneous rate is a quadratic function of the states. In either class of models, successfully matching the dynamics of yields, their correlation and volatility structures, and holding period returns requires both flexibility in the correlation structure of the state variables and in the sources of conditional volatility of yields.

An N-factor affine term structure model (ATSM), introduced in Duffie and Kan (1996), and examined empirically in Dai and Singleton (2002a) (hereafter DS2) and Duffee (2002), describes yield dynamics using the following assumptions:

Assumption A1: The dynamics of the N exogenous factors, F, under the physical measure, P, are formalized in a stochastic differential equation of the form

$$dF_t = (\mathcal{K}\theta - \mathcal{K}F_t)\,dt + \Sigma\sqrt{S_t}dW_t, \tag{1}$$

where $\mathcal{K}$ and Σ are $N \times N$ matrices, $\mathcal{K}\theta$ is an N-vector, S_t is an N-dimensional diagonal matrix with elements on the main diagonal

$$[S_t]_{ii} = \alpha_i + \beta_i' F_t, \tag{2}$$

and W_t is an N-dimensional Brownian motion under P.

Assumption A2: The short-rate is a linear function of the factors:

$$r_t = \delta_0 + \delta_1' F_t. \tag{3}$$

Assumption A3: The market price of factor risk function is

$$\Lambda_t = S_t \lambda_1 + S_t^- \lambda_2 F_t, \tag{4}$$

where λ_1 is an N-vector, λ_2 is an $N \times N$ matrix, and

$$\left[S_t^-\right]_{ii} = \begin{cases} \left(\alpha_i + \beta_i' F_t\right)^{-1/2}, & \text{if } \inf\left(\alpha_i + \beta_i' F_t\right) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

This assumption ensures that the state variable dynamics are affine under both the physical measure and the risk neutral measure, Q.

Assumptions **A1**–**A3** actually define the *essentially affine models* of DS2 and Duffee (2002). When $\lambda_2 = \mathbf{0}$, this formulation simplifies to the *completely* affine models of Duffie and Kan (1996) and Dai and Singleton (2000) (hereafter DS1). In the remainder of the paper, we will not make this distinction, and the term "affine model," without qualification, will be understood to mean an essentially affine model of the form defined by Assumptions **A1** to **A3**.[1]

In models of this type, the date t price of a zero-coupon bond maturing at date $t + \tau$ is of the exponential affine form:

$$P_t(\tau) = \exp\left[A(\tau) - B(\tau)' F_t\right], \tag{6}$$

where $A(\tau)$ and $B(\tau)$ satisfy the ordinary differential equations

$$\frac{dA(\tau)}{d\tau} = -\left(\mathcal{K}\theta^Q\right)' B(\tau) + \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma' B(\tau)\right]_i^2 \alpha_i - \delta_0 \tag{7}$$

and

$$\frac{dB(\tau)}{d\tau} = -\mathcal{K}^{Q\prime} B(\tau) + \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma' B(\tau)\right]_i^2 \beta_i + \delta_1, \tag{8}$$

[1] We focus on the essentially affine class because of evidence examined in Section 4, below.

first specified in Duffie and Kan (1996), where $\mathcal{K}\theta^Q$ and $\mathcal{K}^Q$ refer to the components of the state variable drift under the Q-measure.

Under the risk-neutral measure, the correlations of the state variables are determined by the off-diagonal elements of $\mathcal{K}^Q$, and their conditional volatilities are determined by the form of S. Under the physical measure, yield correlations are also affected by the dynamics of Λ. DS1 introduced the notation $\mathbb{A}_m(N)$ $(m \leq N)$ to denote an affine model with N total state variables, m of which affect the conditional volatility of yields. They also provide a detailed treatment of the necessary parameter restrictions for both admissibility (whether or not the model generates valid bond prices) and identification of an affine model. These parameter restrictions are shown in Figure 1, for the three-factor models examined below.

The canonical form of an N-factor quadratic term structure model (QTSM) of Ahn, Dittmar, and Gallant (2002) (hereafter ADG) is defined by the following assumptions:

Assumption Q1: The factors are Gaussian; so

$$dF_t = (\mathcal{K}\theta - \mathcal{K}F_t)\,dt + \Sigma dW_t, \tag{9}$$

where $\mathcal{K}$ and Σ are $N \times N$ matrices and $\mathcal{K}\theta$ is an N-vector.

Assumption Q2: The short-rate is a quadratic function of the state

$$r_t = \delta_0 + F_t'\Psi F_t, \tag{10}$$

where $\delta_0 \geq 0$ and Ψ is an $N \times N$ positive definite matrix

$$\Psi = \begin{bmatrix} 1 & \Psi_{12} & \cdots & \Psi_{1N} \\ \Psi_{12} & 1 & \cdots & \Psi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{1N} & \Psi_{2N} & \cdots & 1 \end{bmatrix}. \tag{11}$$

Assumpion Q3: The market price of risk is

$$\Lambda_t = \lambda_0 + \lambda_1 F_t, \tag{12}$$

Zero-coupon bond prices in a QTSM are equal to

$$P_t(\tau) = \exp\left[A(\tau) + B(\tau)' F_t + F_t' C(\tau) F_t\right], \tag{13}$$

where $A(\tau)$ (scalar), $B(\tau)$ (N-vector), and $C(\tau)$ ($N \times N$ matrix) are functions defined as the solutions to the sets of ordinary differential equations specified in ADG. In the canonical form, quadratic models guarantee strictly positive variances of the state variables and strictly positive bond prices. Beyond the normalizations embedded in assumptions **Q1-Q3**, there are no further parameter restrictions imposed on the most flexible QTSM model.

3 The Stylized Facts of U.S. Term Structure Data

The fundamental goal of any DTSM is to provide a parsimonious description of a small number of stylized facts of default-free bond yields, but what are these stylized facts? Chapman and Pearson (2001), Dai and Singleton (2002b), and Piazzesi (2002) are all recent reviews of the term structure literature. Although these papers concentrate on different aspects of this large body of work, the following stylized facts emerge as central features of U.S. Treasury data:

Fact 1 *The unconditional mean Treasury yield curve is upward sloping at maturities from 3- to 10-years.*

Fact 2 *The levels of Treasury yields, measured weekly or monthly at maturities from 3-months to 10-years, are very highly correlated.*

Contemporaneous correlations for adjacent maturities are in excess of 0.98.

Fact 3 *Yield levels are highly persistent.*

Monthly first-order autocorrelations are typically in excess of 0.98, and cross-correlations of yields at one lag also remain above 0.98.

These facts are consistent with a small number of common factors determining the movements of all yields. This intuition was formalized by Litterman and Scheinkman (1991) who showed that the first principal component in weekly yield changes account for (roughly) 88 percent of the variation in yield changes, and the first three components account for 99 percent of the variation.[2] After examining the sensitivity of yield changes to these components, Litterman and Scheinkman named these factors the "level", "slope", and "curvature."

Fact 4 *The slope coefficient, ϕ_τ, in the following linear regression*

$$[Y_{t+1}(\tau-1) - Y_t(\tau)] = \phi_c + \phi_\tau \left[(Y_t(\tau) - r_t) / (\tau - 1) \right] + \varepsilon_{t+1}(\tau) \tag{14}$$

is negative and increasingly so for longer maturities, where $Y_t(\tau)$ is the time t yield on a zero-coupon bond that matures at time $t+\tau$, r_t is the yield on the short maturity bond and $\varepsilon_{t+1}(\tau)$ is the error term.

Dai and Singleton (2002b) refer to this fact as "LPY" (for "linear projection of yields"), and it was first examined in the empirical literature in Campbell and Shiller (1991).[3]

[2]Chapman and Pearson (2001) confirm this finding and extend it to a much larger time period than the 1984 to 1988 period examined in Litterman and Scheinkman (1991).

[3]Fama and Bliss (1987) run related regressions of future yield levels on the difference between a maturity matched forward rate and the short rate.

Fact 5 *The term structure of the unconditional volatility of Treasury yield levels is downward sloping. The unconditional volatility of yield changes decreases from 1-month to 6-month maturities, and then it increases up to about 24-months. It is constant for maturities beyond 24-months.*

Fact 6 *The conditional volatilities of changes in yields are time-varying and persistent, and they are positively related to the level of the short rate.*

Fact 7 *The period of the "Federal Reserve experiment" from 1979 to 1982 is a set of influential observations.*

Whether or not this time period is in the sample of data being analyzed can have a significant effect on the conclusions about what features in a term structure model are necessary to adequately characterize the data.

In conclusion, these are the (comparatively) small set of yield moments that have been of consistent interest to financial economists over the history of term structure estimation. The ability of a specific model to capture these facts is the metric that we want to impose directly on our estimates and comparison of DTSMs. However, before we discuss further our estimation approach, we first briefly review the existing evidence.

4 A Brief Literature Review

In examining the prior literature, we focus on what the various studies imply about the ability of ATSMs and QTSMs to explain the basic facts stated in the previous section. The answers to these questions are neither simple nor conclusive for three primary reasons. First, existing studies use data sets that are constructed in different ways and that cover different time periods and data frequencies. Second, the papers in the existing literature use a variety of different estimation methods. To the extent that all of these methods involve approximations to the unobservable continuous-time dynamics, they all introduce different amounts of specification error that can be difficult to assess across model types. Finally, as DS1 point out, direct likelihood-based comparisons even across different classes of N-factor affine models are impossible, since the models are not nested.

DS1 is the first systematic analysis of affine, albeit completely affine, models imposing the minimal admissibility and identification conditions on the estimation. They examine various versions of $\mathbb{A}_m(3)$, $m = \{0, 1, 2, 3\}$ by EMM (using the semi-nonparametric density estimator from Gallant and Tauchen (1989), hereafter SNP, as the auxiliary model). DS1 use weekly data on swap rates of maturities from six months to 10 years for the period from April 1987 to August 1996. They do not produce direct evidence of the different models' abilities to match the basic moments of yields. This is because the EMM estimator uses the scores of the auxiliary model, rather than economic moments directly. Instead, DS1 shows that allowing for negative correlation among the state variables and time-varying conditional volatility are both important for matching the average scores. Different elements of the score vector correspond to parameters that determine nonnormality, conditional first moments, and conditional second moments of the data. Since none of these moments for the flexible $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$ models are significantly different from zero, we can *indirectly* infer that these models match the first and second moment properties of the data.

ADG examine nested versions of a three-factor quadratic model using EMM (with SNP) and monthly observations on zero-coupon yields from December 1946 to February 1991.[4] Consistent with the results in DS1, they find that allowing for correlation among the latent factors results in a dramatic improvement in the fit of the model. However, unlike in the completely affine models examined by DS1, the correlations in the most preferred model are not uniformly negative. The analysis in ADG favors the most general version of the three-factor quadratic model, but overall measures of model fit still reject this specification. In part, this rejection reflects the more stringent test imposed on any DTSM by a sample period that spans the 1979 to 1982 period. ADG conclude that the rejections of the model are due to problems with fitting the conditional volatility and the overall shape of the conditional density of Treasury yields.

Duffee (2002) proposes a natural metric for evaluating the ability of a given multifactor model to explain future movements in the yield curve. He demonstrates that the common choices of completely affine models fail to outperform a simple random walk model in terms of average forecast errors, both in and out of sample. Furthermore – and perhaps of more fundamental importance – he demonstrates that a completely affine model cannot fit, simultaneously, the distribution of yields and the observed patterns of predictability in the excess holding period returns on U.S. Treasury bills and bonds.

The problem appears to be with the form of the market price of risk in a completely affine model. Duffee (2002) examines monthly zero-coupon yields for the period from January 1952 to December 1998 taken from McCulloch and Kwon (1993) and extended using the method in Bliss (1997). He notes that Treasury data suggests that instantaneous excess returns on bonds of different maturities are small, relative to the volatility of bond returns. Given the form of (4), instantaneous excess bond

[4] ADG use the zero-coupon yields computed in McCulloch and Kwon (1993).

returns are bounded below by zero. The only way for an affine model to fit the ratio of expected bond returns to the volatility of bond returns is for the underlying factors to be highly positively skewed. Duffee (2002) goes on to argue that, in completely affine models, parameterizations that match the (relative) first two moments of holding period returns cannot match the observed shapes of the term structure. In essence, the model does not have sufficient flexibility to match the cross-section of both prices and returns.

Duffee (2002) finds that essentially affine models are more successful at forecasting yield changes for six-month, two-year, and ten-year Treasury yields at the three-, six- and twelve-month horizons. A three-factor Gaussian essentially affine model can consistently outperform, both in and out of sample, a simple random walk forecasting model for yield changes at all maturities and forecast horizons. Interestingly, DS2 use monthly zero-coupon yields to demonstrate that a three-factor Gaussian essentially affine model also can match the regression coefficients in LPY, in a qualitative sense.[5] It is important to note that DS2 do not actually estimate the model using the regression-based moments. Instead, they calculate the moments from a long time series of simulated bond yields computed from the model at the estimated parameter values.

The results in DS2 and Duffee (2002) immediately generate an important contradiction: In order to match yield level forecasts and yield change regressions, essentially affine models emphasize flexibility in the correlation structure of the state variables and turn off all conditional variation in yield volatilities. In other words, the results in DS2 and Duffee (2002) suggest that ATSMs can fit the first moment properties of yields or the second moment properties *but not both*.

[5]DS2 use monthly yields from February 1970 to December 1995. Their data is from Fama and Bliss (1987) augmented by Backus, Foresi, Mozumdar, and Wu (2001).

What about direct comparisons *across* model classes? There is only a limited amount of information available. DS1, DS2, and Duffee (2002) do not examine quadratic models. ADG find that the preferred completely affine model from DS1 (*i*) is rejected by an overall measure of model fit and (*ii*) has a hard time fitting the observed volatility of yields at all maturities. While this is not entirely consistent with the findings in DS1, the longer sample period of 1946 to 1991, poses a more stringent test for any multifactor model than the relatively short time period used in DS1. There are currently no direct comparisons in the literature between essentially affine models and quadratic models.

In summary, the data and estimation methods of existing studies of DTSMs make it impossible to state any consistent conclusions about whether ATSMs are better or worse than QTSMs or whether either class of term structure model can produce a variant that adequately reproduces the full range of interesting yield moments.

5 Estimating Multifactor Models

Exact maximum likelihood estimation (MLE) is always the preferred method of estimation, since it ensures consistent and asymptotically efficient estimates, but in the case of DTSMs, MLE is only feasible for relatively few cases; namely multifactor generalizations of the Gaussian model of Vasicek (1977) or the square-root diffusion model of Cox, Ingersoll, and Ross (1985) (CIR).[6] Consequently, the majority of the literature on affine and quadratic models uses either approximate MLE or moments-based

[6] A number of approximations to the likelihood functions for affine diffusion models have been proposed based on numerical solutions of the SDEs (Lo, 1988), numerical inversion of the characteristic functions (Singleton, 2001), simulations (Brandt and Santa-Clara, 2002), and polynomial density approximations (Aït-Sahalia, 2002). Brandt and He (2002) show how to combine these approximations with importance sampling to obtain ML estimates of ATSMs in a state-space form, where all bond yields are assumed to be measured with error. This approach can be extended to QTSMs, but the properties of the estimator in that context are unknown.

estimation. For our purposes, moments-based estimation is more suitable because it naturally lends itself to non-nested model comparison and model diagnostics. Specifically, we use a simulation-based variant of Hansen's (1982) generalized method of moments (GMM), the so-called simulated method of moments (SMM), described by Lee and Ingram (1991) and Duffie and Singleton (1993). We first describe the SMM estimation method in general terms and then discuss in more detail the moments we use in the context of this method.

5.1 SMM Estimation

Any moments-based estimator of a DTSM depends on two fundamental components:

1. A set of *moment conditions* that can be evaluated given the observed data $\{Y_t\}_{t=1}^T$, $G_T = E_T[g(Y_t)] = T^{-1}\sum_{t=1}^T g(Y_t)$, and can also be solved for as a function of the model parameters, $G(\pi) = E_\pi[g(Y_t)]$, where g denotes a vector of $K \geq \dim(\pi)$ functions of the data, π are the model parameters, and E_π denotes an expectation taken under the physical measure implied by the model. The functions g are assumed to be non-redundant, in the sense that each contains independent information that helps identify the parameters. The goal of the moments-based estimator is to find parameters π_T for which the implied theoretical moments $G(\pi_T)$ are as close as possible, in a metric defined more precisely below, to the empirical moments G_T.

   In our case, the data Y_t are default-free zero-coupon bond yields, the model parameters are $\pi = (\mathcal{K}, \mathcal{K}\theta, \Sigma, \alpha, \beta, \delta_0, \delta, \lambda_1, \lambda_2)$ for affine models or $\pi = (\mathcal{K}, \mathcal{K}\theta, \Sigma, \delta_0, \Psi, \lambda_0, \lambda_1)$ for quadratic models, and the theoretical expectations, E_π, are taken with respect to the physical measures characterized by the diffusion models in equations (1) and (9), respectively. The moments are discussed in greater detail below.

When the theoretical moments $G(\pi)$ cannot be solved analytically, which is often the case for DTSMs, they can usually still be computed by simulations (i.e., Monte Carlo integration). In particular, for a given parameter vector π and initial condition Y_0 (which can itself be simulated from the unconditional distribution of Y_t), we simulate $M \gg T$ hypothetical data realizations $\{Y_m^\pi\}_{m=1}^M$ from the model and then approximate the theoretical moments $G(\pi)$ with their Monte Carlo analogues $G_M(\pi) = E_{\pi,M}[g(Y_m^\pi)] = M^{-1}\sum_{m=1}^M g(Y_M^\pi)$.[7] It is this use of simulations to evaluate the theoretical moments that differentiates the SMM approach from the usual GMM estimation.

2. When the number of moment conditions, K, exceeds the number of parameters, $\dim(\pi)$, a *weighting matrix,* denoted W, is needed to determine the relative importance of each moment condition in constructing the moments-based estimates of π.

   Following Hansen (1982), the *statistically* optimal weighting matrix that yields the most efficient estimator is the inverse of the covariance matrix of the sample moments:

$$W = \{\mathrm{cov}\,[G_T]\}^{-1}, \tag{15}$$

   where in practice we replace the covariance with a consistent estimate (e.g., Newey and West, 1987). However, as Cochrane (1996) argues in the context of cross-sectional tests of equity pricing models, this statistically optimal weighting matrix is not always the most *robust* and *economically transparent* weighting matrix. What is the economic meaning of the model failing to fit the W-weighted

[7] We use $10,000$ simulated observations, constructed using the method of antithetic variables.

set of moments? If all moments are equally important from an economic perspective, then equal weighting of the moments, using W equal to an identity matrix is a reasonable alternative.

Given these two building blocks, the SMM estimates, π_T, are constructed as

$$\pi_{T,M} = \arg\min_{\pi \in \Pi} \left[G_T - G_M(\pi) \right]' W_T \left[G_T - G_M(\pi) \right]. \tag{16}$$

Duffie and Singleton (1993) prove that, under certain conditions, the estimator $\pi_{T,M}$ is consistent and asymptotically normal, even when the initial conditions Y_0 are chosen arbitrarily (rather than drawn from the true stationary distribution of Y_t) and given the fact that the simulated $\{Y_m^\pi\}$ depend on the parameters of the model.

Analogous to Hansen (1982) solving for the statistically optimal weighting matrix, Gallant and Tauchen (1996) address the issue of choosing statistically optimal moment conditions. In principle, the solution to this problem is straightforward since MLE can be viewed as a moment-based estimator. The parameters that maximize the log-likelihood function $\mathcal{L}\left(\{Y_t\}_{t=1}^T ; \pi\right)$ also solve the corresponding first-order conditions:

$$\frac{\partial \mathcal{L}}{\partial \pi} = \sum_{t=1}^{T} \frac{\partial \ln f\left(Y_t \mid Y_{t-1}, Y_{t-2}, \ldots ; \pi\right)}{\partial \pi} = 0, \tag{17}$$

where $f\left(Y_t \mid Y_{t-1}, Y_{t-2}, \ldots ; \pi\right)$ is the conditional density of a single observation Y_t. It follows that GMM or SMM estimates based on the moments $G_T = T^{-1}\, \partial \mathcal{L} / \partial \pi$ (the so-called *scores* or *score vector* of the likelihood function) are identical to MLE and are hence statistically optimal.[8] Of course, Cochrane's (1996) point about statistically optimal versus robust and economically transparent inferences applies equally well to the choice of moment conditions. The model or previous empirical work may suggest a

[8] Gallant and Tauchen (1996) proceed to describe a moments-based estimator that is implementable even when the likelihood function is unknown analytically. Their *efficient* method of moments (EMM) estimator is described in more detail below.

set of moments that are economically more transparent than the scores and therefore allow for more direct diagnostics of the empirical successes and failures of the model.

5.2 Moment Conditions

With this trade-off between statistical optimality versus robustness and economic transparency in mind, we focus primarily on SMM estimates with economically motivated moments (using either statistically optimal and equal weighting). However, to facilitate comparison to the literature, we also present score-based results.

Both sets of moments as based on three variables that accurately describe the yield curve. Specifically, we use versions of the "level", "slope", and "curvature" factors identified first by Litterman and Scheinkman (1991) and examined more recently by Chapman and Pearson (2001). Our level factor is the yield on the six-month bond. We define the slope factor as the difference between the yields on the 10-year and six-month bonds. Finally, the curvature factor is $[Y_t\,(6\text{-}m) + Y_t\,(10\text{-}y) - 2Y_t\,(2\text{-}y)]$, where $Y_t\,(\tau)$ denotes the yield to maturity on a τ-period bond (in percent per year).

We choose only three variables to describe the yield curve because, if the data is truly generated by a three-factor model, any three yields (or linear combinations of them) are (approximately) sufficient statistics. The reason for rotating the 6-month, two-year, and ten-year yields into the level, slope and curvature factors is that we know from Litterman and Scheinkman (1991) that these factors are (nearly) orthogonal. This orthogonality allows us to interpret the results for individual moments, as opposed to having to disentangle the information contained jointly in highly correlated moments.

5.2.1 Economic Moments

Our choice of economically motivated moments is guided by the discussion in Section 3. Specifically, the vector G_T consist of: (*i*) the unconditional means and residual standard deviation from a first-order autoregression of the level, slope, and curvature factors (six moments); (*ii*) the contemporaneous and first-order lagged correlations of the same variables (nine moments); (*iii*) the slope coefficients from the LPY regressions (equation(14)), for maturities of six months, two years, and 10 years (three moments); and (*iv*) the slope coefficients from the conditional volatility regressions (LPV regressions):

$$[Y_{t+1}(\tau-1) - Y_t(\tau)]^2 = \mu_{0,\tau} + \sum_{i=1}^{3} \mu_{i,\tau} F_{it} + \xi_{t+1}(\tau) \tag{18}$$

for the same three bonds (i.e., $\tau =$ six months, two years, and 10 years), where F_{it} denotes the level, slope, and curvature factors (nine moments).[9] In total, our results are based on 27 moments.

Before we proceed to describe an alternative estimator based on score-based moments, it is worthwhile to reiterate our motivation for considering economically motivated moments. It is undeniably true that we are sacrificing potential efficiency by using economic moments. However, in return we hope to gain economic intuition about the successes and failures of different DTSMs, to compare them on a "level playing field," and to construct results that are more robust than the (in theory) more efficient estimates.

[9]This regression characterizes the conditional second moment of holding period returns as a function of the current shape of the yield curve. It summarizes the variation in holding period returns in a manner analogous to the LPY regression in equation (14).

5.2.2 Score-Based Moments

As we noted above, MLE can be viewed as a moments-based estimator, applied to the scores of the log-likelihood function. Unfortunately, this way of obtaining an efficient moments-based estimator requires an expression for the likelihood function (or at least its scores), which is typically not available for DTSMs. Gallant and Tauchen (1996) suggest a way to overcome this problem (in a general context). Their so-called *efficient* method of moments (EMM) estimator is based on an approximation of the true likelihood function by the likelihood function of a semi-nonparametric *auxiliary* model that describes the data well in finite samples and converges asymptotically to the true model. In place of the scores of the true likelihood, $\partial\mathcal{L}(\{Y_t\}_{t=1}^T;\pi)/\partial\pi$, the estimator uses the scores of the auxiliary likelihood $\partial\mathcal{L}(\{Y_t\}_{t=1}^T;\varphi)/\partial\varphi$, where φ denotes the parameters of the auxiliary model and $\dim(\varphi) > \dim(\pi)$ for identification. Gallant and Tauchen (1996) prove that, as long as the auxiliary model is flexible enough to approximate the true model arbitrarily well, the EMM estimator shares the asymptotic distribution of the unattainable MLE. Gallant and Tauchen (2000) suggest the semi-nonparametric (SNP) conditional density estimator of Gallant and Tauchen (1989) as a candidate auxiliary model.

The EMM estimator promises efficiency and, as a result, it is not surprising that EMM has been used extensively, including for the estimation of DTSMs (e.g., ADG and DS1). However, the recent Monte Carlo study of Duffee and Stanton (2001) for a single-factor CIR model has cast a shadow on this theoretically appealing estimator, at least in the context of estimating DTSMs. Duffee and Stanton's experiment suggests that the finite sample performance of the EMM/SNP estimators is, quite frankly, dismal. With the best fitting SNP auxiliary model (which involves 48 parameters/moments), as determined by standard in-sample model selection criteria, the overall specification tests rejected the *true* model repeatedly at extremely high levels of significance. Even more troubling, the EMM/SNP estimator, using 20 years of simulated weekly data, could not recover values close to the true parameters, *even when the estimation process was started at the true parameters*![10].

Is EMM just another theoretically appealing estimator with dismal finite sample properties? Our conjecture is that the basic idea of EMM is both sensible and practical. The scores of a model that describes the data well contain a wealth of information about the parameters of the true model. The problem, we believe, lies in the blind use of the SNP density estimator as auxiliary model. To illustrate the problem, suppose we fit an SNP model to the level, slope, and curvature factors, following closely the procedure described in Gallant and Tauchen (2000). The best fitting model, according to Gallant and Tauchen's procedure, involves 63 parameters (resulting in 63 moments for the EMM estimator). However, 27 of these parameters have t-statistics of less than two, suggesting that their role in the auxiliary model is dubious. Furthermore, judging by the significant parameters and by plots of the conditional densities, the preferred SNP model captures essentially the first and second unconditional moments of the factors, the autoregressive nature of the factors, the autoregressive nature of the conditional volatility of the factors, and some excess conditional kurtosis. These are undoubtedly important characteristics of the data that help identity the true parameters. However, does it require 63 parameters to describe these characteristics? If not, what is the finite-sample effect of including 27 or more nuisance moments in the estimation?

[10] Although some of the problems with the EMM/SNP estimator undoubtedly reflect the well-known downward bias in estimating the autoregressive parameter in a near unit root process, the overall performance of the EMM/SNP estimator still seems surprising. It also seems difficult to reconcile with some of the existing Monte Carlo evidence on the finite sample performance of EMM estimators. See, for example, Chumacero (1997), Andersen, Chung, and Sørenson (1999), and Gallant and Tauchen (1999).

Guided by the intuition that scores are informative but that nuisance moments can result in a deterioration of the finite sample properties of a score-based estimator, we follow a different procedure for choosing the scores for an EMM-like estimator. We specify an auxiliary model that captures the same important characteristics of the data (and some additional ones) in a much more compact parameterization. In particular, our auxiliary model of the level, slope, and curvature factors is comprised of the following four parts:

1. The conditional means follow a first-order vector autoregression (VAR):

$$F_t = \begin{bmatrix} b_0 & B \end{bmatrix} \begin{bmatrix} 1 \\ F_{t-1} \end{bmatrix} + \varepsilon_t \tag{19}$$

(12 parameters/moments).

2. The conditional variances of the VAR innovations, ε_{it}, are time-varying and depend on both the level of the factors as well as on separate GARCH(1,1) processes:

$$h_{it} \equiv \text{var}_{t-1}\left(\varepsilon_{it}\right) = \exp\left(F'_{t-1}\gamma_i\right) v_{it}, \tag{20}$$

where

$$v_{it} = \left(1 - \phi_i - \psi_i\right)\omega_i + \phi_i \varepsilon^2_{it-1} + \psi_i v_{it-1} \tag{21}$$

(6 parameters/moments per factor for a total of 18 parameters/moments).

3. The standardized innovations $u_{it} \equiv \varepsilon_{it}/\sqrt{h_{it}}$ have constant correlations:

$$\text{corr}\left(u_t\right) = \begin{bmatrix} 1 & & \\ \rho_{21} & 1 & \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix} \tag{22}$$

(3 parameters/moments).

4. The conditional (on the VAR means and level-GARCH variances) distribution of the factors is multivariate Student's t with ζ degrees of freedom (1 parameter/moment).

This auxiliary model has only 34 parameters but it captures essentially the same features of the data as the preferred SNP model. In addition, it reflects the empirical fact (and the theoretical prediction of certain ATSMs) that the variances of the factors depend on the levels of the factors. Together, the fewer parameters and the level-dependence of the variances leads to substantial improvements of the in-sample model selection criteria, relative to the best-fitting SNP model.

5.3 QML Estimator

The evidence of poor performance of the EMM/SNP estimator leads Duffee and Stanton (2001) to propose a quasi maximum likelihood (QML) estimator for affine term structure models based on an extended version of a Kalman filter-based algorithm. The exact Kalman filter is extended to allow for a nonlinear model by using a first-order Taylor series expansion around a discretized version of the instantaneous dynamics implied by equation (1). Duffee and Stanton (2001) demonstrate that, even though the QML estimator based on the extended Kalman filter is literally misspecified (in all but multivariate versions of the Gaussian model of Vasicek, 1977), it outperforms the EMM/SNP estimator and the EMM estimator that uses the (extended) Kalman filter as the auxiliary model.[11]

Although these results suggest the superiority of a QML approach for estimating a DTSM, there are two important problems that mitigate against our using it. First,

[11] In particular, they find that the (misspecified) QML estimator and the EMM/Kalman filter estimator have roughly the same bias (which is much smaller than the EMM/SNP estimator), but the QML estimator is more efficient.

the QML model cannot be used to estimate quadratic models. The measurement equation (relating yields to factors) is not linear, which would require introducing a second layer of Taylor series approximation in order to implement the estimator. At a minimum, this additional approximation would place the QML estimates of a quadratic model at a relative disadvantage to the QML estimates of an affine model. Second, in a quadratic model yields cannot be uniquely inverted to recover factors.

6 Data

As noted earlier, a variety of data sets have been used in examining the fit of alternative models. We use the same data source as Duffee (2002).[12] This is the McCulloch and Kwon (1993) data augmented with data constructed using the techniques in Bliss (1997). The data is observed monthly from January 1953 to December 1998 ($T = 552$), and it consists of (constant maturity) zero-coupon yields at maturities of 6-months, 1-, 2-, 5-, and 10-years.

Basic summary statistics, consistent with the stylized facts described in Section 3, are contained in Tables 1 through 5. According to the statistics in Table 1, the unconditional mean yields increase monotonically with maturity from 5.77 at 6-months to 6.75 at 10-years.[13] In the second line of Table 1, we see that the unconditional volatilities decrease monotonically with maturity, from 2.93 to 2.71. Tables 2 and 3 document that yield level correlations – both contemporaneously and at lag one – are always in excess of 0.91 for all maturities and usually in excess of 0.95. The linear regression coefficients of yield changes on the slope of the yield curve are presented in Table 4, and they are negative, significant, and increasing (in absolute value) with maturity, as documented in Campbell and Shiller (1991) and discussed in DS2.

[12]We would like to thank Greg Duffee for making his data available on his web page.

[13]All yield moments are reported in percent per year.

The regressions of squared yield changes on lagged values of the level, slope, and curvature factors are summarized in Table 5, and there are interesting patterns in the time variation in conditional volatilities. In particular, the volatilities for all maturities are positively related to the level factor. This dependence is statistically significant, and the magnitude of the slope coefficient is decreasing with maturity.[14] The regression coefficients for the slope factor are negative and significant for the short end of the yield curve, but insignificant (or marginally significant) for the 5- and 10-year maturities. Conditional volatility is not significantly related to the curvature factors at the long and short maturities, but the coefficients on curvature for the 1- and 2-year maturities are positive and marginally significant.

The fitted parameters for the auxiliary model described in Section 5.2.2 are shown in Table 6. There is nothing particularly surprising in the estimates of the VAR parameters. The volatility specification allows for both factor level effects and GARCH effects, and both components of the conditional variance seem to be important. The lagged level factor affects the variance of all three factors, and the volatility of the slope factor is also affected by its own lagged level. GARCH effects are strong in all three factors, which is to be expected, given the manner in which they are constructed. The level residuals are negatively correlated with both the slope and curvature residuals, and the slope and curvature residuals are slightly positively correlated. Finally, the innovations appear to be fat-tailed, relative to a normal distribution.

[14]There is a large literature in single-factor models devoted to examining this "level effect." See Chapman and Pearson (2001) for a review of this issue.

7 Economic Moments-Based Estimates

7.1 The Quadratic Model

Evidence on the performance of the canonical 3-factor quadratic model are presented in Tables 7 and 8.[15] This model has twenty five parameters, and it is estimated using the twenty seven moments identified above. The standard moments based test for the overall performance of an estimated model is equal to:

$$J_T \equiv \left[G_T - G_M(\pi)\right]' W_T \left[G_T - G_M(\pi)\right] \overset{a}{\sim} \chi^2_{K-\dim(\pi)}, \tag{23}$$

where $\overset{a}{\sim}$ denotes the asymptotic distribution of the statistic, K is the number of moment conditions, and $K - \dim(\pi)$ is the number of degrees of freedom for the chi-squared distribution.

Given the set of deviations of model moments from sample moments, $\eta_T(\pi) \equiv G_T - G_M(\pi)$, the covariance matrix of these deviations, $\Sigma_{\eta T}$, and an index set of moments $Ind(J)$ of dimension J, tests of the J parameter restrictions

$$\eta_{jT}(\pi) = 0, \text{ for all } j \in Ind(J) \tag{24}$$

can be implemented using a test statistic of the form

$$\eta'_{Ind(J)T} \left[C\Sigma_{\eta T}C'\right]^{-1} \eta_{Ind(J)T} \overset{a}{\sim} \chi^2_J, \tag{25}$$

where C is a $J \times K$ matrix of zeros and ones and $\eta_{Ind(J)T}$ refers to the specific set of moments being tested. We examine joint tests of the following specific hypotheses:

[15] Appendix Table A-1 presents a complete list of the point estimates (and standard errors) for the SMM models estimated with both the optimal and identity weighting matrices. Although these estimates are not intrinsically interesting in their own right, they are provided to allow comparison to previous work. In fact, these estimates in Table A-1 are quite different from the values reported in ADG, for the same model. These differences reflect both the different time period used in estimating the model (1946 to 1991 in ADG versus 1953 to 1998 here) and in the use of EMM versus SMM in the estimation.

(i) all mean and variance moment conditions are equal to zero; (ii) all correlation moment conditions are equal to zero; (iii) all LPY slope coefficients are equal to zero; and (iv) all LPV regression coefficients are equal to zero.

These statistics for the SMM estimators of the $QTSM3$ model are reported in the last line of Table 7. As in ADG, the $QTSM3$ model is rejected by the data, based on the overall test of model specification. However, the statistic in equation (23) provides little insight into the precise reason for the model's rejection. By contrast, all of the group moment conditions fail to reject the null that the $QTSM3$ model is consistent with the basic facts about the shape of the yield curve, factor autorocorrelations, and the LPY and LPV regression coefficients. Among other things, this result shows that a quadratic model can reproduce the Campbell and Shiller (1991) results emphasized in DS2.

The sample moments and the estimates of the individual moments for the SMM estimators are shown in Table 8. The SMM estimates with optimal weights generates too little volatility in the level factor, and it cannot match the slope coefficients of the squared changes in 2- and 10-year bond yields (the LPV coefficients) on the level factor. While the estimated moments have the right sign (positive), they can be reliably rejected as too small, compared to the data. The quadratic model appears to perform reasonably well at matching the average level of the yield factors, their cross-correlations (with the exception of the correlation between one lag of the slope and curvature factors). It is able to match the signs and magnitudes of the slope coefficients in the LPY regressions with reasonable accuracy and with no statistically significant difference from the sample moments. In this sense, the $QTSM3$ model performs as well as the preferred Gaussian model from DS2.

In comparing the results for the optimal weighting matrix and the identity matrix versions of the estimator, it seems clear that the statistically optimal matrix puts comparatively more weight on the first order autocorrelations of the factors. The point estimates of these moments under optimal W_T are quite close to the sample moments, and they are not statistically significantly different from the sample moments. In contrast, the estimated moments weighting all moments equally generates factor autocorrelations that are reliably different from the sample moments. However, the equally weighted estimator seems to do a better job of matching the volatility at the short end of the yield curve, when compared to the optimal-weight estimator. For example, the conditional volatility for the level factor is much closer to its sample moment, and the LPV regression coefficients for levels are also closer to their sample counterparts. The equally weighted SMM estimates generate too much dependence of the 10-year yield volatility on the matched slope of the yield curve.

7.2 Affine Models

The results for $\mathbb{A}_m(3)$, for $m = 0, 1, \ldots, 3$, are examined in Table 7 and Tables 9 through 12. Table 7 reports evidence on the overall fit of the affine models, using the test statistics in equations (23) and (25). The detailed results of each fitted moment for each estimated model are shown in Tables 9 through 12. The verdict on the general performance of the affine models can only be described as disappointing. None of the models is capable of explaining the complete set of economic moments. The Gaussian model, not surprisingly, fails to capture the patterns observed in the volatility dynamics of yields. However, the mixed Gaussian-square-root models, $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$ fare even worse! Although these models allow for both correlation among the state variables and time-varying volatility, issues that have been emphasized in the prior literature, the results in Table 7 demonstrate that they are incapable of matching either the correlation patterns (auto- and cross-correlations) or the patterns in volatility. The specific moments that the optimal- and identity-weighted estimators trade-off against each other illustrates the specific nature of the problems facing these affine models.

When the Gaussian $\mathbb{A}_0(3)$ model is estimated using equal weighting of the twenty-seven economic moments (see the right two columns in Table 9), the point estimates of the average factor moments (equivalently, the average shape of the yield curve) is virtually identical to the sample moments, as are two of the three factor autocorrelation coefficients. The equal-weighted estimator can also match the pattern of LPY regression coefficients quite closely. However, it cannot match the conditional factor volatility or the pattern of LPV regression coefficients. If anything, when the $\mathbb{A}_0(3)$ model is estimated with the optimal weighting matrix (i.e., taking account of the statistical precision with which each moment is measured), model fit seems to deteriorate. In trying to provide a closer fit to the precisely measured factor variances and the autocorrelation of the level factor, the model misses on the short-end of the LPY regressions.

The results for the $\mathbb{A}_1(3)$ model in Table 10 shows that it performs about as well (or as poorly) as the $\mathbb{A}_0(3)$ model. In particular, the $\mathbb{A}_1(3)$ model effectively matches the average shape of the term structure, and it closely matches the LPY regression coefficients The t-statistics on the factor volatility moments are not significantly different from zero, but this is as much a manifestation of imprecise measurement of the moments as it is of the accuracy of the model generated point estimates. As noted above, allowing for a single-square root factor allows for conditional volatility in yields, but it appears to be of a form that is not consistent with the sample data. The slope coefficients in the LPV regressions frequently have the wrong sign, and they are statistically significantly different from the actual moments. The $\mathbb{A}_1(3)$

model estimated with equal weighting of the moments also wants the factors to be even more persistent than is found in the data. Moving from equal weighting to optimal weighting of the moments does not improve the performance of the $\mathbb{A}_1(3)$ model. The point estimates of the factor volatilities are now farther from their sample counterparts, and they are statistically significantly different from the data. All of the LPY slope coefficients are now substantially less than their sample values, and the ability of the model to rationalize the LPV results does not improve.

The results for the $\mathbb{A}_2(3)$ model, shown in Table 11 are dramatically worse than the results for the $\mathbb{A}_0(3)$ or $\mathbb{A}_1(3)$. The $\mathbb{A}_2(3)$ is not capable of matching *any* of the important dimensions of the economic moments, whether these moments are equally weighted or optimally weighted. This model is particularly bad at capturing the yield regression results. Finally, as might have been anticipated from the existing literature and as documented in Table 12, the correlated square-root model, $\mathbb{A}_3(3)$, is inconsistent with all of the economic moments.[16]

8 Score-Based Estimates

8.1 The Quadratic Model

These results are not available yet.

8.2 Affine Models

These results are not available yet.

8.3 Summary

These results are not available yet.

[16]Currently, this statement is based solely on the $\mathbb{A}_3(3)$ model estimated with equal weighting of the economic moments.

9 Conclusions

The purpose of a reduced-form multifactor term structure model is to provide a consistent (i.e., arbitrage-free) explanation for the dynamics of the term structure. In this paper, we have proposed a broad set of stylized facts that can be used as a benchmark for evaluating the performance of alternative models, and we have incorporated these facts directly into a simulated method of moments estimator that can be applied to the most popular classes of three-factor arbitrage-free models. The factors that we focus on are constant across models, and they are the (more or less) standard choices of level, slope, and curvature, formed from the yields on 6-month, 2-year, and 10-year bonds. Our conclusion is that a simulated moment estimator based on key economic moments provides a substantial amount information about which multifactor model is most consistent with the data.

In particular, we find that matching the average shape of the yield curve is comparatively easy for three-factor affine and quadratic models, but that simultaneously matching both conditional holding period returns and conditional volatility is a powerful discriminator among models. Even though the simulated moment estimator is not asymptotically efficient, we are able to construct tests that strongly reject all models in the classes $\mathbb{A}_m(3)$ for $m = 0, 1, 2, 3$. For the Gaussian model, the obvious problem is matching conditional volatility, but the mixed $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$ models fare even worse at matching the important economic moments from the term structure. Although these models can generate conditional volatility, it is qualitatively different than the volatility dynamics uncovered by regressing squared holding period returns on the level, slope, and curvature factors.

The most flexible three-factor Gaussian-quadratic model is the clear winner in this horse race. While the overall test statistic based on statistically optimal weighting of the moments rejects the model, all of the tests based on groups of similar moment conditions fail to reject the equality of the model and sample moments. Furthermore, even where the model can be rejected formally, the model-implied moments generally conform to the level and patterns found in the sample moment conditions.

Table 1: Yield Level Means and Standard Deviations

	6-m	1-y	2-y	5-y	10-y
Mean	5.7741	5.9751	6.2219	6.5493	6.7491
Std. Dev.	2.9266	2.9023	2.8457	2.7621	2.7134

The source of the yield data is McCulloch and Kwon (1993) spliced to data constructed using the methods in Bliss (1997). Yields are measured monthly from January 1953 to December 1998. They are expressed in percent at an annual rate.

Table 2: Contemporaneous Yield Level Correlations.

	6-m	1-y	2-y	5-y	10-y
6-m	1.000	0.996	0.982	0.946	0.914
1-y		1.000	0.993	0.963	0.935
2-y			1.000	0.987	0.966
5-7				1.000	0.994
10-y					1.000

The source of the yield data is McCulloch and Kwon (1993) spliced to data constructed using the methods in Bliss (1997). Yields are measured monthly from January 1953 to December 1998. They are expressed in percent at an annual rate.

Table 3: Yield Level Correlations at Lag 1

	6-m	1-y	2-y	5-y	10-y
6-m	0.984	0.981	0.971	0.940	0.912
1-y		0.985	0.981	0.956	0.931
2-y			0.987	0.978	0.961
5-y				0.991	0.988
10-y					0.994

The source of the yield data is McCulloch and Kwon (1993) spliced to data constructed using the methods in Bliss (1997). Yields are measured monthly from January 1953 to December 1998. They are expressed in percent at an annual rate.

Table 4: LPY Slope Coefficients and R^2 Statistics

	6-m	1-y	2-y	5-y	10-y
Slope	-0.776	-1.195	-1.678	-2.564	-3.832
t-statistics	-2.842	-2.792	-2.834	-2.995	-3.172
R^2 (in %)	1.45	1.40	1.44	1.61	1.80

The slope coefficients are from ordinary least squares regressions of monthly yield changes, $y_{t+1}(\tau-1)-y_t(\tau)$, on matched maturity term structure slopes, $y_t(\tau)-r_t$. t-statistics are calculated using Newey-West standard errors with 6 lags.

Table 5: LPV Slope Coefficients and R^2 Statistics

Maturity	Level	Slope	Curvature	R^2
6-month	0.145 (7.61)	-0.157 (-3.31)	0.161 (1.50)	15.04
1-year	0.122 (8.01)	-0.123 (-3.24)	0.174 (2.03)	15.64
2-year	0.090 (7.52)	-0.082 (-2.75)	0.130 (1.93)	13.52
5-year	0.050 (9.38)	-0.011 (-0.78)	0.056 (1.85)	15.96
10-year	0.032 (10.21)	0.015 (1.88)	0.012 (0.07)	16.83

The slope coefficients are from ordinary least squares regressions of monthly squared yield changes, $[y_{t+1}(\tau-1)-y_t(\tau)]^2$, the level, slope, and curvature factors. t-statistics are calculated using Newey-West standard errors with 6 lags.

Table 6: Parameters of the Auxiliary Model for Yield Factors.

$$Y_t = \left[\begin{array}{cccc} 0.0561 & 0.9888 & 0.0249 & -0.0658 \\ (13.93) & (269.8) & (1.680) & (-1.992) \\ 0.0017 & 0.0061 & 0.9479 & 0.0764 \\ (0.089) & (1.552) & (75.98) & (2.788) \\ 0.0128 & -0.0065 & 0.0333 & 0.8280 \\ (0.5468) & (-1.540) & (2.847) & (29.67) \end{array}\right] \left[\begin{array}{c} 1 \\ Y_{t-1} \end{array}\right] + \varepsilon_t$$

$$h_{1t} = \exp\left(\left[\begin{array}{ccc} 0.1983 & 0.0750 & 0.0126 \\ (4.384) & (0.9874) & (0.0895) \end{array}\right] \cdot Y_{t-1}\right) \cdot v_{1t}$$

$$h_{2t} = \exp\left(\left[\begin{array}{ccc} 0.1545 & 0.1358 & -0.1246 \\ (4.696) & (1.999) & (-0.8733) \end{array}\right] \cdot Y_{t-1}\right) \cdot v_{2t}$$

$$h_{3t} = \exp\left(\left[\begin{array}{ccc} 0.0630 & -0.0990 & 0.0536 \\ (2.216) & (-1.477) & (0.359) \end{array}\right] \cdot Y_{t-1}\right) \cdot v_{3t}$$

$$v_{1t} = (1 - 0.0130 - 0.8634) \begin{array}{c} 0.0235 \\ (4.393) \end{array} + \begin{array}{c} 0.0130 \\ (1.631) \end{array} \varepsilon^2_{1t-1} \begin{array}{c} 0.8634 \\ (15.82) \end{array} v_{1t-1}$$

$$v_{2t} = (1 - 0.0337 - 0.7584) \begin{array}{c} 0.0219 \\ (4.442) \end{array} + \begin{array}{c} 0.0337 \\ (2.053) \end{array} \varepsilon^2_{2t-1} \begin{array}{c} 0.7584 \\ (10.32) \end{array} v_{2t-1}$$

$$v_{3t} = (1 - 0.1268 - 0.7263) \begin{array}{c} 0.0515 \\ (2.688) \end{array} + \begin{array}{c} 0.1268 \\ (2.605) \end{array} \varepsilon^2_{3t-1} \begin{array}{c} 0.7263 \\ (8.230) \end{array} v_{3t-1}$$

$$\mathrm{corr}(\widetilde{\varepsilon}_t) = \left[\begin{array}{ccc} 1 & & \\ \begin{array}{c} -0.681 \\ (24.89) \end{array} & 1 & \\ \begin{array}{c} -0.365 \\ (8.724) \end{array} & \begin{array}{c} 0.144 \\ (3.095) \end{array} & 1 \end{array}\right]$$

$$\zeta = \begin{array}{c} 4.9533 \\ (7.7088) \end{array}$$

$Y \equiv$ (level, slope, curvature)$'$, as defined in Section 5.2. The model parameters are estimated using maximum likelihood. $\widetilde{\varepsilon}_{it} \equiv \varepsilon_{it}/\sqrt{h_{it}}$, the standardized VAR residual. $\widetilde{\varepsilon}_t$ is distributed multivariate Student's t with ζ degrees of freedom. t-statistics for the null hypothesis that the parameter equals 0 are reported in parentheses.

Table 7: Moment-Based Tests of Model Performance

Model	Overall	Means & Var.	Correlations	LPY	LPV
$\mathbb{A}_0(3)$	21.648	10.781	13.752	4.766	19.842
	(0.001) [5]	(0.095) [6]	(0.131) [9]	(0.190) [3]	(0.019) [9]
$\mathbb{A}_1(3)$	86.719	33.625	26.086	4.575	35.278
	(0.000) [4]	(0.000) [6]	(0.002) [9]	(0.206) [3]	(0.000) [9]
$\mathbb{A}_2(3)$	428.9	1116	1511	17.857	6019
	(0.000) [4]	(0.000) [6]	(0.000) [9]	(0.000) [3]	(0.000) [9]
$\mathbb{A}_3(3)$	NA	NA	NA	NA	NA
$QTSM3$	19.406	4.2535	8.672	0.967	10.261
	(0.000) [2]	(0.642) [6]	(0.192) [9]	(0.809) [3]	(0.330) [9]

The **Overall** test statistic is defined in equation (23), and it has an asymptotic χ^2 distribution, with the number of degrees of freedom equal to the number of moment conditions (27) minus the number parameters in the model. The **Means & Var**, **Correlations**, **LPY**, and **LPV** statistics are all statistics of the form of equation (25) that test, jointly, whether the differences between sets of actual and model-generated moment conditions are equal to zero. NA indicates that these results are not currently available.

Table 8: Actual Moments and SMM Estimates for the QTSM3 Model.

Panel A: Means, Conditional Volatility, and Autocorrelations.

		Optimal W_T		Identity W_T	
	Sample Moment	Population Moment	N-W t-statistic	Population Moment	N-W t-statistic
Means					
Level	5.777	5.594	1.038	5.782	−0.031
Slope	0.976	1.022	−0.659	0.985	−0.121
Curvature	0.079	0.097	−0.591	0.084	−0.156
Cond. Volatility					
Level	0.527	**0.423**	**1.923**	0.540	−0.225
Slope	0.397	0.332	1.781	0.379	0.518
Curvature	0.310	0.280	1.348	0.319	−0.389
First-Order Autocorr.					
Level	0.984	0.989	−1.791	**0.944**	**8.246**
Slope	0.942	0.958	−1.541	**0.913**	**2.865**
Curvature	0.804	0.840	−1.176	**0.873**	**−2.222**
First-Order Cross-corr.					
Level & Slope	−0.341	−0.327	−0.233	−0.319	−0.353
Level & Curv.	−0.328	−0.378	0.849	−0.360	0.537
Slope & Curv.	0.349	**0.501**	**−2.108**	0.345	0.048

The table description is on the next page.

Table 8 (continued):

Panel B: LPY and LPV Regression Slope Coefficients.

	Sample Moment	Optimal W_T Population Moment	Optimal W_T N-W t-statistic	Identity W_T Population Moment	Identity W_T N-W t-statistic
LPY Slope Coeff.					
6-m	−0.776	−0.770	−0.012	−0.821	0.095
2-y	−1.678	−1.740	0.063	−1.662	−0.016
10-y	−3.831	−3.066	−0.414	−3.835	0.002
LPV Slope Coeff.					
6-m on Level	0.145	0.073	1.603	0.166	−0.461
6-m on Slope	−0.157	0.061	−1.697	−0.192	0.272
6-m on Curv.	0.161	0.016	0.835	0.090	0.410
2-y on Level	0.090	**0.036**	**2.118**	0.059	1.219
2-y on Slope	−0.082	0.050	−1.818	−0.103	0.284
2-y on Curv.	0.130	0.010	0.968	0.029	0.818
10-y on Level	0.032	**0.020**	**2.124**	**0.050**	**−3.191**
10-y on Slope	0.015	0.021	−0.480	**0.043**	**−2.087**
10-y on Curv.	0.012	0.006	0.200	0.020	−0.273

The level factor is defined as the yield on the 6-month bond. The slope factor is the difference between the yield on the 120-month and 6-month bond, and the curvature factor is defined as the sum of the 6-month and 120-month yields minus two times the yield on the 24-month bond. Conditional volatility is the standard deviation of the residual in the first-order autoregression for each factor. The first order autocorrelation measures $corr\,(F_{it}, F_{it-1})$ for $i = 1, 2, 3$, and the first-order cross-correlation is defined as $corr\,(F_{it}, F_{jt-1})$ for $i \neq j$. The LPY and LPV regressions are defined in the text in equations (14) and (18), respectively. The Newey-West standard error estimator is constructed using 6 lags.

Table 9: Actual Moments and SMM Estimates for the $\mathbb{A}_0(3)$ Model.

Panel A: Means, Conditional Volatility, and Autocorrelations.

		Optimal W_T		Identity W_T	
	Sample Moment	Population Moment	N-W t-statistic	Population Moment	N-W t-statistic
Means					
Level	5.777	5.609	0.952	5.776	0.001
Slope	0.976	1.017	0.575	0.979	0.042
Curvature	0.079	0.088	0.295	0.076	0.105
Cond. Volatility					
Level	0.527	**0.369**	**2.927**	0.529	0.028
Slope	0.397	**0.309**	**2.420**	0.353	1.204
Curvature	0.310	0.279	1.390	**0.357**	**2.138**
First-Order Autocorr.					
Level	0.984	**0.993**	**2.992**	0.984	0.293
Slope	0.942	**0.966**	**2.354**	**0.997**	**5.303**
Curvature	0.804	0.834	0.968	0.791	0.390
First-Order Cross-corr.					
Level & Slope	−0.341	−0.345	0.070	−0.356	0.254
Level & Curv.	−0.328	−0.267	1.020	−0.331	0.055
Slope & Curv.	0.349	0.343	0.081	0.368	0.270

The table description is on the next page.

Table 9 (continued):

Panel B: LPY and LPV Regression Slope Coefficients.

		Optimal W_T		Identity W_T	
	Sample Moment	Population Moment	N-W t-statistic	Population Moment	N-W t-statistic
LPY Slope Coeff.					
6-m	−0.776	**0.222**	**2.095**	−0.782	0.014
2-y	−1.678	−1.256	0.429	−1.679	0.002
10-y	−3.831	−4.191	0.194	−3.834	0.002
LPV Slope Coeff.					
6-m on Level	0.145	**0.000**	**3.212**	**−0.003**	**3.272**
6-m on Slope	−0.157	0.000	1.221	−0.028	1.006
6-m on Curv.	0.161	−0.006	0.964	0.012	0.861
2-y on Level	0.090	**0.004**	**3.320**	**−0.012**	**3.972**
2-y on Slope	−0.082	0.010	1.269	−0.021	0.838
2-y on Curv.	0.130	−0.011	1.140	−0.032	1.309
10-y on Level	0.032	**0.011**	**3.744**	**−0.0001**	**5.725**
10-y on Slope	0.015	0.002	0.604	−0.0002	1.088
10-y on Curv.	0.012	−0.014	0.945	−0.0004	0.444

The level factor is defined as the yield on the 6-month bond. The slope factor is the difference between the yield on the 120-month and 6-month bond, and the curvature factor is defined as the sum of the 6-month and 120-month yields minus two times the yield on the 24-month bond. Conditional volatility is the standard deviation of the residual in the first-order autoregression for each factor. The first order autocorrelation measures $corr\left(F_{it}, F_{it-1}\right)$ for $i = 1, 2, 3$, and the first-order cross-correlation is defined as $corr\left(F_{it}, F_{jt-1}\right)$ for $i \neq j$. The LPY and LPV regressions are defined in the text in equations (14) and (18), respectively. The Newey-West standard error estimator is constructed using 6 lags.

Table 10: Actual Moments and SMM Estimates for the $\mathbb{A}_1(3)$ Model.

Panel A: Means, Conditional Volatility, and Autocorrelations.

		Optimal W_T		Identity W_T	
	Sample Moment	Population Moment	N-W t-statistic	Population Moment	N-W t-statistic
Means					
Level	5.777	5.171	1.877	5.779	0.009
Slope	0.976	1.216	1.921	0.979	0.021
Curvature	0.079	0.168	1.811	0.078	0.033
Cond. Volatility					
Level	0.527	**0.337**	**2.655**	0.462	0.905
Slope	0.397	**0.288**	**2.315**	0.451	1.121
Curvature	0.310	**0.190**	**4.526**	0.285	0.925
First-Order Autocorr.					
Level	0.984	0.989	1.557	**0.998**	**4.099**
Slope	0.942	0.966	1.915	0.931	0.908
Curvature	0.804	**0.999**	**4.635**	**0.966**	**3.868**
First-Order Cross-corr.					
Level & Slope	−0.341	**−0.207**	**1.332**	−0.346	0.057
Level & Curv.	−0.328	**−0.483**	**1.804**	−0.362	0.404
Slope & Curv.	0.349	**0.734**	**3.525**	0.467	1.079

The table description is on the next page.

Table 10 (continued):

Panel B: LPY and LPV Regression Slope Coefficients.

		Optimal W_T		Identity W_T	
	Sample Moment	Population Moment	N-W t-statistic	Population Moment	N-W t-statistic
LPY Slope Coeff.					
6-m	−0.776	−0.381	0.785	−0.799	0.046
2-y	−1.678	−0.036	1.694	−1.664	0.014
10-y	−3.831	**−0.216**	**2.066**	−3.837	0.003
LPV Slope Coeff.					
6-m on Level	0.145	0.091	1.174	**−0.043**	**4.046**
6-m on Slope	−0.157	−0.028	0.963	−0.084	0.547
6-m on Curv.	0.161	0.316	0.779	0.158	0.019
2-y on Level	0.090	**0.028**	**1.984**	**−0.017**	**3.433**
2-y on Slope	−0.082	0.060	1.621	−0.021	0.702
2-y on Curv.	0.130	0.019	0.721	−0.003	0.862
10-y on Level	0.032	0.021	1.744	**−0.025**	**9.425**
10-y on Slope	0.015	0.035	1.259	**−0.036**	**3.106**
10-y on Curv.	0.012	0.016	0.129	0.001	0.323

The level factor is defined as the yield on the 6-month bond. The slope factor is the difference between the yield on the 120-month and 6-month bond, and the curvature factor is defined as the sum of the 6-month and 120-month yields minus two times the yield on the 24-month bond. Conditional volatility is the standard deviation of the residual in the first-order autoregression for each factor. The first order autocorrelation measures $corr\,(F_{it}, F_{it-1})$ for $i = 1, 2, 3$, and the first-order cross-correlation is defined as $corr\,(F_{it}, F_{jt-1})$ for $i \neq j$. The LPY and LPV regressions are defined in the text in equations (14) and (18), respectively. The Newey-West standard error estimator is constructed using 6 lags.

Table 11: Actual Moments and SMM Estimates for the $\mathbb{A}_2(3)$ Model.

Panel A: Means, Conditional Volatility, and Autocorrelations.

		Optimal W_T		Identity W_T	
	Sample Moment	Population Moment	N-W t-statistic	Population Moment	N-W t-statistic
Means					
Level	5.777	**5.099**	**2.099**	5.353	1.311
Slope	0.976	**0.474**	**4.021**	**0.199**	**6.228**
Curvature	0.079	**0.294**	**4.372**	**0.185**	**2.157**
Cond. Volatility					
Level	0.527	**0.327**	**2.789**	0.656	1.790
Slope	0.397	**0.291**	**2.244**	**0.199**	**4.190**
Curvature	0.310	**0.225**	**3.192**	**0.196**	**4.312**
First-Order Autocorr.					
Level	0.984	**0.992**	**2.565**	**0.996**	**3.748**
Slope	0.942	0.964	1.734	**0.977**	**2.824**
Curvature	0.804	**0.962**	**3.755**	**0.945**	**3.358**
First-Order Cross-corr.					
Level & Slope	−0.341	**−0.588**	**2.469**	**0.717**	**10.57**
Level & Curv.	−0.328	−0.463	1.578	**0.882**	**14.08**
Slope & Curv.	0.349	**0.587**	**2.179**	**0.679**	**3.015**

The table description is on the next page.

Table 11 (continued):

Panel B: LPY and LPV Regression Slope Coefficients.

	Sample Moment	Optimal W_T Population Moment	Optimal W_T N-W t-statistic	Identity W_T Population Moment	Identity W_T N-W t-statistic
LPY Slope Coeff.					
6-m	−0.776	**1.109**	**3.744**	**0.916**	**3.360**
2-y	−1.678	**1.125**	**2.892**	**0.322**	**2.063**
10-y	−3.831	**1.230**	**2.892**	−3.594	0.136
LPV Slope Coeff.					
6-m on Level	0.145	**0.000**	**3.117**	**−0.025**	**3.671**
6-m on Slope	−0.157	−0.050	0.796	**0.125**	**2.097**
6-m on Curv.	0.161	0.094	0.339	**0.880**	**3.623**
2-y on Level	0.090	**−0.000**	**2.0902**	**−0.020**	**3.539**
2-y on Slope	−0.082	−0.016	0.757	0.075	1.790
2-y on Curv.	0.130	0.031	0.641	**0.576**	**2.889**
10-y on Level	0.032	**−0.001**	**5.404**	**−0.029**	**10.18**
10-y on Slope	0.015	−0.008	1.380	**0.100**	**5.250**
10-y on Curv.	0.012	0.014	0.072	**0.786**	**23.67**

The level factor is defined as the yield on the 6-month bond. The slope factor is the difference between the yield on the 120-month and 6-month bond, and the curvature factor is defined as the sum of the 6-month and 120-month yields minus two times the yield on the 24-month bond. Conditional volatility is the standard deviation of the residual in the first-order autoregression for each factor. The first order autocorrelation measures $corr\,(F_{it}, F_{it-1})$ for $i = 1, 2, 3$, and the first-order cross-correlation is defined as $corr\,(F_{it}, F_{jt-1})$ for $i \neq j$. The LPY and LPV regressions are defined in the text in equations (14) and (18), respectively. The Newey-West standard error estimator is constructed using 6 lags.

Table 12: Actual Moments and SMM Estimates for the $\mathbb{A}_3(3)$ Model.

Panel A: Means, Conditional Volatility, and Autocorrelations.

		Optimal W_T		Identity W_T	
	Sample Moment	Population Moment	N-W t-statistic	Population Moment	N-W t-statistic
Means					
Level	5.777	NA	NA	**7.192**	**8.060**
Slope	0.976	NA	NA	**−0.780**	**24.89**
Curvature	0.079	NA	NA	**0.018**	**2.065**
Cond. Volatility					
Level	0.527	NA	NA	0.596	1.264
Slope	0.397	NA	NA	**0.535**	**3.763**
Curvature	0.310	NA	NA	**0.162**	**6.670**
First-Order Autocorr.					
Level	0.984	NA	NA	**0.956**	**9.248**
Slope	0.942	NA	NA	0.953	1.051
Curvature	0.804	NA	NA	0.854	1.622
First-Order Cross-corr.					
Level & Slope	−0.341	NA	NA	−0.954	10.09
Level & Curv.	−0.328	NA	NA	−0.223	1.755
Slope & Curv.	0.349	NA	NA	0.203	2.012

The table description is on the next page.

Table 12 (continued):

Panel B: LPY and LPV Regression Slope Coefficients.

		Optimal W_T		Identity W_T	
	Sample Moment	Population Moment	N-W t-statistic	Population Moment	N-W t-statistic
LPY Slope Coeff.					
6-m	-0.776	NA	NA	**0.837**	**3.388**
2-y	-1.678	NA	NA	**0.377**	**2.090**
10-y	-3.831	NA	NA	**0.455**	**2.319**
LPV Slope Coeff.					
6-m on Level	0.145	NA	NA	**−2.746**	**63.99**
6-m on Slope	-0.157	NA	NA	**−3.344**	**24.72**
6-m on Curv.	0.161	NA	NA	**−0.752**	**5.244**
2-y on Level	0.090	NA	NA	**4.757**	**180.7**
2-y on Slope	-0.082	NA	NA	**5.350**	**74.43**
2-y on Curv.	0.130	NA	NA	**0.609**	**3.860**
10-y on Level	0.032	NA	NA	**0.315**	**50.60**
10-y on Slope	0.015	NA	NA	**0.355**	**24.98**
10-y on Curv.	0.012	NA	NA	0.043	1.101

The level factor is defined as the yield on the 6-month bond. The slope factor is the difference between the yield on the 120-month and 6-month bond, and the curvature factor is defined as the sum of the 6-month and 120-month yields minus two times the yield on the 24-month bond. Conditional volatility is the standard deviation of the residual in the first-order autoregression for each factor. The first order autocorrelation measures $corr\,(F_{it}, F_{it-1})$ for $i = 1, 2, 3$, and the first-order cross-correlation is defined as $corr\,(F_{it}, F_{jt-1})$ for $i \neq j$. The LPY and LPV regressions are defined in the text in equations (14) and (18), respectively. The Newey-West standard error estimator is constructed using 6 lags. NA means that these estimates are currently not available.

	$\mathbb{A}_0(3)$	$\mathbb{A}_1(3)$	$\mathbb{A}_2(3)$	$\mathbb{A}_3(3)$
$\mathcal{K}$	$\begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix}$	$\begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix}$	$\begin{bmatrix} \kappa_{11} & \kappa_{12} & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix}$	$\begin{bmatrix} \kappa_{11} & 0 & 0 \\ 0 & \kappa_{22} & 0 \\ 0 & 0 & \kappa_{33} \end{bmatrix}$
$\mathcal{B}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
α'	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$
Σ	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \Sigma_{11} & 0 & 0 \\ 0 & \Sigma_{22} & 0 \\ 0 & 0 & \Sigma_{33} \end{bmatrix}$
$[\delta_0, \delta]'$	$\begin{bmatrix} \delta_0 & \delta_1 & \delta_2 & \delta_3 \end{bmatrix}$	$\begin{bmatrix} \delta_0 & \delta_1 & \delta_2 & \delta_3 \end{bmatrix}$	$\begin{bmatrix} \delta_0 & \delta_1 & \delta_2 & \delta_3 \end{bmatrix}$	$\begin{bmatrix} \delta_0 & 1 & 1 & 1 \end{bmatrix}$
θ'	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} \theta_1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} \theta_1 & \theta_2 & 0 \end{bmatrix}$	$\begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \end{bmatrix}$
λ_1'	$\begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \end{bmatrix}$	$\begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \end{bmatrix}$	$\begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \end{bmatrix}$	$\begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \end{bmatrix}$
λ_2	$\begin{bmatrix} \lambda_{2(11)} & \lambda_{2(12)} & \lambda_{2(13)} \\ \lambda_{2(21)} & \lambda_{2(22)} & \lambda_{2(23)} \\ \lambda_{2(31)} & \lambda_{2(32)} & \lambda_{2(33)} \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ \lambda_{2(21)} & \lambda_{2(22)} & \lambda_{2(23)} \\ \lambda_{2(31)} & \lambda_{2(32)} & \lambda_{2(33)} \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \lambda_{2(31)} & \lambda_{2(32)} & \lambda_{2(33)} \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
$\dim(\pi)$	22	23	21	13

Figure 1: Three-Factor Affine Model Parameter Restrictions

A Model Parameter Estimates

In this appendix, in the interest of completeness, we provide the model parameter estimates for the affine and quadratic models constructed using the EMM and SMM estimators discussed in the text.

Table A-1: SMM Estimates of the QTSM Model Parameters

	Weighting Matrix	
Parameters	Identity	Optimal
α	0.207 (0.013)	0.141 (0.016)
Ψ_{12}	1.409 (0.121)	1.580 (0.265)
Ψ_{13}	0.780 (0.062)	1.073 (0.216)
Ψ_{23}	1.539 (0.023)	1.328 (0.088)
μ_1	0.229 (0.008)	0.121 (0.131)
μ_2	0.038 (0.000)	0.040 (0.002)
μ_3	0.618 (0.004)	0.963 (0.064)
ξ_{11}	−3.131 (0.109)	−1.759 (2.327)
ξ_{21}	0.018 (0.003)	0.065 (0.035)
ξ_{31}	2.112 (0.060)	1.030 (0.848)
ξ_{22}	−0.105 (0.001)	−0.113 (0.006)
ξ_{32}	−2.316 (0.012)	−3.067 (0.165)
ξ_{33}	−0.142 (0.007)	−0.305 (0.117)
Σ_{11}	0.076 (0.002)	0.068 (0.016)
Σ_{22}	0.004 (0.000)	0.000 (0.004)
Σ_{33}	0.020 (0.002)	0.049 (0.008)
$\mu_1 - \delta_{01}$	0.280 (0.030)	0.272 (0.135)
$\mu_2 - \delta_{02}$	0.271 (0.004)	0.126 (0.086)
$\mu_3 - \delta_{03}$	−0.255 (0.003)	−0.119 (0.031)
$\xi_{11} - \delta_{111}$	−4.248 (0.446)	−2.608 (1.490)
$\xi_{21} - \delta_{121}$	1.402 (0.054)	3.060 (0.899)
$\xi_{31} - \delta_{131}$	−1.306 (0.049)	−0.726 (0.235)
$\xi_{22} - \delta_{122}$	−1.013 (0.010)	−1.122 (0.227)
$\xi_{32} - \delta_{132}$	0.913 (0.009)	0.441 (0.070)
$\xi_{33} - \delta_{133}$	−0.014 (0.003)	−0.051 (0.029)

The parameters are defined in Section 2 in the text. The SMM parameter estimates were computed following an algorithm similar to Duffee (2002): given initial parameter values, the SMM objective function was minimized using a simplex algorithm and then re-optimized with a gradient-based method. Standard errors are reported in parentheses using the Newey-West estimator with 6 lags.

Table A-2: SMM Estimates of ATSM Models, Identity Matrix Weights.

Parameters	$\mathbb{A}_0(3)$	$\mathbb{A}_1(3)$	$\mathbb{A}_2(3)$	$\mathbb{A}_3(3)$
κ_{11}	3.578 (0.115)	0.035 (0.019)	1.759 (0.415)	2.585 (0.339)
κ_{21}	1.940 (0.072)	−1.264 (0.327)	7.204 (4.137)	0
κ_{31}	4.827 (0.451)	−0.850 (0.640)	1.860 (1.476)	0
κ_{12}	0	0	0.074 (0.047)	0
κ_{22}	0.005 (0.021)	1.317 (0.114)	0.347 (0.137)	0.255 (0.044)
κ_{32}	0.862 (0.045)	−1.297 (0.749)	1.728 (0.615)	0
κ_{13}	0	0	0	0
κ_{23}	0	5.127 (0.191)	0	0
κ_{33}	0.232 (0.023)	0.004 (0.007)	1.256 (0.463)	0.023 (0.057)
θ_1	0	7.249 (0.763)	0.280 (0.062)	0.053 (0.015)
θ_2	0	0	0.596 (0.604)	0.010 (0.007)
θ_3	0	0	0	0.016 (0.010)
β_{11}	0	0	0	0
β_{21}	0	0.019 (0.013)	0	0
β_{31}	0	0.001 (0.001)	0.609 (0.8.987)	0
β_{12}	0	0	0	0
β_{22}	0	0	0	0
β_{32}	0	0	0.002 (0.311)	0
β_{13}	0	0	0	0
β_{23}	0	0	0	0
β_{33}	0	0	0	0
δ_0	0.058 (0.018)	0.049 (0.018)	−0.007 (0.002)	0.010 (0.002)
δ_1	0.005 (0.006)	0.002 (0.001)	0.039 (0.011)	1
δ_2	0.016 (0.006)	−0.013 (0.005)	−0.001 (0.001)	1
δ_3	0.005 (0.002)	−0.033 (0.009)	0.004 (0.001)	1
λ_{11}	3.037 (1.473)	0.079 (0.059)	−0.104 (0.044)	0.010 (0.023)
λ_{12}	−2.308 (1.577)	−10.02 (1.442)	−0.039 (0.015)	0.024 (0.003)
λ_{13}	5.911 (1.872)	−2.831 (0.725)	0.387 (0.450)	0.036 (0.003)
$\lambda_{2(11)}$	4.494 (0.865)	0	0	0
$\lambda_{2(12)}$	0.351 (0.077)	0	0	0
$\lambda_{2(13)}$	−0.213 (0.047)	0	0	0
$\lambda_{2(21)}$	0.369 (0.072)	−1.249 (0.230)	0	0
$\lambda_{2(22)}$	−0.468 (0.129)	−2.093 (0.415)	0	0
$\lambda_{2(23)}$	−0.024 (0.015)	0.717 (0.157)	0	0
$\lambda_{2(31)}$	7.955 (0.943)	−0.199 (0.352)	−0.022 (0.108)	0
$\lambda_{2(32)}$	1.685 (1.057)	−6.422 (1.086)	0.006 (0.013)	0
$\lambda_{2(33)}$	−0.055 (0.147)	−7.649 (0.251)	0.021 (0.080)	0
Σ_{11}	1	1	1	0.103 (0.002)
Σ_{22}	1	1	1	0.155 (0.002)
Σ_{33}	1	1	1	0.176 (0.004)

See the next page for notes.

Notes to Table A-2:

The parameters are defined in Section 2 in the text. Parameter restrictions are shown in Figure 1. The SMM parameter estimates were computed following an algorithm similar to Duffee (2002): given initial parameter values, the SMM objective function was minimized using a simplex algorithm and then re-optimized with a gradient-based method. Standard errors are reported in parentheses using the Newey-West estimator with 6 lags.

Table A-3: SMM Estimates of ATSM Models, Optimal Weighting Matrix.

Parameters	$\mathbb{A}_0(3)$	$\mathbb{A}_1(3)$	$\mathbb{A}_2(3)$	$\mathbb{A}_3(3)$
κ_{11}	2.813 (0.058)	0.018 (0.007)	1.826 (0.221)	NA
κ_{21}	−7.339 (0.227)	−1.039 (0.249)	0.521 (0.916)	NA
κ_{31}	16.11 (1.299)	−0.188 (0.117)	3.349 (0.822)	NA
κ_{12}	0	0	0.045 (0.029)	NA
κ_{22}	0.131 (0.056)	−0.166 (0.041)	0.016 (0.009)	NA
κ_{32}	0.621 (0.023)	−0.021 (0.017)	0.345 (0.081)	NA
κ_{13}	0	0	0	NA
κ_{23}	0	4.790 (0.190)	0	NA
κ_{33}	0.484 (0.033)	1.094 (0.208)	0.159 (0.060)	NA
θ_1	0	7.522 (0.556)	0.384 (0.067)	NA
θ_2	0	0	1.029 (0.255)	NA
θ_3	0	0	0	NA
β_{11}	0	0	0	NA
β_{21}	0	0.112 (0.050)	0	NA
β_{31}	0	0.428 (0.102)	1.305 (0.648)	NA
β_{12}	0	0	0	NA
β_{22}	0	0	0	NA
β_{32}	0	0	0.002 (0.037)	NA
β_{13}	0	0	0	NA
β_{23}	0	0	0	NA
β_{33}	0	0	0	NA
δ_0	0.059 (0.018)	0.044 (0.009)	0.001 (0.000)	NA
δ_1	−0.005 (0.003)	−0.001 (0.001)	0.018 (0.009)	NA
δ_2	−0.011 (0.003)	0.005 (0.002)	0.000 (0.000)	NA
δ_3	−0.003 (0.000)	0.019 (0.000)	0.007 (0.001)	NA
λ_{11}	1.797 (0.368)	0.134 (0.038)	−0.082 (0.020)	NA
λ_{12}	−5.278 (0.930)	−8.225 (1.028)	0.000 (0.000)	NA
λ_{13}	12.10 (1.897)	−2.554 (0.802)	−0.141 (0.099)	NA
$\lambda_{2(11)}$	3.374 (0.558)	0	0	NA
$\lambda_{2(12)}$	0.937 (0.178)	0	0	NA
$\lambda_{2(13)}$	0.474 (0.063)	0	0	NA
$\lambda_{2(21)}$	−5.190 (0.680)	−1.675 (0.211)	0	NA
$\lambda_{2(22)}$	−2.212 (0.089)	−2.064 (0.437)	0	NA
$\lambda_{2(23)}$	−0.952 (0.322)	1.489 (0.138)	0	NA
$\lambda_{2(31)}$	9.799 (1.164)	−0.090 (0.102)	−0.002 (0.003)	NA
$\lambda_{2(32)}$	5.819 (1.799)	−4.222 (0.813)	−0.009 (0.008)	NA
$\lambda_{2(33)}$	2.535 (1.037)	−7.077 (0.210)	−0.004 (0.002)	NA
Σ_{11}	1	1	1	NA
Σ_{22}	1	1	1	NA
Σ_{33}	1	1	1	NA

See the notes to **Table A-2** for a description of the results in the Table.

Table A-4: EMM Estimates of the QTSM Model Parameters.

Parameters	
α	NA
Ψ_{12}	NA
Ψ_{13}	NA
Ψ_{23}	NA
μ_1	NA
μ_2	NA
μ_3	NA
ξ_{11}	NA
ξ_{21}	NA
ξ_{31}	NA
ξ_{22}	NA
ξ_{32}	NA
ξ_{33}	NA
Σ_{11}	NA
Σ_{22}	NA
Σ_{33}	NA
$\mu_1 - \delta_{01}$	NA
$\mu_2 - \delta_{02}$	NA
$\mu_3 - \delta_{03}$	NA
$\xi_{11} - \delta_{111}$	NA
$\xi_{21} - \delta_{121}$	NA
$\xi_{31} - \delta_{131}$	NA
$\xi_{22} - \delta_{122}$	NA
$\xi_{32} - \delta_{132}$	NA
$\xi_{33} - \delta_{133}$	NA

The parameters are defined in Section 2 in the text. The SMM parameter estimates were computed following an algorithm similar to Duffee (2002): given initial parameter values, the SMM objective function was minimized using a simplex algorithm and then re-optimized with a gradient-based method. Standard errors are reported in parentheses using the Newey-West estimator with 6 lags.

Table A-5: EMM Estimates of ATSM Model Parameters

Parameters	$\mathbb{A}_0\,(3)$	$\mathbb{A}_1\,(3)$	$\mathbb{A}_2\,(3)$	$\mathbb{A}_3\,(3)$
κ_{11}	NA	NA	NA	NA
κ_{21}	NA	NA	NA	NA
κ_{31}	NA	NA	NA	NA
κ_{12}	NA	NA	NA	NA
κ_{22}	NA	NA	NA	NA
κ_{23}	NA	NA	NA	NA
κ_{13}	NA	NA	NA	NA
κ_{23}	NA	NA	NA	NA
κ_{33}	NA	NA	NA	NA
$\mathcal{K}\theta_1$	NA	NA	NA	NA
$\mathcal{K}\theta_2$	NA	NA	NA	NA
$\mathcal{K}\theta_3$	NA	NA	NA	NA
β_{11}	NA	NA	NA	NA
β_{21}	NA	NA	NA	NA
β_{31}	NA	NA	NA	NA
β_{12}	NA	NA	NA	NA
β_{22}	NA	NA	NA	NA
β_{23}	NA	NA	NA	NA
β_{13}	NA	NA	NA	NA
β_{23}	NA	NA	NA	NA
β_{33}	NA	NA	NA	NA
δ_0	NA	NA	NA	NA
δ_1	NA	NA	NA	NA
δ_2	NA	NA	NA	NA
δ_3	NA	NA	NA	NA
λ_{11}	NA	NA	NA	NA
λ_{12}	NA	NA	NA	NA
λ_{13}	NA	NA	NA	NA
$\lambda_{2(11)}$	NA	NA	NA	NA
$\lambda_{2(12)}$	NA	NA	NA	NA
$\lambda_{2(13)}$	NA	NA	NA	NA
$\lambda_{2(21)}$	NA	NA	NA	NA
$\lambda_{2(22)}$	NA	NA	NA	NA
$\lambda_{2(23)}$	NA	NA	NA	NA
$\lambda_{2(31)}$	NA	NA	NA	NA
$\lambda_{2(32)}$	NA	NA	NA	NA
$\lambda_{2(33)}$	NA	NA	NA	NA
Σ_{11}	NA	NA	NA	NA
Σ_{22}	NA	NA	NA	NA
Σ_{33}	NA	NA	NA	NA

See the notes to **Table A-2** for a description of the results in the Table.

References

[1] Ahn, D-H., R. F. Dittmar, and A. R. Gallant, 2002, "Quadratic Term Structure Models: Theory and Evidence," *Review of Financial Studies*, 15, pages 243-288.

[2] Aït-Sahalia, Y., 2001, "Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, pages 223-262.

[3] Andersen, T. G., H-J Chung, and B. E. Sørensen, 1999, "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Econometrics,* 91, pages 61-87.

[4] Backus, D., S. Foresi, A. Mozumdar, and L. Wu, 2001, "Predictable Changes in Yields and Forward Rates," *Journal of Financial Economics,* 59, pages 281-311.

[5] Bliss, R. R., 1997, "Testing Term Structure Estimation Methods," *Advances in Futures and Options Research,* 9, pages 197-231.

[6] Brandt, M. W., and P. He, 2002, "Simulated Likelihood Estimation of Affine Term Structure Models from Panel Data," Manuscript, The Wharton School and Economics Department, University of Pennsylvania.

[7] Brandt, M. W., and P. Santa-Clara, 2002, "Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets," *Journal of Financial Economics*, 63,161-210.

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

[9] Chapman, D. A., and N. D. Pearson, 2001, ,"Recent Advances in Estimating Term-Structure Models," *Financial Analysts Journal*, 57(4), pages 77-95.

[10] Chumacero, R., 1997, "Finite Sample Properties of the Efficient Method of Moments," *Studies in Nonlinear Dynamics and Econometrics,* 2, pages 35-51.

[11] Cochrane, J. H., 1996, "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy,* 104, pages 572-621.

[12] Cox, J. C., J. Ingersoll, and S. A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica,* 53, pages 385-407.

[13] Dai, Q. and K. J. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," *Journal of Finance,* 55, pages 1943-1978.

[14] ————————, 2002a, "Expectations Puzzles, Time-Varying Risk Premia, and Dynamic Models of the Term Structure," *Journal of Financial Economics,* 63, pages 415-441.

[15] ————————, 2002b, "Term Structure Dynamics in Theory and Reality," Manuscript, New York University and Stanford University.

[16] Duffee, G. R., 2002, "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance,* 57, pages 405-443.

[17] Duffee, G. R., and R. H. Stanton, 2001, "Estimation of Dynamic Term Structure Models," Manuscript, Finance Department, Haas School of Business, University of California at Berkeley.

[18] Duffie, D., and K. J. Singleton, 1993, "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica,* 61, pages 929-952.

[19] Duffie, D., and R. Kan, 1996, “A Yield-Factor Model of Interest Rates,” *Mathematical Finance* 6, pages 379-406.

[20] Fama, E. F. and R. R. Bliss, 1987, “The Information in Long-Maturity Forward Rates,” *American Economic Review,* 77, pages 680-692.

[21] Gallant, A. R., and G. Tauchen, 1989, “Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications,” *Econometrica,* 57, pages 1091-1120.

[22] ———, 1996, “Which Moments to Match?,” *Econometric Theory,* 12, pages 657-681.

[23] ———, 1999, “The Relative Efficiency of Method of Moments Estimators,” *Journal of Econometrics,* 92, pages 149-172.

[24] ———, 2000, “Efficient Method of Moments,” Manuscript, University of North Carolina and Duke University.

[25] Hansen, L. P., 1982, “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica, 50*, pages 1029-1054.

[26] Lee, B., and B. Ingram, 1991, “Simulation Estimation of Time Series Models,” *Journal of Econometrics*, 47, pages 197-205

[27] Litterman, R. and J. Scheinkman, 1991, “Common Factors Affecting Bond Returns,” *Journal of Fixed Income,* 3, pages 54-61.

[28] Lo, A. W., 1988, “Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data,” *Econometric Theory*, 4, pages 231-247.

[29] McCulloch, J. H., and H. Kwon, 1993, "U.S. Term Structure Data, 1947-1991," Manuscript, Economics Department, The Ohio State University.

[30] Piazzesi, M, 2002, "Affine Term Structure Models," Manuscript, Anderson School, UCLA.

[31] Singleton, K. J., 2001, "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, pages 111-141.

[32] Vasicek, O., 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics,* 5, pages 177-188.