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

First Draft: May 2002
This Draft: October 2002

Abstract
There are a large number of reduced-form multifactor term structure models 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 the canonical three-factor Gaussian-quadratic model dominates the three factor essentially affine models at matching the economic moments.

*We thank Andrea Buraschi, Mike Chernov, Greg Duffee, Rob Engle, Tim Johnson, Neil Pearson, Allen Poteshman, Jacob Sagi and seminar participants at Berkeley, Illinois, LBS, LSE, and Rice for comments.
†Philadelphia, PA 19104. Phone: (215) 898-3600. E-mail: brandtm@wharton.upenn.edu.
‡Austin, Texas 78712. Phone: (512) 471-6621. E-mail: chapman@eco.utexas.edu.
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 auto-correlations 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 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 for future research.

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 and the SMM estimator allow for powerful distinctions between affine and quadratic models. In particular, simultaneously matching the conditional mean of holding period returns and conditional return volatilities is impossible for all but the three-factor essentially affine model that has one square-root process driving variations in conditional volatility. Even this model is rejected at conventional significance levels by a standard test of overall model fit.

However, the unconstrained three-factor Gaussian-quadratic model emerges as the clear winner in this “horse race.” The model is not rejected by the overall specification test. Tests based on smaller groups of similar moments, for example the slope coefficients in holding period regressions or all cross- and autocorrelations, also fail to reject the equality of model and sample moments. The individual point estimates from the quadratic model reproduce the magnitude and patterns found in the sample economic moments. Finally, the sample paths of the implied yields in the quadratic model is qualitatively similar to the data.

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 both the SMM estimator and a variant of the efficient method of moments (EMM) are in Section 7, and Section 8 contains our conclusions.


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 (2000) (DS1), Dai and Singleton (2002a) (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

$$
\frac{dF_t}{F_t} = (K\theta - KF_t) \, dt + \Sigma \sqrt{S_t} \, dW_t,
$$

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

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

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

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

$$
r_t = \delta_0 + \delta_1' F_t.
$$

**Assumption A3:** The market price of factor risk function is

$$
\Lambda_t = S_t \lambda_1 + S_t^- \lambda_2 F_t,
$$

where $\lambda_1$ is an $N$-vector, $\lambda_2$ is an $N \times N$ matrix, and $S_t^-$ is a diagonal matrix with nonzero elements

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

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 $\lambda_2 = 0$, this formulation simplifies to the completely affine models of Duffie and Kan (1996) and 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.

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:

$$P_t(\tau) = \exp \left[ A(\tau) - B(\tau)' F_t \right],$$  \hspace{1cm} (5)

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

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

and

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

first specified in Duffie and Kan (1996), and where $K\theta^Q$ and $KQ$ 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 $KQ$, and their conditional volatilities are determined by the form of $S$. Under the physical measure, yield correlations are also affected by the dynamics of $\Lambda$. DS1 introduced the notation $A_m(N)$ ($m \leq N$) 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) (ADG) is defined by the following assumptions:

**Assumption Q1:** The factors are Gaussian; so

$$dF_t = (K\theta - KF_t) dt + \Sigma dW_t,$$  \hspace{1cm} (6)

where $K$ and $\Sigma$ are $N \times N$ matrices and $K\theta$ is an $N$-vector.
**Assumption Q2:** The short-rate is a quadratic function of the state

\[ r_t = \delta_0 + F_t' \Psi F_t, \]  

where \( \delta_0 \geq 0 \) and \( \Psi \) is an \( N \times N \) positive definite matrix

\[ \Psi = \begin{bmatrix} 1 & \Psi_{12} & \cdots & \Psi_{1N} \\ \Psi_{12} & 1 & \cdots & \Psi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{1N} & \Psi_{2N} & \cdots & 1 \end{bmatrix}. \]  

**Assumption Q3:** The market price of risk is

\[ \Lambda_t = \lambda_0 + \lambda_1 F_t, \]  

Zero-coupon bond prices in a QTSM are equal to

\[ P_t(\tau) = \exp \left[ A(\tau) + B(\tau)' F_t + F_t' C(\tau) F_t \right], \]  

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 three-months to ten-years.
Fact 2 The levels of Treasury yields, measured weekly or monthly at maturities from three-months to ten-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 is formalized by Litterman and Scheinkman (1991) who show that the first principal component in yield changes accounts for about 88 percent of the variation in yield changes, and the first three components account for 99 percent of the variation.1 After examining the sensitivity of yield changes to these components, Litterman and Scheinkman name these factors the “level,” “slope,” and “curvature.”

Fact 4 The slope coefficient, $\phi_\tau$, in the following linear regression

$$\left[ Y_{t+1} (\tau - 1) - Y_t (\tau) \right] = \phi_c + \phi_\tau \left[ (Y_t (\tau) - r_t) / (\tau - 1) \right] + \epsilon_{t+1} (\tau) \quad (11)$$

is negative and increasingly so for longer maturities, where $Y_t (\tau)$ is the continuously compounded 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 $\epsilon_{t+1} (\tau)$ 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).2

Fact 5 The term structure of the unconditional volatility of Treasury yield levels is downward sloping. The unconditional volatility of yield changes decreases from one- to six-month maturities, and then it increases up to two-years. It is constant for longer maturities.

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.

---


2 Fama and Bliss (1987) run related regressions of future yield levels on the difference between a maturity matched forward rate and the short rate. See also Cochrane and Piazzesi (2002) for a recent examination of bond return predictability.
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 our estimation approach further, we first briefly review the existing evidence.

4 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 $A_m(3)$, $m = \{0, 1, 2, 3\}$ by EMM (using the seminonparametric density (SNP) estimator from Gallant and Tauchen (1989) as auxiliary model). DS1 use weekly data on swap rates of maturities from six months to ten 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 conditional first moments, conditional second moments, and deviations from normality in the conditional distribution of the data. Since none of these moments for the flexible $A_1(3)$ and $A_2(3)$ models are significantly different from zero, we can indirectly infer that these models successfully match 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. This estimation approach is described in more detail in Section 5.2.2.

ADG use the zero-coupon yields constructed by McCulloch and Kwon (1993).
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 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.

Leippold and Wu (2002) also examine the performance of quadratic models. They examine the relative performance of an unrestricted two-factor quadratic model compared to models that constrain the correlation between the states, require a constant market price of risk, or that restricts the short rate to be a linear function of the states. As in the analysis below, the alternative models are estimated using economic moments that include the shape of the term structure, volatility of yields, and LPY regressions. The models are estimated using GMM applied to the analytical moments from an Euler discretization of the quadratic model, but they only use monthly data from January 1985 to December 1999. They also do not compare affine with quadratic models. Leippold and Wu (2002) conclude that an unrestricted two-factor quadratic model is capable of matching the basic characteristics of the first two moments of yields and returns.

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 completely affine models. 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 the 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 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.\footnote{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).} 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.

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, and Leippold and Wu (2002) do not examine affine 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). Consequently, the majority of the literature on affine and quadratic models uses either approximate MLE or moments-based 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 that we examine.

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 \( K \geq \text{dim} (\pi) \) functions of the data, \( \pi \) are the model parameters, and \( E_\pi \) 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 \( \pi_T \) 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, \( E_\pi \), are taken with respect to

---

\(^6\)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 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.
the physical measures characterized by the diffusion models in equations (1) and (6), 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. In particular, for a given parameter vector \( \pi \) 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 \( \{Y^\pi_m\}_{m=1}^M \) from the model and then approximate the theoretical moments \( G(\pi) \) with their Monte Carlo analogues

\[
G_M(\pi) = E_{\pi,M}[g(Y^\pi_m)] = M^{-1} \sum_{m=1}^M g(Y^\pi_m). \tag{7}
\]

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, \( \text{dim} (\pi) \), a \textit{weighting matrix}, denoted \( W \), is needed to determine the relative importance of each moment condition in constructing the moments-based estimates of \( \pi \).

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

\[
W^* = \text{cov} [G_T]^{-1}, \tag{12}
\]

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 \textit{robust} and \textit{economically transparent} weighting matrix. What is the economic meaning of the model failing to fit the \( W^* \) weighted set of moments if these weights depend on cross-covariances of the model that may lack clear economic intuition? In order to ensure that our conclusions are robust to this potential problem, we also use the weighting matrix

\[
W^a = \text{diag} \left( \text{cov} [G_T]^{-1} \right),
\]

which ensure that the moments are scaled appropriately.

Given these two building blocks, the SMM estimates, \( \pi_T \), are constructed as

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

\(^7\)We use 10,000 simulated observations, constructed using the method of antithetic variates.
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 \( \{Y_n^\pi\} \) 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 \( \mathcal{L}(\{Y_t\}_{t=1}^T ; \pi) \) also solve the corresponding first-order conditions:

\[
\frac{\partial \mathcal{L}}{\partial \pi} = \sum_{t=1}^T \frac{\partial \ln f(Y_t \mid Y_{t-1}, Y_{t-2}, \ldots ; \pi)}{\partial \pi} = 0,
\]

where \( f(Y_t \mid Y_{t-1}, Y_{t-2}, \ldots ; \pi) \) 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 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 or robust 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 ten-year and six-month bonds. Finally, the curvature

\(^8\)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.
factor is \[ Y_t(6-m) + Y_t(10-y) - 2Y_t(2-y) \], where \( Y_t(\tau) \) denotes the yield to maturity on a \( \tau \)-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 six-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 (11)) for maturities of six months, two years, and ten years (three moments); and (iv) the slope coefficients from the conditional volatility regressions (LPV regressions):

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

for the same three bonds (i.e., \( \tau = \) six months, two years, and ten years), where \( F_{it} \) denotes the level, slope, and curvature factors (nine moments) and

\[
E_t(Y_{t+1}(\tau - 1) - Y_t(\tau))
\]

is the fitted value from the LPY regressions.\(^9\) 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

---

\(^9\)This regression characterizes the conditional variance 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 (11).
are more robust than the (in theory) more efficient estimates.

### 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 *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 \( \varphi \) 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, a recent Monte Carlo study contained in 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, using 20 years of simulated weekly data, the EMM/SNP estimator 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.

\(^{10}\)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).
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 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?

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):

$$F_t = \begin{bmatrix} b_0 & B \end{bmatrix} \begin{bmatrix} 1 \\ F_{t-1} \end{bmatrix} + \varepsilon_t$$  \hspace{1cm} (14)

(12 parameters/moments).

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

$$h_{it} \equiv \text{var}_{t-1}(\varepsilon_{it}) = \exp(F_{t-1}'\gamma_i) v_{it},$$  \hspace{1cm} (15)

where

$$v_{it} = (1 - \phi_i - \psi_i) \omega_i + \phi_i \varepsilon_{it-1}^2 + \psi_i v_{it-1}$$

(six 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:

\[
\text{corr} (u_t) = \begin{bmatrix}
1 & \rho_{21} & 1 \\
\rho_{31} & \rho_{32} & 1 \\
\end{bmatrix}
\]

(three parameters/moments).

4. The conditional (on the VAR means and level-GARCH variances) distribution of the factors is multivariate Student’s t with \( \zeta \) degrees of freedom (one 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 Alternative Estimators

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.\(^\text{11}\)

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, 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

\(^{11}\text{In particular, they find that the (misspecified) QML estimator and the EMM/Kalman filter estimator have roughly the same bias (which is much smaller than the EMM/SNP estimator), but the QML estimator is more efficient.}\)
disadvantage to the QML estimates of an affine model. Second, in a quadratic model yields cannot be uniquely inverted to recover factors.

The approximate maximum likelihood estimator (AML) of Duffie, Pedersen, and Singleton (2002) is applicable to an affine model in which the state vector can be partitioned into a set of independent square root diffusions that affect the conditional volatilities in the model and the remaining state vector which has both a more general volatility structure and allows for correlation among this subset of states. The AML estimator works by writing the likelihood function as the product of two components. The exact conditional density for the independent square roots is the product of noncentral chi-squares, and the joint density for the correlated states is approximately normal, conditioned on the paths of the independent states. The problem with the Duffie et al (2002) AML estimator in the current context is that, as in the case of the QML estimator, it requires that a set of model yields be uniquely inverted to recover observable proxies for the states. While this is possible for affine specifications, as noted above, this cannot be done for quadratic models.

6 Data

A variety of data sets have been used in examining the alternative models. We use the same data source as Duffee (2002).\textsuperscript{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 six-months, one-, two-, five-, and ten-years.

Basic summary statistics, consistent with the stylized facts described in Section 3, are contained in Tables 1 through 3. According to the statistics in Panel A of Table 1, the unconditional mean yields increase monotonically with maturity from 5.77 at six-months to 6.75 at 10-years.\textsuperscript{13} In the second line of Panel A of Table 1, we see that the unconditional volatilities decrease monotonically with maturity, from 2.93 to 2.71. Panels B and C of Table 1 document that yield level correlations – both contemporaneously and at the first lag – 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 2, and they are negative, significant, and increasing (in absolute value) with maturity, as documented in Campbell and Shiller (1991) and discussed in DS2.

The regressions of yield change variances on lagged values of the level, slope, and

\textsuperscript{12}We would like to thank Greg Duffee for making his data available on his web page.

\textsuperscript{13}All yield moments are reported in percent per year.
curvature factors are summarized in Table 3, 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.\textsuperscript{14} The regression coefficients for the slope factor are negative and significant for the short end of the yield curve, but insignificant for the five- and ten-year maturities. Conditional volatility is significantly related to the curvature factors at maturities of up to five years.

The fitted parameters for the auxiliary model described in Section 5.2.2 are shown in Table 4. 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 seem 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.

7 Estimation Results

7.1 Affine Models

The results for the $A_0 (3)$ and $A_1 (3)$ models are contained in Tables 5 and 6.\textsuperscript{15} These models have twenty-two and twenty-three parameters, respectively, and they are estimated using both SMM and the twenty-seven economic moments and EMM and the 34 score-based moments.\textsuperscript{16} The problem of finding reasonable starting values for an estimation problem of this magnitude is nontrivial, as is the problem of verifying that the numerical optimization of the objective function did not converge to a local minimum.

In order to ensure that the results are robust to different starting values, we used the following algorithm:

\textsuperscript{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.

\textsuperscript{15}The performance of the $A_2 (3)$ model was significantly worse – on all dimensions – when compared to the $A_0 (3)$ and $A_1 (3)$ models. In the interest of brevity, these results are not presented in the paper, but they are available upon request.

\textsuperscript{16}Point estimates and standard errors of the parameter values of the different fitted models are contained in the tables in Appendix A.
**Step 1:** For each of the affine and quadratic models (and each of the weighting matrices), we generated 1,000 random sets of admissible starting values.

**Step 2:** For each of the starting values, we minimized the SMM objective function for a weak convergence tolerance using a Newton-Raphson algorithm.

**Step 3:** We selected the 25 parameter combinations with the smallest values of the objective function and optimized thoroughly using alternating rounds of a simplex algorithm and Newton-Raphson algorithm until a tight convergence tolerance was achieved.

**Step 4:** We chose the set of parameters that minimized the objective function across the parameterizations considered in Step 3.

The standard moment-based test for overall model performance is equal to:

\[
J_T \equiv [G_T - G_M(\pi)]' W_T [G_T - G_M(\pi)] \overset{a}{\sim} \chi^2_{K - \text{dim}(\pi)}
\]

where \( \overset{a}{\sim} \) denotes the asymptotic distribution of the statistic, \( K \) is the number of moment conditions, and \( K - \text{dim}(\pi) \) 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 \( \text{Ind}(J) \) of dimension \( k \), tests of the \( k \) parameter restrictions

\[
\eta_{kT}(\pi) = 0, \text{ for all } k \in \text{Ind}(k)
\]

(17)
can be implemented using a test statistic of the form

\[
\eta_{\text{Ind}(k)}' [C \Sigma_{\eta_T} C']^{-1} \eta_{\text{Ind}(k)} \overset{a}{\sim} \chi^2_k
\]

(18)

where \( C \) is a \( J \times K \) matrix of zeros and ones and \( \eta_{\text{Ind}(k)} \) refers to the specific set of moments being tested.

Using the economic moments, we examine joint tests of the following specific hypotheses: (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. For the score-based moments, the hypotheses are: (i) all of the VAR moment conditions are zero; (ii) all of the level effect coefficients in the volatility are zero; (iii) all of the GARCH moment conditions are zero; and (iv) all of the standardized residual correlations are zero.
The overall fit statistic of equation (16) for the \( A_0 \) (3) and \( A_1 \) (3) models are reported in Table 5. The Gaussian model, \( A_0 \) (3), is strongly rejected by the data, based on the standard asymptotic approximation to the finite-sample distribution of the test. The tests of subsets of moments indicate that the model completely fails to fit the conditional volatility of the yield factors. This is not a surprise, but the subset tests also yield \( p \)-values of 0.06 and 0.07 on the null hypotheses that the unconditional moments are zero. The \( A_1 \) (3) model fares much better than the \( A_0 \) (3) model, based on the overall test statistic, but it too is rejected, at a 5 percent confidence level, as providing an adequate fit to all of the economic moments. Interestingly, the tests on subsets of moments of the \( A_1 \) (3) model fail to reject that all the different groups of moments are zero (although the correlation moments have a \( p \)-value of 0.099). The overall rejection of the \( A_1 \) (3) model must be due to cross-correlations among the different moment condition groups.

Table 6 allows us to look in more detail at the performance of the affine models at matching each of the 27 economic moments. The table shows the actual moments from the data (along with standard errors) in the first column, and then it shows the \( A_0 \) (3) model fit using the two alternative weighting matrices, followed by the \( A_1 \) (3) fit using the alternative weighting matrices. The results in Table 6 yield the following insights: (i) All variants of the models do a good job at fitting the unconditional moments of the data. (ii) Both models, using both weighting matrices, are capable of matching the LPY regression moments that capture the dynamics of bond holding period premiums. In fact, and contrary to the results in Dai and Singleton (2002) and Duffee (2002), the model that allows for conditional volatility dynamics, \( A_1 \) (3), actually point estimates of the LPY moments that are closer to the actual moments than the Gaussian \( A_0 \) (3) model. (iii) The rejections of the LPV moment conditions for the \( A_0 \) (3) model occur primarily due to the level volatility, since this is the LPV moment that is measured most precisely. (iv) In general, the point estimates based on the diagonal weighting matrix are closer to the actual moments than the estimates based on the full optimal weighting matrix.

7.2 The Quadratic Model

Table 5 also contains the overall fit and subsets of moments tests for the canonical 3-factor quadratic model using the test statistics in equations (16) and (18). In stark contrast to the results in ADG, we find that the QTSM(3) model is not rejected by the data, using the overall test of model specification or any of the subgroups of moments that we examine. As in the case of the affine models, the overall test statistic has a much smaller \( p \)-value than any of the subgroup tests. The issue that we need to be concerned about is whether the
quadratic model looks good based on tests of the form of (16) and (18) because it has small moment differences or because it has very large standard errors associated with the different moment conditions.

This question can be resolved by examining the sample economic moments and their SMM estimates, as shown in Table 7. It seems clear that the model does an excellent job at matching the economic moments. Even a cursory examination of each group of unconditional moments shows that the three factor quadratic model matches the data – not only qualitatively – but quantitatively as well. The LPY regression coefficients – the focus of the DS2 and Duffee (2002) analyses – are matched qualitatively by the SMM estimator that uses the optimal weighting matrix, and the are matched even more closely by the diagonal weighting matrix. The strength of the QTSM(3) model is the ease with which it matches the conditional volatility moments. Where the moments are measured precisely, the model matches them precisely, and where the moments are measured with substantial error, the model is qualitatively correct. As a general observation, it seems that the estimator formed using the diagonal weighting matrix matches the point estimates more closely than the optimal matrix based estimator.\textsuperscript{17}

### 7.3 Reprojecting the Quadratic Model

In the previous subsections, we evaluated the performance of the alternative multifactor models based solely on how well they matched the moments incorporated into the various estimators. Additional insight into the implications of the different models can be developed by examining the model-implied yields, constructed at the estimated parameter values. This type of analysis is called “reprojection” in the EMM framework, and it is extended in a straightforward manner to our SMM estimator as well.

The reprojection of the model implied yields is constructed as follow:

**Step 1:** At the SMM (EMM) parameter estimates, simulate (from equation (6)) a long time series of the latent factors. Construct a matched time series of model implied yields (again, using the point estimates of the model parameters, simulated factors and the pricing equation (10)). In these long simulations, compute the functional relationship between yields and factors using a flexible functional form.\textsuperscript{18}

**Step 2:** Using the inverse of the computed relationship between yields and factors, construct estimates of the factor realizations from the actual data on yields.

\textsuperscript{17}Of course, such comparisons must remain heuristic, since formal tests of the differences in the moments would require that a common weighting matrix be used in constructing the estimates.

\textsuperscript{18}We use orthonormal polynomials in estimating the relationship between model yields and factors.
Step 3: Given the implied factor realizations from the actual data and the parameter estimates in the pricing functions, we compute the “reprojected” yields implied by the point estimates of the model parameters.

In the case of the essentially affine models, the process of reprojection is equivalent to inverting the yields for the factors, since the yields are linear functions of the factors. This simple inversion property guarantees that the implied factors from an affine model can match any three yields exactly, whereas there is no unique mapping for the implied quadratic factors. Whether or not this should be viewed as an inherent weakness in the form of the quadratic model, it clearly makes comparisons between affine and quadratic models on the basis of reprojected yields problematic. Therefore, we only report reprojected quadratic factors.

In Figure 2, we examine the reprojected short rate and the 1-, 3-, and 5-year yield levels in the QTSM(3) model based on the SMM point estimates. The sample paths of all of the reprojected series are qualitatively similar. They generally match the corresponding actual series quite well through the first half of the sample period. They also capture the run-up in yield levels associated with the first part of the period associated with the Federal Reserve experiment of the late-1970s and early-1980s. However, the fitted yields from the model have a difficult time matching the decline in rates in the early to mid-1980s. The implied yields decline too quickly and remain below the actual series for an extended period of time.

8 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 reject the most general forms of the $A_0 (3)$ and $A_1 (3)$
models. For the Gaussian model, the obvious problem is matching conditional volatility, and
while test statistics fail to reject the null hypothesis that the mixed $A_1 (3)$ model matches
the volatility moments, the point estimates from this model do not seem to capture the
qualitative features of the data.

The most flexible three-factor Gaussian-quadratic model is the clear winner in this horse
race. The overall test statistic, based on statistically optimal weighting of the moments, fails
to reject the null that the model can match all of the economic moments simultaneously,
and the tests based on groups of similar moment conditions also fail to reject the equality of
the model and the sample moments. Furthermore, the point estimates of the model-implied
moments conform quite closely to the levels and patterns found in the moments constructed
from the actual data.

In the end, the QTSM(3) model seems to provide a parsimonious way of summarizing
the dynamics of the yield moments and the bond holding period return moments that have
consistently proven to be of interest to financial economists. The next challenge is to use this
comparatively simple structure to guide the construction of structural models of Treasury
bond markets that simultaneously match the important features of the data and provide
insight into the effects of investor decisions and Federal Reserve policy choices.
References


Table 1: Yield Level Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>6-m</th>
<th>1-y</th>
<th>2-y</th>
<th>5-y</th>
<th>10-y</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Means and Standard Deviations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.7741</td>
<td>5.9751</td>
<td>6.2219</td>
<td>6.5493</td>
<td>6.7491</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>2.9266</td>
<td>2.9023</td>
<td>2.8457</td>
<td>2.7621</td>
<td>2.7134</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>6-m</th>
<th>1-y</th>
<th>2-y</th>
<th>5-y</th>
<th>10-y</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel B: Contemporaneous Yield Correlations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-m</td>
<td>1.000</td>
<td>0.996</td>
<td>0.982</td>
<td>0.946</td>
<td>0.914</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.072)</td>
<td>(0.067)</td>
<td>(0.064)</td>
<td></td>
</tr>
<tr>
<td>1-y</td>
<td>1.000</td>
<td>0.993</td>
<td>0.963</td>
<td>0.935</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.071)</td>
<td>(0.067)</td>
<td>(0.064)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-y</td>
<td>1.000</td>
<td>0.987</td>
<td>0.966</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.063)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-y</td>
<td>1.000</td>
<td>0.994</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>6-m</th>
<th>1-y</th>
<th>2-y</th>
<th>5-y</th>
<th>10-y</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel C: Lag-1 Yield Autocorrelations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-m</td>
<td>0.984</td>
<td>0.981</td>
<td>0.971</td>
<td>0.940</td>
<td>0.912</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.072)</td>
<td>(0.070)</td>
<td>(0.067)</td>
<td>(0.064)</td>
</tr>
<tr>
<td>1-y</td>
<td>0.985</td>
<td>0.981</td>
<td>0.956</td>
<td>0.931</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.071)</td>
<td>(0.069)</td>
<td>(0.066)</td>
<td>(0.063)</td>
<td></td>
</tr>
<tr>
<td>2-y</td>
<td>0.987</td>
<td>0.978</td>
<td>0.961</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.065)</td>
<td>(0.063)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-y</td>
<td>0.991</td>
<td>0.988</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.061)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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: LPY Slope Coefficients and $R^2$ Statistics

<table>
<thead>
<tr>
<th></th>
<th>6-m</th>
<th>1-y</th>
<th>2-y</th>
<th>5-y</th>
<th>10-y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>-0.776</td>
<td>-1.195</td>
<td>-1.678</td>
<td>-2.564</td>
<td>-3.832</td>
</tr>
<tr>
<td>$R^2$ (in %)</td>
<td>1.45</td>
<td>1.40</td>
<td>1.44</td>
<td>1.61</td>
<td>1.80</td>
</tr>
</tbody>
</table>

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 3: LPV Slope Coefficients and $R^2$ Statistics

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Level</th>
<th>Slope</th>
<th>Curvature</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-month</td>
<td>0.140</td>
<td>-0.164</td>
<td>0.198</td>
<td>15.65</td>
</tr>
<tr>
<td>(7.72)</td>
<td>(-3.63)</td>
<td>(1.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-year</td>
<td>0.119</td>
<td>-0.133</td>
<td>0.206</td>
<td>15.49</td>
</tr>
<tr>
<td>(7.85)</td>
<td>(-3.53)</td>
<td>(2.42)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>0.089</td>
<td>-0.092</td>
<td>0.151</td>
<td>13.23</td>
</tr>
<tr>
<td>(7.29)</td>
<td>(-3.03)</td>
<td>(2.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-year</td>
<td>0.049</td>
<td>-0.016</td>
<td>0.059</td>
<td>15.43</td>
</tr>
<tr>
<td>(9.05)</td>
<td>(-1.17)</td>
<td>(1.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-year</td>
<td>0.031</td>
<td>0.012</td>
<td>0.013</td>
<td>16.07</td>
</tr>
<tr>
<td>(9.90)</td>
<td>(1.53)</td>
<td>(0.75)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The slope coefficients are from ordinary least squares regressions of monthly squared yield change deviations from their expected value, $[Y_{t+1}(\tau - 1) - Y_t(\tau) - E_t(Y_{t+1}(\tau - 1) - Y_t(\tau))]^2$, on the level, slope, and curvature factors. $t$-statistics are calculated using Newey-West standard errors with 6 lags.
Table 4: Parameters of the Auxiliary Model for Yield Factors

\[
\begin{align*}
Y_t &= \begin{bmatrix}
0.0561 & 0.9888 & 0.0249 & -0.0658 \\
0.0017 & 0.0061 & 0.9479 & 0.0764 \\
0.089 & 1.552 & 75.98 & 2.788 \\
0.0128 & -0.0065 & 0.0333 & 0.8280 \\
0.5468 & -1.540 & 2.847 & 29.67
\end{bmatrix} + \begin{bmatrix} 1 \\ Y_{t-1} \end{bmatrix} + \varepsilon_t
\end{align*}
\]

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

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

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

\[
\begin{align*}
v_{1t} &= (1 - 0.0130 - 0.8634) 0.0235 (4.393) + 0.0130 0.8634 (15.82) \varepsilon^{2}_{1t-1} v_{1t-1}
\end{align*}
\]

\[
\begin{align*}
v_{2t} &= (1 - 0.0337 - 0.7584) 0.0219 (4.442) + 0.0337 0.7584 (10.32) \varepsilon^{2}_{2t-1} v_{2t-1}
\end{align*}
\]

\[
\begin{align*}
v_{3t} &= (1 - 0.1268 - 0.7263) 0.0515 (2.688) + 0.1268 0.7263 (8.230) \varepsilon^{2}_{3t-1} v_{3t-1}
\end{align*}
\]

\[
\text{corr}(\tilde{\varepsilon}_t) = \begin{bmatrix} 1 & -0.681 & -0.365 \\
-0.681 & 1 & 0.144 \\
-0.365 & 0.144 & 1
\end{bmatrix}
\]

\[
\zeta = 4.9533 \\
(7.7088)
\]

\(Y \equiv \text{(level, slope, curvature)'}\), as defined in Section 5.2. The model parameters are estimated using maximum likelihood. \(\tilde{\varepsilon}_{it} \equiv \varepsilon_{it}/\sqrt{h_{it}}\), the standardized VAR residual. \(\tilde{\varepsilon}_t\) is distributed multivariate Student’s t with \(\zeta\) degrees of freedom. t-statistics for the null hypothesis that the parameter equals 0 are reported in parentheses.
Table 5: Moment-Based Tests of Model Performance

Panel A: Economic Moments

<table>
<thead>
<tr>
<th>Model</th>
<th>Overall</th>
<th>Means &amp; Var.</th>
<th>Correlations</th>
<th>LPY</th>
<th>LPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_0 (3)$</td>
<td>45.539</td>
<td>11.493</td>
<td>16.237</td>
<td>2.326</td>
<td>24.057</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.074)</td>
<td>(0.062)</td>
<td>(0.508)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$A_1 (3)$</td>
<td>11.166</td>
<td>5.809</td>
<td>14.701</td>
<td>4.574</td>
<td>11.637</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.445)</td>
<td>(0.099)</td>
<td>(0.206)</td>
<td>(0.235)</td>
</tr>
<tr>
<td>QTSM(3)</td>
<td>4.418</td>
<td>1.412</td>
<td>1.821</td>
<td>2.589</td>
<td>6.341</td>
</tr>
<tr>
<td></td>
<td>(0.110)</td>
<td>(0.965)</td>
<td>(0.994)</td>
<td>(0.459)</td>
<td>(0.705)</td>
</tr>
</tbody>
</table>

Panel B: Score-Based Moments

<table>
<thead>
<tr>
<th>Model</th>
<th>Overall</th>
<th>VAR</th>
<th>Vol. Level</th>
<th>Vol. GARCH</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_0 (3)$</td>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
</tr>
<tr>
<td></td>
<td>(0.??)</td>
<td>(0.??)</td>
<td>(0.??)</td>
<td>(0.??)</td>
<td>(0.??)</td>
</tr>
<tr>
<td>$A_1 (3)$</td>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
</tr>
<tr>
<td></td>
<td>(0.??)</td>
<td>(0.??)</td>
<td>(0.??)</td>
<td>(0.??)</td>
<td>(0.??)</td>
</tr>
<tr>
<td>QTSM(3)</td>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
</tr>
<tr>
<td></td>
<td>(0.??)</td>
<td>(0.??)</td>
<td>(0.??)</td>
<td>(0.??)</td>
<td>(0.??)</td>
</tr>
</tbody>
</table>

The Overall test statistic is defined in equation (16), and it has an asymptotic $\chi^2$ distribution, with the number of degrees of freedom equal to the number of moment conditions (27) minus the number parameters in the model. In Panel A, the Means & Var, Correlations, LPY, and LPV statistics are all statistics of the form of equation (18) that test, jointly, whether the differences between sets of actual and model-generated moment conditions are equal to zero. In Panel B, the same tests are used to examine the score-based moments associated with the VAR, the “level effects” in volatility, the GARCH components of volatility, and the factor correlations.
Table 6: Actual Moments and SMM Estimates for the $A_0 (3)$ and $A_1 (3)$ Models

Panel A: Means, Volatility, and Autocorrelations.

<table>
<thead>
<tr>
<th></th>
<th>Sample Mom. (Std. Err)</th>
<th>$A_0 (3)$ Optimal $W_T$ Fitted Moment</th>
<th>Diagonal $W_T$ Fitted Moment</th>
<th>$A_1 (3)$ Optimal $W_T$ Fitted Moment</th>
<th>Diagonal $W_T$ Fitted Moment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Means</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>5.777 (0.323)</td>
<td>5.718 (0.182)</td>
<td>5.776 (0.001)</td>
<td>5.662 (0.355)</td>
<td>5.777 (0.001)</td>
</tr>
<tr>
<td>Slope</td>
<td>0.976 (0.125)</td>
<td>0.936 (0.317)</td>
<td>0.975 (0.006)</td>
<td>0.997 (0.166)</td>
<td>0.976 (0.002)</td>
</tr>
<tr>
<td>Curvature</td>
<td>0.079 (0.049)</td>
<td>0.068 (0.227)</td>
<td>0.080 (0.004)</td>
<td>0.094 (0.304)</td>
<td>0.080 (0.004)</td>
</tr>
<tr>
<td><strong>Volatility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>0.528 (0.072)</td>
<td><strong>0.387 (1.957)</strong></td>
<td>0.526 (0.022)</td>
<td>0.434 (1.301)</td>
<td>0.510 (0.238)</td>
</tr>
<tr>
<td>Slope</td>
<td>0.398 (0.047)</td>
<td>0.323 (1.570)</td>
<td>0.421 (0.490)</td>
<td>0.331 (1.395)</td>
<td>0.407 (0.194)</td>
</tr>
<tr>
<td>Curvature</td>
<td>0.310 (0.027)</td>
<td>0.3284 (0.993)</td>
<td>0.298 (0.445)</td>
<td>0.287 (0.860)</td>
<td>0.310 (0.015)</td>
</tr>
<tr>
<td><strong>Cross-Correlations:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level &amp; Slope</td>
<td>−0.376 (0.102)</td>
<td>−0.409 (0.315)</td>
<td>−0.368 (0.084)</td>
<td>−0.365 (0.107)</td>
<td>−0.378 (0.020)</td>
</tr>
<tr>
<td>Level &amp; Curv.</td>
<td>−0.352 (0.082)</td>
<td>−0.363 (0.139)</td>
<td>−0.340 (0.146)</td>
<td>−0.316 (0.440)</td>
<td>−0.343 (0.111)</td>
</tr>
<tr>
<td>Slope &amp; Curv.</td>
<td>0.378 (0.106)</td>
<td>0.381 (0.027)</td>
<td>0.384 (0.057)</td>
<td>0.394 (0.150)</td>
<td>0.400 (0.206)</td>
</tr>
<tr>
<td><strong>First-order Autocorrelations:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>0.984 (0.003)</td>
<td><strong>0.992 (2.441)</strong></td>
<td>0.983 (0.147)</td>
<td>0.988 (1.412)</td>
<td>0.984 (0.000)</td>
</tr>
<tr>
<td>Slope</td>
<td>0.942 (0.012)</td>
<td>0.963 (1.707)</td>
<td>0.942 (0.065)</td>
<td>0.961 (1.480)</td>
<td>0.945 (0.179)</td>
</tr>
<tr>
<td>Curvature</td>
<td>0.804 (0.042)</td>
<td>0.803 (0.504)</td>
<td>0.809 (0.121)</td>
<td>0.832 (0.684)</td>
<td>0.806 (0.045)</td>
</tr>
<tr>
<td><strong>First-order Cross-Autocorrelations:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level &amp; Slope</td>
<td>−0.341 (0.100)</td>
<td>−0.409 (0.680)</td>
<td>−0.370 (0.291)</td>
<td>−0.338 (0.030)</td>
<td>−0.337 (0.041)</td>
</tr>
<tr>
<td>Level &amp; Curv.</td>
<td>−0.328 (0.086)</td>
<td>−0.361 (0.382)</td>
<td>−0.340 (0.140)</td>
<td>−0.319 (0.101)</td>
<td>−0.339 (0.133)</td>
</tr>
<tr>
<td>Slope &amp; Curv.</td>
<td>0.349 (0.109)</td>
<td>0.320 (0.261)</td>
<td>0.329 (0.182)</td>
<td>0.417 (0.623)</td>
<td>0.317 (0.291)</td>
</tr>
</tbody>
</table>

The table description is on the next page.
Panel B: LPY and LPV Regression Slope Coefficients.

<table>
<thead>
<tr>
<th></th>
<th>$\hat{A}_0$ (3)</th>
<th>$\hat{A}_1$ (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample Moment</td>
<td>Optimal $W_T$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fitted Moment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Optimal $W_T$</td>
</tr>
<tr>
<td>LPY Slopes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-m</td>
<td>$-0.776 (0.503)$</td>
<td>$-0.490 (0.567)$</td>
</tr>
<tr>
<td>2-y</td>
<td>$-1.678 (0.969)$</td>
<td>$-0.892 (0.811)$</td>
</tr>
<tr>
<td>10-y</td>
<td>$-3.832 (1.750)$</td>
<td>$-3.802 (0.017)$</td>
</tr>
<tr>
<td>LPV Slopes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-m on Level</td>
<td>$0.140 (0.045)$</td>
<td><strong>0.000 (3.100)</strong></td>
</tr>
<tr>
<td>6-m on Slope</td>
<td>$-0.164 (0.129)$</td>
<td>$0.001 (1.275)$</td>
</tr>
<tr>
<td>6-m on Curv.</td>
<td>$0.198 (0.186)$</td>
<td>$-0.001 (1.068)$</td>
</tr>
<tr>
<td>6-m on Level</td>
<td>$0.089 (0.032)$</td>
<td><strong>0.000 (2.774)</strong></td>
</tr>
<tr>
<td>6-m on Slope</td>
<td>$-0.092 (0.091)$</td>
<td>$0.001 (1.024)$</td>
</tr>
<tr>
<td>6-m on Curv.</td>
<td>$0.151 (0.152)$</td>
<td>$-0.001 (1.000)$</td>
</tr>
<tr>
<td>10-y on Level</td>
<td>$0.031 (0.006)$</td>
<td><strong>0.003 (4.466)</strong></td>
</tr>
<tr>
<td>10-y on Slope</td>
<td>$0.012 (0.017)$</td>
<td>$0.002 (0.560)$</td>
</tr>
<tr>
<td>10-y on Curv.</td>
<td>$0.013 (0.030)$</td>
<td>$-0.004 (0.555)$</td>
</tr>
</tbody>
</table>

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 twice 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 $\text{corr}(F_{it}, F_{it-1})$ for $i = 1, 2, 3$, and the first-order cross-correlation is defined as $\text{corr}(F_{it}, F_{jt-1})$ for $i \neq j$. The LPY and LPV regressions are defined in the text in equations (11) and (13), respectively. $t$-statistics of the absolute parameter differences, based on the Newey-West estimator using 6 lags, are reported in parentheses.
## Table 7: Actual Moments and SMM Estimates for the QTSM(3) Model

Panel A: Means, Conditional Volatility, and Autocorrelations.

<table>
<thead>
<tr>
<th></th>
<th>Sample Moment</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Optimal $W_T$</td>
<td>Fitted Moment $t$-statistic</td>
<td>Diagonal $W_T$</td>
<td>Fitted Moment $t$-statistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>5.777</td>
<td>0.323</td>
<td>5.782</td>
<td>0.017</td>
<td>5.777</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>0.976</td>
<td>0.125</td>
<td>0.972</td>
<td>0.034</td>
<td>0.977</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curvature</td>
<td>0.079</td>
<td>0.049</td>
<td>0.079</td>
<td>0.006</td>
<td>0.078</td>
<td>0.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond. Volatility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>0.528</td>
<td>0.072</td>
<td>0.555</td>
<td>0.390</td>
<td>0.545</td>
<td>0.244</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>0.398</td>
<td>0.047</td>
<td>0.423</td>
<td>0.534</td>
<td>0.392</td>
<td>0.125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curvature</td>
<td>0.310</td>
<td>0.027</td>
<td>0.297</td>
<td>0.483</td>
<td>0.311</td>
<td>0.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-Correlations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level &amp; Slope</td>
<td>-0.376</td>
<td>-0.102</td>
<td>-0.377</td>
<td>0.002</td>
<td>-0.380</td>
<td>0.036</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level &amp; Curv.</td>
<td>-0.352</td>
<td>-0.082</td>
<td>-0.349</td>
<td>0.041</td>
<td>-0.359</td>
<td>0.080</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope &amp; Curv.</td>
<td>0.378</td>
<td>0.106</td>
<td>0.375</td>
<td>0.032</td>
<td>0.403</td>
<td>0.236</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-Order Autocorrelations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>0.984</td>
<td>0.003</td>
<td>0.982</td>
<td>0.588</td>
<td>0.983</td>
<td>0.118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>0.942</td>
<td>0.012</td>
<td>0.933</td>
<td>0.748</td>
<td>0.941</td>
<td>0.114</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curvature</td>
<td>0.804</td>
<td>0.042</td>
<td>0.814</td>
<td>0.249</td>
<td>0.806</td>
<td>0.048</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-Order Cross-Autocorrelations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level &amp; Slope</td>
<td>-0.341</td>
<td>-0.100</td>
<td>-0.328</td>
<td>0.126</td>
<td>-0.338</td>
<td>0.028</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level &amp; Curv.</td>
<td>-0.328</td>
<td>-0.086</td>
<td>-0.295</td>
<td>0.378</td>
<td>-0.318</td>
<td>0.119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope &amp; Curv.</td>
<td>0.349</td>
<td>0.109</td>
<td>0.316</td>
<td>0.302</td>
<td>0.339</td>
<td>0.090</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table description is on the next page.
Panel B: LPY and LPV Regression Slope Coefficients.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Optimal $W_T$</th>
<th>Diagonal $W_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fitted Moment</td>
<td>N-W Fitted Moment</td>
</tr>
<tr>
<td>LPY Slope Coeff.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-m</td>
<td>-0.776 0.503</td>
<td>-0.598 0.353</td>
</tr>
<tr>
<td>2-y</td>
<td>-1.678 0.969</td>
<td>-0.228 1.496</td>
</tr>
<tr>
<td>10-y</td>
<td>-3.832 1.750</td>
<td>-1.843 1.136</td>
</tr>
<tr>
<td>LPV Slope Coeff.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-m on Level</td>
<td>0.140 0.045</td>
<td>0.124 0.361</td>
</tr>
<tr>
<td>6-m on Slope</td>
<td>-0.164 0.129</td>
<td>-0.184 0.156</td>
</tr>
<tr>
<td>6-m on Curv.</td>
<td>0.198 0.186</td>
<td>0.245 0.257</td>
</tr>
<tr>
<td>2-y on Level</td>
<td>0.089 0.032</td>
<td>0.089 0.022</td>
</tr>
<tr>
<td>2-y on Slope</td>
<td>-0.092 0.091</td>
<td>-0.108 0.184</td>
</tr>
<tr>
<td>2-y on Curv.</td>
<td>0.151 0.152</td>
<td>0.097 0.355</td>
</tr>
<tr>
<td>10-y on Level</td>
<td>0.031 0.006</td>
<td>0.029 0.307</td>
</tr>
<tr>
<td>10-y on Slope</td>
<td>0.012 0.017</td>
<td>0.017 0.325</td>
</tr>
<tr>
<td>10-y on Curv.</td>
<td>0.013 0.030</td>
<td>-0.007 0.674</td>
</tr>
</tbody>
</table>

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 (11) and (13), respectively. The absolute value of the $t$-statistic, computed using the Newey-West standard error estimate constructed using 6 lags, is reported in the column N-W $t$-statistic.
\[ \begin{array}{cc}
\kappa & \begin{bmatrix}
\kappa_{11} & 0 & 0 \\
\kappa_{21} & \kappa_{22} & 0 \\
\kappa_{31} & \kappa_{32} & \kappa_{33}
\end{bmatrix} \\
\beta & \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
\end{array} \]

\[ \begin{array}{c}
\alpha' = [1, 1, 1] \\
\Sigma & \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} \\
[\delta_0, \delta]' & \begin{bmatrix}
\delta_0 & \delta_1 & \delta_2 & \delta_3
\end{bmatrix} \\
\theta' & \begin{bmatrix}
0 & 0 & 0 \\
\theta_1 & 0 & 0
\end{bmatrix} \\
\lambda_1' & \begin{bmatrix}
\lambda_{11} & \lambda_{12} & \lambda_{13}
\end{bmatrix} \\
\lambda_2 & \begin{bmatrix}
\lambda_{2(11)} & \lambda_{2(12)} & \lambda_{2(13)} \\
\lambda_{2(21)} & \lambda_{2(22)} & \lambda_{2(23)} \\
\lambda_{2(31)} & \lambda_{2(32)} & \lambda_{2(33)}
\end{bmatrix}
\end{array} \]

\[ \begin{array}{cc}
\dim(\pi) & 22 \\
& 23
\end{array} \]

Figure 1: Parameter restrictions in the \( A_0 (3) \) and \( A_1 (3) \) models that are required for identification and admissibility.
Figure 2: The Reprojected Yields in Quadratic Model Using the SMM Estimates.