Simulated Likelihood Estimation of Affine Term Structure Models from Panel Data*

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

Ping He
Department of Economics
University of Pennsylvania‡

First Draft: December 2000
This Draft: July 2002

Abstract

We show how to estimate affine term structure models from a panel of noisy bond yields using simulated maximum likelihood based on importance sampling. We approximate the likelihood function of the state-space representation of the model by correcting the likelihood function of a Gaussian first-order approximation for the non-normalities introduced by the affine factor dynamics. Depending on the accuracy of the correction, which is computed through simulations, the quality of the estimator ranges from quasi-maximum likelihood (no correction) to exact maximum likelihood as the simulation size grows.

*We thank Frank Diebold, Qiang Kang, and Ken Singleton for their comments. Financial support from the Rodney L. White Center for Financial Research at the Wharton School and the Department of Economics is gratefully acknowledged.

†Philadelphia, PA 19104-6367. Phone: (215) 898-3609. E-mail: brandtm@wharton.upenn.edu.

‡Philadelphia, PA 19104-6297. Phone: (215) 898-0658. E-mail: heping@ssc.upenn.edu.

1 Introduction

The term structure of interest rates describes the relationship between the yield of a bond and its maturity. Given the high correlation among bond yields of different maturities, most models rely on a small number of factors to explain the time-series variation of the term structure. Starting with the classic Vasicek (1977) model, the term structure is typically characterized by the current realization and the dynamics of a set of exogenous state variables. Furthermore, most empirical implementations of these factor models have focused on the special cases of *affine* term structure models (ATSMs) of Duffie and Kan (1996) and Dai and Singleton (2000).[1] ATSMs accommodate time-varying moments of the state variables through affine specifications of the risk-neutral drift and variance functions. At the same time, they deliver essentially closed-form expressions for bond prices.

The empirical research on particular specifications of ATSMs is extensive but, despite the popularity of ATSMs, the tools available for estimation are limited. Existing econometric methods can be roughly categorized as follows:

- *Bond returns.* Given the closed-form expressions for bond prices, we can derive a set of cross-sectional moments that involve only observed returns of bonds with different maturities and that can be used for GMM estimation (Brown and Dybvig, 1986; Gibbons and Ramaswamy, 1993). However, much information in the data is potentially lost by estimating the model from noisy unconditional moments of bond returns as opposed to estimating the model from more precise conditional bond prices.

- *State variable proxies.* If we assume that the state variables are observable or can be proxied for by observable variables, estimation of the model with ML, QML, or different forms of GMM is straightforward. A common example is to assume that the one-month interest rate is a proxy for the instantaneous short rate (the state variable) (Marsh and Rosenfeld, 1983; Chan, Karolyi, Longstaff, and Sanders, 1992; Nowman, 1997). But, as Chapman, Long, and Pearson (1999) illustrate, differences between the proxies and state variables can lead to economically significant biases in the inferences.

- *Implicit state variables.* Given the closed-form expressions for bond prices, we can invert any n bond prices, which are assumed to be error-free, into the n state variables and then use these implicit state variables in the estimation as if they were directly

[1] ATSMs include the factor models of Vasicek (1977), Langetieg (1980), Cox, Ingersoll, and Ross (1985), Jamshidian (1989, 1991, 1992), Heston (1991), Chen and Scott (1992), Longstaff and Schwartz (1992), and Chen (1996). Prominent exceptions are the quadratic term structure models of Beaglehole and Tenney (1992), Constantinides (1992), and Ahn, Dittmar, and Gallant (2001).

observable (Pearson and Sun, 1994; Chen and Scott, 1993; Fisher and Gilles, 1996; Dai and Singleton, 2000). The problem is that, in practice, different choices of the reference bonds imply different state variable realizations, which is of course inconsistent with the model. To not immediately reject the common factor structure of the model, we must therefore assume that all bonds except for the n bonds used in the inversion are priced or measured with error. This dichotomy in the error assumptions for the reference bonds versus all other bonds is obviously difficult to rationalize.

- *Panel of noisy bond yields.* If we accept that all bonds are priced or measured with error, we can rewrite the ATSMs in state-space form:

$$\begin{aligned} &\text{measurement equation} && y_t = A + Bz_t + \varepsilon_t \\ &\text{transition equation} && z_t = \kappa(\theta - z_t)dt + \Sigma\sqrt{S_t}dW_t, \end{aligned} \tag{1}$$

  where S_t is a diagonal matrix with $\{S_t\}_{ii} = \alpha_i + \beta_i' z_t$, y_t is a vector of the bond yields at time t with different maturities, z_t is a vector of state variables, ε_t are normally distributed errors, and the remaining quantities are model parameters. If we then discretize the transition equation and *pretend* that the innovations to the discretized transition equation are normally distributed, we can use the Kalman filter to estimate the model (Chen and Scott, 1995; Jegadeesh and Pennacchi, 1996; Babbs and Nowman, 1999; Duan and Simonato, 1999; De Jong, 2000; Duffee and Stanton, 2001). Unfortunately, the innovations to the discretized transition equation are only normally distributed if $\beta_i = 0$ for all i, which corresponds to a multivariate Vasicek (1977) model. Otherwise, for the majority of ATSMs, we are dealing with a non-Gaussian state-space model, which means that the Kalman filter-based estimates are only QML estimates that are potentially inconsistent (see Lund, 1997) and certainly inefficient.

We propose a simulated maximum likelihood (SML) method for estimating affine term structure models from noisy panel data that overcomes the problems of the Kalman filter-based QML estimates. In particular, we use the importance-sampling approach of Durbin and Koopman (1997, 2001a) to correct the likelihood function of the QML estimator for the non-normalities introduced by the affine factor dynamics. Depending on the accuracy of the correction, which is determined by the number of simulations, the quality of the estimator ranges from QML (no correction) to exact ML.

Despite the obvious appeal of the panel data approach, there has been relatively little work on overcoming the problems with QML. Lund (1997) and Baadsgaard, Nielsen, and Madsen (2000) use a non-linear filter, instead of the Kalman filter, to obtain consistent

estimates of the level of the latent factors, but still end up with inconsistent QML estimates of the model parameters (because they assume that the filtering errors are Gaussian). Our SML estimator is most closely related to the Bayesian Markov chain Monte Carlo (MCMC) approach of Frühwirth-Schnatter and Geyer (1998). They use the QML likelihood as a proposal density to obtain draws from the posterior density of the model parameters, while we use the QML likelihood as an importance-sampling density to evaluate the likelihood function of the model (which we then maximize to obtain the SML parameter estimates). Given sufficiently flat priors, the two methods should lead to similar inferences.

The paper proceeds as follows: Section 2 explains the SML method and relates it to the more standard QML approach; Section 3 demonstrates the finite sample properties of our estimator through Monte Carlo experiments; Section 4 offers an empirical application for one- and two-factor Cox, Ingersoll, and Ross (CIR,1985) models; and Section 5 concludes.

2 Simulated Likelihood Estimation

2.1 Affine Term Structure Models

We first provide a brief review of affine term structure models (Duffie and Kan, 1996; Dai and Singleton, 2000). Absent arbitrage opportunities, the time t price of a zero-coupon bond that matures at time $t+\tau$, denoted P_t^τ, is given by:

$$P_t^\tau = \mathrm{E}_t^Q\Big[\exp\Big\{-\int_t^{t+\tau} r_s\, ds\Big\}\Big], \tag{2}$$

where $\mathrm{E}_t^Q[\cdot\,]$ denotes a conditional expectation under the risk-neutral measure Q. An n-factor affine term structure model is obtained under the assumption that the instantaneous short rate r_t is an affine function of a vector of unobserved state variables $z_t = [z_{1t}, z_{2t}, ..., z_{nt}]$:

$$r_t = \delta_0 + \sum_{i=1}^{n} \delta_i z_{it} = \delta_0 + \delta_z z_t \tag{3}$$

and that z_t follows an *affine diffusion*:

$$dz_t = \kappa^Q(\theta^Q - z_t)dt + \Sigma\sqrt{S_t}dW_t^Q, \tag{4}$$

where W_t^Q is an n-dimensional independent standard Brownian motion under the risk-neutral measure, θ^Q is an n-vector, κ^Q and Σ are $n \times n$ matrices, which may be non-diagonal and

asymmetric, and S_t is an $n \times n$ diagonal matrix with the ith diagonal element:

$$\{S_t\}_{ii} = \alpha_i + \beta_i' z_t. \tag{5}$$

The key feature of affine diffusions is that both the drift and the conditional variance are affine functions of the state variables z_t.

Provided a parameterization is admissible, we know from Duffie and Kan (1996) that:

$$P_t^\tau = \exp\left\{A(\tau) + B(\tau)' z_t\right\} \tag{6}$$

or, denoting the yield of the bond by y_t^τ, that:

$$y_t^\tau = -\frac{1}{\tau} \ln P_t^\tau = -\frac{1}{\tau}\big[A(\tau) + B(\tau)' z_t\big], \tag{7}$$

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

$$\begin{aligned} \frac{dA(\tau)}{d\tau} &= \theta^{\mathcal{Q}\prime} \kappa^{\mathcal{Q}\prime} B(\tau) + \frac{1}{2} \sum_{i=1}^{n} \big[\Sigma' B(\tau)\big]_i^2 \alpha_i - \delta_0 \\ \frac{dB(\tau)}{d\tau} &= -\kappa^{\mathcal{Q}\prime} B(\tau) + \frac{1}{2} \sum_{i=1}^{n} \big[\Sigma' B(\tau)\big]_i^2 \beta_i - \delta_z. \end{aligned} \tag{8}$$

These ODEs, which can be solved easily through numerical integration starting from the initial conditions $A(0) = 0$ and $B(0) = 0_{n\times 1}$, are completely determined by the specification of the risk-neutral dynamics of the state variables z_t.

In order to empirically implement ATSMs, we also need to know the dynamics of z_t under the actual measure P. For this we need to make an assumption about the market price of risks Λ_t, such as:

$$\Lambda_t = \sqrt{S_t}\lambda, \tag{9}$$

where λ is an n-vector of constants. Under this particular assumption about the market prices of risk, the process for z_t under the actual measure also has an affine form:

$$dz_t = \kappa(\theta - z_t)dt + \Sigma\sqrt{S_t}dW_t, \tag{10}$$

where W_t is an n-dimensional vector of independent standard Brownian motions under the

actual measure and:

$$\begin{aligned}\kappa &= \kappa^Q - \Sigma\Phi \\ \theta &= \kappa^{-1}(\kappa^Q\theta^Q + \Sigma\psi).\end{aligned} \tag{11}$$

The ith row of the $n \times n$ matrix Φ is $\lambda_i\beta_i'$ and ψ is an n-vector whose ith element is $\lambda_i\alpha_i$.

Any sensible parameterization of ATSMs must be both theoretically admissible and econometrically identified. Dai and Singleton (2000) discuss the parameter restrictions that these two requirements place on the general class of models described above. They propose a classification scheme that assigns any admissible ATSM to a unique *canonical* representation with identical econometric implications for the instantaneous short rate and hence for bond prices. Throughout this paper, we assume that we are dealing with such canonical ATSMs, so that the admissibility and identification requirements are automatically satisfied.

2.2 State-Space Representation

The fact that in ATSMs the bond yields are linear functions of the state variables implies that the covariance matrix of k yields for maturities $\{\tau_1, \tau_2, \ldots, \tau_k\}$ is only of rank n. For reasonable values of n (e.g., $n \leq 3$), this implication is decisively rejected by the data. To reconcile the data with low-dimensional ATSMs, we therefore assume that *each* bond is priced or measured with a random error. In particular, we assume normally distributed additive errors in yields (or log-normally distributed multiplicative errors in prices) that may be correlated across bonds but are iid through time.[2] Formally, we write:

$$y_t = A + Bz_t + \varepsilon_t, \quad \text{with } \varepsilon_t \overset{\text{iid}}{\sim} \text{MVN}[0, H_{t-1}], \tag{12}$$

where $y_t = [y_t^{\tau_1}, y_t^{\tau_2}, \ldots, y_t^{\tau_k}]'$, $A = [A(\tau_1), A(\tau_2), \ldots, A(\tau_k)]'$, $B = [B(\tau_1), B(\tau_2), \ldots, B(\tau_k)]'$, $\varepsilon_t = [\varepsilon_t^{\tau_1}, \varepsilon_t^{\tau_2}, \ldots, \varepsilon_t^{\tau_k}]'$, and H_{t-1} is the potentially time-varying covariance matrix of the errors.

ATSMs are formulated in continuous time but we only observe data at discrete points in time. To facilitate econometric inferences based on discrete time data, we work with the discrete time (e.g., daily or weekly) factor dynamics implied by the continuous time process (10). We denote this discrete time process:

$$z_t = m(z_{t-1}) + v(z_{t-1})^{1/2}\,\eta_t, \tag{13}$$

[2]Although the assumption of additive errors in yields is prominent in the literature, Lund (1997) argues that in certain settings the assumption of additive errors in prices is more appropriate.

where $m(z_{t-1}) = \mathrm{E}[z_t|z_{t-1}]$, $v(z_{t-1}) = \mathrm{Var}[z_t|z_{t-1}]$, and the standardized innovations η_t are iid with zero mean and unit variance, but are generally *not* normally distributed.[3] Due to the affine form of the continuous time process, the discrete time conditional mean and variance functions are both linear in z_{t-1}, so that $m(z_{t-1}) = m_0 + m_1{}'z_{t-1}$ and $v(z_{t-1}) = v_0 + v_1{}'z_{t-1}$, with coefficients that are expressed analytically in the Appendix.

Equations (12) and (13) make up a state-space model. In the terminology of state-space models, equation (12) is the measurement or observation equation and equation (13) is the transition or state equation. To draw inferences about the parameters of the model and the realizations of the latent factors from the observed yields, we need to solve a set of filtering and smoothing problems. Filtering generates the one-step-ahead forecasts of the factors $\widehat{z}_{t|t-1} \equiv \mathrm{E}[z_t|y_1, \ldots, y_{t-1}]$ and the corresponding forecast covariance matrix $\Omega_{t|t-1} \equiv \mathrm{Var}[\widehat{z}_{t|t-1}|y_1, \ldots, y_{t-1}]$, which in a linear Gaussian state-space model are used to construct the likelihood function. Smoothing yields the full-information forecasts of the factors $\widehat{z}_{t|T} \equiv \mathrm{E}[z_t|y_1, \ldots, y_T]$ and the corresponding forecast covariance matrix $\Omega_{t|T} \equiv \mathrm{Var}[\widehat{z}_{t|t-1}|y_1, \ldots, y_T]$. Unfortunately, the filtering and smoothing problems for ATSMs are non-standard because the state equation innovations η_t are generally not normally distributed. This means that the standard Kalman filter, designed for linear Gaussian state-space models, cannot be used directly to construct the likelihood function of ATSMs.[4]

2.3 Quasi-Maximum Likelihood Estimation

Consider a multifactor Vasicek (1977) model with $\beta = 0$, and hence with normally distributed discrete time factor innovations η_t, which is the only ATSM that corresponds to a linear Gaussian state-space model. For this special case, the assumptions of the Kalman filter are satisfied and ML estimation is straightforward. Intuitively, the likelihood function should be based on the normally distributed yield errors $[y_t - A - Bz_t] \sim \mathrm{MVN}[0, H_{t-1}]$, but these errors are not measurable at time t because they depend on the latent factors z_t. To make the errors measurable, we replace z_t with the filtered estimates $\widehat{z}_{t|t-1}$ from the Kalman filter, which, under the assumptions of the Kalman filter, are distributed $\mathrm{MVN}[z_t, \Omega_{t|t-1}]$. Since the yield and filtering errors are independent, the distribution of the measurable errors $[y_t - A - B\widehat{z}_{t|t-1}]$

[3] For example, in a one-factor CIR model, the innovations η_t have a non-central χ^2 distribution.

[4] The Kalman filter is a recursive algorithm for generating conditional (on the data) forecasts of the latent state vector in linear Gaussian state-space models. See Harvey (1989) or Hamilton (1994) for details.

is MVN$[0, H_{t-1} + B\Omega_{t|t-1}B']$. We use this distribution to form the likelihood function:

$$\mathcal{L}(\psi) = \prod_{t=1}^{T}(2\pi)^{k/2}\left|H_{t-1} + B\Omega_{t|t-1}B'\right|^{-1/2}\times \\ \exp\Big\{-\frac{1}{2}\big(y_t - A - B\widehat{z}_{t|t-1}\big)'\big(H_{t-1} + B\Omega_{t|t-1}B'\big)^{-1}\big(y_t - A - B\widehat{z}_{t|t-1}\big)\Big\}, \tag{14}$$

where $\widehat{z}_{1|0}$ and $\Omega_{1|0}$ correspond to the unconditional estimates of the initial z_1. Finally, ML estimates of the model parameters are obtained by maximizing $\mathcal{L}(\psi)$ with respect to ψ.

For general ATSMs with $\beta \neq 0$, and hence with innovations η_t that are not normally distributed, Kalman filter-based ML estimation breaks down for two reasons. First, the Kalman filter estimates of z_t do not correspond to the conditional expectations of z_t given the observed yields because the Kalman filter relies on a *linear* projection of z_t onto the *linear sub-space* of yields (which happens to coincide with the conditional expectations under multivariate normality). Therefore, we cannot use the filtered estimates to evaluate the likelihood function. Second, in a non-Gaussian model, the filtering errors, the differences between z_t and the linear projections, are not normally distributed. This means that the form of the likelihood function evaluated at the filtered estimates is not Gaussian.

Nonetheless, to the extent that the discrete time factor dynamics are almost Gaussian, perhaps because the data is sampled at high frequency, we might expect the Kalman filter-based ML estimates to be close to the true ML estimates (based on the conditional expectations of z_t and the corresponding non-Gaussian likelihood function). This naturally leads to the idea of QML estimation. Beyond the intuitive appeal of QML, there exists an extensive theory for misspecified inference problems. In fact, in some settings QML estimators can be shown to be consistent (White, 1982; Bollerslev and Wooldridge, 1992), but unfortunately, ATSMs do not satisfy the conditions required for consistency (Lund, 1997). Furthermore, the Monte Carlo evidence is mixed. For highly persistent single-factor models (which typically have nearly Gaussian factor dynamics even at the monthly frequency), QML performs reasonably well (Duan and Simonato, 1999); but for multifactor models it exhibits significant biases (Frühwirth-Schnatter and Geyer, 1998; Duffee and Stanton, 2001).

2.4 Simulated Maximum Likelihood Estimation

We therefore turn to SML estimation. Specifically, we show how to numerically construct the likelihood function of the non-Gaussian state-space model of ATSMs using an importance-sampling approach similar to the one used by Durbin and Koopman (1997), Sandmann and Koopman (1998), and Brandt and Kang (2001) in different contexts. The basic idea of the

estimator is to correct the likelihood function of the QML estimator for the non-normalities introduced by the affine factor dynamics.

2.4.1 Correcting the QML Likelihood Function

Let y and z denote the $T \times k$ and $T \times n$ matrices $[y_1', y_2', \ldots, y_T']'$ and $[z_1', z_2', \ldots, z_T']'$, respectively. The exact likelihood function of the model is:

$$\begin{aligned} \mathcal{L}(\psi) &= p(y|\psi) \\ &= \int p(y, z|\psi)\, dz \\ &= \int p(y|z, \psi)\, p(z|\psi)\, dz. \end{aligned} \tag{15}$$

The first line of equation (15) defines the likelihood function as the density of the observed bond yields y given the parameter ψ. The second line expresses this density as the marginal density of the yields obtained from the joint density of y and the latent factors z. Finally, in the third line we write the joint density of y and z as the product of the conditional density of y given z and the marginal density of z.

The point of this algebra is that the likelihood function $\mathcal{L}(\psi)$ can be viewed as an expectation of the conditional density $p(y|z, \psi)$ with respect to the marginal density $p(z|\psi)$. With T data points and n latent factors z, this expectation involves an $n \times T$ dimensional integral. The high dimensionality of the integral and the nonlinearities of the conditional density $p(y|z, \psi)$ make it practically impossible to compute this expectation analytically. We therefore rely on an importance-sampling scheme, based on the misspecified QML likelihood, to evaluate the exact likelihood function numerically.[5]

The SML estimator is based on the relationship between the exact likelihood function in equation (15) and the likelihood function of the QML estimator described in Section 2.3.

[5] In principle, the latent factors z can be sampled directly from the density $p(z|\psi)$ and Monte Carlo integration can then be used to solve the integral in the third line of equation (15). In practice, however, this brute-force simulation approach is grossly inefficient. We can greatly improve the efficiency of the Monte Carlo integration through the use of importance sampling.

The QML likelihood function can be written as:

$$\begin{aligned}\mathcal{L}_{QML}(\psi) &= p_{QML}(y|\psi) \\ &= \frac{p_{QML}(y,z|\psi)}{p_{QML}(z|y,\psi)} \\ &= \frac{p_{QML}(y|z,\psi)\, p_{QML}(z|\psi)}{p_{QML}(z|y,\psi)} \\ &= \frac{p(y|z,\psi)\, p_{QML}(z|\psi)}{p_{QML}(z|y,\psi)}.\end{aligned} \tag{16}$$

The first line of equation (16) defines the likelihood function, the second line follows from the definition of a conditional probability $P(A|B) = P(A,B)/P(B)$, and in the third line we again write the joint density of y and z as the product of the conditional density of y given z and the marginal density of z. The most important step is replacing the QML conditional density $p_{QML}(y|z,\psi)$ in the third line with the true conditional density $p(y|z,\psi)$ in the fourth line. This replacement is possible because the true model and the approximate QML model are based on the *same* linear measurement equation (12), which means that for the same z and ψ, the conditional densities of y are also the same.

Solving the last line of equation (16) for:

$$p(y|z,\psi) = \mathcal{L}_{QML}(\psi)\, \frac{p_{QML}(z|y,\psi)}{p_{QML}(z|\psi)} \tag{17}$$

and substituting it into equation (15) yields:

$$\begin{aligned}\mathcal{L}(\psi) &= \int p(y|z,\psi)\, p(z|\psi)\, dz \\ &= \mathcal{L}_{QML}(\psi) \int \frac{p(z|\psi)}{p_{QML}(z|\psi)}\, p_{QML}(z|y,\psi)\, dz \\ &= \mathcal{L}_{QML}(\psi)\, \mathrm{E}_{QML}\left[\left.\frac{p(z|\psi)}{p_{QML}(z|\psi)}\right| y,\psi\right].\end{aligned} \tag{18}$$

In summary, the true likelihood function $\mathcal{L}(\psi)$ can be expressed as the product of the QML likelihood function $\mathcal{L}_{QML}(\psi)$ and a correction factor $\mathrm{E}_{QML}[w(z)|y,\psi]$, where for notational convenience we define:

$$w(z) = \frac{p(z|\psi)}{p_{QML}(z|\psi)}. \tag{19}$$

The correction factor characterizes the departure of the true likelihood from the QML

likelihood as the average distance between the marginal densities $p(z|\psi)$ and $p_{QML}(z|\psi)$. If the two densities are close to each other, the function $w(z)$ is close to one and so is its expectation with respect to the QML inferences about the latent factors $p_{QML}(z|y,\psi)$.

To evaluate the true likelihood function $\mathcal{L}(\psi)$, we need to evaluate the QML likelihood function $\mathcal{L}_{QML}(\psi)$, which is straightforward, and the corresponding likelihood correction, given by the expectation of $w(z)$ with respect to the smoothed QML inferences about the latent factors $p_{QML}(z|y,\psi)$. However, even if we can evaluate $w(z)$ for a given factor history z (an issue we discuss below), its expectation cannot generally be computed analytically. We therefore use simulations to evaluate the correction factor. Specifically, we use Durbin and Koopman's (2001b) simulation smoother algorithm to simulate L factor histories $\{z^l\}_{l=1}^L$ from the conditional density $p_{QML}(z|y,\psi)$. We then evaluate $w(z^l)$ for each z^l and average across all L simulations to obtain the following *estimate* of the correction factor $\mathrm{E}_{QML}[w(z)|y,\psi]$:

$$\bar{w} = \frac{1}{L}\sum_{l=1}^{L} w(z^l). \tag{20}$$

Finally, we use equation (18) to construct an estimate of the likelihood function:

$$\widehat{\mathcal{L}}(\psi) = \mathcal{L}_{QML}(\psi)\,\bar{w}. \tag{21}$$

Since $\bar{w} \to \mathrm{E}_{QML}[w(z)|y,\psi]$ as $L \to \infty$, it follows that $\widehat{\mathcal{L}}(\psi)$ converges to $\mathcal{L}(\psi)$. Furthermore, as long as the convergence of the likelihood function is sufficiently smooth, the SML estimator $\widehat{\psi}_{SML}$ that maximizes $\widehat{\mathcal{L}}(\psi)$ converges to the ML estimator $\widehat{\psi}_{ML}$ that maximizes $\mathcal{L}(\psi)$.

However, it is typically more convenient to work with the *log* likelihood function. Taking logs of equation (21) gives:

$$\ln\widehat{\mathcal{L}}(\psi) = \ln\mathcal{L}_{QML}(\psi) + \ln\overline{w}. \tag{22}$$

But, $\ln\widehat{\mathcal{L}}(\psi)$ is a slightly biased estimator of $\ln\mathcal{L}(\psi)$ because $\mathrm{E}[\ln\bar{w}] \neq \ln\mathrm{E}[w]$. Durbin and Koopman (1997) and Shephard and Pitt (1997) suggest adding a term to the log likelihood function to correct the bias (up to order $O(L^{-3/2})$) from the log transformation:

$$\ln\widehat{\mathcal{L}}(\psi) = \ln\mathcal{L}_{QML}(\psi) + \ln\bar{w} + \frac{\eta_w^2}{2L\bar{w}^2}, \tag{23}$$

where:

$$\eta_w^2 = \frac{1}{L-1}\sum_{l=1}^{L}(w^l - \bar{w})^2. \tag{24}$$

2.4.2 Evaluating the Ratio of Marginal Densities

A critical step in SML estimation is to evaluate the ratio of densities $w(z) = p(z|\psi)/p_{QML}(z|\psi)$ for a given factor history z. Since the factor dynamics under both the true and QML models are first-order Markov, the numerator and denominator of $w(z)$ can be expressed as products of one-step-ahead transition densities, $p(z|\psi) = p(z_0|\psi)\prod_{t=1}^{T} p(z_t, t|z_{t-1}, t-1, \psi)$ and $p_{QML}(z|\psi) = p_{QML}(z_0|\psi)\prod_{t=1}^{T} p_{QML}(z_t, t|z_{t-1}, t-1, \psi)$, respectively. For the QML model, the unconditional distribution of z_0 and the one-step-ahead transition densities are all Gaussian by assumption. Evaluating $p_{QML}(z|\psi)$, the denominator of $w(z)$, is therefore straightforward. However, for the true model, these densities are not only generally non-Gaussian, which is the source of the problems with QML, but are also often analytically unknown. This makes evaluating $p(z|\psi)$, the numerator of $w(z)$, more problematic.

To focus the discussion, it is useful to differentiate between three cases: (i) multifactor Vasicek models with $\alpha \neq 0$ but $\beta = 0$, (ii) multifactor CIR models with $\alpha = 0$ but $\beta \neq 0$, and (iii) general ATSMs with $\alpha \neq 0$ and $\beta \neq 0$. For Vasicek models, $p(z_0|\psi) = p_{QML}(z_0|\psi)$ and $p(z_t, t|z_{t-1}, t-1, \psi) = p_{QML}(z_t, t|z_{t-1}, t-1, \psi)$, so that $w(z) = 1$ for all z and $\mathcal{L}(\psi) = \mathcal{L}_{QML}(\psi)$. For CIR models, $p(z_0|\psi)$ and $p(z_t, t|z_{t-1}, t-1, \psi)$ are non-Gaussian, which means $w(z) \neq 1$ for some z and $\mathcal{L}(\psi) \neq \mathcal{L}_{QML}(\psi)$, but can be evaluated analytically because they are Gamma and non-central χ^2 densities, respectively. In this case, evaluating the numerator of $w(z)$ is only slightly more complicated than evaluating the denominator. Finally, for general ATSMs, $p(z_0|\psi)$ and $p(z_t, t|z_{t-1}, t-1, \psi)$ are unfortunately analytically *unknown*.

There are at least two ways to implement the estimator when the densities are unknown. First, $p(z_0|\psi)$ and $p(z_t, t|z_{t-1}, t-1, \psi)$ can be evaluated numerically using the characteristic functions method of Singleton (2001) . Second, we can use the Hermite polynomial expansion method of Aït-Sahalia (2001) to obtain arbitrarily precise analytical approximations of the densities. In practice, these analytical approximations are more appealing in our context because of their computational speed and the fact that in each iteration of the likelihood maximization the $T+1$ densities need to be evaluated L times. Even if it takes less than a second to solve for the densities numerically, with $T = 500$ and $L = 100$, it takes 14 hours for just one evaluation of the likelihood function. In contrast, once the polynomial expansion coefficients have been computed as a function of the model parameters, which is done only once, evaluating the analytical density approximations is virtually instantaneous.

2.4.3 Extracting the Latent Factors

Analogous to the reformulation of the likelihood function, we can obtain an expression for the smoothed estimates $\widehat{z}_{|T} = [\widehat{z}_{1|T}', \widehat{z}_{2|T}', \ldots, \widehat{z}_{T|T}']'$ of the latent factors z. We substitute:

$$p(y|z,\psi) = \frac{p(y,z|\psi)}{p(z|\psi)} = \frac{p(z|y,\psi)\mathcal{L}(\psi)}{p(z|\psi)} \tag{25}$$

into equation (17) and solve for:

$$p(z|y,\psi) = \frac{\mathcal{L}_{QML}(\psi)}{\mathcal{L}(\psi)} \frac{p(z|\psi)}{p_{QML}(z|\psi)} p_{QML}(z|y,\psi) = \frac{\mathcal{L}_{QML}(\psi)}{\mathcal{L}(\psi)} w(z)\, p_{QML}(z|y,\psi). \tag{26}$$

The smoothed estimates of z are then given by:

$$\begin{aligned} \widehat{z}_{|T} &= \mathrm{E}\big[z|y,\psi\big] \\ &= \int z\, p(z|y,\psi)\, dz \\ &= \frac{\mathcal{L}_{QML}(\psi)}{\mathcal{L}(\psi)} \int z\, w(z)\, p_{QML}(z|y,\psi)\, dz \\ &= \frac{\mathcal{L}_{QML}(\psi)}{\mathcal{L}(\psi)} \mathrm{E}_{QML}[w(z)z|y,\psi]. \end{aligned} \tag{27}$$

Finally, we replace the expectation with respect to the smoothed QML inferences with an average over the simulated factor histories:

$$\widehat{\widehat{z}}_{|T} = \frac{\mathcal{L}_{QML}(\psi)}{\mathcal{L}(\psi)} \frac{1}{L} \sum_{l=1}^{L} w(z^l) z^l. \tag{28}$$

3 Monte Carlo Study

To explore the properties of our SML approach, we conduct a Monte Carlo study for a single-factor CIR model. Like CIR, we assume that the instantaneous interest rate is $r_t = z_t$ (i.e., $n = 1$, $\delta_0 = 0$, and $\delta_1 = 1$) and the factor follows a square-root diffusion:

$$dz_t = \kappa(\theta - z_t)dt + \Sigma\sqrt{z_t}dW_t \tag{29}$$

(i.e., $\alpha = 0$ and $\beta = 1$). Zero-coupon bond prices are then given by equation (6) with:

$$\begin{aligned} A(\tau) &= \nu \ln \left[\frac{2\gamma \exp\left\{\frac{1}{2}(\kappa+\lambda+\gamma)\tau\right\}}{(\kappa+\lambda+\gamma)\big(\exp\{\gamma\tau\}-1\big)+2\gamma} \right], \\ B(\tau) &= -\frac{2\big(\exp\{\gamma\tau\}-1\big)}{(\kappa+\lambda+\gamma)\big(\exp\{\gamma\tau\}-1\big)+2\gamma} \end{aligned} \tag{30}$$

where $\nu = 2\kappa\theta/\Sigma^2$ and $\gamma = \sqrt{(\kappa+\lambda)^2 + 2\Sigma^2}$.

The advantage of the CIR model is that the unconditional distribution of z_0 is Gamma:

$$p(z_0|\psi) = \frac{1}{\Gamma[\nu]\phi^\nu}\, z_0^{\nu-1}\, \exp\{-z_0/\phi\}, \tag{31}$$

where $\phi = \Sigma^2/(2\kappa)$, and the discrete time transition density of z_t is non-central χ^2:

$$p(z_t, t|z_{t-1}, t-1, \psi) = \frac{\Delta\big[\Psi(z_t)/\Phi(z_s)\big]^{(\nu-1)/2}}{\exp\big\{\Psi(z_t)+\Phi(z_s)\big\}}\, \mathcal{I}_{\nu-1}\Big[2\sqrt{\Phi(z_s)\Psi(z_t)}\Big], \tag{32}$$

where $\Delta = 1/[\phi(1-\exp\{-\kappa(t-s)\})]$, $\Phi(z_s) = z_s\,\Delta\,\exp\{-\kappa(t-s)\}$, $\Psi(z_s) = z_t\,\Delta$, and $\mathcal{I}_{\nu-1}[\,\cdot\,]$ denotes the modified Bessel function of the first kind of order $\nu-1$. These closed-form expressions for the densities allow us to directly simulate discrete time data from the model using the inverse-CDF method (as opposed to using a discretization scheme) and speed up the SML estimation because we can evaluate the ratio of marginal densities $w(z)$ without having to resort to numerical solutions or analytical approximations.

The experiment proceeds as follows. Setting $\kappa = 0.8$, $\theta = 0.03$, $\sigma = 0.1$, and $\lambda = -0.5$, which corresponds roughly to our estimates in the next section, we simulate a time-series of $T = 520$ weekly factor realizations and generate for each realization a set of $k = 5$ zero-coupon bond yields y_t^τ for maturities τ of one, three, 12, 60, and 120 months. We then add to each yield an iid measurement error ϵ_t^τ with standard deviation $h = 0.005$ (i.e., $H_t = h^2 I_{k\times k}$) and apply both the QML and SML estimators to the resulting noisy panel data. Finally, we repeat the procedure 100 times to obtain sampling distributions of the estimates.

Table 1 reports the across-simulations means and standard deviations of the QML and SML estimates, where for SML we consider simulation sizes L ranging from 50 to 1000. The differences between the estimators are striking. For κ, θ, and λ, the SML estimates are less biased and much less variable than the QML estimates. For example, for the market price of interest rate risk λ, with a true value of -0.5, the average QML estimate is -0.531 with a standard deviation of 0.185. The average SML estimate with $L = 1000$ is -0.503 with a

standard deviation of only 0.014 (a 13-fold reduction). The relative performance of SML is equally impressive for the mean-reversion coefficient κ and long-run mean θ.

The roles are reversed for the standard deviation coefficient Σ. Compared to the unbiased QML estimates, the SML estimates are slightly upward-biased and about twice as variable across simulations. Intuitively, the reason for this difference in relative performance is that the simulation-induced noise in filtering the current value of the factor makes the factor look more volatile that it actually is. Consistent with this explanation, both the bias and sampling variation of the SML estimator diminish as we increase the simulation size. Finally, for the standard deviation of the yield errors h, the estimators perform comparably.

4 Empirical Application

4.1 Data

Following Dai and Singleton (2000), we collect weekly data on U.S. dollar LIBOR and swap rates from Data Stream. The rates are sampled every Wednesday from April 1987 through December 1999 (666 observations). The LIBOR rates are for three- and six- month maturities and the swap rates are for one-, two-, three-, five-, seven-, and 10-year maturities. From this raw data, we infer the zero-coupon yields for 12 points on the zero-coupon yield curve (three, six, 12, 24, 36, ... 108, and 120 months to maturity) using the bootstrapping method outlined in James and Webber (2001). We plot the resulting panel of yields in Figure 1.

4.2 Single-factor CIR model

We first estimate the single-factor CIR model assuming, as in the Monte Carlo study, that the measurement errors are iid and homogeneous in the cross-section with standard deviation h. We use $L = 500$ simulations to compute the likelihood correction for the SML estimator, which in our view represents a sensible compromise between accuracy of the estimates and computational speed. Panel A of Table 2 presents both the QML and SML estimates.

The two sets of estimates are very similar. They are also consistent with comparable estimates of the single-factor CIR model in the literature. The mean-reversion speed κ is 0.491, implying a weekly autocorrelation of 0.991 and a half-life of 1.413 years. The long-run mean of the factor is 1.96 percent and the steady state volatility $\sqrt{\Sigma\theta}$ is 5.21 percent. The market price of interest rate risk is negative at -0.456, which together with the other parameters, implies an asymptotic yield $y_t^\infty = 2\kappa\theta/(\gamma + \kappa + \lambda)$ of 8.24 percent. Finally, the standard deviation of the pricing or measurement errors is fairly large at 43 basis points.

The differences between the QML and SML estimates are in line with the biases of the two estimators documented in Table 1. On one hand, the biased QML estimates of κ and λ are larger in magnitude than the corresponding SML estimates. On the other hand, the biased SML estimate of Σ is larger than the corresponding QML estimate. The estimates of θ and h are roughly the same. Notice also that the asymptotic standard errors of the SML estimates are of the same magnitude as the finite sample standard errors. The asymptotic standard errors of the QML estimates, in contrast, are much smaller than the finite sample standard errors. For example, the asymptotic standard error on the market price of risk is 0.011, compared to the finite sample counterpart of 0.185. This observation serves as a caution against relying on asymptotic inferences in QML estimation and, at the same time, illustrates the strength of our SML approach.

As a model specification diagnostic, we report in Panel B of Table 2 the means, standard deviations, and MSEs of the yield errors $y_t^\tau - A(\tau) - B(\tau)\widehat{z}_t$, where $A(\tau)$ and $B(\tau)$ are evaluated at the estimated parameters and $\widehat{z}_t$ denotes the smoothed estimate of the latent factor from the Kalman filter (QML) or from the estimator described in Section 2.4.3 (SML). The striking feature of the pricing errors is the pronounced u-shape of the standard deviation across maturities. Bonds with two to four years to maturity are priced much more precisely than bonds with less than one year or more than five years to maturity. If the data is generated by a single-factor CIR model, the assumption of iid errors in the cross-section is clearly false. (An obvious alternative is that the data is not generated from a single-factor CIR model. We consider this alternative in Section 4.3.)

To accommodate this u-shape in the standard deviation of the errors, we parameterize the variance of the errors as a function of the time-to-maturity of the bond:

$$h^2(\tau) = \exp\left\{a_0 + a_1\tau + a_2\tau^2\right\}. \tag{33}$$

The obvious advantage of parameterizing the standard deviation is that it allows for some heterogeneity in the measurement errors without requiring a separate parameter for each bond. In particular, with the parameterization (33), the log standard deviation is a quadratic function of τ that can be u-shaped. We work with logs to ensure positivity.

Table 3 shows in Panel A the QML and SML estimates with heterogeneous measurement errors, and describes in Panel B the corresponding yield errors. The parameter estimates are somewhat different from the ones with homogeneous measurement errors. In particular, the estimated long-run mean θ is smaller at 1.75 percent, and the estimates of κ and λ are larger in magnitude, implying a weekly autocorrelation of 0.98 and an asymptotic yield of 8.32 percent. These differences illustrate the importance of correctly modeling the measurement

errors for drawing inferences about the other parameters of the model.

The estimates of a_0, a_1, and a_2 are all statistically significant at the five-percent level. The signs and magnitudes of the estimates are such that $\ln h^2(\tau)$ is indeed a u-shaped function of τ that obtains a minimum at approximately $\tau = 3$. This is consistent with Panel B, where the standard deviation of the pricing errors is smallest for the three-year bond.

The yield errors are not particularly informative about which estimates are closer to the true parameters because they ignore the time-series dimension of the model. It is possible for a model to fit well the cross-sections of yields but for the estimated factor dynamics to be a poor description of the true factor dynamics (Backus, Foresi, and Zin, 1998). To get a better sense for the quality of the estimates, we conduct the following forecasting experiment. For each date t, we obtain the smoothed estimates of the latent factor $\widehat{z}_{t|T}$, either from the Kalman filter (QML) or from the estimator described in Section 2.4.3 (SML), and use these estimates as an anchor to generate forecasts of the factors m periods in the future $\widehat{z}_{t+m|t}$ through the estimated factor dynamics. We then use these forecasts of the factors to construct forecasts of the entire cross-section of bond yields and compute the yield forecast errors $y^{\tau}_{t+m} - A(\tau) - B(\tau)\widehat{z}_{t+m|t}$.

Table 4 presents the means, standard deviations, and MSEs of the yield forecast errors for bonds with one, two, five and 10 years to maturity and forecast horizons m ranging from one week to one quarter (12 weeks). Judging by the magnitude of the MSEs across horizons, it is obviously more difficult to forecast yields at long horizons than at short horizons. As a result, the differences between the QML and SML estimates are most pronounced at the one-quarter horizon, for which a clear pattern emerges. Except for the one-year bond, the SML estimates produce less biased but more variable forecasts.

This pattern makes sense. The QML estimator produces more biased forecasts for two reasons. First, the smoothed estimates of the factors from the Kalman filter are biased. Second, the QML parameter estimates are biased, which means that the forecasts are potentially even more biased than the Kalman filter-based estimates to which these forecasts are anchored. The SML estimator produces more variable forecasts because the smoothed estimates of the factors are contaminated with some simulation error. This illustrates that, in the end, the choice between QML and SML boils down to a standard bias-versus-variance tradeoff. It is important to realize, however, that the bias of the QML estimator is set in stone, while the simulation-induced variance of the SML estimator can always be reduced by increasing the simulation size L and through variance reduction techniques.

4.3 Two-Factor CIR Model

The sheer magnitudes of the pricing errors in Tables 2–4 confirm the consensus in the literature that more than one factor is needed to explain the dynamics of the term structure (Stambaugh, 1988; Litterman and Scheinkman, 1991). Although a careful analysis of multifactor term structure models, along the lines of Dai and Singleton (2000) and Brandt and Chapman (2002), is outside the scope of this paper, we at least consider the two-factor CIR model of Chen and Scott (1992). Specifically, we assume that the instantaneous interest rate is $r_t = z_{1,t} + z_{2,t}$ (i.e., $n = 2$, $\delta_0 = 0$, $\delta_1 = 1$, and $\delta_2 = 2$) and each factor follows an independent square-root diffusion:

$$dz_{i,t} = \kappa_i(\theta_i - z_{i,t})dt + \Sigma_i\sqrt{z_{i,t}}dW_{i,t}, \tag{34}$$

(i.e., $\alpha_i = 0$ and $\beta_i = 1$). Zero-coupon bond prices are then given by equation (6) with:

$$\begin{aligned} A_i(\tau) &= \nu_i \ln\left[\frac{2\gamma_i \exp\left\{\frac{1}{2}(\kappa_i+\lambda_i+\gamma_i)\tau\right\}}{(\kappa_i+\lambda_i+\gamma_i)\big(\exp\{\gamma_i\tau\}-1\big)+2\gamma_i}\right] \\ B_i(\tau) &= -\frac{2\big(\exp\{\gamma_i\tau\}-1\big)}{(\kappa_i+\lambda_i+\gamma_i)\big(\exp\{\gamma_i\tau\}-1\big)+2\gamma_i}. \end{aligned} \tag{35}$$

Due to the independence of the factors, the unconditional density of z_0 and the transition density of z_t are simply *products* of the corresponding univariate densities:

$$p(z_0|\psi) = \prod_{i=1}^{2} \frac{1}{\Gamma[\nu_i]\phi_i^{\nu_i}}\, z_{i,0}^{\nu_i-1}\, \exp\{-z_{i,0}/\phi_i\} \tag{36}$$

and

$$p(z_t, t|z_{t-1}, t-1, \psi) = \prod_{i=1}^{2} \frac{\Delta_i\big[\Psi_i(z_{i,t})/\Phi_i(z_{i,s})\big]^{(\nu_i-1)/2}}{\exp\big\{\Psi_i(z_{i,t})+\Phi_i(z_{i,s})\big\}}\, \mathcal{I}_{\nu_i-1}\Big[2\sqrt{\Phi_i(z_{i,s})\Psi_i(z_{i,t})}\Big], \tag{37}$$

where the subscripted coefficients are straightforward factor-specific versions of the coefficients defined in conjunction with equations (30)–(32).

Table 5 presents in Panel A the QML and SML estimates of the two-factor model with heterogeneous measurement errors and describes in Panel B the corresponding yield errors (the results for homogeneous errors are available on request). The estimators identify two distinctly different factors. The first factor mean-reverts very slowly at a rate of $\kappa_1 = 0.154$, which implies a weekly autocorrelation of 0.997 and a half-life of 4.5 years, while the

second factor mean-reverts much faster at a rate of $\kappa_2 = 0.571$, which implies a weekly autocorrelation of 0.989 and a half-life of 1.2 years. Both factors have a negative market price of risk, with λ_2 being about twice as large (in magnitude) as λ_1 (so the second factor co-varies more strongly with the stochastic discount factor). Finally, the estimates of a_0, a_1, and a_2 imply again that the function $h^2(\tau)$ is u-shaped in τ with a minimum at $\tau = 6$, which is consistent with the standard deviations of the measurement errors in Panel B.

Judging by the statistical significance of the parameters corresponding to the second factor, by the differences between the log likelihoods in Tables 3 and 5, or by the differences in the corresponding MSEs of the yield errors, it is obvious that the second factor substantially improves the fit of the model. In particular, the MSEs are uniformly reduced by factors ranging from 1.5 for the two-year bond to 136 for the eight-year bond.

Comparing the two estimators, the differences between the results are noticeably larger for the two-factor model than for the single-factor model in Tables 2 and 3. For example, the SML estimate of κ_1 is 0.154 while the QML estimate is 0.246 (a difference of 6.4 standard deviations). The SML estimate of λ_1 is -0.133 while the QML estimate is -0.228 (a difference of 8.1 standard deviations). These differences between the two sets of estimates are consistent with the Monte Carlo study of Frühwirth-Schnatter and Geyer (1998), which documents increasingly severe biases of QML as the number of factors increases.

Finally, Table 6 reports the means, standard deviations, and MSEs of the yield forecast errors of the two-factor model. Consistent with the improved fit of the model, the magnitude of the MSEs is reduced substantially at all horizons, relative to the forecast errors of the single-factor model in Table 4. Furthermore, the pattern in the results is the same as for the single-factor model. Except for the one-year bond, the SML estimates produce less biased but more variable (due to the simulation-induced noise) forecasts than the QML estimates. The resulting MSEs are slighly larger for the one- and ten-year bonds but smaller for the two- and five-year bonds. This illustrates, once again, that the biases of the QML approach are noticable in practical applications.

5 Conclusion

In this paper, we proposed a simulated maximum likelihood (SML) method for estimating affine term structure model from noisy panel data that overcomes the problems of the Kalman filter based QML estimates. In particular, we used the importance sampling approach of Durbin and Koopman (1997,2001a) to correct the likelihood function of the QML estimator for the non-normalities introduced by the affine factor dynamics. Depending on the accuracy

of the correction, which is determined by the number of simulations, the quality of the estimator ranges from QML (no correction) to exact ML. We used both a Monte Carlo study and an empirical application to demonstrate the merrits of our approach.

A natural extention of our estimation approach is to consider the class of quadratic term structure models (QTSMs) of Beaglehole and Tenney (1992), Constantinides (1992), and Ahn, Dittmar, and Gallant (2001), among others. In the state-space representation of a QTSM, the measurement equation is non-linear but the state-equation is linear and Gaussian. Except for the form of the non-linearities in the measurement equation, this setting is identical to the application of importance sampling-based SML to stochastic volatility models by Sandmann and Koopman (1998) and Brandt and Kang (2001).

A Appendix

This appendix provides analytical expressions for the conditional mean $m(z_{t-1})$ and variance $v(z_{t-1})$ of the discrete time factor dynamics (13) used in the Kalman filter recursions. The results here are taken from Fisher and Gilles (1996) and Duffee (2001).

The continuous time factor dynamics are given by the affine diffusion (10). Assume that the $n \times n$ the matrix of mean-reversion coefficients κ can be diagonalized as follows:

$$\kappa = MDM^{-1}, \tag{A.1}$$

where D is an $n \times n$ diagonal matrix with $\{d_1, d_2, \ldots, d_n\}$ on the central diagonal.[6] Define the transformed factors:

$$z_t^* = M^{-1} z_t, \tag{A.2}$$

which, by Ito's lemma, have the continuous time dynamics:

$$dz_t^* = D(\theta^* - z_t^*)dt + \Sigma^* S_t^* dW_t, \tag{A.3}$$

where $\theta^* = M^{-1}\theta$, $\Sigma^* = M^{-1}\Sigma$, and S_t^* is a diagonal matrix with ith diagonal element:

$$\{S_t^*\}_{ii} = \sqrt{\alpha_i + \beta_i^{*\prime} z_t^*} \tag{A.4}$$

with $\beta^* = \beta M$. The essential property of these transformed factors is that the drift of each factor is only affected by its own level.

Define the following notation. If x is an n-vector, diag$[x]$ is an $n \times n$ diagonal matrix with elements $\{x_1, x_2, \ldots, x_n\}$ on the central diagonal. If x is an $n \times n$ diagonal matrix, $\exp\{x\}$ is also an $n \times n$ diagonal matrix with elements $\{\exp\{x_{1\,1}\}, \exp\{x_{2\,2}\}, \ldots, \exp\{x_{n\,n}\}\}$ on the central diagonal. Finally, the vector $x_{\cdot i}$ is the ith column of the matrix x.

The conditional mean of the transformed factors is simply:

$$\begin{aligned} \mathrm{E}[z_t^* | z_s^*] &= \theta^* + \exp\{-D(t-s)\}(z_s^* - \theta^*) \\ &= \big[I_{n\times n} - \exp\{-D(t-s)\}\big]\theta^* + \exp\{-D(t-s)\} z_s^*, \end{aligned} \tag{A.5}$$

so that the conditional mean of the original factors is given by:

$$\begin{aligned} \mathrm{E}[z_t | z_s] &= M\,\mathrm{E}[z_t^* | z_s^*] \\ &= M\big[I_{n\times n} - \exp\{-D(t-s)\}\big]\theta^* + M\exp\{-D(t-s)\}M^{-1} z_s. \end{aligned} \tag{A.6}$$

Deriving an expression for the conditional variance is somewhat more complicated. Write

[6] See Fisher and Gilles (1996) for the more general case when κ cannot be diagonalized.

the instantaneous covariance matrix of the tranformed factors as:

$$\begin{aligned}\Sigma^* S_t^* S_t^{*\prime} \Sigma^{*\prime} &= \Sigma^* \mathrm{diag}[\alpha] \Sigma^{*\prime} + \sum_{i=1}^{n} \Sigma^* \mathrm{diag}[\beta_{\cdot i}^*] \Sigma^{*\prime} z_{t\,i}^* \\ &\equiv G_0 + \sum_{i=1}^{n} G_i z_{t\,i}^*,\end{aligned} \tag{A.7}$$

where $G_0 = \Sigma^* \mathrm{diag}[\alpha] \Sigma^{*\prime}$, $G_i = \Sigma^* \mathrm{diag}[\beta_{\cdot i}^*] \Sigma^{*\prime}$, and then define the $n \times n$ matrix:

$$\begin{aligned}F(t,s) &= G_0 + \sum_{i=1}^{n} G_i \mathrm{E}[z_{t\,i}^* | z_s^*] \\ &= G_0 + \sum_{i=1}^{n} G_i \big[\theta_i^* + \exp\{-d_i(t-s)\}(z_{s\,i}^* - \theta_i^*)\big].\end{aligned} \tag{A.8}$$

Fisher and Gilles (1996) shows that the conditional variance of the transformed factors can then be written as:

$$\begin{aligned}\mathrm{Var}[z_t^* | z_s^*] &= \int_s^t \exp\{-2D(t-\tau)\} F(t,\tau) d\tau \\ &= \int_s^t \exp\{-2D(t-\tau)\} G_0 d\tau + \sum_{i=1}^{n} \theta_i^* \int_s^t \exp\{-2D(t-\tau)\} G_i d\tau \\ &\quad + \sum_{i=1}^{n} \left[(z_{s\,i}^* - \theta_i^*) \int_s^t \exp\{-2D(t-\tau)\} G_i \exp\{-d_i(t-\tau)\} d\tau \right]\end{aligned} \tag{A.9}$$

and

$$\begin{aligned}\mathrm{Cov}[z_{t\,j}^*, z_{t\,k}^* | z_s^*] =& \frac{G_{0,j\,k}\big[1 - \exp\{-(t-s)(d_j + d_k)\}\big]}{d_j + d_k} + \\ & \sum_{i=1}^{n} \frac{\theta_i^* G_{i,\,j,k}\big[1 - \exp\{-(t-s)(d_j + d_k)\}\big]}{d_j + d_k} + \\ & \sum_{i=1}^{n} \frac{(z_{s\,i}^* - \theta_i^*) G_{i,j\,k}\big[\exp\{-d_i(t-s)\} - \exp\{-(t-s)(d_j + d_k)\}\big]}{d_j + d_k - d_i}.\end{aligned} \tag{A.10}$$

Collecting terms, we have:

$$\mathrm{Var}[z_t^* | z_s^*] = b_0 + \sum_{i=1}^{n} b_i z_{s\,i}^*, \tag{A.11}$$

so that the conditional variance of the original factors is given by:

$$\begin{aligned}\mathrm{Var}[z_t | z_s] &= M \,\mathrm{Var}[z_t^* | z_s^*] M' \\ &= M b_0 M + \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} M b_j M' M_{j_1}^{-1} \Big) z_{s\,i}.\end{aligned} \tag{A.12}$$

References

Ahn, D., R.F. Dittmar, and A.R. Gallant, 2001, Quadratic Term Structure Models: Theory and Evidence, *Review of Financial Studies*, forthcoming.

Aït-Sahalia, Y., 2001, Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach, *Econometrica*, forthcoming.

Baadsgaard, M., J.N. Nielsen, and H. Madsen, 2000, Estimating Exponential-Affine Term Structure Models from Coupon Bond Prices Using Nonlinear Filtering, Working Paper, Technical University of Denmark.

Babbs, S.K., and K.B. Nowman, 1999, Kalman Filtering of Generalized Vasicek Term Structure Models, *Journal of Financial and Quantitative Analysis* 34, 115–130.

Backus, D.K., S. Foresi, and S. Zin, 1998, Arbitrage Opportunities in Arbitrage-Free Models of Bond Pricing, *Journal of Business and Economic Statistics* 16, 13–26.

Beaglehole, D., and M. Tenney, 1992, Corrections and Additions to "A Nonlinear Equilibrium Model of the Term Structure of Interest Rates," *Journal of Financial Economics* 32, 345–353.

Bollerslev, T., and J.M. Wooldridge, 1992, Quasi-Maximum Likelihood Estimation of Dynamic Models with Time-Varying Covariances, *Econometric Review* 11, 143–172.

Brandt, M.W., and D.A. Chapman, 2002, Comparing Multifactor Models of the Term Structure, Working Paper, University of Pennsylvania.

Brandt, M.W., and Q. Kang, 2001, On the Relationship between the Conditional Mean and Volatility of Stock Returns: A Latent VAR Approach, Working Paper, University of Pennsylvania.

Brown, S.J., and P.H. Dybvig, 1986, The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates, *Journal of Finance* 41, 617–630.

Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A.B. Sanders, 1992, An Empirical Comparison of Alternative Models of the Short-Term Interest Rate, *Journal of Finance* 47, 1209–1227.

Chapman, D.A., J. Long, and N.D. Pearson, 1999, Using Proxies for the Short Rate: When Are Three Months Like an Instant?, *Review of Financial Studies* 12, 763–806.

Chen, R., and L. Scott, 1992, Pricing Interest Rate Options in a Two-Factor Cox-Ingersoll-Ross Model of the Term Structure, *Review of Financial Studies* 5, 613–636.

Chen, R., and L. Scott, 1993, Maximum Likelihood Estimation for a Multi-Factor Equilibrium Model of the Term Structure of Interest Rates, *Journal of Fixed Income* 4, 14–31.

Chen, R., and L. Scott, 1995, Multifactor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from a Kalman Filter Model, Working Paper, University of Georgia.

Chen, L., 1996, Stochastic Mean and Stochastic Volatility – a Three-Factor Model of the Term Structure of Interest Rates and its Application in Derivatives Pricing and Risk Management, *Financial Markets, Institutions, and Instruments* 5, 1–88.

Constantinides, G.M., 1992, A Theory of the Nominal Term Structure of Interest Rates, *Review of Financial Studies* 5, 531–552.

Cox, J.C., J.E. Ingersoll Jr., and S.A. Ross, 1985, A Theory of the Term Structure of Interest Rates, *Econometrica* 53, 385–408.

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

De Jong, F., 2000, Time-Series and Cross-Section Information in Affine Term Structure Models, *Journal of Business and Economic Statistics* 18, 300–314.

Duan, J.C., and J.G. Simonato, 1999, Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter, *Review of Quantitative Finance and Accounting* 13, 111–135.

Duffee, G.R., 2000, Term Premia and Interest Rate Forecasts in Affine Models, *Journal of Finance*, forthcoming.

Duffee, G.R., and R.H. Stanton, 2001, Estimation of Dynamic Term Structure Models, Working Paper, University of California at Berkeley.

Duffie, D., and R. Kan, 1996, A Yield-Factor Model of Interest Rates, *Mathematical Finance* 6, 379–406.

Durbin, J., and S.J. Koopman, 1997, Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models, *Biometrika* 84, 669–684.

Durbin, J., and S.J. Koopman, 2001a, *Time Series Analysis by State Space Models*, Oxford University Press: New York, NY.

Durbin, J., and S.J. Koopman, 2001b, A Simple and Efficient Simulation Smoother for State Space Time Series Analysis, Working Paper, London School of Economics.

Fisher, M., and C. Gilles, 1996, Estimating Exponential-Affine Models of the Term Structure, Working Paper, Federal Reserve Bank of Atlanta.

Frühwirth-Schnatter, S., and A.L.J. Geyer, 1998, Bayesian Estimation of Econometric Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure of Interest Rates Via MCMC Methods, Working Paper, Vienna University of Economics and Business Administration.

Gibbons, M.R., and K. Ramaswamy, 1993, A Test of the Cox, Ingersoll, and Ross Model of the Term Structure, *Review of Financial Studies* 6, 619–658.

Hamilton, J.D., 1994, *Time Series Analysis*, Princeton University Press: Princeton, NJ.

Harvey, A., 1989, *Forecasting, Structural Time Series Models, and the Kalman Filter*, Cambridge University Press: New York, NY.

Heston, S.L., 1991, Testing Continuous-Time Models of the Term Structure of Interest Rates, Working Paper, Yale University.

James, J., and N. Webber, 2001, *Interest Rate Modeling*, John Wiley & Sons: New York, NY.

Jamshidian, F., 1989, An Exact Bond Option Formula, *Journal of Finance* 44, 205–209.

Jamshidian, F., 1991, Bond and Option Evaluation in the Gaussian Interest Rate Model and Implementation, *Research in Finance* 9, 131–170.

Jamshidian, F., 1992, A Simple Class of Square-Root Interest Rate Models, Working Paper, Fuji International.

Jegadeesh, N., and G. Pennacchi, 1996, The Behavior of Interest Rates Implied by the Term Structure of Eurodollar Futures, *Journal of Money, Credit, and Banking* 28, 426–446.

Langetieg, T.C., 1980, A Multivariate Model of the Term Structure of Interest Rates, *Journal of Finance* 35, 71–97.

Lund, J., 1997, Non-Linear Kalman Filtering Techniques for Term Structure Models, Working Paper, Aarhus School of Business.

Litterman, R., and J.A. Scheinkman, 1991, Common Factors Affecting Bond Returns, *Journal of Fixed Income* 1, 54–61.

Longstaff, F.A., and E.S. Schwartz, 1992, Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model, *Journal of Finance* 47, 1259–1283.

Marsh, T.A., and E. Rosenfeld, 1983, Stochastic Processes for Interest Rates and Equilibrium Bond Prices, *Journal of Finance* 38, 635–646.

Nowman, K.B., 1997, Gaussian Estimation of Single-Factor Continuous Time Models of the Term Structure of Interest Rates, *Journal of Finance* 52, 1695–1706.

Pearson, N.D., and T. Sun, 1994, Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model, *Journal of Finance* 49, 1279–1304.

Sandmann, G., and S.J. Koopman, 1998, Estimation of Stochastic Volatility Models Via Monte Carlo Maximum Likelihood, *Journal of Econometrics* 87, 271–301.

Shephard, N., and M.K. Pitt, 1997, Likelihood Analysis of Non-Gaussian Measurement Time Series, *Biometrika* 84, 653–667.

Singleton, K.J., 2001, Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function, *Journal of Econometrics* 102, 111–141.

Stambaugh, R.F., 1988, The Information in Forward Rates: Implications for Models of the Term Structure, *Journal of Financial Economics* 21, 41–70.

Vasicek, O., 1977, An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics* 5, 177–188.

White, H., 1982, Maximum Likelihood Estimation of Misspecified Models, *Econometrica* 50, 1–25.

Figure 1: Zero-Coupon Yields

Yields-to-maturity of zero-coupon libor bonds with three-months to ten-years to maturity. Weekly data from April 1987 through December 1999 (666 observations).

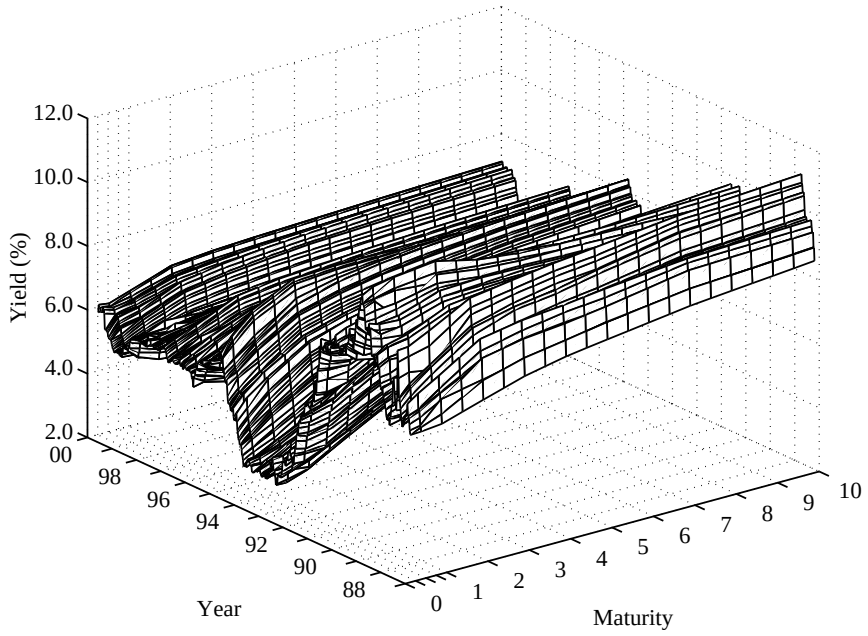

Table 1: Sampling Distribution of QML and SML Estimates

Means and standard deviations of SML estimates across 100 Monte Carlo simulations. The true parameter values are $\kappa = 0.8$, $\theta = 0.03$, $\Sigma = 0.1$, $\lambda = -0.5$, $h = 0.005$.

	QML		SML ($L = 50$)		SML ($L = 500$)		SMLE ($L = 1000$)	
Params	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev
κ	0.8341	0.1859	0.7919	0.0317	0.7909	0.0261	0.7943	0.0143
θ	0.0304	0.0070	0.0300	0.0013	0.0301	0.0012	0.0301	0.0007
Σ	0.0983	0.0067	0.1331	0.0140	0.1274	0.0127	0.1155	0.0113
λ	-0.5313	0.1852	-0.5065	0.0310	-0.4993	0.0260	-0.5029	0.0138
h	0.00550	0.00008	0.00497	0.00009	0.00496	0.00008	0.00496	0.00007

Table 2: Single-Factor CIR Model with Homogeneous Errors

QML and SML estimates of single-factor CIR model with homogeneous measurement errors in Panel A. Means, standard deviations, and MSEs of corresponding yield pricing errrors in Panel B. SML estimates are based on 500 simulations.

Panel A: Parameter Estimates

Params	QML		SML	
	Estimate	t-Stat	Estimate	t-Stat
κ	0.4964	33.39	0.4907	28.75
θ	0.0202	10.05	0.0196	12.43
Σ	0.1226	6.91	0.1387	7.61
λ	-0.4835	-42.14	-0.4563	-38.10
h	0.00424	4.93	0.00425	4.97
Log Likelihood	42384		42992	

Panel B: Yield Errors

Maturity (Months)	QML			SML		
	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)
3	8.2741	6.5074	4.3031	3.9931	6.3738	4.0785
6	11.0920	5.6446	3.3092	7.7432	5.5138	3.1002
12	9.9494	4.3836	2.0206	8.2235	4.2880	1.9063
24	-12.1735	2.2588	0.6584	-11.5483	2.1108	0.5789
36	-12.7002	1.0234	0.2660	-10.7491	1.0585	0.2276
48	-10.4110	1.6156	0.3694	-7.9317	1.7417	0.3663
60	-6.6199	2.5833	0.7112	-4.1883	2.6978	0.7454
72	-2.2090	3.2292	1.0477	-0.1962	3.3698	1.1356
84	-0.3525	3.9406	1.5529	1.0409	4.1209	1.6993
96	3.9509	4.2884	1.8546	4.6572	4.5108	2.0565
108	6.2805	4.6543	2.2057	6.3275	4.9216	2.4622
120	6.9144	5.0346	2.5826	6.3919	5.3441	2.8968

Table 3: Single-Factor CIR Model with Heterogeneous Errors

QML and SML estimates of single-factor CIR model with heterogeneous measurement errors in Panel A. Means, standard deviations, and MSEs of corresponding yield pricing errrors in Panel B. SML estimates are based on 500 simulations.

Panel A: Parameter Estimates

Params	QML		SML	
	Estimate	t-Stat	Estimate	t-Stat
κ	0.5955	38.36	0.5705	33.26
θ	0.0143	10.58	0.0175	12.08
Σ	0.1240	6.86	0.1412	6.86
λ	-0.5618	-47.40	-0.5336	-42.20
a_0	-10.8422	-29.87	-11.1532	-28.58
a_1	-0.4915	-5.00	-0.1731	-1.94
a_2	0.0572	4.58	0.0286	2.60
Log Likelihood	42869		43262	

Panel B: Yield Errors

Maturity (Months)	QML			SML		
	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)
3	10.0303	6.7171	4.6126	3.4890	5.7418	3.3089
6	12.9710	5.8266	3.5631	7.4725	4.8326	2.3912
12	12.0552	4.4976	2.1681	8.3804	3.5691	1.3440
24	-9.6845	2.2576	0.6035	-10.6833	1.5081	0.3416
36	-9.9177	0.7754	0.1585	-9.3453	1.2274	0.2380
48	-7.4130	1.3927	0.2489	-6.1280	2.2914	0.5626
60	-3.4722	2.4174	0.5964	-2.0887	3.2834	1.0824
72	1.0369	3.1016	0.9631	2.1246	3.9698	1.5805
84	2.9526	3.8417	1.4846	3.5331	4.7143	2.2350
96	7.2878	4.2057	1.8219	7.2900	5.0890	2.6429
108	9.6316	4.5863	2.1962	9.0836	5.4813	3.0870
120	10.2701	4.9798	2.5853	9.2631	5.8836	3.5475

Table 4: Single-Factor CIR Model Forecasts

Means, standard deviations, and MSEs of one-week to one-quarter ahead forecasts of one- to ten-year yields from QML and SML estimates of single-factor CIR model with heterogeneous measurement errors.

	QML: One-Week Forecast			SML: One-Week Forecast		
Maturity (Months)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)
12	4.7084	4.5075	2.0539	2.6589	4.4065	1.9488
24	-16.9411	2.7474	1.0418	-16.2620	2.6330	0.9577
60	-10.2043	2.9205	0.9571	-7.0309	3.0496	0.9795
120	4.6970	5.2163	2.7430	5.4600	4.5624	3.1239
	QML: One-Month Forecast			SML: One-Month Forecast		
Maturity (Months)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)
12	-7.9381	5.2161	2.7838	-9.4852	5.1067	2.6978
24	-29.3404	4.0523	2.5029	-28.0002	3.9660	2.3569
60	-21.6894	3.8301	1.9374	-17.4110	3.9392	1.9549
120	-4.7512	5.6842	3.2536	-2.4661	6.0207	3.6310
	QML: One-Quarter Forecast			SML: One-Quarter Forecast		
Maturity (Months)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)
12	-38.9910	7.1252	6.5972	-39.1512	7.0286	6.4730
24	-59.6321	6.6506	7.9790	-56.5123	6.5978	7.5468
60	-49.8392	5.8230	5.8746	-42.6791	5.9013	5.3040
120	-27.8595	6.7739	5.3647	-21.6730	7.0778	5.4793

Table 5: Two-Factor CIR Model with Heterogeneous Errors

QML and SML estimates of two-factor CIR model with heterogeneous measurement errors in Panel A. Means, standard deviations, and MSEs of corresponding yield pricing errrors in Panel B. SML estimates are based on 500 simulations.

Panel A: Parameter Estimates

Params	QML		SML	
	Estimate	t-Stat	Estimate	t-Stat
κ_1	0.2456	14.13	0.1538	10.70
θ_1	0.0085	8.24	0.0175	10.50
Σ_1	0.0648	4.25	0.0734	4.82
λ_1	-0.2275	-23.09	-0.1326	-11.25
κ_2	0.7976	8.83	0.5709	8.45
θ_2	0.0289	2.97	0.0380	2.20
Σ_2	0.1911	6.99	0.1874	6.21
λ_2	-0.3092	-10.97	-0.2808	-9.98
a_0	-13.3951	-35.22	-13.4132	-45.82
a_1	-0.2405	-3.05	-0.1016	-3.02
a_2	0.0223	1.32	0.0089	1.10
Log Likelihood	52996		53317	

Panel B: Yield Errors

Maturity (Months)	QML			SML		
	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)
3	-4.3561	2.2929	0.5447	-2.8280	2.2819	0.5287
6	3.2325	1.2419	0.1647	4.1716	1.2148	0.1649
12	9.2688	1.0234	0.1906	9.2833	0.9525	0.1769
24	-5.4229	1.4946	0.2528	-6.4644	1.3864	0.2340
36	-4.2570	1.2391	0.1716	-5.6248	1.1025	0.1532
48	-3.2562	0.9518	0.1012	-4.4996	0.8079	0.0855
60	-2.0720	0.7777	0.0647	-2.9243	0.6584	0.0519
72	-0.6116	0.5200	0.0274	-0.9250	0.4159	0.0181
84	-1.5093	0.3538	0.0148	-1.2127	0.3880	0.0165
96	0.5041	0.2695	0.0075	1.4345	0.4172	0.0194
108	1.1098	0.3402	0.0128	2.6687	0.5663	0.0392
120	0.5894	0.5398	0.0295	2.7547	0.7963	0.0710

Table 6: Two-Factor CIR Model Forecasts

Means, standard deviations, and MSEs of one-week to one-quarter ahead forecasts of one- to ten-year yields from QML and SML estimates of two-factor CIR model with heterogeneous measurement errors.

	QML: One-Week Forecast			SML: One-Week Forecast		
Maturity (Months)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)
12	8.6117	1.7558	0.3824	10.002	1.7140	0.3938
24	-6.1233	2.3088	0.5706	-5.9462	2.2194	0.5279
60	-2.7669	1.6390	0.2763	-2.6817	1.5633	0.2516
120	-0.0335	1.4193	0.2014	2.8641	1.5005	0.2334
	QML: One-Month Forecast			SML: One-Month Forecast		
Maturity (Months)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)
12	6.9206	3.3713	1.1845	12.3732	3.3417	1.2698
24	-7.8159	3.8197	1.5201	-4.0389	3.7517	1.4238
60	-4.4716	2.9947	0.9168	-1.6118	2.9576	0.8774
120	-1.5694	2.6918	0.7270	3.4980	2.7186	0.7513
	QML: One-Quarter Forecast			SML: One-Quarter Forecast		
Maturity (Months)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)	Mean ($\times 10^{-4}$)	StdDev ($\times 10^{-3}$)	MSE ($\times 10^{-5}$)
12	3.3046	6.1880	3.8401	19.1201	6.1172	4.1075
24	-11.2950	6.6153	4.5038	1.6969	6.5519	4.2956
60	-8.1219	5.4704	3.0585	1.8691	5.4938	3.0216
120	-4.8582	4.7860	2.3142	5.8087	4.8177	2.3547