Itô Conditional Moment Generator and the Estimation of Short Rate Processes[1]

Hao Zhou

Mail Stop 91
Federal Reserve Board
Washington, DC 20551

First Draft: January 1998
This Version: March 2003

[1]This paper grows out of an essay of my PhD dissertation, and is previously distributed under the title "Jump-Diffusion Term Structure and Itô Conditional Moment Generator." George Tauchen gave me valuable advice. I thank Ravi Bansal, David Bates, Tim Bollerslev, Peter Christoffersen, Sanjiv Das, Lars Hansen, Michael Hemler, Nour Meddahi, and Kenneth Singleton for their helpful suggestions. I am also grateful to the editor René Garcia, the associate editor, and two anonymous referees for their constructive suggestions. Comments from the seminar participants at Duke, Brown, Virginia economics departments, SNDE 1999, FMA 1999, ES 2000 annual meetings, and Duke Risk Conference 2000 are greatly appreciated. The views presented here are solely those of the author and do not necessarily represent those of the Federal Reserve Board or its staff. For questions and comments, please contact Hao Zhou, Trading Risk Analysis Section, Division of Research and Statistics, Federal Reserve Board, Washington DC 20551 USA; Phone 1-202-452-3360; Fax 1-202-728-5887; e-mail hao.zhou@frb.gov.

Abstract

This paper exploits the Itô's formula to derive the conditional moments vector for the class of interest rate models that allow for nonlinear volatility and flexible jump specifications. Such a characterization of continuous-time processes by the Itô Conditional Moment Generator noticeably enlarges the admissible set beyond the affine jump-diffusion class. A simple GMM estimator can be constructed based on the analytical solution to the lower order moments, with natural diagnostics of the conditional mean, variance, skewness, and kurtosis. Monte Carlo evidence suggests that the proposed estimator has desirable finite sample properties, relative to the asymptotically efficient MLE. The empirical application singles out the nonlinear quadratic variance as the key feature of the U.S. short rate dynamics.

Keywords: Itô Conditional Moment Generator, Short Term Interest Rate, Jump-Diffusion Process, Quadratic Variance, Generalized Method of Moments, Monte Carlo Study.
JEL classification: C51, C52, G12

1 Introduction

In modeling the short-term interest rate, researchers face the challenge of accommodating all relevant features in a single model specification. Those features, include but are not limited to, (1) short-term persistence, (2) long-run mean reversion, (3) nonlinear state-dependence in volatility, (4) non-Gaussian features in skewness and kurtosis. The celebrated CIR model (Cox et al., 1985) and its various extensions, although appealing in their general equilibrium nature and closed-form solution, have difficulty in fitting all these features simultaneously for the US interest rate data (Brown and Dybvig, 1986).[1] Rigorous specification tests using historical data tend to reject the square-root model (Aït-Sahalia, 1996b; Conley et al., 1997; Gallant and Tauchen, 1998). Although having an inherent advantage in fitting features (1) and (2) as indicated in the literature, the CIR-type model fails to capture the rich volatility dynamics and the nonlinear non-Gaussian features.

Consequently, efforts to modify the square-root model largely concentrate on more flexible specifications of the volatility dynamics. It is clear that the CIR model is just one special case of so-called linear CEV (constant elasticity of volatility) specification, where the elasticity equals one half. Recent comparative studies (Chan et al., 1992; Conley et al., 1997; Tauchen, 1997; Christoffersen and Diebold, 2000) found that an elasticity around one and one-half is more desirable. Alternatively, one can estimate the volatility function nonparametrically (Aït-Sahalia, 1996a; Stanton, 1997; Jiang and Knight, 1997; Jiang, 1998; Bandi, 2002; Bandi and Phillips, 2003). The empirical findings along this line suggest that the square-root model fits reasonably well for the medium range of interest rates, but the estimated nonlinearity at both the high and low ends is neither accurate nor conclusive. A pertinent approach is to introduce an unobserved stochastic volatility factor into the diffusion function, which finds considerable support in empirical studies (Andersen and Lund, 1996, 1997).[2] The jump-diffusion approach to interest-rate modeling (and bond pricing exercise) is of more recent

[1] The bivariate extensions of CIR specification (Gibbons and Ramaswamy, 1993; Chen and Scott, 1993; Pearson and Sun, 1994) also meet with poor empirical performance. Duffie and Singleton (1997) found favorable evidence for a two-factor CIR model with serially-correlated error structure. Dai and Singleton (2000) estimated more flexible three-factor affine specifications similar to Chen (1996) and Balduzzi et al. (1996) for interest rate swap data after 1987.

[2] This paper focuses on the maximum flexibility in the univariate setting, and the extension to multivariate or stochastic volatility is deferred to future research.

origin (Baz and Das, 1996; Das, 1998), and its general equilibrium formulation is explored by Ahn and Thompson (1988).[3]

The innovation of this paper is to generate the parametric conditional moments only using the Itô's formula and to construct a computationally efficient GMM estimator. Maximum Likelihood Estimation (MLE) is available only for a very restricted class of jump-diffusion models (Lo, 1988). Our method differs with the infinitesimal generator of Hansen and Scheinkman (1995) (GMM) in that it fully exploits the conditional information, does not rely on simulations as do Duffie and Singleton (1993) (SMM), uses model-dependent moments instead of data-dependent moments (Gallant and Tauchen, 1996) (EMM), generalizes to an arbitrary number of moments rather than only to conditional mean and variance (Fisher and Gilles, 1996) (QML), and has reliable small sample properties in comparison with the nonparametric approach (Aït-Sahalia, 1996a) (NP). As shown below, our method reduces a complicated task of solving a stochastic differential equation (SDE) to a simple matrix solution of an ordinary differential equation (ODE) system. The computational burden is reduced to a minimum of elementary algebra.[4] Another important advantage is that the characterization of short rate processes by the Itô's approach allows for nonlinear volatility and semiparametric jump specifications. Within the univariate paradigm, nonlinearity is indispensable to successful modeling of the U.S. short term interest rate. In the literature, the most closely related method is to identify the stochastic differential equations with an orthogonal series representation (Hansen et al., 1998), which is attributed to the generalized eigenvalue-eigenfunction technique (Wong, 1964).

We further justify the aforementioned methodology by a numerical exercise and illustrate by an empirical application. Monte Carlo evidence suggests that the finite sample efficiency of the proposed GMM estimator is comparable to the asymptotically efficient MLE, the

[3]Recently there is a growing literature on jump-diffusion interest rate modeling (see Chacko and Das, 1999; Johannes, 1999; Piazzesi, 2000, among many others), which ranges from short-rate dynamics to fixed-income derivatives, from market-implied jumps to macroeconomic announcements, and from parametric to nonparametric specifications.

[4]Alternatively, an equivalent spectral method of moments is developed by exploiting the closed-form conditional characteristic functions for the affine jump-diffusion model (Chacko and Viceira, 1999; Singleton, 2001; Jiang and Knight, 2002; Carrasco et al., 2002). However, the selection of spectral moments remains as a difficult challenge, whereas in the classical method of moments, a natural choice is the lower-order moments. Moreover, a strategy to derive moments using the Itô's formula alone and not relying on the characteristic function or moment generating function may be more desirable for certain non-standard processes, e.g., the quadratic variance model discussed in this paper.

sampling t-statistics of individual parameter is not far away from the normal reference distribution, and the GMM test of over-identifying restrictions has a typical upward bias but with reasonable magnitude. When applied to the U.S. short term interest rate from 1954 to 2002, both the square-root model and the restricted CEV model are rejected outright. Adding jumps shows some improvement but only the quadratic variance model cannot be rejected at the 1% significance level. U-shaped volatility and nonlinear higher order moments seem to be the main challenges of fitting the U.S. short rate dynamics, in additional to the well-known linear mean persistence.

The remainder of this paper is organized as follows: Section 2 derives the conditional moments for an admissible class of processes including the square-root, the restricted CEV, the jump-diffusion, and the quadratic variance; Section 3 builds an easy-to-implement GMM estimator and provides some finite sample evidence; Section 4 applies the estimating procedure to the four models mentioned above and contrasts the specification differences using the conditional moment profiles; and Section 5 concludes.

2 Itô Conditional Moment Generator

This section outlines a strategy to derive the conditional moments simultaneously for certain continuous-time processes, relying only on the Itô's formula and the specifications of drift, diffusion, and jump functions. The resulting characterization not only nests the popular affine jump-diffusion class, but also features nonlinear quadratic variance and semiparametric flexible jumps.

2.1 A General Characterization of Admissible Processes

Suppose that the evolution of the state variable (i.e., the short rate) is governed by a reduced-form jump-diffusion process

$$dr_t = \mu_t dt + \sigma_t dW_t + J_t dN(\rho_t t), \tag{1}$$

where W_t is a standard Brownian motion, $N(\rho_t t)$ is a Poisson driving process with an intensity function ρ_t, and J_t is the jump size with distribution $\Pi(J_t)$. Note that both the jump rate and jump size are allowed to be state-dependent but conditionally independent of each other and with respect to the Brownian motion. Process (1) must satisfy certain regularity conditions and the critical ones are: (a) both μ_t and σ_t are Lipschitz continuous, (b) ρ_t and $\Pi(J_t)$ are $\mathcal{F}_{t_-}$ measurable.

The strategy is to solve all the conditional moments up to the K'th order simultaneously, by first applying the Generalized Itô's lemma (Merton, 1971; Lo, 1988) to each r_T^k for $k = 1, 2, \cdots, K$, and then take the conditional expectation

$$E_t(r_T^k) = r_t^k + E_t\left[\int_t^T \left(\mu_u k r_u^{k-1} + \frac{1}{2}\sigma_u^2 k(k-1) r_u^{k-2} + \rho_u E_J[(r_u + J_u)^k - r_u^k]\right) du\right] \tag{2}$$

Interchanging the expectation and integration operators, and taking the derivative with respect to time T, we arrive at a differential equation system

$$\frac{dE_t(r_s^k)}{ds} = E_t\left[\mu_s k r_s^{k-1} + \frac{1}{2}\sigma_s^2 k(k-1) r_s^{k-2} + \rho_s \sum_{i=1}^{k} \binom{k}{i} r_s^{k-i} E_J(J_s^i)\right]. \tag{3}$$

with boundary condition $E_t(r_t^k) = r_t^k$. The following proposition characterizes the class of jump-diffusion processes that sustain a closed-form solution to equation (3),

Proposition 1 (Characterization) *The sufficient condition for the K-dimensional ordinary differential equation system (3) to have a first order linear solution, is to restrict the drift, diffusion, and jump functions in the following forms*

(1) $\mu_t = \kappa(\theta - r_t)$*; and*

(2) $\sigma_t = \sqrt{\sigma_0 + \sigma_1 r_t + \sigma_2 r_t^2}$*; and*

(3) $\rho_t E_J(J_t^k) = \sum_{j=0}^{k} J_{kj} r_t^j$.

Many linear or nonlinear restrictions need to be imposed to ensure existence and identification, for example, the sign constraints on $\kappa, \theta, \sigma_0, \sigma_1, \sigma_2$ and the zero constraints on some J_{kj}. The proof only involves a straightforward verification, hence omitted.[5]

For the admissible process under Proposition 1, the K-vector of its conditional moments $E_t(R_s) = [E_t(r_s), E_t(r_s^2), \cdots, E_t(r_s^K)]'$ is characterized by a linear differential equation system,

$$\frac{dE_t(R_s)}{ds} = A(\beta)E_t(R_s) + g(\beta), \tag{4}$$

[5]The necessary conditional for the K-dimensional ordinary differential equation system (3) to have a first order linear solution, is to require the term $\left[\mu_s k r_s^{k-1} + \frac{1}{2}\sigma_s^2 k(k-1) r_s^{k-2} + \rho_s \sum_{i=1}^{k} \binom{k}{i} r_s^{k-i} E_J(J_s^i)\right]$ to be a k'th order polynomial of r_s, which is trivial and not as informative as the sufficient condition.

where $A(\cdot)$ is a $K \times K$ lower-triangular matrix and $g(\cdot)$ is a $K \times 1$ vector. Both $A(\cdot)$ and $g(\cdot)$ are nonlinear functions of the parameter vector $\beta = [\kappa, \theta, \sigma_0, \sigma_1, \sigma_2, J_{10}, \cdots, J_{KK}]'$, defined by the underlying the jump-diffusion process (1). Since the coefficients of such a non-homogeneous linear first-order differential equation do not depend on time, one obtains the following closed-form solution,

$$E_t(R_T) = e^{(T-t)A(\beta)} R_t + A^{-1}(\beta) \left(e^{(T-t)A(\beta)} - I \right) g(\beta), \tag{5}$$

where I is the $K \times K$ identity matrix and $e^{(\cdot)}$ denotes the matrix exponential.

There are some advantages in using the "Itô Transformation" to generate the conditional moments. From the perspective of richer dynamics, although the drift function has to be restricted as linear, the diffusion function can be nonlinear, and the jump function only requires the specification of its moments. More detailed examples are examined in the next subsection to illustrate the enhanced flexibility of such an Itô characterization. From the perspective of easier implementation, the calculation of moments in a typical matrix programming language remains a one-line code as equation (5), and the computation of each entry of $A(\cdot)$ and $g(\cdot)$ in equation (3) does not require differentiation; whereas using the conditional moment generating function involves messy high order derivatives. Once computed, a moment-based estimator (like GMM) is readily available, while a likelihood-based method requires the Fourier inversion of the characteristic function. It is also possible to apply the Itô transformation to processes that lack analytical solution to the moment generating function. The major disadvantage of relying on a potentially limited set of moments, is the possible loss of estimation efficiency relative to MLE. To address this concern, the next section designs a GMM estimator and quantifies its adequate finite sample performance.

2.2 Leading Empirical Examples

To illustrate the applicability of the proposed methodology, here we present several specifications that are useful to model the short term interest rate. Only the solutions to the first four moments are spelled out, as the higher order moments are trivial extensions.

2.2.1 Flexible Jump-Diffusion Process

We start with a simple jump-diffusion process

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t + J_t dN(\rho_t t) \tag{6}$$

where $\rho_t = \rho$ and J_t is specified by its four moments. Although the diffusion part of this model is affine, the state variable may not be affine if the jump-size moments are state-dependent as in Proposition 1. The solution to its first four conditional moments in the form of equation (5), can be characterized by the matrix $A(\beta)$

$$\begin{bmatrix} -\kappa + \rho E(J) & 0 & 0 & 0 \\ 2\kappa\theta + \sigma^2 & -2\kappa + \rho\sum_{i=1}^{2}\binom{2}{i}E(J^i) & 0 & 0 \\ 0 & 3\kappa\theta + 3\sigma^2 & -3\kappa + \rho\sum_{i=1}^{3}\binom{3}{i}E(J^i) & 0 \\ 0 & 0 & 4\kappa\theta + 6\sigma^2 & -4\kappa + \rho\sum_{i=1}^{4}\binom{4}{i}E(J^i) \end{bmatrix}$$

and the vector $g(\beta)$

$$\begin{bmatrix} \kappa\theta \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

If we specialize to the case of uniform distribution ($J_t \sim \mathrm{U}[ar_t, br_t]$), the moments of the jump-size are, respectively, $E(J) = \frac{(b^2-a^2)}{2(b-a)}r_t$, $E(J^2) = \frac{(b^3-a^3)}{3(b-a)}r_t^2$, $E(J^3) = \frac{(b^4-a^4)}{4(b-a)}r_t^3$, and $E(J^4) = \frac{(b^5-a^5)}{5(b-a)}r_t^4$.

Two important points are worth noting here. First, the model is not affine as the conditional variance is not linear but quadratic in the state variable, which is qualitatively similar to the quadratic variance diffusion model discussed next. Second, the particular state-dependence of the jump-size rules out the possibility of negative interest rate, under the mild restriction that $-1 \leq a < b < +\infty$. Negative short rate level is difficult to dealt with for certain affine specifications and is conceptually problematic in a nominal economic environment.

2.2.2 Quadratic Variance Diffusion Model

An important alternative to the affine variance model is the "quadratic variance" process defined as

$$dr_t = \kappa(\theta - r_t)dt + \sqrt{\sigma_0^2 - \sigma_1^2 r_t + \sigma_2^2 r_t^2}dW_t \tag{7}$$

No sign restrictions are imposed in the GMM estimation procedure, but are adopted here in line with the actual result to highlight some nice properties—non-zero volatility when rate approaches zero, high volatility when rate is high, and comparable scale of the local variance parameter (as in the square-root model).

For this quadratic variance model, the conditional moments are characterized by the equation (5) in terms of

$$A(\beta) = \begin{bmatrix} -\kappa & 0 & 0 & 0 \\ 2\kappa\theta - \sigma_1^2 & -2\kappa + \sigma_2^2 & 0 & 0 \\ 3\sigma_0^2 & 3\kappa\theta - 3\sigma_1^2 & -3\kappa + 3\sigma_2^2 & 0 \\ 0 & 6\sigma_0^2 & 4\kappa\theta - 6\sigma_1^2 & -4\kappa + 6\sigma_2^2 \end{bmatrix}$$

and

$$g(\beta) = \begin{bmatrix} \kappa\theta \\ \sigma_0^2 \\ 0 \\ 0 \end{bmatrix}$$

Note that the solution structure is similar for both the jump-diffusion and the quadratic variance models, and that the only difference is in each entry. This feature makes the numerical calculation of the moments straightforward and fast.

The quadratic variance model has several important advantages. First, the model is not affine hence its moment generating function or characteristic function may not be easy to derive. Then the Itô conditional moment generator may be the only choice among all the non-simulation-based methods to calculate the moments. Second, there is a great deal of debate about whether the volatility is linear or nonlinear, e.g., the "U" shaped volatility pattern reported by Aït-Sahalia (1996a). Here we can provide a simple parametric nonlinear alternative and a feasible GMM estimator with conditional moment based diagnostics. Third, the quadratic variance specification seems to nest several famous short rate models, namely, log-linear ($\sigma_0 = \sigma_1 = 0$), Ornstein-Uhlenbeck ($\sigma_1 = \sigma_2 = 0$), and square-root ($\sigma_0 = \sigma_2 = 0$ and reversing the sign of σ_1^2). Of course, the obvious disadvantage is that the bond pricing solution is not easily obtained except for using Monte Carlo simulation. Nevertheless, the empirical evidence of Section 4 seems to suggest that the quadratic variance function is indispensable in modeling the univariate short rate dynamics.

2.2.3 Cubic or Transformable CEV Model

Some models are not directly solvable by the Itô conditional moment generator, but can be "reduced" to the tractable cases by appropriate transformations. For a detailed discussion on the reducibility technique, see Chapter 4 of Kloeden and Platen (1992). Consider the following nonlinear drift and constant-elasticity-of-volatility (CEV) specification

$$dr_t = \kappa(\theta r_t^{2\gamma-1} - r_t)dt + \sigma r_t^{\gamma} dW_t, \tag{8}$$

where under appropriate parameter restrictions $r_t \in (0, +\infty)$ and has a positive starting value. Note that the cross restriction on parameter γ between drift and diffusion is required for the reducibility, and may prove to be empirically too restrictive relative to the standard linear drift CEV model. Marsh and Rosenfeld (1983) first proposed such a modeling strategy and estimated with maximum likelihood for distinct values of $\gamma = 0, 0.5, 1$. Eom (1997) studied the distributional properties and the optimal GMM instruments for $\gamma \in [0, 1)$. Ahn and Gao (1999) examined the term structure implications for the case of $\gamma = 1.5$ and estimated with GMM. The adopted GMM estimators were based on time discretization and approximate first and second moments.

Using the transformation $x_t = r_t^{\alpha}$, which is a *state-preserving* transformation when $r_t \in (0, +\infty)$ and $\gamma \in [0, 1)$ or $\gamma \in (1, +\infty)$, one arrives at the familiar square-root model[6]

$$dx_t = a(b - x_t)dt + c\sqrt{x_t}dW_t. \tag{9}$$

The above transformation can be characterized by the following proposition

Proposition 2 (Transformation) *The mappings between the CEV process (8) and the square-root model (9) are*

$$\begin{aligned} \alpha &= 2(1-\gamma) \\ a &= 2(1-\gamma)\kappa \\ b &= \theta + \tfrac{(1-2\gamma)\sigma^2}{2\kappa} \\ c &= 2(1-\gamma)\sigma \end{aligned} \tag{10}$$

The proof is a straightforward application of the Itô's lemma, and is available from the author upon request. The solution to conditional moment of the transformed process x_t is a special

[6]When $\gamma = 1$ the square-root process (8) is reduced to the log-normal process $dr_t = \kappa(\theta - 1)r_t dt + \sigma r_t dW_t$, and the parameters κ and θ are not separately identifiable.

case of the jump-diffusion process (6) without jumps (letting $\rho_t = 0$ would be sufficient). The fourth parameter γ in the nonlinear drift CEV model (8) is identified through the nonlinear but monotonic transformation $x_t = r_t^{\alpha}$, given that $r_t \in (0, +\infty)$.

3 Estimation Strategy and Monte Carlo Evidence

Deriving the conditional moment restrictions (5) only achieves half of the task for estimating the underlying continuous-time model. The other half rests on designing an appropriate estimator with desirable large and small sample properties. The purpose of the this section is to outline an easy-to-implement GMM estimator based on the moment condition solution (5), and to assure the readers that the estimator performs reasonably well for the benchmark CIR model under empirically plausible scenarios.

3.1 The GMM Estimator

The condition moments solution (5) can be spelled out as a vector-auto-regressive (VAR) formula

$$E_t[h_{t+1}(\beta)] = \begin{bmatrix} E_t(r_{t+1}) \\ E_t(r_{t+1}^2) \\ E_t(r_{t+1}^3) \\ E_t(r_{t+1}^4) \end{bmatrix} - \begin{bmatrix} d_{11} & 0 & 0 & 0 \\ d_{21} & d_{22} & 0 & 0 \\ d_{31} & d_{32} & d_{33} & 0 \\ d_{41} & d_{42} & d_{43} & d_{44} \end{bmatrix} \begin{bmatrix} r_t \\ r_t^2 \\ r_t^3 \\ r_t^4 \end{bmatrix} - \begin{bmatrix} d_{01} \\ d_{02} \\ d_{03} \\ d_{04} \end{bmatrix} = 0 \qquad (11)$$

which is a recursive simultaneous equation system and its unrestricted version can be estimated by the ordinary least square (OLS). To form a generalized method of moments (GMM) estimator, a natural choice of instruments is the constant one and the lagged variables, hence the moment condition vector (with a total of fourteen equations)

$$f_t(\beta) \equiv \begin{bmatrix} (E(r_{t+1}) - r_{t+1})(1, r_t)' \\ (E(r_{t+1}^2) - r_{t+1}^2)(1, r_t, r_t^2)' \\ (E(r_{t+1}^3) - r_{t+1}^3)(1, r_t, r_t^2, r_t^3)' \\ (E(r_{t+1}^4) - r_{t+1}^4)(1, r_t, r_t^2, r_t^3, r_t^4)' \end{bmatrix} \qquad (12)$$

By construction $E[f_t(\beta_0)] = 0$, and the corresponding GMM or minimum chi-square estimator is defined by $\hat{\beta}_T = \arg\min g_T(\beta)' W g_T(\beta)$, where $g_T(\beta)$ refers to the sample mean of the moment conditions, $g_T(\beta) \equiv 1/T \sum_{t=1}^{T-1} f_t(\beta)$, and W denotes the asymptotic covariance matrix of $g_T(\beta_0)$ (Hansen, 1982). An iterative estimator of W is adopted here; and since the error is not serially correlated, only the heteroscedasticity need to be accounted for. Under standard regularity conditions, the minimized value of the objective function (normalized by the sample size) is asymptotically distributed a chi-square random variable, which allows for an omnibus test of the overidentifying restrictions. Moreover inference regarding individual parameters is readily available from the standard formula of the asymptotic variance-covariance matrix, $(\partial f_t(\beta)/\partial\beta' W \partial f_t(\beta)/\partial\beta)/T$.

3.2 Considerations for Identification and Efficiency

Identification, or global identification, is equivalent to the assumption that the GMM estimator achieves a unique minimum at some $\beta_0 \in \mathcal{B}$, where $\mathcal{B}$ is a compact set. In the unrestricted recursive VAR model (11), the total number of identifiable parameters is fourteen, which can be easily verified by the standard *order* and *rank* conditions. Since the underlying jump-diffusion model is nonlinear, the identification issue becomes more complicated—on the one hand, the restricted nonlinear dynamics may not be able to identify as many as fourteen parameters; on the other hand a nonlinear structure usually helps to identity more parameters than a linear structure. There is not much theoretical guidance in literature on how to verify the identification condition in a nonlinear model before the model is actually estimated. However, there is a sufficient condition—$\text{plim}(\partial g_T/\partial\beta' W_T \partial g_T/\partial\beta)/T$ being nonsingular—that can be numerically verified with the estimation result from a given sample data set. It is equivalent to the more primitive condition of local identification that the gradient is of full column rank and the Hessian is negative definite. In practice, all the empirical examples seem not to violate this sufficient condition, except a variation of the jump-diffusion model where both the jump-rate and jump-size parameters are state-independent constants.

Following Hansen (1985) and Hansen et al. (1988), the conditional moment restriction $E_t[h_{t+1}(\beta)] = 0$ indicated by equation (11) implies an efficient choice of instruments as $E_t[\partial h_{t+1}(\beta_0)/\partial\beta] Var_t[h_{t+1}(\beta_0)]^{-1}$. In theory such a choice of instruments should be ideal, but in practice other considerations may favor the natural choice of (12). First, the optimal instruments involve unknown true distribution parameter β_0, which has to be approximated in the GMM estimation procedure. Second, to calculate the optimal instruments one needs to solve for eight lower order moments if one uses only four lower order moments in the estimation, which is trivial analytically using the Itô approach but may be numerically unstable for real world data sets. Further, there is a logical inconsistency—one has the knowledge of eight order moments but does not use it in the moment condition restriction. Meddahi and Renault (1997) proposed an interesting treatment that reduces the information of the third and fourth conditional moments to the unconditional skewness and kurtosis, and achieves the efficient estimates of conditional mean and variance. The GMM estimator implemented here explicitly incorporates the conditional third and fourth moments and is conceptually related to their efficient estimation of the first two conditional moments. The relative efficiency of the proposed GMM estimator can be judged in a Monte Carlo setting, against the asymptotically efficient MLE (which is theoretically superior to GMM for a given set of moment restrictions with optimal instruments).

3.3 Monte Carlo Evidence

To assess the finite sample performance of the proposed GMM estimator, a limited Monte Carlo study is conducted for the benchmark square-root model $dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t$, in comparisons with the MLE estimates reported by Durham and Gallant (2002). There are six scenarios chosen in their paper, with varying degrees of persistence and volatility, and a fixed long-run mean of 6 percent. I adopted the exact same setup with 1000 observations in each random sample and a total of 512 Monte Carlo replications. To avoid the discretization bias, I simulate the square-root model from the exact non-central Chi-square distribution

$$f(r_{t+\Delta}|r_t;\kappa,\theta,\sigma) = ce^{-u-v}\left(\frac{v}{u}\right)^{\frac{q}{2}} I_q\left(2\sqrt{uv}\right), \tag{13}$$

where $q = 2\kappa\theta/\sigma^2 - 1$, $c = 2\kappa/\sigma^2(1 - e^{-\kappa\Delta})$, $u = cr_t e^{-\kappa\Delta}$, $v = cr_{t+\Delta}$, and $I_q(\cdot)$ is a modified Bessel function of the first kind with a fractional order q (Oliver, 1972). A *composite method* of generating random number (Devroye, 1986) is adopted here after transforming the above density function into,

$$f(y) = \sum_{j=0}^{\infty} \frac{y^{j+\lambda-1}e^{-y}}{\Gamma(j+\lambda)} \cdot \frac{u^j e^{-u}}{j!} = \sum_{j=0}^{\infty} \text{Gamma}(y|j+\lambda, 1) \cdot \text{Poisson}(j|u) \tag{14}$$

with $y = v$ and $\lambda = q + 1$. In practice, one first draws a random number j from the $Poisson(j|u)$ distribution; then draws another random number y from the $Gamma(y|j+\lambda, 1)$

distribution; and finally calculates the target state variable $r_{t+\Delta} = y/c$. See Zhou (2001) for implementation detail.

Table 1 compares the parameter estimates of the proposed GMM estimator in this paper with those of the MLE estimator, under the six scenarios (a-f) in Durham and Gallant (2002). In terms of bias, only the mean-reversion parameter κ has a sizable upward bias when the persistence level is high (scenario a, b, and d)—about 10% of the parameter value for MLE and about 20% for GMM; while for the less persistent scenarios (c, e, and f), the bias is noticeably reduced. This is a classical case of finite sample bias in estimating the AR(1) coefficient for near-unit-root processes. For the long-run mean parameter θ and the local variance parameter σ, MLE has negligible positive bias and GMM has negligible negative biases. In terms of relative efficiency, the proposed GMM estimator is remarkably close to the asymptotically efficient MLE. The root-mean-squared-error of GMM is no more 10% larger than that of MLE for most parameters in the persistent cases (scenarios a, b, and d), and is practically indistinguishable for most parameters in the less persistent cases (scenarios c, e, and f). Figure 1 reports the GMM test of the overidentifying restrictions, which exhibits a typical over-rejection bias but with a reasonable size comparing with the reference level. The sampling distribution of the t-test statistics is graphed in Figure 2, and indicates that the finite sample distortion is rather small comparing with the reference Normal(0,1) distribution.

4 Empirical Application

In this section, the Itô moment generator and the related GMM estimator are applied to the empirical U.S. interest rate data. The weekly 3-month t-bill rate from January 1954 to July 2002, totaling 2504 observations, is obtained from the Federal Reserve Bank of St. Louis public website. The time series plot is given in Figure 3 and the summary statistics are reported in Table 2. The short rate exhibits the typical features found in literature—high persistence (auto-regressive coefficients close to one), high volatility (standard deviation 277 basis points), moderately high skewness (1.14) and kurtosis (4.87). I will focus on the estimation result of the four empirical examples (including the benchmark CIR model) presented in Section 2, and illustrate how to use the conditional moment functions to further compare different model specifications.

4.1 Estimation Result

The GMM estimator designed in Section 3 is applied to the four candidate models discussed in Section 2: square-root, restricted CEV, jump-diffusion, and quadratic variance.[7] The results are summarized in Table 3.

The standard square-root model is strongly rejected by the GMM specification test, with a chi-square (df=11) of 49.36. The long-run mean parameter (0.0497) is about 50 basis points lower than the sample average (0.0542), the mean-reversion parameter is very low (0.0020) and imprecisely estimated with standard error 0.0013, and the local variance parameter is also lower (0.0062 which implies a unconditional standard deviation 0.0219 versus the sample standard deviation 0.0277).

The nonlinear drift CEV model is also strongly rejected with a chi-square (df=10) of 48.57. Although most parameters are accurately estimated, the model also has difficulty in nailing down the mean reversion parameter κ (0.0017 with standard error 0.0010). The restricted CEV model accurately estimates elasticity parameter as 0.4825 with standard error 0.0028, which confirms the empirical finding by Eom (1997). All other parameter estimates are close to and/or slightly lower than the square-root estimates.[8]

The jump-diffusion model is implemented here with a constant jump-rate ρ and a uniform jump-size $(-ar_t, ar_t)$. The symmetry restriction on jump size is to ensure identification. The result predicts roughly two jumps per year; with a state-dependent jump-size of plus or minus 119 basis points at the sample average (0.0542), plus or minus 13 basis points at sample minimum (0.0058), and plus or minus 368 basis points at sample maximum (0.1676). Such a jump pattern is more realistic than the constant jump-size distribution, and can rule

[7] As pointed out by a referee, one could estimate a comprehensive model nesting both time-varying jumps and conditional quadratic variance. I found out that such a specification is not empirically identifiable by the GMM estimator. My intuition is that the particular jump and diffusion specifications adopted here produce the similar quadratic conditional variance. Therefore they are substituting for each other instead of being complementary. This can be easily seen from the diagnostic conditional moment graphs in the next subsection.

[8] This result differs from the typical empirical finding for the linear drift CEV model, in that the elasticity coefficient there is found to be in the range of 1.0-1.5 (Chan et al., 1992; Conley et al., 1997; Tauchen, 1997; Christoffersen and Diebold, 2000), possibly because that the nonlinear drift CEV model imposes an unrealistic restriction across the drift and diffusion functions.

out the negative interest rates, which is quite troublesome in a nominal economic environment. Nevertheless, the model is rejected at the p-value 0.0002, and the parameter θ is unconvincingly large (0.0995).

The quadratic variance model performs the best, and is not rejected at the 1% significance level (p-value 0.0121). All the parameter estimates are highly significant. The estimates of the drift parameters fall between the square-root model (similarly the CEV model) and the jump-diffusion model. The parameter estimates of the diffusion function guarantee that (a) instant variance does not admit negative value, (b) minimum volatility is achieved at a positive short rate level, and (c) volatility increases more when short rate level is high than when short rate level is low (see the conditional moment graphs bellow). Such a result from a parametric perspective seems to be confirm the finding of Aït-Sahalia (1996a) from a nonparametric perspective.

4.2 Conditional Moment Graphs

The conditional moment vector (11) not only serves as the basis for constructing a GMM estimator, but also provides intuitive diagnostics in conditional mean, volatility, skewness, and kurtosis. The conditional mean and conditional variance in discrete sampling intervals, are equivalent to the drift and volatility functions in instant times for the pure diffusion processes, but more general in covering also the jump-diffusion processes. The conditional skewness and kurtosis provide natural assessment on how much the implied transitional density deviates from the conditional normality. Higher order conditional moments are especially informative about the jump impact when the time horizon is longer than zero, but the instantaneous higher order moments cannot provide any new information than the instantaneous drift and volatility.

Figure 4 plots the conditional mean (top panel) and conditional variance (bottom panel). It is clear that the square-root model has the least persistence in level. Although the restricted CEV model has a potential nonlinear drift, the estimated mean function is mostly linear and close to the square-root model. On the other hand, the jump-diffusion model is the most persistent case, suggesting an observational equivalence between occasional jumps and near unit-root in interest rate processes. Our preferred quadratic variance model has a linear mean function with a moderate persistence among the four models. Turning to the conditional volatility, both the square-root model and the restricted CEV model produce nearly identical linear volatility profiles, underpinning the clear rejection by the GMM specification tests. Jump-diffusion process provides a slightly nonlinear quadratic variance function, due to the state-dependent jump-size specification ($J_t \sim U[-ar_t, ar_t]$) that differs from the standard affine jump-diffusion models. Of course the most dramatic result comes from the U-shaped quadratic variance model, which partially confirms the nonparametric finding of the nonlinear volatility by Aït-Sahalia (1996a) and the parametric finding of CEV elasticity between 1.0 and 1.5 (Chan et al., 1992; Conley et al., 1997; Tauchen, 1997; Christoffersen and Diebold, 2000). The post-war U.S. history suggests that the interest rate volatility is certainly high when the short rate level is high, but the volatility is also none trivial when the rate is close to zero. Therefore a nonlinear dependence of short rate volatility on its level may be better captured by a quadratic variance model than by a standard affine model.

Figure 5 depicts the conditional skewness and kurtosis functions and offers some assessment of the departures from the conditional normality. From the top panel we can see that both the square-root and the nonlinear CEV model give a virtually same hyperbolic skewness function—shooting up at the lower end and approaching zero at the higher end. The jump-diffusion process has a similar profile but a uniformly higher skewness once the short rate level reaches above 2 percent. The quadratic volatility model is unique in presenting a nonlinear increasing skewness function that approaches -0.1 at the low end and +0.1 at the high end. Turning to the bottom panel, again, both the square-root and the nonlinear CEV models give a virtually same hyperbolic kurtosis function—shooting up at the lower end and approaching three at the higher end. Note that the jump-diffusion model gives an extraordinarily high kurtosis, ranging from 9 at the lower end to 44 at the higher end (outside and above the picture range). Usually introducing jumps helps to increase the model skewness and kurtosis, but an unusually high kurtosis of 9-44 must be caused by the restrictive jump specification (constant jump rate $\rho_t = \rho$ and uniform jump size $J_t \sim U[-ar_t, ar_t]$), which is imposed to identify all the model parameters. The preferred quadratic model has a nonlinear V-shaped kurtosis function for the short rate level between 0 and 8 percent and then mostly a constant. In short, the quadratic variance model produces unique nonlinear conditional skewness and kurtosis, which are dramatically different from all other candidate models.

5 Conclusion

This paper proposes an Itô's approach to generate the conditional moments for continuous time Markov processes and gives a characterization of the class of admissible models. The resulting conditional moment vector forms the basis of a natural GMM estimator. Monte Carlo evidence suggests that such a moment generator and the related estimator behave reasonably well for a benchmark square-root model. When applied to the empirical U.S. short rate data, the procedure singles out the quadratic variance model as the only unrejected specification at the one percent level. The benchmark square-root model, the state-dependent jump-diffusion process, and the nonlinear drift CEV model all fail in the GMM tests of overidentifying restrictions. Further diagnostics suggests that the U-shaped conditional variance and non trivial conditional skewness and kurtosis are important in modeling the short rate dynamics in the univariate setting. One important extension is to estimate a multivariate asset return model with possible quadratic volatility components.

References

Ahn, Chang Mo and Howard E. Thompson (1988), "Jump-Diffusion Process and the Term Structure of Interest Rates," *Journal of Finance*, vol. 43, 155–174.

Ahn, Dong-Hyun and Bin Gao (1999), "A Parametric Nonlinear Model of Term Structure Dynamics," *Review of Financial Studies*, vol. 12, 721–762.

Aït-Sahalia, Yacine (1996a), "Nonparametric Pricing of Interest Rate Derivatives," *Econometrica*, vol. 64, 527–560.

Aït-Sahalia, Yacine (1996b), "Testing Continuous-Time Models of the Spot Interest Rate," *The Review of Financial Studies*, vol. 9, 385–426.

Andersen, Torben G. and Jesper Lund (1996), "Stochastic Volatility and Mean Drift in the Short Rate Diffusion: Source of Steepness, Level, and Curvature of the Yield Curve," *Working Paper*, Kellogg School of Management, Northwestern University.

Andersen, Torben G. and Jesper Lund (1997), "Estimating Continuous Time Stochastic Volatility Models of the Short Term Interest Rate," *Journal of Econometrics*, vol. 77, 343–378.

Balduzzi, Pierluigi, Sanjiv R. Das, Silverio Foresi, and Rangarajan K. Sundaram (1996), "A Simple Approach to Three Factor Affine Term Structure Models," *Journal of Fixed Income*, vol. 6, 43–53.

Bandi, Federico M. (2002), "Short-term Interest Rate Dynamics: A Spatial Approach," *Journal of Financial Economics*, vol. 65, 73–110.

Bandi, Federico M. and Peter C. B. Phillips (2003), "Fully Nonparametric Estimation of Scalar Diffusion Models," *Econometrica*, vol. 71.

Baz, Jamil and Sanjiv Ranjan Das (1996), "Analytical Approximation of the Term Structure for Jump-Diffusion Process: A Numerical Analysis," *Journal of Fixed Income*, vol. 6, 78–86.

Bollerslev, Tim and Hao Zhou (2002), "Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility," *Journal of Econometrics*, vol. 109, 33–65.

Brown, Stephen J. and Philip H. Dybvig (1986), "The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates," *Journal of Finance*, vol. 41, 617–630.

Carrasco, Marine, Mikhail Chernov, Jean-Pierre Florens, and Eric Ghysels (2002), "Efficient Estimation of Jump Diffusions and General Dynamic Models with a Continuum of Moment Conditions," *Working Paper*, University of North Carolina.

Chacko, George and Sanjiv Das (1999), "Pricing Interest Rate Derivatives: A General Approach," *Working Paper*, Graduate School of Business, Harvard University.

Chacko, George and Luis M. Viceira (1999), "Spectral GMM Estimation of Continuous-Time Processes," *Working Paper*, Graduate School of Business, Harvard University.

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

Chen, Lin (1996), "Stochastic Mean and Stochastic Volatility - A Three-Factor Model of the Term Structure of Interest Rates and Its Applications in Derivatives Pricing and Risk Management," *Financial Markets, Institution and Instruments*, vol. 5, 1–88.

Chen, Ren-Raw and Louis Scott (1993), "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates," *Journal of Fixed Income*, vol. 3, 14–31.

Christoffersen, Peter F. and Francis X. Diebold (2000), "How Relevant Is Volatility Forecasting for Financial Risk Management?" *Review of Economics and Statistics*, vol. 82, 12–22.

Conley, Tim, Lars Peter Hansen, Erzo Luttmer, and Jose Scheinkman (1997), "Short Term Interest Rates as Subordinated Diffusions," *Review of Financial Studies*, vol. 10, 525–578.

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross (1985), "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, vol. 53, 363–384.

Dai, Qiang and Kenneth J. Singleton (2000), "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, vol. 55, 1943–1978.

Das, Sanjiv Ranjan (1998), "Poisson-Gaussian Process and the Bond Markets," *Working Paper*, Harvard Business School.

Devroye, Luc (1986), *Non-Uniform Random Variate Generation*, Spinger-Verlag.

Duffie, Darrell and Kenneth Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, vol. 61, 929–952.

Duffie, Darrell and Kenneth Singleton (1997), "An Econometric Model of the Term Structure of Interest-Rate Swap Yields," *Journal of Finance*, vol. 52, 1287–1321.

Durham, Garland B. and A. Ronald Gallant (2002), "Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes," *Journal Business and Economic Statistics*, vol. 20.

Eom, Young Ho (1997), "An Efficient GMM Estimation of Continuous-Tim Asset Dynamics: Implications for the Term Structure of Interest Rates," *Working Paper*, Federal Reserve Bank of New York.

Fisher, Mark and Christian Gilles (1996), "Estimating Exponential Affine Models of the Term Structure," *Working Paper*, Finance and Economics Discussion Series, Federal Reserve Board.

Gallant, A. Ronald and George Tauchen (1996), "Which Moment to Match?" *Econometric Theory*, vol. 12, 657–681.

Gallant, A. Ronald and George Tauchen (1998), "Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions," *Journal of the American Statistical Association*, vol. 93, 10–24.

Gibbons, Michael R. and Krishna Ramaswamy (1993), "A Test of the Cox, Ingersoll, and Ross Model of the Term Structure," *Review of Financial Studies*, vol. 6, 619–658.

Hansen, Lars Peter (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, vol. 50, 1029–1054.

Hansen, Lars Peter (1985), "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics*, vol. 30, 203–238.

Hansen, Lars Peter, John C. Heaton, and Masao Ogaki (1988), "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *Journal of the American Statistical Association*, vol. 83, 863–871.

Hansen, Lars Peter and José Alexandre Scheinkman (1995), "Back to the Future: Generalized Moment Implications for Continuous Time Markov Process," *Econometrica*, vol. 63, 767–804.

Hansen, Lars Peter, José Alexandre Scheinkman, and Nizar Touzi (1998), "Spectral Methods for Identifying Scalar Diffusions," *Journal of Econometrics*, vol. 86, 1–32.

Jiang, George and John Knight (2002), "Estimation of Continuous Time Stochastic Processes via the Empirical Characteristic Function," *Journal of Business and Economic Statistics*, vol. 20, 198–212.

Jiang, George J. (1998), "Nonparametric Modeling of U.S. Interest Rate Term Structure Dynamics and Implications on the Prices of Derivative Securities," *Journal of Financial and Quantitative Analysis*, vol. 33, 465–497.

Jiang, George J. and John Knight (1997), "A Nonparametric Approach to the Estimation of Diffusion Processes, with an Application to a Short-Term Interest Rate Model," *Econometric Theory*, pages 615–645.

Johannes, Michael S. (1999), "Jumps in Interest Rates: A Nonparametric Approach," *Working Paper*, Department of Economics, University of Chicago.

Kloeden, Perter E. and Eckhard Platen (1992), *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics, Springer-Verlag.

Lo, Andrew W. (1988), "Maximum Likelihood Estimation of Generalized Itô Process with Discretely Sampled Data," *Econometric Theory*, vol. 4, 231–247.

Marsh, Terry A. and Eric R. Rosenfeld (1983), "Stochastic Processes for Interest Rates and Equilibrium Bond Prices," *Journal of Finance*, vol. 38, 635–646.

Meddahi, Nour and Éric Renault (1997), "Quadratic M-Estimators for ARCH-Type Processes," *Working Paper*, CRDE, Université de Montréal.

Merton, Robert C. (1971), "Optimum Consumption and Portfolio Rules in a Continuous Time Model," *Journal of Economic Theory*, vol. 3, 373–413.

Oliver, F. W. J. (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, John Wiley & Sons.

Pearson, Neil D. and Tong-Sheng Sun (1994), "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model," *Journal of Finance*, vol. 49, 1279–1304.

Piazzesi, Monika (2000), "An Econometric Model of the Yield Curve with Macroeconomic Jump Effects," *Working Paper*, Graduate School of Business, Stanford University.

Pritsker, Matt G. (1998), "Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models," *Review of Financial Studies*, vol. 11, 449–487.

Singleton, Kenneth (2001), "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, forthcoming.

Stanton, Richard (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *The Journal of Finance*, vol. 52, 1973–2002.

Tauchen, George (1997), "New Minimum Chi-Square Methods in Empirical Finance," in "Advances in Econometrics, Seventh World Congress," (edited by Kreps, D. and K. Wallis), Cambridge University Press, Cambridge UK.

Wong, Eugene (1964), "The Construction of a Class of Stationary Markoff Processes," in "Sixteenth Symposium in Applied Mathematics - Stochastic Process in Mathematical Physics and Engineering," (edited by Belleman, R.), pages 264–276, Providence, RI.

Zhou, Hao (2001), "Finite Sample Properties of EMM, GMM, QMLE, and MLE for a Square-Root Interest Rate Diffusion Model," *Journal of Computational Finance*, vol. 5, 89–122.

Table 1: Monte Carlo Experiment

This table compares the finite sample performance of the GMM estimator proposed in this paper with that of the MLE estimator provided by Durham and Gallant (2002). Δ stands for the discrete sampling interval and df for the degree of freedom of the implied non-central chi-square distribution. The random sample size is chosen as 1000 and the number of Monte Carlo replicates is 512. Here we report the mean bias and root-mean-squared-error.

True Value	MLE Mean Bias	GMM Mean Bias	MLE Root-MSE	GMM Root-MSE
Scenario (a), $\Delta = 1/12$, $df = 5.33$				
$\kappa = 0.50$	0.0489	0.0828	0.1344	0.1473
$\theta = 0.06$	0.0006	-0.0027	0.0080	0.0086
$\sigma = 0.15$	0.0002	-0.0028	0.0034	0.0046
Scenario (b), $\Delta = 1/12$, $df = 2.48$				
$\kappa = 0.50$	0.0597	0.1036	0.1413	0.1697
$\theta = 0.06$	-0.0003	-0.0052	0.0114	0.0118
$\sigma = 0.22$	0.0001	-0.0045	0.0054	0.0075
Scenario (c), $\Delta = 1/12$, $df = 133.33$				
$\kappa = 0.50$	0.0438	-0.0042	0.1299	0.1221
$\theta = 0.06$	0.0001	-0.0000	0.0016	0.0017
$\sigma = 0.03$	0.0002	-0.0003	0.0007	0.0008
Scenario (d), $\Delta = 1/12$, $df = 4.27$				
$\kappa = 0.40$	0.0458	0.0892	0.1210	0.1446
$\theta = 0.06$	0.0008	-0.0046	0.0102	0.0102
$\sigma = 0.15$	0.0001	-0.0029	0.0035	0.0048
Scenario (e), $\Delta = 1/12$, $df = 53.33$				
$\kappa = 5.00$	0.0151	0.0169	0.4630	0.4580
$\theta = 0.06$	0.0001	-0.0002	0.0008	0.0009
$\sigma = 0.15$	0.0000	-0.0040	0.0043	0.0059
Scenario (f), $\Delta = 2$, $df = 53.33$				
$\kappa = 0.50$	0.0013	0.0283	0.0430	0.0483
$\theta = 0.06$	0.0004	-0.0022	0.0018	0.0029
$\sigma = 0.15$	0.0004	-0.0041	0.0056	0.0066

Table 2: Summary Statistics of Three Month T-Bill Rates

The following table summarizes the weekly U.S. 3-month t-bill rates from January 1954 to July 2002 with a total of 2504 observations. The data is obtained from the public website of the Federal Reserve Bank of St. Louis.

Moments and Quantiles		j^{th} Order Autocorrelations	
Mean	0.0542	ρ_0	1.0000
Std. Dev.	0.0277	ρ_1	0.9964
Skewness	1.1414	ρ_2	0.9912
Kurtosis	4.8728	ρ_3	0.9856
Minimum	0.0058	ρ_4	0.9798
5%-qntl.	0.0171	ρ_5	0.9734
25%-qntl.	0.0347	ρ_6	0.9667
Medium	0.0504	ρ_7	0.9600
75%-qntl.	0.0689	ρ_8	0.9537
95%-qntl.	0.1045	ρ_9	0.9477
Maximum	0.1676	ρ_{10}	0.9418

Table 3: Empirical Estimation Results

This table presents the main empirical results of the four model specifications discussed in Section 2 and estimated by the GMM estimator outlined in Section 3.

	Square-Root	Nonlinear CEV	Jump-Diffusion	Quadratic Variance
$\kappa =$	0.0020	0.0017	0.0005	0.0010
	(0.0013)	(0.0010)	(0.0001)	(0.0002)
$\theta =$	0.0497	0.0458	0.0995	0.0669
	(0.0131)	(0.0128)	(0.0186)	(0.0016)
$\sigma =$	0.0062	0.0059	0.0031	
	(0.0002)	(0.0002)	(0.0012)	
$\gamma =$		0.4825		
		(0.0028)		
$\rho =$			0.0381	
			(0.0025)	
$a =$			0.2196	
			(0.0265)	
$\sigma_0 =$				0.0015
				(0.0002)
$\sigma_1 =$				0.0097
				(0.0005)
$\sigma_2 =$				0.0412
				(0.0006)
Chi-Square =	49.3617	48.5714	31.6703	21.1222
d.o.f =	11	10	9	9
p-value =	0.0000	0.0000	0.0002	0.0121

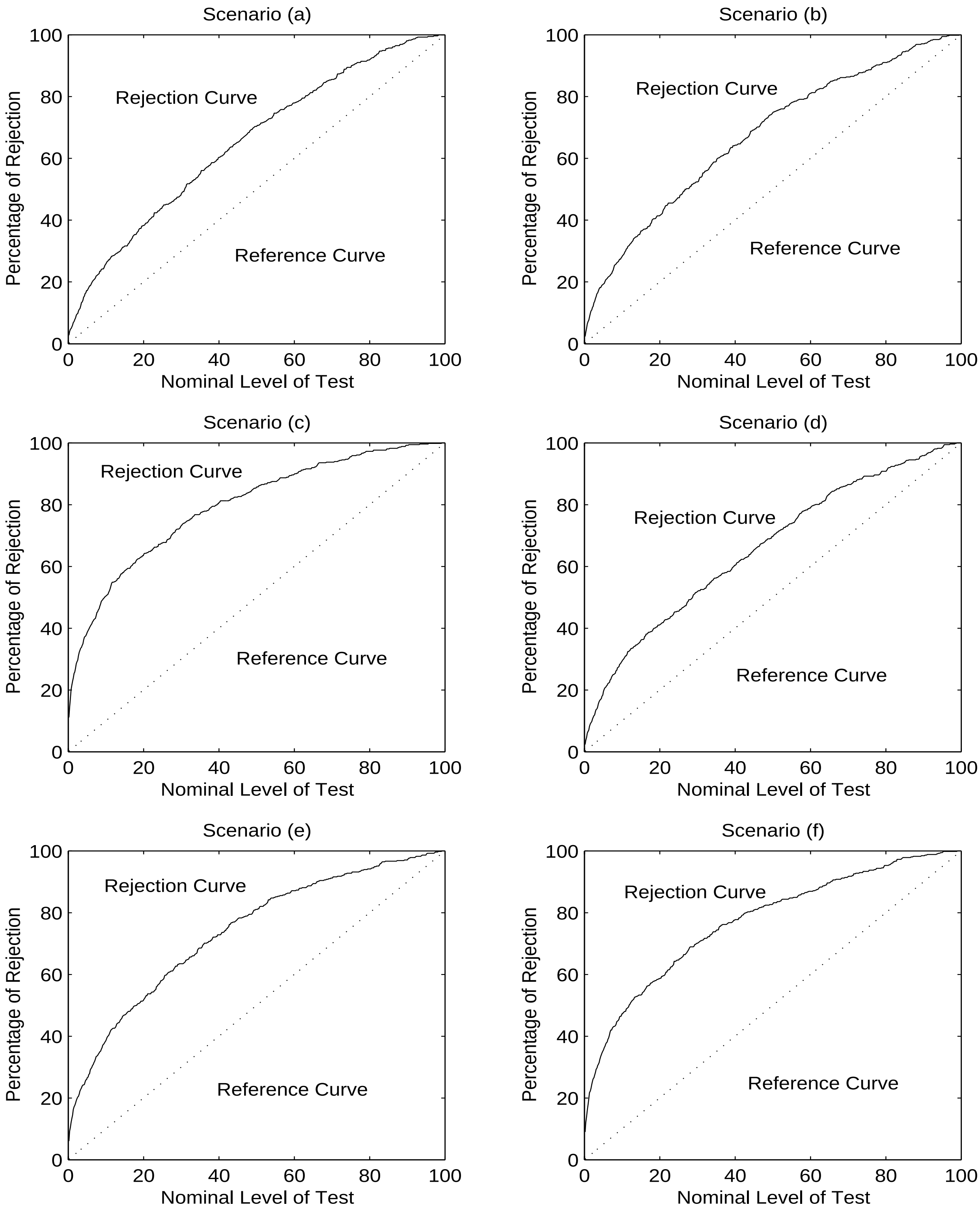


Figure 1: GMM Specification Test of Overidentifying Restrictions.

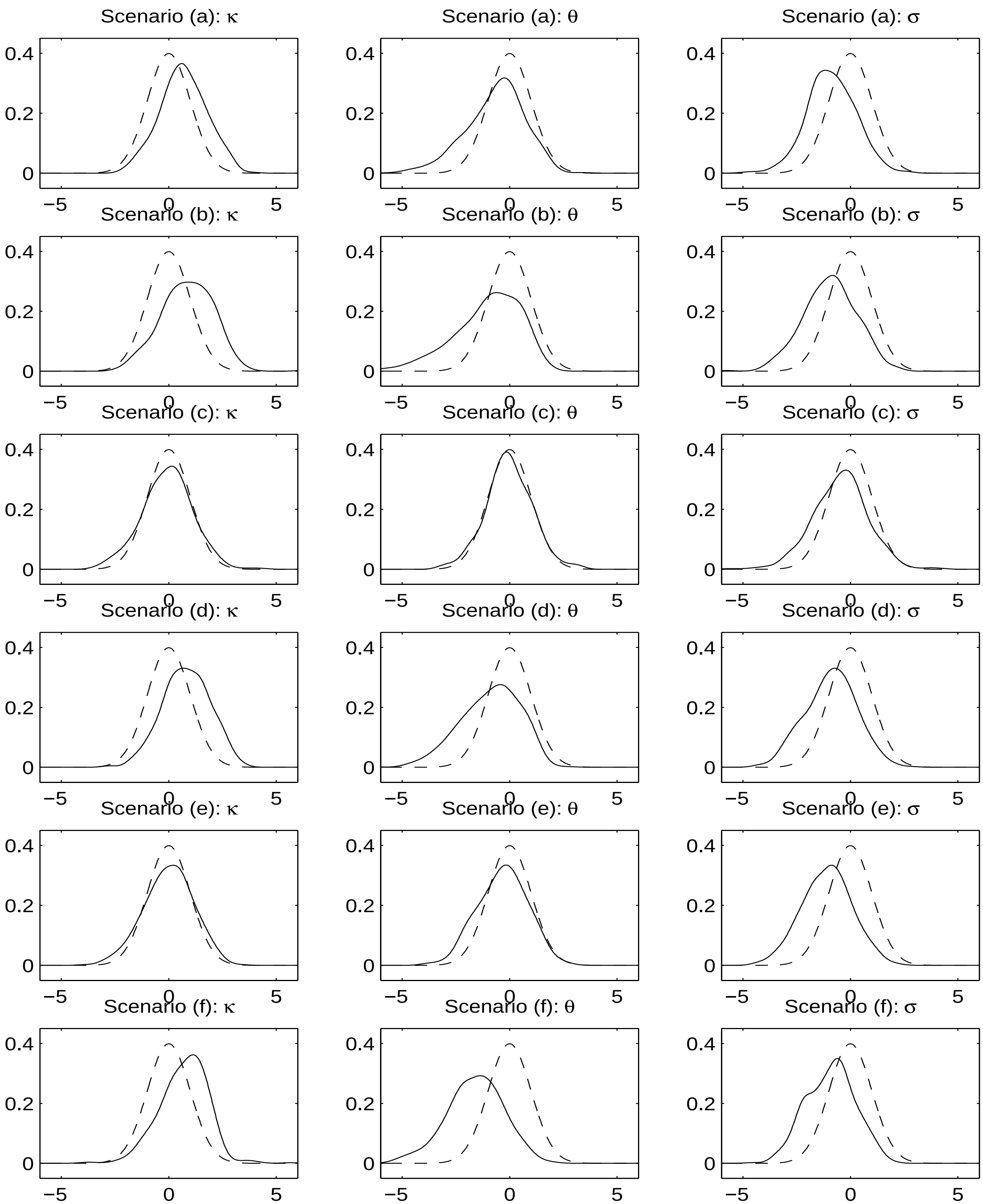


Figure 2: "- -" Normal (0,1) reference density; "—" t-test statistics.

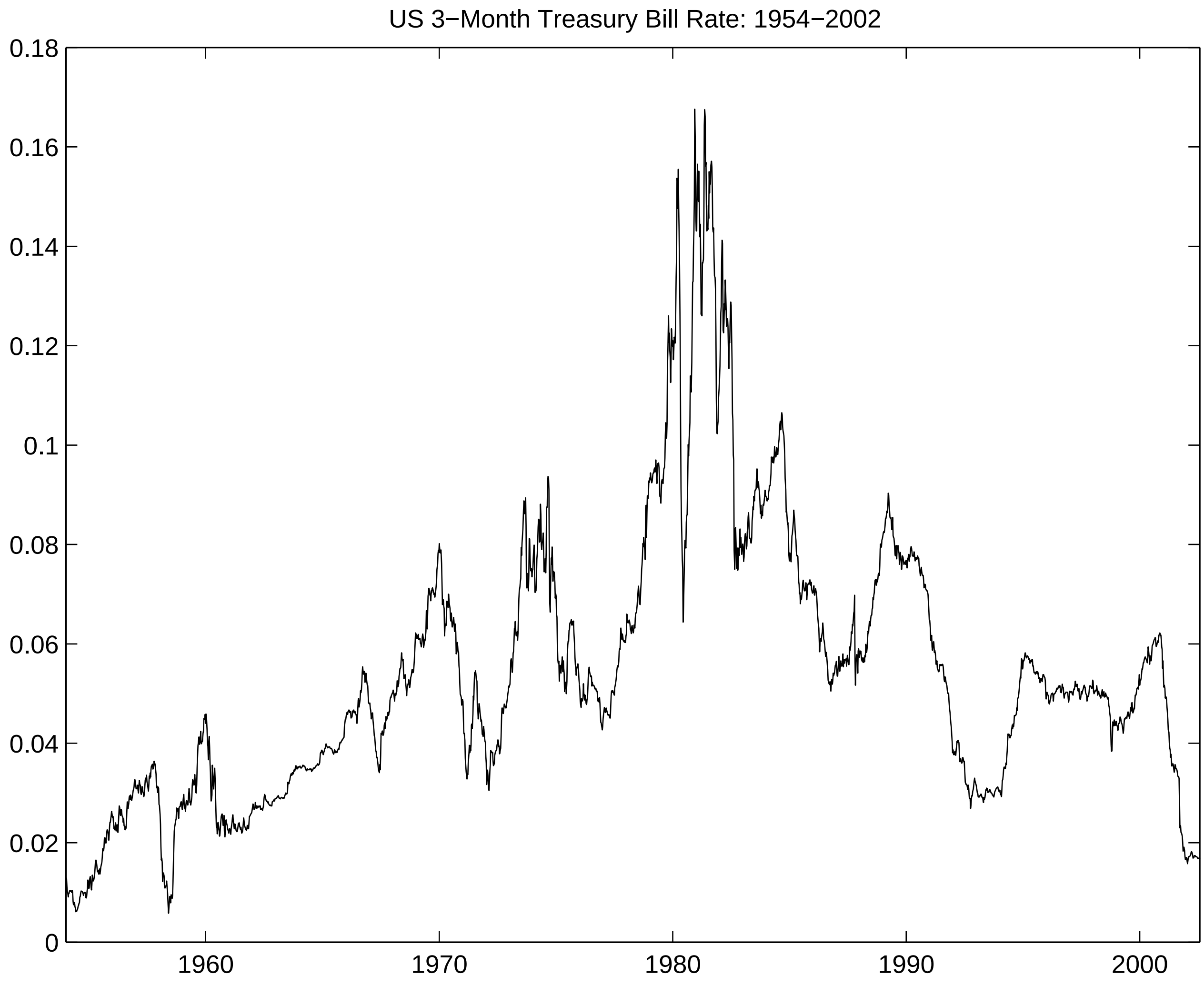


Figure 3: Time Series Plot of the Short Term Interest Rate.

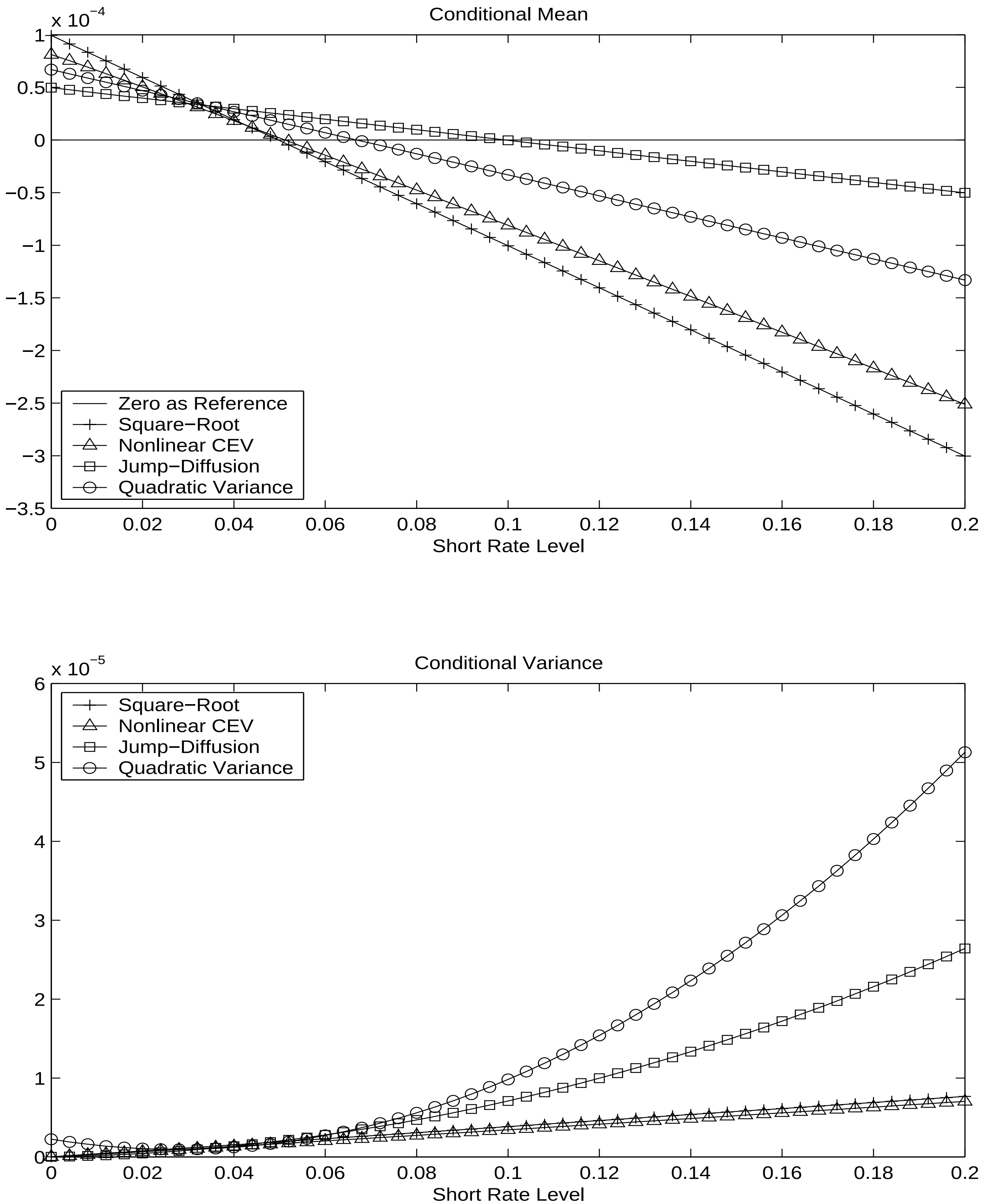
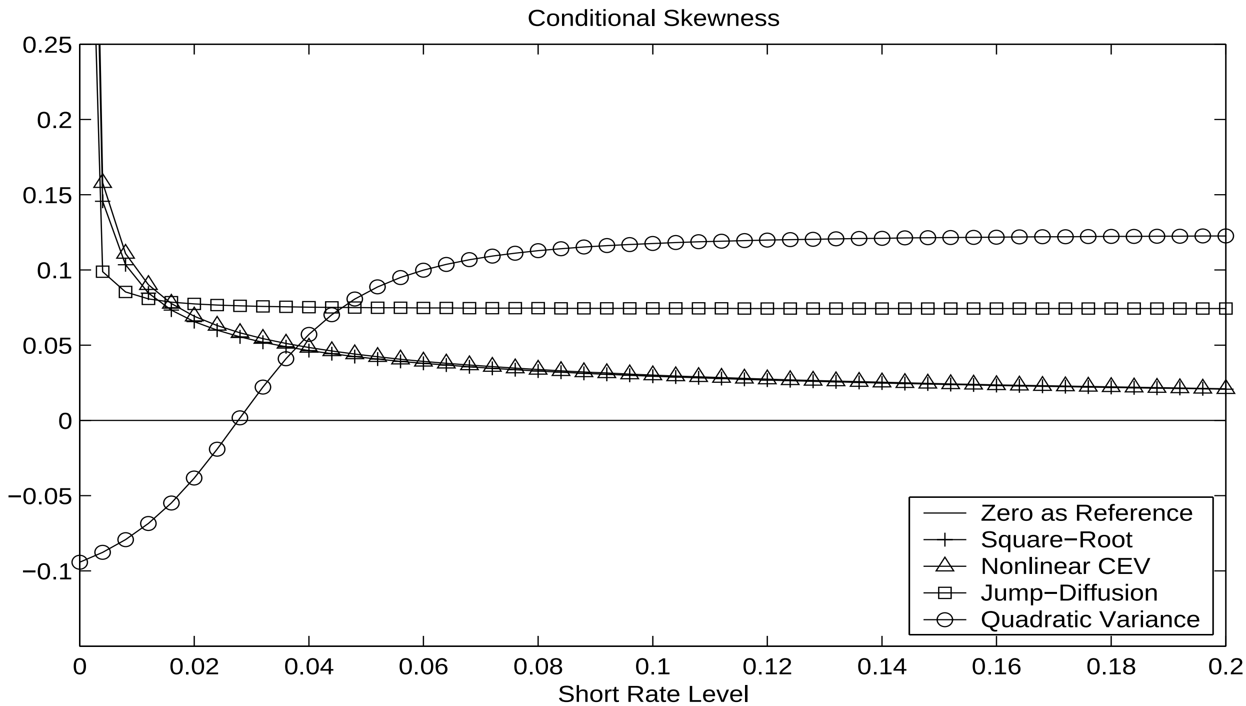


Figure 4: Conditional Mean and Variance.



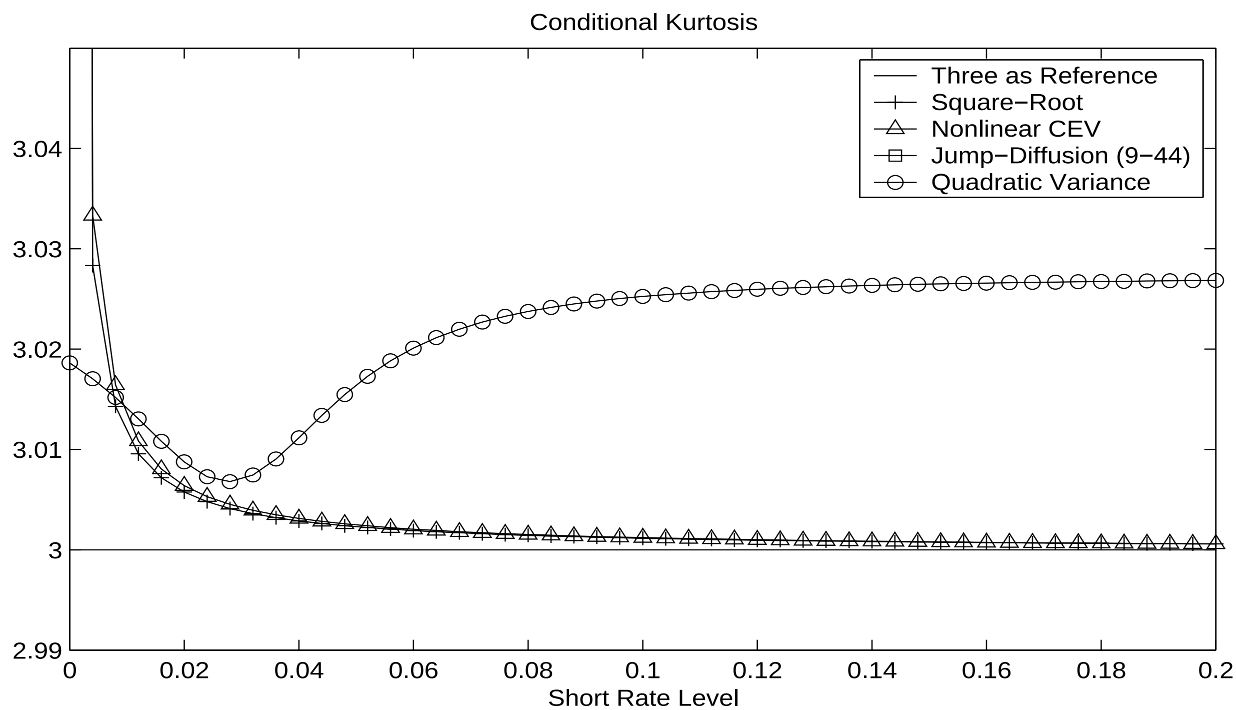


Figure 5: Conditional Skewness and Kurtosis.