MPRA

Munich Personal RePEc Archive

A Damped Diffusion Framework for Financial Modeling and Closed-form Maximum Likelihood Estimation

Li, Minqiang

30. July 2008

A Damped Diffusion Framework for Financial Modeling and Closed-form Maximum Likelihood Estimation

Minqiang Li [*][†]

Georgia Institute of Technology, GA 30308, USA

July 30, 2008

ABSTRACT

Asset price bubbles can arise unintentionally when one uses continuous-time diffusion processes to model financial quantities. We propose a flexible damped diffusion framework that is able to break many types of bubbles and preserve the martingale pricing approach. Damping can be done on either the diffusion or drift function. Oftentimes, certain solutions to the valuation PDE can be ruled out by requiring the solution to be a limit of martingale prices for damped diffusion models. Monte Carlo study shows that with finite time-series length, maximum likelihood estimation often fails to detect the damped diffusion function while fabricates nonlinear drift function.

Keywords: Damped diffusion, asset price bubbles, martingale pricing, maximum likelihood estimation

JEL Classification: C60, G12, G13

i

Starting from Bachelier's 1900 thesis on the theory of speculation, it has become standard practice to model financial quantities such as stock prices and short-term interest rates using continuous-time processes, in particular diffusion processes. Famous models include the geometric Brownian motion model frequently used for stock prices and the Vasicek (1977) and Cox-Ingersoll-Ross (1985) (henceforth CIR) models for short rates. These models share the feature that strong and unique solutions exist. That strong solutions exist is an attractive property for financial modeling since we would want a financial quantity at a future time to depend only on the shocks in the process specification up to that time.

Another popular model that nests these models as special cases is the constant elasticity of variance (CEV) model in which the diffusion coefficient is a power law function of the state variable with any non-negative exponent. The CEV process has been used extensively by researchers to model different financial quantities. For example, Cox (1975, 1996) and Cox and Ross (1976) use the CEV process to model stock prices as an alternative to geometric Brownian motion. Beckers (1980), Emanuel and MacBeth (1982), Schroder (1989), among others, have also used the CEV process to model stock prices. Chacko and Viceira (2003), Jones (2003a), and Aït-Shahalia and Kimmel (2007), among others, study stochastic volatility models in which the stock price volatility follows a CEV process. Chan, Karolyi, Longstaff and Sanders (1992), among others, study CEV models for short rates. Andersen and Andreasen (2000), among others, considers forward LIBOR market models with a CEV volatility.

However, the CEV process has a variety of drawbacks, including: (i) with linear mean-reverting drift and an exponent between 1 and 3/2, the second and higher-order moments of the invariant distribution are infinite; (ii) consistent with this, simulated sample paths from the process take on extremely large values with non-trivial probabilities; and (iii) with this process for either the underlying stock price or stochastic volatility, the martingale approach for option pricing can fail. These drawbacks will be discussed in more detail in the next section.

To overcome the first two drawbacks, Aït-Sahalia (1996a) proposes a flexible specification in which the drift is a non-linear function of the state variable. The nonlinear drift helps produce global stationarity of the process if the exponent of the power law diffusion is greater than 1. On the other hand, Conley, Hansen, Luttmer and Scheinkman (1997) point out that if the existence of a stationary density is what one wants in the model specification, then a mean-reverting drift is not necessary. In fact, when interest rates are high, the increased volatility of the diffusion process could be a mechanism to induce stationarity. Conley et al. call this phenomenon "volatility-induced stationarity." Also, by using various methods, Chapman and Pearson (2000), Jones (2003b), and Li, Pearson and Poteshman (2004) all present evidence that the nonlinearity in the drift function might be spurious for the short-rate data.

Instead of modifying the linear drift as in Aït-Sahalia (1996a), this paper takes a new unified approach by proposing a damped diffusion framework. The idea is to damp the drift function and/or the diffusion

function of a diffusion process so that the damped diffusion process is better behaved. As we will see, the choice of damping functions is quite flexible. For example, in the case of short-rate modeling, we can keep the linear drift while damping the diffusion function so that it approaches a more sensible model as the interest rate goes to infinity. Another nice feature of this framework is that the solution to any model in this framework is strong unique with modest conditions on the damping functions.

The damping idea can be applied to a broad class of financial models, for example, models of stock prices, the term structure of interest rates, forward rates, and stochastic volatility. In the case of derivative pricing, one important application of the damped diffusion framework is to preserve the usual martingale pricing approach. Through concrete examples, we show that damping either the drift or the diffusion function can help establish equivalence of measure changes and break many types of asset price bubbles, such as stock future bubbles, bond price bubbles and option price bubbles. Furthermore, in two cases where the martingale pricing approach fails and the asset valuation partial differential equation (PDE) has infinitely many solutions, introducing successively weaker damping allows us to rule out certain solutions if we require the solution be a limit of well-behaved martingale prices. In the case of CIR short-rate process and linear premium function $\psi_0 + \psi_1 r$, while Heston, Loewenstein and Willard (2007) show that Cox, Ingersoll and Ross's (1985) conjecture that a nonzero ψ_0 always leads to arbitrage is in general not true, our result shows that in some weaker sense, any positive ψ_0 can be allowed.

In contrast to the Vasicek and CIR models, models in the damped diffusion framework usually do not have explicit transition densities. Fortunately, this shortcoming is partially overcome by Aït-Sahalia's (1999, 2002) series expansion method. We take a closer look at this approximation and point out some nice properties of the expansion coefficients in this method, including the symmetry, differentiability and invariance properties. In Appendix B, we provide the expansion coefficients for the log-densities of many frequently used processes in financial modeling. It should be very helpful to researchers who need to perform maximum likelihood estimation since these processes are frequently used to model short rates, stock prices, and many other financial quantities. We then demonstrate that the damped diffusion model we proposed can be easily estimated using Aït-Sahalia's series approximation method. However, through a Monte Carlo experiment, we show that with finite time-series length, maximum likelihood estimation often fails to detect the damped diffusion function while fabricates nonlinear drift function.

The paper is organized as follows. Section I describes some drawbacks of the CEV process in modeling financial quantities and introduces the damped diffusion framework as a way to overcome those drawbacks. In Section II, we show through several examples that by damping either the drift or diffusion function, we can preserve the martingale pricing approach in asset pricing and break many types of asset price bubbles. Section III points out some nice properties in Aït-Sahalia's approximation method. This is followed by a maximum likelihood estimation of common short-rate models and a Monte-Carlo experiment of the finite sample bias of the nonlinear drift model and the damped diffusion model. Section IV concludes.

I. The CEV Process and the Damped Diffusion Framework

A. Drawbacks of the CEV process

The CEV process $\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma X_t^\rho \mathrm{d}W_t$ with $\mu(x) = \kappa(m - x)$ is quite popular in financial modeling Despite its popularity, the CEV process has some drawbacks. We now discuss those in some detail.

First, the CEV model does not necessarily have the same mean-reverting behavior as the CIR model. For the special case $\mu(x) = \kappa x$ with $\kappa > 0$, Emanuel and MacBeth (1982) (p.536) shows that the mean of X_t does not grow exponentially as expected.[1] This undermines the very attractiveness of using a linear mean-reverting drift.

Another issue is that if $1 < \rho \le 1.5$, the steady-state variance is infinite. In fact, in the steady state, $\mathbb{E}X_t^\nu$ is infinite for any $\nu \ge 2\rho - 1$. Different authors have obtained values of ρ larger than 1 using different approaches, including nonparametric estimation, generalized method of moments and Bayesian analysis. Jones (2003a) uses the CEV process to model stochastic volatility and finds an exponent of 1.33. The infinite variance introduces some difficulties in econometric estimation. For example, the usual generalized method of moments estimation requires the moments to exist in the first place to perform the estimation. Also, much of the asymptotic analysis in econometrics requires that the Fisher information matrix be well-defined in order to guarantee convergence.

The unusually large unconditional probability for the interest rate to be at a very high level translates to large probabilities for sample paths to reach very high levels during a finite period. To examine this, we do a Monte Carlo study of the maximum of sample paths. We use the Euler scheme with time step 1/3000 (roughly corresponding to every two hours) to simulate 200,000 sample paths of length 20 years starting at $X_0 = 7.5\%$. The parameters used are those from Table I. That is, $\kappa = 0.0886$, $m = 0.0842$, $\sigma = 0.7792$ and $\rho = 1.4812$. Out of these sample paths, 2.90% reach a maximum higher than 50%, 0.82% reach a maximum higher than 100%, and 0.27% reach a maximum higher than 200%. In fact, there are 14 paths that have a maximum higher than 5000%! Durham (2003) considers the CEV1 model $\mathrm{d}X(t) = \alpha \mathrm{d}t + \beta_1 X(t)^{\beta_2} \mathrm{d}W(t)$, and shows that adding additional flexibility to the drift function beyond a constant term provides negligible benefit (in terms of the value of the likelihood function). Not surprisingly, the CEV1 model is more ill-behaved than the mean-reverting CEV model. In fact, with his estimates, there are now 103 paths that have a maximum higher than 5000%.[2]

Another severe drawback of the CEV model is that if we use the CEV process for the underlying stock price, there can exist multiple solutions for the call option price when the exponent is greater than one. This leads to the possibility of arbitrage and the breakdown of the martingale pricing approach.

---

[1] Interestingly, they also discusses an "immeasurable" modification of tail method very similar to the damped diffusion idea in this paper to overcome the explosion problem. See also Davydov and Linetsky (2001).

[2] Using other higher-order schemes as in discussed in Kloeden and Platen (1992) does not change the quantitative results at all.

As shown in Cox and Hobson (2005) and Heston, Loewenstein and Willard (2007), when this happens, many standard results can fail. For example, put-call parity can be violated, option prices are no longer convex in the strike price, call prices may not approach zero when strikes approach infinity, etc. Cox and Hobson (2005) and Heston, Loewenstein and Willard (2007) interpret the breakdown of martingale pricing as bubbles in the financial markets. Similar complications can also occur when one uses the CEV process for the volatility process in a stochastic volatility model, as discussed in Sin (1998) and Lewis (Chapter 9, 2000). In the words of Heston, Loewenstein and Willard (2007), "(t)hese counterfactual implications for option values provide a persuasive rationale for specifying models without bubbles in many applications."

B. The Damped diffusion framework

The idea of the damped diffusion framework is to modify the drift and/or the diffusion function of a continuous-time diffusion process with a damping function so that the modified process is more appropriate than the original process. In this section, we will only consider modifying the diffusion function since we are working on a linear-drift CEV process. Damping the diffusion function is in some ways similar to Aït-Sahalia's (1996a) nonlinear drift approach to regularizing the CEV short-rate process. However, the drift and diffusion in the nonlinear model do not satisfy the usual linear growth and Lipschitz conditions. Also, there is no convincing evidence that the short rate in actual data has a nonlinear drift. As Durham (2003) and others have argued, the diffusion rather than the drift is the critical component. On the other hand, a simple mean-reverting linear drift is more appealing than a constant drift. Thus we keep the linear drift but apply a damping function to the diffusion function somewhere above the maximal observed level of the interest rate path. The philosophy of using a damped diffusion function is that from the actual sample path only, it is very hard to infer precisely what the true diffusion function is at interest rate levels much higher than the realized maximum.[3]

The most important application of the damped diffusion framework is to rule out asset price bubbles arising from the failure of martingale pricing approach, as discussed in detail in Heston, Loewenstein, and Willard (2007) and others. We devote the entire Section II to demonstrate this claim. The reason is that damping functions can help to make the solution of a PDE unique and to make a local martingale a genuine martingale by satisfying the Novikov condition. This is a very strong economic rationale for the damped diffusion framework as Proposition 2.1 of Heston, Loewenstein, and Willard (2007) shows that

---

[3]Another approach to regularizing the CEV process is to introduce two reflecting boundaries at two regular points of the process. However, this approach requires careful handling of the boundary conditions because those boundaries can be reached with positive probabilities. In the damped diffusion framework, positive infinity is a natural boundary and we can handle it just as we normally do. Also, there is no easy extension of Aït-Sahalia's series approximation for the transition density in the reflecting boundary approach. Finally, Goldstein and Keirstead (1997) show that while it is sometimes convenient to put reflecting boundaries on the spot rate process, reflecting boundaries can not be put on the forward rate dynamics without generating arbitrage opportunities.

having equivalent martingale measures is equivalent to the nonexistence of bond and stock price bubbles.

The damped diffusion framework also has another more explicit economic rationale. As a thought experiment, suppose that the interest level becomes extraordinarily high. In this case the Federal Reserve Board has two ways to help it go down. One possibility would be to make people to believe that the (expected) future interest level will decline quickly. This translates roughly to the nonlinear drift function. Another possibility, which might be operationally easier, is to limit very short-time fluctuations. For example, the Federal Reserve might not allow the ratio of interest rate volatility to the interest rate level to become unbounded (which happens in the CEV process). This roughly translates to a damped diffusion function, since within a short time changes in the interest rate level are dominated by the diffusion term.

The power of the damping idea can be clearly seen in the following proposition. It says that for any time-homogeneous univariate diffusion process, we can perform a "minimal" modification on its drift and diffusion functions so that the new process has a strong unique solution and does not explode near boundaries. In other words, as far as only regularizing a process is concerned, any diffusion process with smooth drift and diffusion functions can be regularized by the damping idea.

Proposition 1 *Let* $\mathrm{d}X_t = f(X_t)\mathrm{d}t + g(X_t)\mathrm{d}W_t$ *be a diffusion process defined on domain* (a, b) *with smooth drift and diffusion functions. Here either* a *or* b *could be infinite. For any* $A > a$, $B < b$, *and* $\epsilon > 0$, *there exist smooth functions* $\widetilde{f}$ *and* $\widetilde{g}$ *which are small modifications of* f *and* g *in the sense that*

$$\int_A^B |f(u) - \widetilde{f}(u)| + |g(u) - \widetilde{g}(u)| \ \mathrm{d}u < \epsilon, \tag{1}$$

and the stochastic differential equation $\mathrm{d}\widetilde{X}_t = \widetilde{f}(\widetilde{X}_t)\mathrm{d}t + \widetilde{g}(\widetilde{X}_t)\,\mathrm{d}W_t$ *has a strong unique solution on (possibly slightly smaller) domain* $(\widetilde{a}, \widetilde{b})$. *Furthermore,* $\widetilde{X}_t$ *does not explode to either* $\widetilde{a}$ *or* $\widetilde{b}$.

If the original drift and diffusion functions are nonzero except on a few discrete points, then $D_f \equiv \widetilde{f}/f$ and $D_g \equiv \widetilde{g}/g$ are well-defined damping functions. Proof of the above proposition is in Appendix A and contains a construction of the new functions $\widetilde{f}$ and $\widetilde{g}$ based on mollification. However, in real applications where the explicit forms of f and g are most likely known, the search for the damping function might be much easier, as the following important special case shows. We will use this damped diffusion process to estimate the Federal fund rate process in Section III. Specifically, let

$$\mathrm{d}X_t = \kappa(m - X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad \sigma(x) = \sigma_1 x^{1/2} D_1(x) + \sigma_2 x^\rho D_2(x), \tag{2}$$

where $\rho \geq 1$, $\sigma_i \geq 0$, and the $D_i(x)$ are two continuously differentiable damping functions.[4] Proposition 2 shows that this model is very well behaved with suitable $D_i(x)$. Proof is in Appendix A.[5]

---

[4] There is not much benefit from considering a model in which the first exponent is between $1/2$ and 1 because for the usual values of x between 0 and 0.25, the function x^ν with $1/2 < \nu < 1$ can be expressed almost exactly as a linear combination of $x^{1/2}$ and x^ρ with some $\rho \geq 1$.

[5] The same damping idea can also be applied to $\sigma(x) = \sqrt{\beta_0 + \beta_1 x + \beta_2 x^{\beta_3}}$ proposed by Aït-Sahalia (1996b), with essentially the same proof.

Proposition 2 *Assume that $x^\rho D_2(x) \geq 0$ is globally Lipschitz and there exists constant $M > 0$ such that $0 \leq D_1(x) \leq M$ and $xD_1'(x) \leq M$ for $0 \leq x < \infty$. Then for the damped superimposed CEV process in equation (2) with exponent $\rho \geq 1$, we have the following statements:*

1. the solution, if it exists, is strong unique in the sense that if $X_t^{(1)}$ and $X_t^{(2)}$ are two solutions, then they are indistinguishable, i.e., $P[\, X_t^{(1)} = X_t^{(2)}$ for $\forall t \in [0, \infty)\,] = 1$.

Further assuming that the diffusion function $\sigma(x)$ satisfies the linear growth condition, then the unique strong solution exists, and

2. $\mathbb{E}X_t = m + (x_0 - m)e^{-\kappa t}$ starting from any $x_0 > 0$ and X_t^n is integrable for any positive integer n.

3. The process is nonexplosive: 1) there exists $C > 0$ depending only on model parameters, such that $\limsup_{t \to \infty} \log X_t / t \leq C$, a. s.; 2) let $\xi = \inf\{t > 0 : X_t = \infty\}$, then $P[\xi = \infty] = 1$.

Notice Proposition 2 does not require the damping functions be monotone decreasing. As an example, consider $D_1 = 0$ and $D_2 = \exp(-\alpha x^\beta)$ with $\alpha > 0$ and $\beta > 1$. In this process the diffusion approaches 0 as $x \to \infty$. As another example, a diffusion $\sigma_1 x^{1/2} + \sigma_2 x^\rho \exp(-\alpha x^\beta)$ with $\rho \geq 1$ will approach the CIR model as $x \to \infty$. The third example that might be of interest is a damped diffusion function $\sigma x^\rho / (1 + \alpha x^\rho)$ with $\alpha > 0$ which approaches a constant value, thus mimicking the behavior of the Vasicek model when the interest rate is high. All the above examples satisfy the conditions in the above proposition.

To understand the effect of damping, we consider a particular model with $D_1(x) = D_2(x) = e^{-8x^4}$ and $\rho = 3/2$. The damping function e^{-8x^4} is very close to one when $0 \leq x < 0.2$ and very close to 0 when $x > 0.8$. In the middle, it gradually decreases from a value close to one to close to zero. We perform the same Monte Carlo study that was carried out for the CEV process. The parameters used are those estimated from maximum likelihood in Table I. Out of 200,000 sample paths, 6.99% reach a maximum higher than 30% before 20 years, and 2.04% reach a maximum higher than 50%. There is no path that has a maximum greater than 80%. We see that damping has the effect of regularizing the CEV process.

The idea of damping could be applied in other cases. For example, we could introduce a damping function when performing a Monte Carlo study by simulating sample paths from a certain diffusion process. Introducing a damping function can prevent extreme values to be reached just by pure chance and thus acting as a safeguard mechanism. The damped diffusion framework should also be useful in stochastic volatility models in order to eliminate wild behavior similar to that discussed for short-rate CEV models. If one specifies a CEV process for the volatility process in either the real world or the risk-neutral world, and if the estimated exponent ρ is larger than one, there might be a need to regularize the volatility process. Thus, it might be useful to apply a damping function to the diffusion function of the variance or volatility process. The damped diffusion framework might also be useful with multi-factor term structure models if we use the CEV process for some of the factors.

II. Asset Price Bubbles and the Damped Diffusion Framework

The martingale pricing approach is one of the cornerstones of modern asset pricing theory. However, sometimes the martingale pricing approach can fail if one models financial quantities as diffusion processes. The damped diffusion framework introduced in the last section is very useful to preserve the martingale pricing approach and rule out asset price bubbles. We consider two different types of failures of martingale pricing approaches, both of them arising from process explosions when performing a measure change.

A. Explosion from measure $\mathbb{Q}$ to $\mathbb{Q}^S$

One risk-neutral stock price process people often consider as an alternative to the usual geometric Brownian motion process is the following CEV process:

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sigma S_t^\alpha \mathrm{d}W_t^\mathbb{Q}, \tag{3}$$

where r and σ are constants. We will focus on the case $\alpha > 1$. The above SDE has a strong unique solution and does not explode to infinity. In addition, letting the current time be 0, the transition density for stock price τ-time ahead is as follows (Emanuel and MacBeth (1982):

$$p(\tau, S_\tau | S_0) = 2(\alpha - 1)k^{\frac{1}{2(1-\alpha)}}\left(x_0 x_\tau^{1-4\alpha}\right)^{\frac{1}{4-4\alpha}}e^{-x_0-x_\tau} \cdot I_{\frac{1}{2(\alpha-1)}}\left(2\sqrt{x_0 x_\tau}\right), \tag{4}$$

where

$$k = \frac{r}{\sigma^2(1-\alpha)(e^{2r(1-\alpha)\tau} - 1)}, \quad x_0 = kS_0^{2(1-\alpha)}e^{2r(1-\alpha)\tau}, \quad x_\tau = kS_\tau^{2(1-\alpha)}, \tag{5}$$

and $I_q(\cdot)$ is the modified Bessel function of the first kind with order q.

An immediate shortcoming of this model is that although locally the rate of return is always r, the forward stock price, defined as the expected future stock price under the above transition density, is not $F^1 \equiv S_0 e^{r\tau}$. Instead, we have

$$F^2 \equiv \int_0^\infty p(\tau, S_\tau | S_0)S_\tau \mathrm{d}S_\tau = F^1\left(1 - \frac{\Gamma(v, x_0)}{\Gamma(v, 0)}\right), \tag{6}$$

where $\Gamma(v, u) = \int_u^\infty e^{-z}z^{v-1}\mathrm{d}z$ is the incomplete gamma function. This was realized by Emanuel and MacBeth (1982, p.536), but they did not obtain the above explicit expression. A proof of the above equation is in Appendix A. That F^1 and F^2 are different shows that $\int_0^t \sigma S_u^\alpha \mathrm{d}W_u^\mathbb{Q}$ is not a true martingale, but rather a strict local martingale. Related with this problem, the martingale pricing approach using the stock price deflated measure $\mathbb{Q}^S$ breaks down. Fix $\tau > 0$. Let us examine the asset-or-nothing component of the call option price $e^{-r\tau}\mathbb{E}^\mathbb{Q}[S_\tau \cdot 1_{\{S_\tau \geq K\}}]$. Since the measure $\mathbb{Q}^S$ is defined by $\frac{\mathrm{d}\mathbb{Q}^S}{\mathrm{d}\mathbb{Q}}\big|_T \equiv \frac{e^{-rT}S_T}{S_0}$, by Girsanov theorem, the $\mathbb{Q}^S$-Brownian motion is given by $\mathrm{d}W_t^{\mathbb{Q}^S} = \mathrm{d}W_t^\mathbb{Q} - \sigma S_t^{\alpha-1}\mathrm{d}t$. This gives the $\mathbb{Q}^S$-

dynamics of the stock price as $dS_t = (rS_t + \sigma S_t^{2\alpha-1})dt + \sigma S_t^\alpha dW_t^{\mathbb{Q}^S}$. However, if $\alpha > 1$, the measures $\mathbb{Q}$ and $\mathbb{Q}^S$ are *not* equivalent,[6] so it is no longer true that $e^{-r\tau}\mathbb{E}^{\mathbb{Q}}[S_\tau \cdot 1_{\{S_\tau \geq K\}}] = S_0\mathbb{E}^{\mathbb{Q}^S}[1_{\{S_\tau \geq K\}}]$.

There are many awkward consequences of the above results. For example, since the forward price is no longer $S_0 e^{r\tau}$, it is difficult to interpret r globally. Most importantly, Heston, Loewenstein and Willard (2007) show that there will be multiple solutions to the valuation PDE of the call option price G (page 366).[7] They give two explicit solutions G^1 and G^2 and show that both solutions have weird behavior. We refer the reader to Heston, Loewenstein and Willard (2007) for more details. For our purpose, it is important to notice that $\lim_{S_0\to\infty} G^1/S_0 = 1$ while $\lim_{S_0\to\infty} G^2/S_0 = 0$, and $\lim_{K\uparrow\infty} G^2(K) = 0$ while $\lim_{K\uparrow\infty} G^1(K) = S_0\mathbb{E}^{\mathbb{Q}^S}[1_{\{S_\tau=\infty\}}] > 0$.[8]

In this paper, we are more interested in possible ways to fix these problems. It turns out that the damped diffusion framework can be used to preserve the forward price (and thus the usual put-call parity) and also preserve the martingale option pricing. The following proposition shows the measure change from $\mathbb{Q}$ to $\mathbb{Q}^S$ is equivalent with suitably chosen damping functions. Furthermore, there exists a well-defined call option price as the damping gets weaker and weaker.

Proposition 3 *Assume that the stock price under the risk-neutral measure $\mathbb{Q}$ follows the damped diffusion process $dS = rSdt + \sigma S^\alpha D_\lambda(S)dW^{\mathbb{Q}}$, where $D_\lambda(S)$ is a smooth damping function for any fixed $\lambda > 0$. Assume further that $g_\lambda(S) \equiv \sigma S^\alpha D_\lambda(S)$ is globally Lipschitz in S, and there exists $M > 0$ such that $g_\lambda(S) < MS$ for all $S > 0$. Then, the following statements hold:*

1. The stock price process under $\mathbb{Q}$ has a strong unique solution and is nonexplosive.

2. For any $\lambda > 0$, the time τ forward stock price is $S_0 e^{r\tau}$ and the put-call parity holds.

3. Let $\mathbb{Q}^S$ be a new measure induced by using the stock price as the numeraire. Then the measure change from $\mathbb{Q}$ to $\mathbb{Q}^S$ is equivalent for any $\lambda > 0$. In particular, the stock price process under $\mathbb{Q}^S$ has a strong unique solution and is nonexplosive.

4. Fix $\lambda > 0$ and a strike K. Let G_λ be the call option price with damping parameter λ. Then G_λ is the unique solution to the valuation PDE with at most polynomial growth in S. Similarly for the put option price P_λ.

5. Assume further that $g_\lambda(S)$ is convex in S for any $\lambda > 0$ and for fixed $S > 0$, $D_\lambda(S)$ is decreasing in λ. For fixed t, S and K, G_λ increases to a fixed limit $\widetilde{G}$ as $\lambda \downarrow 0^+$. Similarly P_λ increases to a limit $\widetilde{P}$. In addition, $(S - Ke^{-r\tau})^+ \leq \widetilde{G} \leq S$, $\partial\widetilde{G}/\partial S \geq 0$, $\lim_{S\to\infty}\widetilde{G}/S = 1$ and $\widetilde{G} - \widetilde{P} = S - Ke^{-r\tau}$.

---

[6]This follows from the fact that $Y_t \equiv S_t^{2-2\alpha}$ follows a CIR process $dY_t = k_Y(\theta_Y - Y_t)dt - \sigma_Y\sqrt{Y_t}dW_t^{\mathbb{Q}^S}$ with $k_Y = (2\alpha - 2)r$, $\theta_Y = (2\alpha - 3)\sigma^2/2r$ and $\sigma_Y = (2\alpha - 2)\sigma$, which explodes to 0 by Feller's condition.

[7]Notice that equation (7) in Heston, Loewenstein and Willard (2007) contains two minor typos in their expression $(xz^{1-\alpha})^{1/4-\alpha}$ and does not agree with Emanuel and MacBeth (1982). The parameter u in Heston, Loewenstein and Willard (2007) should be equal to x_0 instead of $2x_0$ as it is now. Also, equation (10) in Heston, Loewenstein and Willard (2007) has an extraneous term rp in it.

[8]In an earlier version of this paper, we show that even if we restrict the solution to satisfy $\lim_{K\to\infty} G = 0$, it is still not unique. However, the solution becomes unique if we fix its tail behavior. Statements and proofs of the above claims are available upon request.

The above proposition is interesting. It shows that one can use the CEV model without damping as long as one keeps in mind that the CEV model should be interpreted as a limit of nicely-behaved martingale pricing models. To reflect this implicit limit procedure, the forward price should be taken as F^1 instead of F^2. In addition, out of the linear combinations of G^1 and G^2, the only possible solution which can serve as a limit of martingale prices is G^1 since otherwise we would have $\lim_{S\to\infty} \widetilde{G}/S < 1$. This limiting procedure also solves the anomaly when $K = \infty$. For $K = \infty$, instead of taking the the limit of $G^1(K)$ as $K \to \infty$, we should take the limit of $G_\lambda(\infty)$ as $\lambda \to 0^+$, which is 0. The order of taking limits is important here.

One question remains on whether we can actually find a damping function $D_\lambda(S)$ so that $g_\lambda(S)$ has the stated properties because otherwise the above proposition would be vacuous.[9] The answer is yes. One can easily show that the following particular damping function will satisfy the conditions in the above proposition (including the one in statement 5):

$$D_\lambda \equiv \left(\frac{1}{1 + \lambda S^{(\alpha-1)/\beta}} \right)^\beta, \tag{7}$$

where the parameter β is any fixed constant such that $\beta > \alpha - 1$. The parameter $\lambda > 0$ controls the strength of damping. As $\lambda \to 0^+$, the damping gets weaker and weaker. Furthermore, as $S \to +\infty$, the diffusion function of the damped diffusion model approaches that of a geometric Brownian motion model. We want to emphasize that it is not necessary to choose the above particular parametric form of the damping function for the following proposition to work, although this particular choice does allow us to prove the proposition through explicit computation.

While the damping idea proposed here is largely on theoretical grounds, in practice one often implicitly incorporates a damping function. For example, in solving PDE's numerically through finite difference methods, one often truncates the state space so that one only discretizes a bounded region, based on the belief that the truncation will introduce negligible effect. The truncation can be thought of as an extreme form of damping the diffusion function. Our preliminary analysis using extreme parameters shows that the numerical solution one gets for the CEV model often approximates G^1 instead of G^2.

B. Explosion from measure $\mathbb{P}$ to $\mathbb{Q}$

B.1. Specifying risk-preference directly

The failure of martingale pricing approach can also happen during a measure change from the real world measure $\mathbb{P}$ to the risk-neutral measure $\mathbb{Q}$. This comes about because under the risk-neutral measure, the drift of the real-world process will be modified after taking into account investors' risk preferences.

---

[9]The requirement that $g_\lambda(S)$ is convex is important for statement 5. While the usual comparison theorem compares diffusion processes with different drift functions, we need to compare diffusion processes with different diffusion functions. The comparison theorem in Hajek (1985) requires convexity on the diffusion functions. However, if one is only interested in exclude asset price bubbles with a fixed damping function, then convexity is not needed.

The following example in Lewis (2000) illustrates this point. Suppose the stock price process follows a stochastic volatility model $dS_t = rS_t dt + \sqrt{V_t}S_t dB_t^{\mathbb{P}}$ with the real world variance process given by

$$dV_t = b(V_t)dt + a(V_t)dW_t^{\mathbb{P}}, \tag{8}$$

where $W^{\mathbb{P}}$ is a Brownian motion under the real world measure $\mathbb{P}$. We assume that the real world variance process is well-behaved and in particular, nonexplosive. Lewis (2000) shows that under usual economic assumptions (for details, see Lewis, Chapter 7) who is facing a pure investment problem (no consumption) with a distant horizon, then under the risk-neutral measure $\mathbb{Q}$, the variance process becomes $dV_t = \widetilde{b}(V_t)dt + a(V_t)dW_t^{\mathbb{Q}}$, where the drift function is changed to $\widetilde{b}(V) = b(V) - (1 - \gamma)\varrho\sqrt{V}a(V) + a^2(V)u'(V)/u(V)$, and where ϱ is the correlation between the two Brownian motions driving the stock and variance processes, γ is the parameter in the power utility $U(W) = W^\gamma/\gamma$, and $u(V)$ is the first eigenfunction of the operator $\mathcal{L}$ defined as $\mathcal{L}u \equiv -a^2(V)u''/2 - b(V)u' - \gamma(1 - \gamma)Vu/2$. For the CIR process, Lewis shows that the effect of risk adjustment is to change the mean reverting parameter and long run mean parameter while keeping their product constant (p.234). With suitable risk parameter values so that the Feller condition for the new process is satisfied, the new process is nonexplosive. However, Lewis shows that for many well-behaved real world variance processes, the corresponding risk-neutral variance process will become explosive. In these cases, the usual martingale pricing approach fails.

The damped diffusion framework can be applied immediately to preserve the martingale pricing formula, thus avoiding the calculation of the correction term due to explosion. For example, we could introduce a damping function for the real world variance process so that the asymptotic behavior of $a(V)$ as $V \to 0$ or ∞ is that of the CIR model. Then the risk adjustment on the real-world process should approach the adjustment for the CIR process when $V \to 0$ or ∞, which is well-behaved. This implies that there will be no volatility explosion in the risk-neutral variance process.

If we are concerned that V might explode to infinity under $\mathbb{Q}$, another approach is to damp $a(V)$ using an exponential damping function. If the large V behavior of $u'(V)/u(V)$ is mild relative to the damping, then the large V behavior of $\tilde{b}$ is similar to that of $b(V)$, thus ensuring the risk-neutral process is nonexplosive too. Lewis worked out $u'(V)/u(V)$ for the operator $\mathcal{L}$ of many processes, including geometric Brownian motion, the GARCH diffusion, the CIR process, and the inverse Feller process. For all these processes, the large V behavior of $u'(V)/u(V)$ is polynomial growth at most so an exponential damping will work. For the damped CEV variance process where $b(V) = \kappa(m - V)$ and $a(V) = \xi V^\rho \exp(-\alpha V^\beta)$, due to the damping, in the large V limit the eigenvalue problem $\mathcal{L}u = \zeta u$ is roughly $\kappa(m - V)u' = (\zeta + \gamma(1 - \gamma)V/2)u$ if we ignore the $-a^2(V)u''/2$ term. Solving this equation gives a large V behavior of $u(V) \sim V^{\eta_1}e^{\eta_2 V \log V}$ for some constants η_1 and η_2, implying slower than polynomial growth for $u'(V)/u(V)$. If the damping is strong in that $2\beta \gg \eta_2$, the large V behavior of $u(V)$ reinforces our ansatz of throwing away the u'' term in the analysis. Thus an exponential damping is able to prevent the

corresponding risk-neutral processes from being explosive for all the above real world processes. Similar methods can be used to prevent volatility explosion to 0.

B.2. Specifying the market price of risk function directly

Another more common way to connect the measure $\mathbb{P}$ and $\mathbb{Q}$ is through specifying the market price of risk function Υ, which is directly related to the exponent process L of the Radon-Nikodym derivative Λ. That is, $\Lambda_T \equiv \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}|_T = \exp(L_T - \langle L, L\rangle_T/2)$, where L is defined by $\mathrm{d}L = -\Upsilon \mathrm{d}W^{\mathbb{P}}$. This puts restrictions on the form Υ can take in order to make $\mathbb{P}$ and $\mathbb{Q}$ equivalent. The following example is discussed in Heston, Loewenstein and Willard (2007).

Let the interest rate process under $\mathbb{P}$ be

$$\mathrm{d}r = \alpha(\beta - r)\mathrm{d}t + \sigma\sqrt{r}\mathrm{d}W^{\mathbb{P}}, \tag{9}$$

where $2\alpha\beta > \sigma^2$ so that both boundaries are unattainable. Now if we specify the market price of risk as $\Upsilon(r) = \Psi(r)/(\sigma\sqrt{r})$, where the "risk premium" function $\Psi(r) = \psi_0 + \psi_1 r$ is linear and $\psi_0, \psi_1 > 0$, then under $\mathbb{Q}$, the interest rate process becomes $\mathrm{d}r = \widehat{\alpha}(\widehat{\beta} - r)\mathrm{d}t + \sigma\sqrt{r}\mathrm{d}W^{\mathbb{Q}}$, where $\widehat{\alpha} = \alpha + \psi_1$, $\widehat{\beta} = (\alpha\beta - \psi_0)/(\alpha + \psi_1)$ and $\mathrm{d}W^{\mathbb{Q}} = \mathrm{d}W^{\mathbb{P}} + \Upsilon \mathrm{d}t$.

However, if ψ_0 and ψ_1 are such that $2\widehat{\alpha}\widehat{\beta} < \sigma^2$, the interest rate process under $\mathbb{Q}$ explodes to 0. This shows that $\mathbb{P}$ and $\mathbb{Q}$ fail to be equivalent. Thus the usual martingale pricing approach breaks down. Indeed, Example 1.1 in Heston, Loewenstein and Willard (2007) gives two solutions to the valuation partial differential equation of a bond maturing at time T below:

$$\frac{\sigma^2}{2}r\frac{\partial^2 P}{\partial r^2} + \widehat{\alpha}(\widehat{\beta} - r)\frac{\partial P}{\partial r} + \frac{\partial P}{\partial t} - rP = 0, \tag{10}$$

with boundary conditions $P(+\infty, t) = 0$ and $P(r, T) = 1$. To conform to our notation, we call the solutions P^1 and P^2. Heston, Loewenstein and Willard (2007) interpret $P^1 - P^2$ as a bond price bubble. In fact, because the valuation PDE is linear, any linear combination $(1 - \mu)P^1 + \mu P^2$ is also a solution for any $\mu \in \mathbb{R}$. Besides this family of solutions, there are actually infinitely many other solutions of the valuation PDE satisfying the boundary condition $P(r, T) = 1$ and $\lim_{r\to\infty} P(r, t) = 0$. These solutions are characterized by their boundary behaviors at $r = 0$. In addition, the solution $P^2(r, t)$ above can be thought of as the "smallest nonnegative solution" in that it stipulates $P(0, t) = 0$. This amounts to stipulate that if the interest rate process ever hits zero, it will automatically jump to and stay in the cemetery point $r = +\infty$.[10]

Again, in this paper, we are interested in ways to solve the above problems. As discussed above Theorem 1 of Cheridito, Filipovic and Kimmel (2007), the nonexistence of an equivalent martingale

---

[10]In an earlier version of this paper, we make the above discussions precise by proving that P^1 and P^2 become unique solutions if we fix the boundary and terminal conditions. The exact statements and proofs are available upon request.

measure is closely related to the behavior of the market price of risk function. With an affine risk premium $\Psi(r)$, the market price of risk grows without bound at both lower and upper ends. Cheridito, Filipovic and Kimmel (2007) point out that it is not necessarily a problem for the market price of risk to grow without bound. The problem comes about when the market price of risk grows too quickly for us to have the Novikov or similar conditions. When $r \to +\infty$, from the model of Cox, Ingersoll and Ross (1985), we know that although the market price of risk grows like $\sqrt{r}$, it poses no problem. However, when $r \to 0^+$, in which case the market price of risk blows up like $1/\sqrt{r}$, it turns out that the Novikov condition is no longer satisfied to guarantee the equivalence of the measure change from $\mathbb{P}$ to $\mathbb{Q}$. Another more mathematical way to see the above point is the following. From the definitions of $\widehat{\alpha}$ and $\widehat{\beta}$, while a positive ψ_1 is harmless, a positive ψ_0 might produce explosion if $2\widehat{\alpha}\widehat{\beta} < \sigma^2$ by Feller's condition.

Knowing the source of the problem, it is immediately clear that the damping idea can be again used to remedy this problem and preserve the martingale pricing approach. More specifically, we should damp the term ψ_0. The following proposition shows that with suitable damping, regardless of how weak it is, the corresponding risk-neutral process is nonexplosive and there is a unique martingale bond price. The result is true regardless of whether $2\widehat{\alpha}\widehat{\beta} > \sigma^2$ or not. Thus, there is no longer a bond price bubble. Furthermore, as $\lambda \to 0^+$, these bond prices approach a unique limit $\widetilde{P}$. Proof is in Appendix A.

Proposition 4 *Consider the real world interest rate process* $dr = \alpha(\beta - r)dt + \sigma\sqrt{r}dW^{\mathbb{P}}$ *which satisfies* $2\alpha\beta > \sigma^2$. *Consider the risk premium function* $\Psi_\lambda(r) = \psi_0 D_\lambda(r) + \psi_1 r$, *where* $\psi_0, \psi_1, \lambda > 0$. *We know that under the risk-neutral measure* $\mathbb{Q}$, $dr = (\alpha(\beta - r) - \Psi_\lambda(r))dt + \sigma\sqrt{r}dW^{\mathbb{Q}}$, *where* $W^{\mathbb{Q}}$ *is a* $\mathbb{Q}$-*Brownian motion. Assume that for fixed* $\lambda > 0$, $D_\lambda(r) > 0$ *and has a bounded derivative for any* $r \geq 0$. *Assume also that for fixed* $\lambda > 0$, *the speed measure* $v_c(r)$ *for the above risk-neutral process satisfies Feller's explosion test*[11] *at 0:* $\lim_{r\downarrow 0^+} v_c(r) = +\infty$. *Then, the following statements hold:*

1. For any $\lambda > 0$, *this process has a strong unique solution and is nonexplosive.*

2. For any $\lambda > 0$, *there is a unique price* $P_\lambda(r, T)$ *for a bond maturing at time* T *with time 0 interest rate* r. $P_\lambda(r, T)$ *is decreasing in both* r *and* T. *The partial derivative* $\partial P_\lambda(r, T)/\partial T$ *is bounded above and below with the bounds independent of* λ.

3. Assume further that for any fixed $r > 0$, $D_\lambda(r)$ *increases monotonically to 1 as* $\lambda \to 0^+$. *Then, as* $\lambda \to 0^+$, $P_\lambda(r, T)$ *increases to a unique limit* $\widetilde{P}(r, T)$. $\widetilde{P}(r, T)$ *is decreasing in both* r *and* T *with bounded derivative* $\partial\widetilde{P}(r, T)/\partial T$.

Finally, a damping function satisfying all the assumptions above exists. In particular, it can be chosen to be $D_\lambda(r) = r/(r + \lambda)$.

Notice that with the choice of the damping function $D_\lambda(r) = r/(r + \lambda)$, the market price of risk

---

[11] Readers unfamiliar with the speed measure and Feller's test are referred to Karlin and Taylor (1981). Feller's explosion test is also used as one of the main techniques in the analysis of Heston, Loewenstein and Willard (2007).

function Υ approaches 0 instead of infinity as r decreases to 0. The nice thing is that we can choose λ to be any positive number, no matter how small it is. As $\lambda \to 0^+$, the damping gets weaker and weaker and Ψ_λ approaches the prescription $\Psi \equiv \psi_0 + \psi_1 r$ in Cox, Ingersoll and Ross (1985) pointwise on $(0, \infty)$. Thus, the proposition shows that although the specification $\Psi(r) = \psi_0 + \psi_1 r$ might induce infinitely many solutions to the bond price PDE, the bond price $\widetilde{P}$ is a limit of martingale bond prices where the limit is such that the damping on the drift gets progressively weaker. In addition, the limit $\widetilde{P}$ can not equal $(1 - \mu)P^1 + \mu P^2$ for any $\mu \neq 0$ since otherwise $\widetilde{P}$ will fail to be a decreasing function in r. Also, while Heston, Loewenstein and Willard (2007) show that Cox, Ingersoll and Ross's (1985) conjecture that a nonzero ψ_0 will always lead to arbitrage is not generally true, the above proposition points out that in some weaker sense any positive value of ψ_0 can be allowed. That is, for any positive ψ_0, there exists a solution to the PDE which is a limit of well-behaved martingale bond prices. This is important as Cheridito, Filipovic and Kimmel (2007) show that ψ_0 is empirically important and we need a broader set of parameters in the market price of risk function.

III. Maximum Likelihood Estimation and Finite Sample Bias

A damped diffusion process usually does not have explicit transition densities. To estimate the parameters in the damped diffusion framework using maximum likelihood, we need to approximate the transition density. In this section, we show that it is straightforward to perform a maximum likelihood estimation using Aït-Sahalia's series expansion method. However, we show that finite sample size tends to fabricate nonlinear drift function while fail to detect the damped diffusion function even if the data are generated from a damped diffusion model.

A. Maximum likelihood estimation

Consider the stochastic differential equation of interest: $\mathrm{d}X_t = \mu(X_t; \theta)\mathrm{d}t + \sigma(X_t; \theta)\mathrm{d}W_t$, where W_t is a standard Brownian motion and the drift μ and diffusion σ^2 are known continuous functions except for an unknown parameter vector θ in a parameter space $\Theta \subset \mathbb{R}^d$. To perform a maximum likelihood estimation, we need to know the transition density. Methods to obtain the transition density include numerical solutions of PDE's using finite difference techniques, analytical approximations, and Monte Carlo methods. One particularly popular analytical approximation is the Euler approximation. In terms of accuracy, the approximation method developed by Aït-Sahalia (1999, 2002) is a significant improvement over the Euler approximation. The idea is to first transform the process into one with a unit diffusion via the transformation $Y_t = \int^{X_t} \mathrm{d}u/\sigma(u; \theta)$ and then approximates the transition density for Y_t using Hermite polynomials. Aït-Sahalia (2008) shows that the same approximation can also be obtained by considering the PDE's the transition density satisfies.

Bakshi, Ju, and Ou-Yang (2006) further solve the expansion coefficients explicitly to fourth order in terms of explicit one-dimensional integrals. In an earlier version of this paper, we also obtained similar result independently, but only to second order. The following proposition is based on Proposition 2 in their paper with several modifications. First, we consider the process X_t directly so a sometimes cumbersome transformation is not needed. Second, we only consider a second-order approximation in Δ. As shown in Aït-Sahalia (2002), Jensen and Poulsen (2002), and others, a second-order approximation is quite accurate for Δ within one month, which is the case for most financial applications. Finally, we work out the degenerate case $x = x_0$ since equation (13) does not applies directly.

Proposition 5 *Consider the process* $\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t$ *with* $\sigma(\cdot) > 0$ *except possibly at the boundaries. Define* $\hat{\mu}(x) \equiv \frac{\mu(x)}{\sigma(x)} - \frac{1}{2}\sigma'(x)$ *and* $\lambda(x) \equiv -\frac{1}{2}(\hat{\mu}^2(x) + \hat{\mu}'(x)\sigma(x))$.

1. The approximate transition density $p_X^{(K)}(\Delta, x|x_0)$ *for the process* X_t *to order* $K = 2$ *in* Δ *is given by*

$$p_X^{(2)}(\Delta, x|x_0) = p_X^{(0)}(\Delta, x|x_0)(1 + c_1(x|x_0)\Delta + c_2(x|x_0)\Delta^2/2), \tag{11}$$

where

$$p_X^{(0)}(\Delta, x|x_0) = \frac{1}{\sqrt{2\pi\Delta}}\sqrt{\frac{\sigma(x_0)}{\sigma^3(x)}}\exp\left(\int_{x_0}^x \frac{\mu(u)}{\sigma^2(u)}\mathrm{d}u - \frac{1}{2\Delta}\left(\int_{x_0}^x \frac{1}{\sigma(u)}\mathrm{d}u\right)^2\right), \tag{12}$$

and for $x \neq x_0$,

$$c_1(x|x_0) = \frac{\int_{x_0}^x \lambda(u)/\sigma(u)\mathrm{d}u}{\int_{x_0}^x 1/\sigma(u)\mathrm{d}u}, \qquad c_2(x|x_0) = c_1(x|x_0)^2 + \frac{\lambda(x) + \lambda(x_0) - 2c_1(x|x_0)}{\left(\int_{x_0}^x 1/\sigma(u)\mathrm{d}u\right)^2}, \tag{13}$$

and for $x = x_0$, $c_1(x_0|x_0) = \lambda(x_0)$ *and* $c_2(x_0|x_0) = \lambda(x_0)^2 + (\sigma(\sigma\lambda')')(x_0)/6$.

2. The series approximation of $\log p_X(\Delta, x|x_0)$ *to order* $K = 2$ *in* Δ *is given by*

$$\log p_X^{(2)}(\Delta, x|x_0) = \log p_X^{(0)}(\Delta, x|x_0) + C_1(x|x_0)\Delta + \frac{1}{2}C_2(x|x_0)\Delta^2, \tag{14}$$

where $C_1(x|x_0) = c_1(x|x_0)$ *and* $C_2(x|x_0) = c_2(x|x_0) - c_1(x|x_0)^2$.

In Appendix B, we provide the approximations for the logarithms of the transition densities for six models to second order in Δ in the original variable by applying the proposition above. The models we consider include the Vasicek, exponentiated Vasicek, CIR, inverse Feller models, the linear drift and CEV diffusion model, and Aït-Sahalia's nonlinear model. These models are frequently used by researchers to model financial quantities. For the first four models, closed-form transition densities exist and are also given. The results in Appendix B should be useful to researchers who want to use these processes in modeling financial quantities.

In the proposition below we point out some nice properties that the expansion coefficients enjoy. The first statement, which we call the invariance property, can be used to derive the series approximation for

14

the exponentiated Vasicek model whose log follows the Vasicek model and for the inverse Feller model whose reciprocal follows the CIR model. The second statement says that the expansion coefficients are symmetric. The third statement says that the expansion coefficients are infinitely differentiable given nice drift and diffusion functions. Proof is in Appendix A.

Proposition 6 *Let* $\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t$ *where* $\sigma(\cdot) > 0$ *except possibly at the boundaries. Let* $Z_t = f(X_t)$ *where* $f(\cdot)$ *is an infinitely differentiable function with* $f'(\cdot) > 0$ *except possibly at the boundaries. Let* $p_X^{(0)}$ *and* $c_i^X(x|x_0)$ *be the zeroth-order approximate density and i-th order expansion coefficients in Aït-Sahalia's series approximation. Similarly for* $p_Z^{(0)}$ *and* $c_i^Z(z|z_0)$. *Then*

1. For any regular points x *and* x_0, *and for any positive integer* i,, *we have*

$$p_Z^{(0)}(\Delta, f(x)|f(x_0)) = \frac{p_X^{(0)}(\Delta, x|x_0)}{f'(x)}, \tag{15}$$

$$c_i^Z(f(x)|f(x_0)) = c_i^X(x|x_0). \tag{16}$$

2. $c_i^X(x|x_0)$ *is symmetric in* x *and* x_0. *That is,* $c_i^X(x|x_0) = c_i^X(x_0|x)$.

3. $c_i^X(x|x_0)$ *is infinitely differentiable in* x *provided that* $1/\sigma(x)$ *and* $\lambda(x)$ *are infinitely differentiable in a connected open set containing* x. *Similarly for* x_0.

The data set we use consists of monthly data on the effective Federal funds rate between July 1954 and June 2008. Thus, it is much longer than the one used in Aït-Sahalia (1999) and Bali and Wu (2006). We report the estimation results both for the period used in Aït-Sahalia (1999) (from January 1963 to December 1998) and for the whole period. Estimating both time periods allows us to compare the results and gain additional insights. The source is the H-15 Federal Reserve Statistical Release (Selected Interest Rate Series). As in Aït-Sahalia (1999), we convert the original data to continuous compounding since the rates are quoted using a 360 day-count convention. In Figure 1, we plot the time series of the Federal funds rate. The sample mean over the whole period is 5.68% and the sample variance is 0.0011. For comparison, the sample mean for the data in Aït-Sahalia (1999) is 6.98% and the sample variance is 0.0010. The fact that our sample mean is smaller can be easily seen from Figure 1, which shows that the Federal funds rates are considerably lower at the beginning and end of our sample period.

As in Aït-Sahalia (1999), we estimate the parameters by maximizing the approximate log-likelihood function

$$\ell(\theta) \equiv \frac{1}{N} \sum_{i=1}^{N} \log p_X(\Delta, X_{i\Delta}|X_{(i-1)\Delta}; \theta) \tag{17}$$

over some parameter space Θ. Here, X_t denotes the interest rate level at time t, and $\Delta = 1/12$ since we have monthly data. We estimate the six models in Appendix B as well as the damped diffusion model introduced in this paper. For each model, we estimate the parameters from its true transition density, the Euler approximation, and the second-order series approximation for $\log p_X$. Closed-form transition

15

densities only exist for four of the models, namely, the Vasicek model, the exponentiated Vasicek model, the CIR model, and the inverse Feller model. For the last three models, the true transition density is computed numerically by solving the backward PDE.[12] For the damped diffusion model, the expansion coefficients C_1 and C_2 for $\log p_X$ are not in closed-form so numerical integration is used.

The results are reported in Table I for the shorter period and Table II for the whole period. Notice that in both tables, for the first four models for which explicit transition densities are available, the parameter estimates obtained from the second-order approximation are almost identical to those obtained from the true densities.[13] This confirms that the second-order series approximation developed by Aït-Sahalia is extremely accurate. On the other hand the Euler approximation is not very accurate, especially when the drift is nonlinear, as expected. In general, by looking at the asymptotic standard errors one can see that the diffusion function is estimated more precisely than the drift function. The CEV and nonlinear drift models have similar diffusion functions and give very similar log-likelihood values, although their drift functions are quite different from each other. The damped diffusion model has a more flexible diffusion function, and it seems to be the case that it is more profitable to specify the diffusion function more precisely than the drift function. This is true for both of the two sample periods, but more so when the whole sample period is used. While it's difficult to compare the performance of nonnested models, the damped diffusion model seems to perform well, at least from the Akaike's information criterion. The nonlinear terms in Aït-Sahalia's nonlinear drift model are both insignificant, as can also be seen from comparing Akaike's information criterion of the CEV and nonlinear drift models. For both of tables, it seems important to allow for more flexibility for the diffusion functions. The first three models, for which the diffusion functions have more restrictive parametric forms, seem to not perform as well.

Comparing Table I with Table II gives some additional information. First, all the estimates for m in

---

[12] We comment here a little bit on the PDE approach, which could be tricky for someone less experienced. First, the terminal boundary condition is a Dirac delta function which has to be approximated. Jensen and Poulsen (2002) solve this problem by approximating the transition density for the last time step with an Euler approximation. We find that the following approximation works better. Specifically, let y be a fixed future interest rate which lies within two adjacent spatial grid points y_{m-1} and y_m where $y_m > y_{m-1}$. Then, the Dirac delta function $\delta(x - y)$ is approximated as two columns:

$$\delta(x - y) = \frac{1}{y_m - y_{m-1}} \left(\frac{y_m - y}{y_m - y_{m-1}} 1_{x=y_{m-1}} + \frac{y - y_{m-1}}{y_m - y_{m-1}} 1_{x=y_m} \right). \tag{18}$$

Like the Euler approximation, this choice of approximation keeps the property that the Dirac delta function should integrate to 1. But it is usually more focused than the Euler approximation and thus mimics the Dirac delta function better. Second, since near the terminal boundary the spatial derivative could be extremely large near the peak of the Dirac delta function, we subdivide the last time step into 10 subintervals and use a fully implicit finite difference scheme to propagate the solution. The rest of the time steps uses the Crank-Nicholson scheme. Similar method is used in Li, Pearson and Poteshman (2004). With these improvements, the accuracy of our PDE approach seems to be in the order of 10^{-3} when tested on models with explicit transition densities, one order of magnitude more accurate than reported in Jensen and Poulsen (2002). However, it is still a little bit less accurate than the second-order series approximation (done for $\log p_X$). Also, searching the optimal parameters for each model with the PDE approach can take well over one day. This highlights the usefulness of Aït-Sahalia's series approximation method.

[13] In Aït-Sahalia (1999), the parameter estimates from the series approximation sometimes are not very close to those obtained from the true densities. This is because that paper used a second-order approximation for the transition densities, rather than for the log of the transition densities. The latter approach, which we use here, is also used in Aït-Sahalia (2002).

the drift function for the longer sample period are smaller than their counterparts for the shorter sample period. This is to be expected since the value of m is closely related to the long-run mean of the processes. Second, the results for the strength of mean-reverting parameter κ (excluding the nonlinear drift model) are mixed. For some of the models, κ becomes larger when we use the longer sample, while for other models, κ becomes smaller. Since the CEV model and the damped diffusion model seem to perform better than other models, the strength of mean-reverting probably increases when we use longer sample. However, we must caution ourselves that the drift functions as a rule are not estimated as accurately as the diffusion functions. Last, but the most important difference, is that the diffusion functions seem to change behavior significantly when we use the longer sample period. This can be most easily seen from the CEV model. The estimate of ρ in Table I is about 1.48, while ρ is estimated to be about 0.62 in Table II. This finding is consistent with previous research, which suggests that the large estimate of ρ probably comes from the unusual period in the early 1980's. Related with this change of magnitude for ρ, the relative performance of the CIR model is much improved for the whole sample period, while that of the inverse Feller model deteriorates dramatically.

B. Finite sample bias – A Monte-Carlo experiment

We now show through a Monte-Carlo experiment that with finite sample sizes, the estimation method one uses often fails to pick up the damping function even if the actual data are generated from a damped diffusion process. Furthermore, often one will find a significant nonlinear drift function even if the actual data are generated from a linear drift process. This has the implication that it is difficult to detect damped diffusions empirically with short time series.

Specifically, we simulate 500 sample paths from a damped CEV process with $\mu(x) = \kappa(m - x)$ and $\sigma(x) = \sigma x^\rho \exp(-\alpha x^\beta)$. The parameters are taken from the CEV model in Table I, that is, $\kappa = 0.0886$, $m = 0.0842$, $\sigma = 0.7792$ and $\rho = 1.4812$. We set $\alpha = 8$ and $\beta = 4$ for the damping function. We use an Euler scheme with time interval about one hour to generate monthly time series with length 36 years, the same data structure as the federal funds rate we used. We then use the series approximation method to estimate the six parameters for those 500 sample paths. The median of the parameter estimates for α is 7.2324 while the median for β is 3.4973. However, out of the 500 sample paths and at 90% confidence level, only 17 paths report a significantly nonzero α and only 13 paths report a significantly nonzero β. The reason for this is that the damping function is only effective when interest rate reaches relatively high level. Thus if the maximum of a sample path happens to be small, the damping function is not revealed by the maximum likelihood estimation.

However, while the estimation fails to detect the damped diffusion function for the majority of the sample paths, it often mistakes the damped diffusion function for a nonlinear drift function. Specifically, we take the same 500 sample paths and estimate them using Aït-Sahalia's nonlinear drift model with

CEV diffusion function and no damping. At 90% confidence level, 67 paths report a significant α_2 while 68 paths report a significant α_{-1}.

The above results show that finite sample size tends to fabricate nonlinear drift function while fail to detect the damped diffusion function. To make this clearer, we simulate 500 new sample paths using the same parameters. However, now the length of the sample paths is 360 years, 10 times longer than before. We again estimate them using the damped CEV model. Now 136 sample paths report a significantly nonzero α and 189 sample paths report a significantly nonzero β. While still more than half of the paths fail to detect the damping function, the number of sample paths that do is much more than the previous case of shorter sample paths. On the other hand, if we estimate the longer time series using the nonlinear drift model with CEV diffusion and no damping, only 6 paths report significantly nonzero α_2 and only 8 paths report significantly nonzero α_{-1}.

That the maximum likelihood estimation often fails to detect the damping function but instead fabricates the nonlinear drift is a drawback of the maximum likelihood estimation. Since by the construction, the damping function usually only kicks in when one has unusual observations, it will generally be difficult to estimate the damping function if the sample size is small. There are a few possible solutions. First, it might be easier to detect the damping function with some other estimation procedures.[14] A generalized method of moments method with suitably chosen moments might work because intuitively, the empirical moments for any finite sample would be finite. Take the CEV model in Table I for example, in the steady state the process has an infinite variance without the damping function. If we could force the steady state variance to match the sample variance by some sort of ergodicity argument, then there would be a need for a damping function. Another method might be to utilize the marginal density of the process (see Aït-Sahalia 1996b). The marginal density of the process would be very different with and without the damping function, at least near the tails. Arapis and Gao (2006) perform a nonparametric estimation of the 7-day Eurodollar deposit rates from June 1973 to February 1995 by utilizing the marginal density. Indeed, Figure 5 of their paper seems to indicate that the diffusion function is damped if one fixes a linear drift function. Second, while it is not easy to detect the damping function using the discretely observed data of the process itself, it might be possible to detect the damping function from other related data. For example, a CEV interest rate process with and without damping would have different implications for the pricing of interest rate derivatives, in particular, interest rate options such as caplets and floorlets. Thus, we might be able to better detect the damping function if we estimate using both the interest rate data and the interest rate derivatives data. We hope that further research along these directions would shed more light on the damped diffusion framework.

---

[14]Of course, one might also argue that a procedure which detects the damping functions better might also tend to fabricate the damping function when the true data generating process has no damping but rather a nonlinear drift.

IV. Conclusion

In this paper, we propose a damped diffusion framework for financial modeling. We first motivate this framework by considering some of the drawbacks of the popular CEV model. First, the invariant distribution of the CEV model has an infinite variance if the exponent is greater than 1, which is the case for many financial applications. Second, we perform a Monte Carlo study of the sample paths for this process generated using parameters estimated from actual short-rate data. We find that many sample paths reach unreasonably high interest levels during a finite period. Contrary to Aït-Sahalia's (1996a) approach of changing the drift function from linear to nonlinear, we modify the diffusion function through damping while keeping the appealing linear drift function. We show that the choice of the damping function is very flexible by considering a damped superimposed CEV model. With suitable choices of the damping function, the damped diffusion can always be made to have a strong unique solution and nonexplosive.

Another drawback of the CEV process has significant economic implications. Heston, Loewenstein and Willard (2007), Sin (1998), and Lewis (2000) have shown that if one uses the CEV process for the underlying stock or the instantaneous stochastic volatility, the usual martingale pricing approach can fail. This failure is not limited to the CEV process and can happen quite unintentionally when one models financial quantities using diffusion processes. When this happens, asset prices often contain bubbles. We show that the damped diffusion framework can be used to preserve the martingale pricing approach and eliminate many types of bubbles, including stock future bubbles, option bubbles and bond bubbles. In addition, we show that although sometimes the asset valuation PDE has multiple solutions, many solutions can be ruled out by requiring that the solution be the limit of martingale prices in successively weakly damped diffusion models. In the case of CIR short-rate process and linear premium function $\psi_0 + \psi_1 r$, while Heston, Loewenstein and Willard (2007) show that Cox, Ingersoll and Ross's (1985) conjecture that a nonzero ψ_0 always leads to arbitrage is in general not true, our result shows that in some weaker sense, any positive ψ_0 can be allowed.

Finally, we carry out maximum likelihood estimation using Aït-Sahalia's (1999) series expansion method. We point out some nice properties of this method, including the symmetry, differentiability and the invariance property of the expansion coefficients. In Appendix B, we provide the expansion coefficients to second order for the log-densities of many diffusion models commonly used in finance. Through a Monte Carlo experiment, we show that with finite time-series length, maximum likelihood estimation often fails to detect the damped diffusion function while fabricates nonlinear drift function.

Appendix A

Proof of Proposition 1: We first prove the following mathematical fact. Let $f(x) : \mathbb{R} \to \mathbb{R}$ be any smooth function. Then, for any $M > 0$ and $\epsilon > 0$, there exists a smooth function $\widetilde{f}(x) : \mathbb{R} \to \mathbb{R}$ such that

$$\int_{-M}^{M} |f(x) - \widetilde{f}(x)| \, \mathrm{d}x < \epsilon, \tag{19}$$

and $\widetilde{f}(x)$ satisfies the global Lipschitz and linear growth conditions on $\mathbb{R}$. To prove this claim, we use the method of mollification (see, for example, Appendix C of Evans 1998). First we construct the truncation of the derivative $f'(x)$. Let $h(x) = f'(x)$ on $[-M, M]$, $h(x) = f'(-M)$ if $x < -M$, and $h(x) = f'(M)$ if $x > M$. Then $h(x)$ is a bounded continuous function since $h(x) \le K \equiv \sup_{u \in [-M,M]} f'(u)$. Now let $\eta(x)$ be the standard mollifier. That is, $\eta(x) = C \exp(1/(x^2 - 1))$ if $|x| < 1$ and zero otherwise. The constant C is chosen so that $\eta(x)$ integrates to 1 on $\mathbb{R}$. Let $\delta > 0$ and $h^\delta(x)$ be the δ-mollification of $h(x)$:

$$h^\delta(x) = \int_{\mathbb{R}} \frac{1}{\delta} \eta\left(\frac{x-y}{\delta}\right) h(y) \mathrm{d}y. \tag{20}$$

By the standard argument of mollification, $h^\delta(x)$ is infinitely differentiable in x and $h^\delta \to h$ uniformly on any compact subset of $\mathbb{R}$ (in particular, $[-M, M]$) as $\delta \to 0$. Furthermore, $h^\delta(x)$ is bounded on $\mathbb{R}$ by K since $h(x)$ is. Pick δ small enough so that $\sup_{u \in [-M,M]} |h(u) - h^\delta(u)| < \epsilon/(2M^2)$. For this fixed δ, define a new function $\widetilde{f}(x)$ by

$$\widetilde{f}(x) = f(0) + \int_0^x h^\delta(u) \mathrm{d}u. \tag{21}$$

Then,

$$\int_{-M}^{M} |f(x) - \widetilde{f}(x)| \, \mathrm{d}x = \int_{-M}^{M} \left| \int_0^x h(u) - h^\delta(u) \mathrm{d}u \right| \mathrm{d}x \le 2M \cdot M \cdot \frac{\epsilon}{2M^2} = \epsilon. \tag{22}$$

Furthermore, the derivative of $\widetilde{f}(x)$ (which equals $h^\delta(x)$) is bounded by K. Thus, $\widetilde{f}(x)$ is globally Lipschitz on $\mathbb{R}$ and has a linear growth.

Without loss of generality, assume that the domain of X_t is $\mathbb{R}$ since we can always perform a monotone transformation to transform the domain (a, b) to $\mathbb{R}$ and later use the inverse transformation on $\widetilde{X}_t$. Thus, in particular, we could assume $g > 0$ on $\mathbb{R}$. By the mathematical fact established above, we can always perform a "small" modification to f and g so that $\widetilde{f}$ and $\widetilde{g}$ are globally Lipschitz and have linear growth. By the standard result in the theory of stochastic differential equations (see, for example, Chapter 5 of Karatzas and Shreve 1991), the stochastic differential equation $\mathrm{d}\widetilde{X}_t = \widetilde{f}(\widetilde{X}_t)\mathrm{d}t + \widetilde{g}(\widetilde{X}_t)\mathrm{d}W_t$ has a strong unique solution. The domain $(\widetilde{a}, \widetilde{b})$ of $\widetilde{X}_t$ depends on the zeroes of $\widetilde{g}$ and could possibly be smaller than $\mathbb{R}$. In any case, the solution $\widetilde{X}_t$ does not explode to $\widetilde{a}$ or $\widetilde{b}$ by the linear growth and global Lipschitz conditions on $\widetilde{f}$ and $\widetilde{g}$.

Remark: Since g and $\widetilde{g}$ can be made extremely close on any compact subset of $\mathbb{R}$, $\widetilde{g}$ can be made positive on any compact subset of $\mathbb{R}$. Thus, although the domain of $\widetilde{X}_t$ could be smaller than $\mathbb{R}$, it could be made as large as one wants to contain any pre-specified compact subset of $\mathbb{R}$.

Proof of Proposition 2: For statement 1, we first show that there exists $C > 0$ such that for all $x, y \geq 0$ $|x^{1/2}D_1(x) - y^{1/2}D_1(y)| \leq C\sqrt{|y - x|}$. The statement is obvious for $x = 0$ so without loss of generality, assume $y \geq x > 0$. Let $g(y) \equiv C^2(y - x) - (x^{1/2}D_1(x) - y^{1/2}D_1(y))^2$. We want to show $g(y) \geq 0$ for all $y \geq x$. By the assumptions that $D_1(y)$ and $yD_1(y)$ are bounded, for sufficiently large C, we have $g'(y) \geq 0$ for all $y \geq x$. Since $g(x) = 0$, we have $g(y) \geq 0$ for all $y \geq x$. Now let $\sigma(x) = \sigma_1 x^{1/2}D_1(x) + \sigma_2 x^\rho D_2(x)$. Since $\sigma_2 x^\rho D_2(x)$ is global Lipschitz, by triangular inequality, we have $|\sigma(x) - \sigma(y)| \leq h(|x - y|) \equiv C'(\sqrt{|y - x|} + |y - x|)$, for some C' large. Now $h(\cdot)$ is strictly increasing, $h(0) = 0$ and for any $\epsilon > 0$, we have $\int_{(0,\epsilon)} h^{-2}(u)\mathrm{d}u = \infty$. The statement now follows from a version of Yamada and Watanabe (1971) (see Karatzas and Shreve (1991)).

Statements 2 and 3 are standard results in stochastic differential equation theory (see, for example, Chapter 5 of Karatzas and Shreve (1991) and Chapters 2 to 4 of Mao (1997)).

Proof of Equation (6): Noticing $S_\tau = (x_\tau/k)^{-v}$, we have

$$F^2 = \int_0^\infty p(\tau, S_\tau|S_0)S_\tau = \int_0^\infty \left(\frac{x_\tau}{x_0}\right)^{v/2} e^{-x_0 - x_\tau} I_v(2\sqrt{x_0 x_\tau}) \left(\frac{x_\tau}{k}\right)^{-v} \mathrm{d}x_\tau. \tag{23}$$

Using the series expansion of $I_v(z)$

$$I_v(z) = \sum_{n=0}^\infty \frac{1}{n!} \frac{1}{\Gamma(v + n + 1)} \left(\frac{z}{2}\right)^{2n+v} \tag{24}$$

and interchanging the order of summation and integration, we get

$$F^2 = \left(\frac{x_0}{k}\right)^{-v} \cdot \sum_{n=0}^\infty \frac{x_0^{n+v} e^{-x_0}}{\Gamma(v + n + 1)} = \left(\frac{x_0}{k}\right)^{-v} \cdot \left(1 - \frac{\Gamma(v, x_0)}{\Gamma(v, 0)}\right) = S_0 e^{r\tau}\left(1 - \frac{\Gamma(v, x_0)}{\Gamma(v, 0)}\right). \tag{25}$$

Proof of Proposition 3: For statement 1, notice that $g(S)$ is Lipschitz and satisfies the growth condition $g^2(S) \leq C(1 + S^2)$ for some constant C. Statement 1 now follows from Proposition 2.

For statement 2, notice that the fact that S_τ^{2n} is integrable implies that $\int_0^t g(S)\mathrm{d}W^\mathbb{Q}$ is a genuine $\mathbb{Q}$-martingale since $g(S) \leq MS$. Thus $\mathbb{E}^\mathbb{Q} S_\tau - S_0 = \int_0^\tau \mathbb{E}^\mathbb{Q} S_u \mathrm{d}u$, giving $\mathbb{E}^\mathbb{Q} S_\tau = F^1 \equiv S_0 e^{r\tau}$. This in turn implies the usual put-call parity holds.

For statement 3, let the Radon-Nikodym derivative be Λ. Then $\mathrm{d}\Lambda = \Lambda\mathrm{d}L$, where the dynamics of the exponent process L is $\mathrm{d}L = \sigma S^{\alpha-1} D_\lambda \mathrm{d}W^\mathbb{Q} = g(S)S^{-1}\mathrm{d}W^\mathbb{Q}$ by Girsanov's theorem. Since $\mathbb{E}^\mathbb{Q}[\exp(\langle L, L\rangle_\tau/2)] = \mathbb{E}^\mathbb{Q}[\exp(\int_0^\tau g(S_u)^2 S_u^{-2}\mathrm{d}u/2)] \leq e^{M\tau^2/2} < \infty$ for any fixed τ, Novikov's condition guarantees Λ is a genuine $\mathbb{Q}$-martingale and the measure change is equivalent. Thus the stock price under $\mathbb{Q}^S$ is nonexplosive.

For statement 4, apply Feynman-Kac theorem such as Theorem 5.7.6 in Karatzas and Shreve (1991).

For statement 5, we need a comparison theorem for processes with different diffusion functions. Theorem 4.1 in Hajek (1985) suits us well. For reader's convenience, we copy this theorem below:

Theorem 4.1. (Hajek (1985)) *Suppose that m and ρ are each convex Lipschitz continuous functions on $\mathbb{R}$, and suppose that μ and σ are Borel measurable functions on $\mathbb{R} \times \mathbb{R}_+$ such that for some constant K and all θ, θ' in*

$\mathbb{R}$ and $t \geq 0$,

$$|\mu(\theta, t) - \mu(\theta', t)| + |\sigma(\theta, t) - \sigma(\theta', t)| \leq K|\theta - \theta'|, \tag{26}$$

$$|\sigma(\theta, t)| + |\mu(\theta, t)| \leq K(1 + |\theta|). \tag{27}$$

Let X and Y be solutions to the stochastic differential equations:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t, \tag{28}$$

$$dY_t = m(Y_t)dt + \rho(Y_t)dB_t, \tag{29}$$

where W and B are Wiener processes. Suppose that for all θ, t,

$$\mu(\theta, t) \leq m(\theta), \quad 0 \leq \sigma(\theta, t) \leq \rho(\theta), \tag{30}$$

and that X_0 and Y_0 are constants with $X_0 \leq Y_0$. Then

$$\mathbb{E}\Phi(X_t) \leq \mathbb{E}\Phi(Y_t) \tag{31}$$

for any nondecreasing convex function Φ on $\mathbb{R}$.

Now take two damped diffusion models with damping functions D_{λ_1} and D_{λ_2} where $\lambda_1 > \lambda_2$. Denote the corresponding diffusion functions by g_1 and g_2 and the corresponding stock price processes under $\mathbb{Q}$ by S_1 and S_2. Then $g_1(\cdot) < g_2(\cdot)$. Let S_1 and S_2 start from the common value S_0. Consider the nondecreasing convex function $\Phi(x) \equiv e^{-r\tau}(x - K)^+$ for fixed $K > 0$. Then Theorem 4.1 tells us that $\mathbb{E}^{\mathbb{Q}}\Phi(S_{1_\tau}) \leq \mathbb{E}^{\mathbb{Q}}\Phi(S_{2_\tau})$, i.e., $G_{\lambda_1} \leq G_{\lambda_2}$. Thus G_λ is a decreasing function in λ. Now take a sequence $\lambda_n \downarrow 0^+$. Then G_{λ_n} is an increasing sequence on the real line which is bounded below by 0 and above by $e^{-r\tau}\mathbb{E}^{\mathbb{Q}}S_\tau = S_0$. Thus G_{λ_n} approaches a fixed limit. Notice that put-call parity $G_\lambda - P_\lambda = S - Ke^{-r\tau}$ holds and P_λ is bounded by $Ke^{-r\tau}$. Since G_λ increases to $\widetilde{G}$, P_λ also increases to a limit $\widetilde{P}$ which is bounded by $Ke^{-r\tau}$. Thus $\widetilde{G} - \widetilde{P} = S - Ke^{-r\tau}$. The result $\lim_{S \to \infty} \widetilde{G}/S = 1$ follows immediately from the put-call parity. The put-call parity of G_λ and P_λ also gives $(S - Ke^{-r\tau})^+ \leq G_\lambda \leq S$. Taking limit on λ gives $(S - Ke^{-r\tau})^+ \leq \widetilde{G} \leq S$. Since each G_λ is increasing in S, $\partial \widetilde{G}/\partial S \geq 0$.

Proof of Proposition 4: For statement 1, notice that $\alpha(\beta - r) - \Psi_\lambda(r)$ is global Lipschitz since it has a bounded derivative. Thus the conditions of Yamada and Watanabe (1971) are satisfied with $h(r) = \sigma\sqrt{r}$. This shows $dr = (\alpha(\beta - r) - \Psi_\lambda(r))dt + \sigma\sqrt{r}dW^{\mathbb{Q}}$ has a strong unique solution. Since $\Psi_\lambda(\cdot) > 0$ and $2\alpha\beta > \sigma^2$, a comparison theorem (for example, Proposition 5.2.18 in Karatzas and Shreve (1991)) tells us that the above process does not explode to infinity. Now the assumption on the Feller's explosion test at $r \to 0^+$ tells us that the process does not explode to 0 either.

For statement 2, from theorem A.1 in Heston, Loewenstein and Willard (2007), the measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent. Thus there is a unique martingale bond price $P_\lambda(r, T)$. If $r_1(t)$ and $r_2(t)$ are two solutions to the above process with initial interest rate $r_1(0) < r_2(0)$, then by the comparison theorem 5.2.18 in Karatzas and Shreve (1991), we have $P[r_1(t) < r_2(t), \forall \, 0 \leq t \leq \infty] = 1$. Since $P_\lambda(r, T) = \mathbb{E}^{\mathbb{Q}}[e^{-\int_0^T r(t)dt}]$, it is a decreasing function of r. Although not need in the proof, it can be further shown by mimicking the proof of Proposition 5.2.13 in Karatzas and Shreve (1991) that for any positive λ, $P_\lambda(r, T)$ has a bounded partial derivative in r. However, the bounds depend on λ. $P_\lambda(r, T)$ is decreasing in T since

$$\frac{\partial P_\lambda(r, T)}{\partial T} = -\mathbb{E}^{\mathbb{Q}}\left[r(T)\exp\left(-\int_0^T r(t)dt\right)\right] < 0. \tag{32}$$

22

Now let $R(t)$ follow the CIR process $\mathrm{d}R = \alpha(\beta - R)\mathrm{d}t + \sigma\sqrt{R}\mathrm{d}W^{\mathbb{Q}}$ with $R(0) = r(0) \equiv r$. By comparison theorem, $\mathbb{E}^{\mathbb{Q}}[r(T)] \leq \mathbb{E}^{\mathbb{Q}}[R(T)]$. Since

$$\frac{\partial P_\lambda(r, T)}{\partial T} \geq -\mathbb{E}^{\mathbb{Q}}[r(T)] \geq -\mathbb{E}^{\mathbb{Q}}[R(T)] = (\beta - r)e^{-\alpha T} - \beta, \tag{33}$$

the partial derivative $\partial P_\lambda(r, T)/\partial T$ is bounded above and below with the bounds independent of λ.

For statement 3, notice that $\Psi_\lambda(r)$ is decreasing in λ. The comparison theorem 5.2.18 in Karatzas and Shreve (1991) with different λ thus shows that $P_\lambda(r, T)$ is an increasing function in λ for fixed r and T. Since $P_\lambda(r, T)$ are bounded by 1, $P_\lambda(r, T)$ converges pointwise as $\lambda \to 0^+$. $\widetilde{P}(r, T)$ is decreasing in both r and T since for each λ, $P_\lambda(r, T)$ is decreasing in both r and T. The derivative $\partial\widetilde{P}/\partial T$ is bounded above by 0 and below by $(\beta - r)e^{-\alpha T} - \beta$.

Now the existence of such a damping function is the most difficult part of the proof. All other assumptions are easy to check except for the condition on Feller's explosion test. Let $D_\lambda = r/(r + \lambda)$. Choose an arbitrary $c \in (0, \infty)$. Using Karatzas and Shreve's (1991) notation, the scale density of the above process is given by

$$p'_c(x) = \exp\left(2\int_x^c \frac{\alpha(\beta - z) - \Psi_\lambda(z)}{\sigma^2 z}\mathrm{d}z\right) = \left(\frac{c}{x}\right)^{\frac{2\alpha\beta}{\sigma^2}}\left(\frac{x + \lambda}{c + \lambda}\right)^{\frac{2\psi_0}{\sigma^2}}\exp\left(\frac{2(\alpha + \psi_1)}{\sigma^2}(x - c)\right). \tag{34}$$

Thus for $y \leq c$, there exists constant $M > 0$ such that

$$\int_y^c \frac{2\,\mathrm{d}z}{p'_c(z)\sigma^2 z} \geq \int_y^c \frac{2}{\sigma^2}\left(\frac{1}{c}\right)^{\frac{2\alpha\beta}{\sigma^2}}z^{\frac{2\alpha\beta}{\sigma^2} - 1}\mathrm{d}z \equiv M\left(c^{\frac{2\alpha\beta}{\sigma^2}} - y^{\frac{2\alpha\beta}{\sigma^2}}\right). \tag{35}$$

Notice that the damping function is essential here because we need λ to be strictly positive in this proof. We can now verify the condition in Feller's explosion test as follows:

$$\lim_{x\downarrow 0} v_c(x) = \lim_{x\downarrow 0}\int_x^c p'_c(y)\mathrm{d}y\int_y^c \frac{2\,\mathrm{d}z}{p'_c(z)\sigma^2 z} \tag{36}$$

$$\geq \int_0^c \left(\frac{c}{y}\right)^{\frac{2\alpha\beta}{\sigma^2}}\left(\frac{y + \lambda}{c + \lambda}\right)^{\frac{2\psi_0}{\sigma^2}}\exp\left(\frac{2(\alpha + \psi_1)}{\sigma^2}(y - c)\right)M\left(c^{\frac{2\alpha\beta}{\sigma^2}} - y^{\frac{2\alpha\beta}{\sigma^2}}\right)\mathrm{d}y \tag{37}$$

$$\geq c^{\frac{2\alpha\beta}{\sigma^2}}\left(\frac{\lambda}{c + \lambda}\right)^{\frac{2\psi_0}{\sigma^2}}\exp\left(-\frac{2(\alpha + \psi_1)}{\sigma^2}c\right)M\int_0^c \left((c/y)^{\frac{2\alpha\beta}{\sigma^2}} - 1\right)\mathrm{d}y = +\infty. \tag{38}$$

Proof of Proposition 5: The bulk of this proposition is proved in Bakshi, Ju and Ou-Yang (2006). To change to the original variable, notice that $y \equiv f(x) = \int^x 1/\sigma(u)\mathrm{d}u$ and $\mathrm{d}y = \mathrm{d}x/\sigma(x)$. The degenerate case $x = x_0$ can be obtained through an application of L'Hospital's rule.

Proof of Proposition 6: For statement 1, notice that by Ito's lemma, $\mathrm{d}Z \equiv \mathrm{d}f = (f'\mu + f''\sigma^2/2)\mathrm{d}t + f'\sigma\mathrm{d}W$. Consider $\hat{\mu}(\cdot)$ and $\lambda(\cdot)$ defined in Proposition 5. Simple algebra shows that $\hat{\mu}^Z \circ f = \hat{\mu}^X$ and $\lambda^Z \circ f = \lambda^X$, where "$\circ$" is the function composition. By using the fact that $\mathrm{d}z = f'(x)\mathrm{d}x$, the expression for c_1 in Proposition 5 then gives $c_1^Z(f(x)|f(x_0)) = c_1^X(x|x_0)$. The relation for $p_Z^{(0)}$ and $p_X^{(0)}$ can be obtained similarly. The recursive relation for c_i then inductively gives $c_i^Z(f(x)|f(x_0)) = c_i^X(x|x_0)$ for all i.

For statement 2, by statement 1, we need only consider a unit diffusion process by letting $Y_t \equiv \int^{X_t} 1/\sigma(x) \mathrm{d}x$. Notice we have the following differential forms of the recursive relations for the $c_k^Y(y|y_0)$'s:

$$c_{k+1}^Y + \frac{y - y_0}{k+1} \frac{\partial c_{k+1}^Y}{\partial y} = \lambda(y) c_k^Y + \frac{1}{2} \frac{\partial^2 c_k^Y}{\partial y^2}, \quad c_{k+1}^Y - \frac{y - y_0}{k+1} \frac{\partial c_{k+1}^Y}{\partial y_0} = \lambda(y_0) c_k^Y + \frac{1}{2} \frac{\partial^2 c_k^Y}{\partial y_0^2}. \tag{39}$$

Aït-Sahalia (2004) showed that the integral representation in Proposition 5 is the only solution that satisfies both differential recursive relations. The claim that $c_k^Y(y|y_0) = c_k^Y(y_0|y)$ follows from the observation that those differential recursive relations are formally symmetric in y and y_0 and mathematical induction.

For statement 3, again by statement 1, we need only consider the process Y_t since the infinite differentiability of c_i^X follows from the infinite differentiability of c_i^Y and the chain rule. Notice

$$c_1^Y(y|y_0) = \int_0^1 \lambda_Y(y_0 + t(y - y_0)) \, \mathrm{d}t, \tag{40}$$

$$c_k^Y(y|y_0) = \int_0^1 kt^{k-1} \left(\lambda_Y(w) c_{k-1}^Y(w|y_0) + \frac{\partial^2 c_{k-1}^Y(w|y_0)}{2 \, \partial w^2} \right) \Bigg|_{w = y_0 + t(y - y_0)} \mathrm{d}t. \quad (k > 1) \tag{41}$$

Now take the derivatives inside the integral by chain rules and use mathematical induction on the smoothness of c_{k-1}^Y to conclude that $c_k^Y(y|y_0)$ is infinitely differentiable in y. The infinite differentiability of c_k^Y with respect to y_0 is now obvious from the symmetry of $c_k^Y(y|y_0)$ shown in statement 2.

Appendix B

Below we list the expansion coefficients and $p_X^{(0)}$ for some commonly used processes in finance. For the first four processes, closed-form transition densities exist and are also given below. Recall that the series approximation of $\log p_X(\Delta, x|x_0)$ to order $K = 2$ in Δ is given by

$$\log p_X^{(2)}(\Delta, x|x_0) = \log p_X^{(0)}(\Delta, x|x_0) + C_1(x|x_0)\Delta + \frac{1}{2}C_2(x|x_0)\Delta^2. \tag{42}$$

Vasicek model $\mathrm{d}X_t = \kappa(m - X_t)\mathrm{d}t + \sigma\mathrm{d}W_t$:

$$p_X^{(0)}(\Delta, x|x_0) = \frac{1}{\sqrt{2\pi\Delta}\sigma}\exp\left(-\frac{(x-x_0)^2}{2\sigma^2\Delta} + \frac{mx\kappa}{\sigma^2} - \frac{x^2\kappa}{2\sigma^2} - \frac{mx_0\kappa}{\sigma^2} + \frac{x_0^2\kappa}{2\sigma^2}\right), \tag{43}$$

$$C_1(x|x_0) = \frac{\kappa}{2} - \frac{m^2\kappa^2}{2\sigma^2} + \frac{mx\kappa^2}{2\sigma^2} + \frac{m\kappa^2 x_0}{2\sigma^2} - \frac{x^2\kappa^2}{6\sigma^2} - \frac{x\kappa^2 x_0}{6\sigma^2} - \frac{\kappa^2 x_0^2}{6\sigma^2}, \tag{44}$$

$$C_2(x|x_0) = -\kappa^2/6. \tag{45}$$

Let $n(z; \mu, V)$ denote the probability density function of a normally distributed random variable Z with mean μ and variance V. Then, the true transition density of the Vasicek model is given by

$$p_X(\Delta, x|x_0) = n\left(x; m + (x_0 - m)e^{-\kappa\Delta}, \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa\Delta})\right). \tag{46}$$

Exponentiated Vasicek model $\mathrm{d}X_t = \kappa X_t\big(\log m - \log X_t\big)\mathrm{d}t + \sigma X_t\mathrm{d}W_t$:

$$p_X^{(0)}(\Delta, x|x_0) = \sqrt{\frac{x_0}{2\pi\sigma^2\Delta x^3}}\exp\left(-\frac{(\log(x/x_0))^2}{2\sigma^2\Delta} + \kappa\frac{2\log m\log(x/x_0) + (\log x_0)^2 - (\log x)^2}{2\sigma^2}\right), \tag{47}$$

$$C_1(x|x_0) = \frac{\kappa}{2} - \frac{\sigma^2}{8} + \frac{\kappa}{4}\log\frac{m^2}{xx_0} + \frac{\kappa^2\log m}{2\sigma^2}\log\frac{xx_0}{m} - \frac{\kappa^2}{6\sigma^2}\big((\log x)^2 + \log x\log x_0 + (\log x_0)^2\big), \tag{48}$$

$$C_2(x|x_0) = -\kappa^2/6. \tag{49}$$

Letting $m' = \log m - \sigma^2/(2\kappa)$, the true transition density of the exponentiated Vasicek model is given by

$$p_X(\Delta, x|x_0) = \frac{1}{x}n\left(\log x; m' + (\log x_0 - m')e^{-\kappa\Delta}, \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa\Delta})\right). \tag{50}$$

Cox-Ingersoll-Ross model $\mathrm{d}X_t = \kappa(m - X_t)\mathrm{d}t + \sigma\sqrt{X_t}\mathrm{d}W_t$:

$$p_X^{(0)}(\Delta, x|x_0) = \frac{1}{\sqrt{2\pi\Delta x}\sigma}\left(\frac{x}{x_0}\right)^{\kappa m/\sigma^2 - 1/4}\exp\left(-\frac{\kappa(x-x_0)}{\sigma^2} - \frac{2(\sqrt{x} - \sqrt{x_0})^2}{\sigma^2\Delta}\right), \tag{51}$$

$$C_1(x|x_0) = \frac{\kappa^2\left(6m - x - \sqrt{xx_0} - x_0\right)}{6\sigma^2} + \frac{16\kappa m - 3\sigma^2}{32\sqrt{xx_0}} - \frac{\kappa^2 m^2}{2\sigma^2\sqrt{xx_0}}, \tag{52}$$

$$C_2(x|x_0) = \frac{48m\kappa\sigma^2 - 9\sigma^4 - 16\kappa^2\left(3m^2 + xx_0\right)}{384xx_0}. \tag{53}$$

If we let $\chi^2(z; f, \theta)$ to denote the probability density function of a noncentral chi-square distributed random variable Z with degree of freedom f and noncentrality θ, then the true transition density for the

CIR process is given by

$$p_X(\Delta, x|x_0) = 2c\,\chi^2(2cx; 2q+2, 2cx_0 e^{-\kappa\Delta}), \tag{54}$$

with

$$q = \frac{2\kappa m}{\sigma^2} - 1, \quad c = \frac{2\kappa}{\sigma^2(1 - e^{-\kappa\Delta})}. \tag{55}$$

Inverse of Feller's square root model: $\mathrm{d}X_t = \kappa(m - X_t)X_t\mathrm{d}t + \sigma X_t^{3/2}\mathrm{d}W_t$

$$p_X^{(0)}(\Delta, x|x_0) = \frac{(x_0/x)^{\kappa/\sigma^2 + 3/4}}{\sqrt{2\pi\Delta}\sigma x^{3/2}}\,\exp\left(\frac{km(x - x_0)}{\sigma^2 x x_0} - \frac{2\left(\sqrt{x} - \sqrt{x_0}\right)^2}{\Delta\sigma^2 x x_0}\right), \tag{56}$$

$$C_1(x|x_0) = m\kappa - \frac{\left(16\kappa + 3\sigma^2\right)\sqrt{xx_0}}{32} + \frac{\kappa^2\left(2m - \sqrt{xx_0}\right)}{2\sigma^2} - \frac{m^2\kappa^2\left(x + x_0 + \sqrt{xx_0}\right)}{6\sigma^2 x x_0}, \tag{57}$$

$$C_2(x|x_0) = -\frac{16m^2\kappa^2 + 48\kappa^2 x x_0 + 48\kappa\sigma^2 x x_0 + 9\sigma^4 x x_0}{384}. \tag{58}$$

The true transition density for the inverse Feller process is given by

$$p_X(\Delta, x|x_0) = \frac{2c'}{x^2}\,\chi^2\left(\frac{2c'}{x}; 2q'+2, \frac{2c}{x_0}e^{-\kappa m\Delta}\right), \tag{59}$$

with

$$q' = \frac{2\kappa}{\sigma^2} + 1, \quad c' = \frac{2\kappa m}{\sigma^2(1 - e^{-\kappa m\Delta})}. \tag{60}$$

Linear drift CEV diffusion model $\mathrm{d}X_t = \kappa(m - X_t)\mathrm{d}t + \sigma X_t^\rho\mathrm{d}W_t$:

$$p_X^{(0)}(\Delta, x|x_0) = \frac{(x_0/x)^{\rho/2}}{\sqrt{2\pi\Delta}\sigma x^\rho}\exp\left(\kappa m\Phi_0 - \kappa\Phi_1 - \frac{(x^{1-\rho} - x_0^{1-\rho})^2}{2\sigma^2\Delta(1-\rho)^2}\right), \tag{61}$$

$$C_1(x|x_0) = -\frac{\kappa^2}{2\sigma^2}\Psi_3 + \rho\kappa m\Psi_{2\rho} + \frac{\kappa(1-2\rho)}{2} + \frac{\kappa^2 m}{\sigma^2}\Psi_2 + \frac{(\rho-2)\rho\sigma^2}{8(xx_0)^{1-\rho}} - \frac{\kappa^2 m^2}{2\sigma^2}\Psi_1,$$

$$C_2(x|x_0) = \frac{(\sigma - \rho\sigma)^2(\lambda(x_0) + \lambda(x) - 2C_1(x|x_0))}{(x^{1-\rho} - x_0^{1-\rho})^2}, \tag{62}$$

where

$$\lambda(x) = \frac{\rho\sigma^4 x^{4\rho}(\rho-2) - 4\kappa^2 x^2(x-m)^2 - 4\kappa\sigma^2 x^{1+2\rho}(x(2\rho-1) - 2m\rho)}{8x^{2(1+\rho)}\sigma^2}, \tag{63}$$

$$\Phi_n \equiv \frac{x^{n+1-2\rho} - x_0^{n+1-2\rho}}{\sigma^2(n+1-2\rho)}, \quad \Psi_\eta \equiv \frac{\rho-1}{3\rho-\eta} \cdot \frac{x^{3\rho}x_0^\eta - x^\eta x_0^{3\rho}}{(xx_0)^{2\rho}(x^\rho x_0 - xx_0^\rho)}. \tag{64}$$

Aït-Sahalia's nonlinear drift model $\mathrm{d}X_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_{-1}/X_t)\mathrm{d}t + \sigma X_t^\rho\mathrm{d}W_t$:

$$p_X^{(0)}(\Delta, x|x_0) = \frac{(x_0/x)^{\rho/2}}{\sqrt{2\pi\Delta}\sigma x^\rho}\exp\left(-\frac{(x^{1-\rho} - x_0^{1-\rho})^2}{2\Delta(\sigma - \sigma\rho)^2} + \Phi_0\alpha_0 + \Phi_1\alpha_1 + \Phi_2\alpha_2 + \Phi_{-1}\alpha_{-1}\right), \tag{65}$$

$$C_1(x|x_0) = \frac{(\rho - 2)\,\rho\sigma^2}{8(xx_0)^{1-\rho}} + \rho\Psi_{2\rho}\alpha_0 + \frac{2\rho - 1}{2}\alpha_1 + (\rho - 1)\Psi_{2+2\rho}\alpha_2 + \frac{(1 + 2\rho)\Psi_{2\rho-1}}{2}\alpha_{-1}$$

$$- \frac{\Psi_{-1}}{2\sigma^2}\alpha_{-1}^2 - \frac{\Psi_0}{\sigma^2}\alpha_0\alpha_{-1} - \frac{\Psi_1}{2\sigma^2}(\alpha_0^2 + 2\alpha_1\alpha_{-1}) - \frac{\Psi_2}{\sigma^2}(\alpha_2\alpha_{-1} + \alpha_0\alpha_1) \qquad (66)$$

$$- \frac{\Psi_3}{2\sigma^2}(\alpha_1^2 + 2\alpha_0\alpha_2) - \frac{\Psi_4}{\sigma^2}\alpha_1\alpha_2 - \frac{\Psi_5}{2\sigma^2}\alpha_2^2 \ ,$$

$$C_2(x|x_0) = \frac{(\sigma - \rho\sigma)^2\,(\lambda(x_0) + \lambda(x) - 2C_1(x|x_0))}{\left(x^{1-\rho} - x_0^{1-\rho}\right)^2}, \qquad (67)$$

where

$$\lambda(x) = \frac{x^{2\rho}(\rho^2 - \rho)\sigma^2 + 2x\rho\alpha_0 + 2x^2\,(\rho - 1)\,\alpha_1 + 2x^3\,(\rho - 2)\,\alpha_2 + 2(\rho + 1)\alpha_{-1}}{4x^2}$$

$$- \frac{1}{2}\left(\frac{x^{\rho-1}\rho\sigma}{2} - \frac{\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_{-1}/x}{x^\rho\sigma}\right)^2. \qquad (68)$$

REFERENCES

Andersen, Leif, and Jesper Andreasen, 2000, Volatility skews and extensions of the LIBOR market model, *Applied Mathematical Finance* 7, 1–32.

Aït-Sahalia, Yacine, 1996a, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385–426.

Aït-Sahalia, Yacine, 1996b, Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, 527–560.

Aït-Sahalia, Yacine, 1999, Transition densities for interest rate and other nonlinear diffusions, *Journal of Finance* 54, 1361–1395.

Aït-Sahalia, Yacine, 2002, Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach, *Econometrica* 70, 223–262.

Aït-Sahalia, Yacine, 2008, Closed-form likelihood expansions for multivariate diffusions, *Annals of Statistics* 36, 906–937.

Aït-Sahalia, Yacine and Robert L. Kimmel, 2007, Maximum likelihood estimation of stochastic volatility models, *Journal of Financial Economics* 83, 413–452.

Arapis, Manuel, and Jiti Gao, 2006, Empirical comparisons in short-term interest rate models using nonparametric methods, *Journal of Financial Econometrics* 4, 310–345.

Bachelier, Louis, 1900, Théorie de la spéculation, *Annales de l'Ecole Normale Superiure* 17, 21–86. Reprinted in Cootner (1964).

Bakshi, Gurdip, Nengjiu Ju, and Hui Ou-Yang, 2006, Estimation of continuous-time models with an application to equity volatility dynamics, *Journal of Financial Economics* 82, 227–249.

Bali, Turan, and Liuren Wu, 2006, A comprehensive analysis of the short-term interest-rate dynamics, *Journal of Banking and Finance* 30, 1269–1290.

Beckers, Stan, 1980, The constant elasticity of variance model and its implications for option pricing, *Journal of Finance* 35, 661–673.

Boyle, Phelim P., and Yisong (Sam) Tian, 1999, Pricing lookback and barrier options under the CEV process, *Journal of Financial and Quantitative Analysis* 34, 241–264.

Chacko, George, and Luis M. Viceira, 2003, Spectral GMM estimation of continuous-time processes, *Journal of Econometrics* 116, 259–292.

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* 47, 1209–1227.

Chapman, David, and Neil D. Pearson, 2000, Is the short rate drift actually nonlinear? *Journal of Finance* 55, 355–399.

Cheridito, Patrick, Damir Filipović, Robert L. Kimmel, 2007, Market price of risk specifications for affine models: Theory and evidence, *Journal of Financial Economics* 83, 123–170.

Conley Timothy G., Lars Peter Hansen, Erzo G. J. Luttmer, and José A. Scheinkman, 1997, Short-term interest rates as subordinated diffusions, *Review of Financial Studies* 10, 525–577.

Cox, John C., 1975, Notes on option pricing I: Constant elasticity of variance diffusions. mimeo, Standard University.

Cox, John C., 1996, The constant elasticity of variance option pricing model, *The Journal of Portfolio Management*, Special Issue.

Cox, John C., and Stephen A. Ross, 1976, The valuation of options for alternative stochastic processes, *Journal of Financial Economics* 3, 145–166.

Cox, John C., Jonathan E. Ingersoll Jr., and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica*, Vol. 53, No. 2, 385–407.

Cox, Alexander M. G., and David G. Hobson, 2005, Local martingales, bubbles and option prices, *Finance and Stochastics*, Vol. 9, No. 4, 477–492.

D. Davydov, and V. Linetsky, 2001, The valuation and hedging of barrier and lookback options under the CEV process, *Management Science* 47, 949-965.

Durham, Garland B., 2003, Likelihood-based specification analysis of continuous-time models of the short-term interest rate, *Journal of Financial Economics* 70, 463–487.

Emanuel, David C., and James D. MacBeth, 1982, Further results on the constant elasticity of variance call option pricing model, *Journal of Financial and Quantitative Analysis* 17, 533–554.

Evans, Lawrence C., 1998. Partial Differential Equations (American Mathematical Society.).

Goldstein, Robert S., and Willaim P. Keirstead, 1997, On the term structure of interest rates in the presence of reflecting and absorbing boundary conditions, Working paper.

Hajek, Bruce, 1985, Mean stochastic comparison of diffusions, *Probability Theory and Related Fields* 68, No. 3, 315–329.

Heston, Steve, Mark Loewenstein, and Gregory A. Willard, 2007, Options and bubbles, *Review of Financial Studies*, 20, 359–390.

Jensen, Bjarke, and Rolf Poulsen, 2002, Transition densities of diffusion processes: Numerical comparison of approximation techniques, *Journal of Derivatives* 9, 18–32.

Jones, Christopher S., 2003a, The dynamics of stochastic volatility evidence from underlying and option markets, *Journal of Econometrics* 116, 181–224.

Jones, Christopher S., 2003b, Nonlinear mean reversion in the short-term interest rate, *Review of Financial Studies* 16, 793–843.

Karlin, Samuel, and Howard M. Taylor, 1981. A Second Course in Stochastic Processes (Academic Press, New York.).

Karatzas, Ioannis, and Steven E. Shreve, 1991. Brownian Motion and Stochastic Calculus (Springer.).

Kloeden, Peter E., and Eckhard Platen, 1992. Numerical Solution of Stochastic Differential Equations (Springer.).

Lewis, Alan L., 2000. Option Valuation Under Stochastic Volatility (Finance Press, Newport Beach, California.).

Li, Minqiang, Neil D. Pearson, and Allen M. Poteshman, 2004, Conditional estimation of diffusion processes, *Journal of Financial Economics* 74, 31–66.

Mao, Xuerong, 1997. Stochastic Differential Equations and Applications (Horwood Publishing.).

Schroder, Mark, 1989, Computing the constant elasticity of variance option pricing formula, *Journal of Finance* 44, 211–219.

Sin, Carlos A., 1998, Complications with stochastic volatility models, *Advances in Applied Probability* 30, 256–268.

Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177–188.

Yamada, Toshio, and Shinzo Watanabe, 1971, On the uniqueness of solutions of stochastic differential equations, *Journal of Mathematics of Kyoto University* 11, 155-167.

Table I
Maximum Likelihood Estimate for the Monthly Federal Funds Rate, 1963–1998

This table reports the parameters estimations for seven short-rate models using maximum likelihood. We use the same data set in Aït-Sahalia (1999) which consists of monthly data on the Federal funds rate between January 1963 and December 1998. The source is the H-15 Federal Reserve Statistical Release (Selected Interest Rate Series). The transition densities needed in the likelihood functions are computed using the true density, Euler approximation, and Aït-Sahalia's series approximation with $K = 2$. The true density is computed either from the explicit transition density (if available) or through a PDE approach. The log-likelihood values are denoted by ℓ. The asymptotic standard error is computed by estimating the inverse of Fisher information matrix using an outer product of gradients method at the estimated parameter values. The last column reports the Akaike Information Criterion (AIC) for series approximation method, computed as $-2\ell + 2\dim(\theta)/N$, where $\dim(\theta)$ is the dimension of the parameter vector θ and $N = 431$ is number of observations.

Models	Explicit density/PDE	Euler approxi.	Series approx.	Asymptotic std. error	AIC
Vasicek:	$\kappa = 0.2612$	0.2584	0.2612	0.10	
$\mu(x) = \kappa(m - x)$	$m = 0.0717$	0.0717	0.0717	0.026	
$\sigma(x) = \sigma$	$\sigma = 0.0224$	0.0221	0.0224	0.00020	
	$\ell = 3.6345$	3.6345	3.6345		-7.2551
Exponentiated Vasicek:	$\kappa = 0.1729$	0.1375	0.1733	0.07	
$\mu(x) = \kappa x(\log m - \log x)$	$m = 0.0788$	0.0829	0.0788	0.017	
$\sigma(x) = \sigma x$	$\sigma = 0.2199$	0.2147	0.2199	0.0036	
	$\ell = 4.0971$	4.1050	4.0971		-8.1803
CIR:	$\kappa = 0.2189$	0.1452	0.2189	0.08	
$\mu(x) = \kappa(m - x)$	$m = 0.0721$	0.0732	0.0721	0.017	
$\sigma(x) = \sigma\sqrt{x}$	$\sigma = 0.0667$	0.0652	0.0667	0.00075	
	$\ell = 3.9182$	3.9309	3.9182		-7.8225
Inverse Feller:	$\kappa = 2.0815$	2.0097	2.0823	1.14	
$\mu(x) = \kappa(m - x)x$	$m = 0.0874$	0.0881	0.0874	0.024	
$\sigma(x) = \sigma x^{3/2}$	$\sigma = 0.8211$	0.8059	0.8211	0.018	
	$\ell = 4.1581$	4.1710	4.1581		-8.3023
CEV:	$\kappa = 0.0875$	0.0971	0.0886	0.10	
$\mu(x) = \kappa(m - x)$	$m = 0.0849$	0.0808	0.0842	0.052	
$\sigma(x) = \sigma x^{\rho}$	$\sigma = 0.7785$	0.7224	0.7792	0.077	
	$\rho = 1.4802$	1.4607	1.4812	0.037	
	$\ell = 4.1584$	4.1720	4.1582		-8.2978
Nonlinear Drift:	$\alpha_0 = -0.0347$	-0.0568	-0.0345	0.080	
$\mu(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_{-1}/x$	$\alpha_1 = 0.6702$	0.9621	0.6676	1.31	
$\sigma(x) = \sigma x^{\rho}$	$\alpha_2 = -4.0023$	-5.017	-4.0069	6.15	
	$\alpha_{-1} = 0.000695$	0.00116	0.000697	0.0015	
	$\sigma = 0.7842$	0.7072	0.7834	0.081	
	$\rho = 1.4836$	1.4533	1.4828	0.038	
	$\ell = 4.1588$	4.1730	4.1587		-8.2896
Damped Diffusion:	$\kappa = 0.0954$	0.0957	0.0965	0.096	
$\mu(x) = \kappa(m - x)$	$m = 0.0815$	0.0812	0.0814	0.040	
$\sigma(x) = (\sigma_1 x^{1/2} + \sigma_2 x^{3/2})e^{-8x^4}$	$\sigma_1 = 0.00359$	0.00451	0.00368	0.0020	
	$\sigma_2 = 0.7283$	0.7278	0.7572	0.041	
	$\ell = 4.1596$	4.1735	4.1593		-8.3000

Table II

Maximum Likelihood Estimate for the Monthly Federal Funds Rate, 1954–2008

This table reports the parameters estimations for seven short-rate models using maximum likelihood. The data set consists of monthly data on the Federal funds rate between July 1954 and June 2008. The source is the H-15 Federal Reserve Statistical Release (Selected Interest Rate Series). The transition densities needed in the likelihood functions are computed using the true density, Euler approximation, and Aït-Sahalia's series approximation with $K = 2$. The true density is computed either from the explicit transition density (if available) or through a PDE approach. The log-likelihood values are denoted by ℓ. The asymptotic standard error is computed by estimating the inverse of Fisher information matrix using an outer product of gradients method at the estimated parameter values. The last column reports the Akaike Information Criterion (AIC) for series approximation method, computed as $-2\ell + 2\dim(\theta)/N$, where $\dim(\theta)$ is the dimension of the parameter vector θ and $N = 647$ is number of observations.

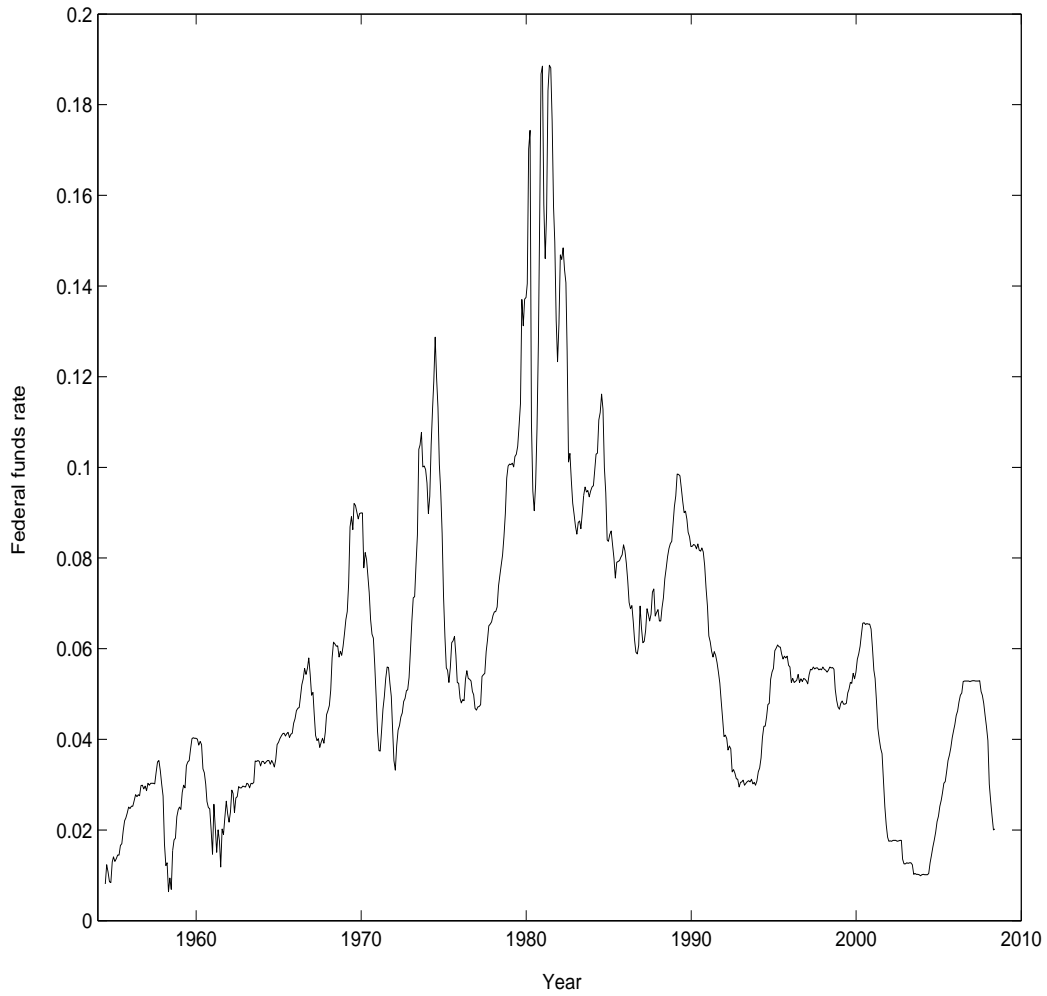| Models	Explicit density/PDE	Euler approxi.	Series approxi.	Asymptotic std. error	AIC
Vasicek:	$\kappa = 0.1685$	0.1673	0.1685	0.06	
$\mu(x) = \kappa(m - x)$	$m = 0.0582$	0.0582	0.0582	0.025	
$\sigma(x) = \sigma$	$\sigma = 0.0186$	0.0185	0.0186	0.00012	
	$\ell = 3.8145$	3.8145	3.8145		-7.6198
Exponentiated Vasicek:	$\kappa = 0.2038$	0.3209	0.2046	0.05	
$\mu(x) = \kappa x(\log m - \log x)$	$m = 0.0684$	0.0604	0.0684	0.023	
$\sigma(x) = \sigma x$	$\sigma = 0.3363$	0.3593	0.3363	0.0030	
	$\ell = 3.9622$	3.8888	3.9622		-7.9151
CIR:	$\kappa = 0.1382$	0.1446	0.1384	0.04	
$\mu(x) = \kappa(m - x)$	$m = 0.0585$	0.0584	0.0585	0.014	
$\sigma(x) = \sigma\sqrt{x}$	$\sigma = 0.0648$	0.0643	0.0648	0.00067	
	$\ell = 4.0864$	4.0880	4.0864		-8.1636
Inverse Feller:	$\kappa = 5.4690$	13.8034	5.4704	4.67	
$\mu(x) = \kappa(m - x)x$	$m = 0.0947$	0.0623	0.0947	0.070	
$\sigma(x) = \sigma x^{3/2}$	$\sigma = 2.5696$	3.1102	2.5696	0.015	
	$\ell = 3.4633$	3.2514	3.4633		-6.9173
CEV:	$\kappa = 0.1251$	0.1591	0.1276	0.049	
$\mu(x) = \kappa(m - x)$	$m = 0.0585$	0.0576	0.0585	0.017	
$\sigma(x) = \sigma x^{\rho}$	$\sigma = 0.0917$	0.0820	0.0915	0.0029	
	$\rho = 0.6181$	0.5822	0.6175	0.010	
	$\ell = 4.0992$	4.0949	4.0990		-8.1856
Nonlinear Drift:	$\alpha_0 = -0.00004$	-0.0540	-0.00004	0.010	
$\mu(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_{-1}/x$	$\alpha_1 = 0.0407$	1.2676	0.0405	0.29	
$\sigma(x) = \sigma x^{\rho}$	$\alpha_2 = -0.8978$	-7.6449	-0.8976	1.64	
	$\alpha_{-1} = 0.00007$	0.00058	0.00007	0.00008	
	$\sigma = 0.0912$	0.0879	0.0913	0.0029	
	$\rho = 0.6165$	0.6129	0.6166	0.011	
	$\ell = 4.0994$	4.1192	4.0993		-8.1800
Damped Diffusion:	$\kappa = 0.1121$	0.1558	0.1139	0.056	
$\mu(x) = \kappa(m - x)$	$m = 0.0596$	0.0568	0.0593	0.022	
$\sigma(x) = (\sigma_1 x^{1/2} + \sigma_2 x^{3/2})e^{-8x^4}$	$\sigma_1 = 0.0480$	0.0497	0.0484	0.0010	
	$\sigma_2 = 0.2485$	0.2208	0.2462	0.014	
	$\ell = 4.1317$	4.1284	4.1316		-8.2508

Figure 1. Federal funds rate, monthly frequency, July 1954 – June 2008. This is the monthly data on the Federal funds rate between July 1954 and June 2008. The source is the H-15 Federal Reserve Statistical Release (Selected Interest Rate Series).