UCLA
Recent Work

Title

Simulated Likeliehood Estimation of Diffusions With an Application to the Short Tem Interest Rate

Permalink

https://escholarship.org/uc/item/8zz2d0q8

Author

Santa-Clara, Pedro

Publication Date

1997

Simulated Likelihood Estimation of Diffusions
With an Application to the Short Term Interest Rate

Pedro Santa-Clara*

January 1996

Abstract

This paper develops a new econometric method to estimate continuous time processes from discretely sampled data. This method extends the maximum likelihood technique to cases where the transition density of the process cannot be computed in closed form but can nevertheless be computed by simulation. The asymptotic properties of the estimator are obtained, showing it to have the same behavior in large samples of the (unknown) true likelihood estimator. That is, the simulated likelihood estimator is consistent and asymptotically normal. The econometric method is used to estimate the parameters of a broad family of processes for the short term interest rate and test some restrictions to well known models of the term structure.

1 Introduction

For most time series, the process generating the observations can be seen as continuous, even though the observations are only made at discrete intervals. In fact, modelling the data generating process in continuous time has important advantages, since continuous time models are, in some ways, more fundamental than discrete time models of statistical time series. First, a particular continuous time process is compatible with (infinitely) many different discrete time processes, in the sense that it generates the same joint distribution for any given sample. Second, continuous time models are not tied to the time interval at which the observations happen to be made. This characteristic is particularly important for nonlinear models, since these models do not preserve their distributional characteristics at different time intervals[1].

Discrete time statistical models have nevertheless been widely used in the literature since, when the time interval of the statistical model coincides with the observation interval, these models usually lead to likelihood functions or moment conditions that are easily computed. On the contrary, the likelihood or moment conditions of nonlinear continuous time processes observed at discrete intervals are for most cases unknown.

Among Markovian continuous time processes, diffusions[2] are a particularly interesting class. First, modelling a data generating process as a diffusion is very parsimonious since it involves specifying only its local mean and variance and yields a very precise statistical specification for estimation. Furthermore, the class of Ito processes is closed under a variety of nonlinear transformations and there is available a calculus that allows the explicit computation of the processes followed by (nonlinear) transformations of any given process. Finally, diffusions approximate well a wide variety of commonly used discrete time processes.

---

[1]Consider a simple example. Suppose that we observe a time series with a given interval, and model it by a nonlinear but conditionally Gaussian process, such as GARCH, with time interval equal to the periodicity of observations. For a given sample, it is easy to construct the likelihood function and optimize it numerically. But, if we choose a frequency for the model that is different from the periodicity of the data, if we encounter missing data, or if the observations are irregularly spaced, the conditional distribution of the process between observations will not be normal anymore.

[2]We will loosely refer to diffusions as Ito stochastic differential equations (SDE).

Diffusions are one of the major building blocks in Finance theory, where a number of methods are available for pricing derivatives given the processes followed by the underlying assets or state variables. In order to use these models it is, however, necessary to be able to estimate the parameters of the process followed by the underlying asset or state variables.

In this paper, we consider the problem of parametric estimation of diffusions from discretely sampled data. This problem has received some attention in the literature, but, so far, there does not seem to be any completely satisfactory method available.

The traditional literature on the estimation of diffusions has dealt with the few cases for which the transition density (and thus the likelihood) or the (conditional) moments of the process are known. As an interesting recent example in the term structure literature, Pearson and Sun (1994) estimate the parameters of the (unobservable) processes followed by the factors in Cox, Ingersoll and Ross' (1985) model. They make use of the fact that the transition density of the state variables is known, since they follow square root processes, to construct the likelihood function of a sample of bond prices. For the same model, Gibbons and Ramaswamy (1993) use unconditional moment conditions to estimate the factor processes' parameters with GMM.

Hansen and Scheinkman (1994) have recently extended the class of models for which we can use econometric methods based on the true moments of the data generating process. They show how to obtain moment conditions from the infinitesimal generators of a variety of stationary Markov processes. These moment conditions can then be used in GMM[3]. This method requires that the process be stationary and ergodic.

The rest of the literature has dealt with the problem of unknown moments and conditional densities by developing methods to approximate these quantities.

The simplest approch in this literature has been to approximate the moments of the continuous time process by the moments of its Euler discretization, with discretization interval equal to the sample frequency. Chen, Karolyi,

---

[3]Conley, Hansen, Luttmer and Scheinkman (1995) estimate a semi-parametrized diffusion for the spot interest rate based on these moment conditions.

Longstaff and Saunders (1992) provide an illustration of this method, estimating a constant elasticity of variance model of the spot interest rate. The problem with this method is that the estimator is biased because of time aggregation.

Lo (1988) has proposed approximating the transition density of the diffusion by a numerical solution of the forward Kolmogorov partial differential equation (PDE) with appropriate boundary condition. Note that, to approximate the likelihood function, this PDE has to be solved for every point in the sample. This is clearly very demanding computationally. Nevertheless, this method has been applied by Mella-Barral and Perraudin (1993). Unfortunately, there is no analysis available of the discretization error induced by the numeric methods used in the solution of the PDE.

Duffie and Singleton (1993) have proposed an estimation method that approximates the unconditional moments of the process by moments obtained from simulations of its sample paths. These simulated moments can then be used with GMM. An important advantage of this method is that it is usable to estimate the processes followed by unobservable state variables that drive the uncertainty of the variables in the sample. Additionally, the paper shows how to control the bias induced by the simulation. However, when the method is applied to diffusions, an additional error is made in the discretization of the process, which the paper does not show how to control. Another problem is that the method imposes very strong requirements in terms of stationarity of the process to be estimated.

Gourieroux, Monfort and Renault (1992) have proposed an indirect method that extends Duffie and Singleton's and that has been specialized to the estimation of diffusions by Broze, Scaillet and Zakoian (1994). This approach involves the use of an auxiliary criterium[4] and an auxiliary parameter for which an estimation method with good properties exists. Then, this criterium is "incorrectly" maximized for values simulated according to the true model. Finally, the estimator for the true model parameters is calibrated by the difference between the estimators of the auxiliary parameter obtained by the appropriate method and the simulated method.

---

[4]Which can be the exact likelihood of an approximate model or an approximation of the exact likelihood of the model.

He (1990) proposes approximating the conditional sample moments by a binomial discretization. With this approach there is only a discretization error, which is controlled for, and no simulation error. However, this method is computationally heavy, since it requires that a binomial tree be evaluated for every observation in the sample, every time the objective function is computed.

Finally, Ait-Sahalia (1995) explores the relationships that exist between the marginal and conditional densities of a diffusion and its drift and diffusion, using a nonparametric kernel estimator of the densities to conduct inference on the parameters defining the diffusion. This method again requires stationarity of the processes to estimate.

Our method has elements of several of these papers, being closest in spirit to the papers by He and Lo. We propose a method for approximating the likelihood function by simulation that is both feasible computationally and that provides an estimator that inherits from the true likelihood estimator the right asymptotic properties.

Making use of the fact that the processes we want to estimate are Markovian, we write the likelihood function as the product of the transition densities between consecutive observations. These transition densities are not known. We approximate them by the transition densities of an Euler discretization of the diffusion. That is, we split the interval between two consecutive observations in many small intervals and consider the Euler discretization of the diffusion, which is a particularly simple, conditionally Gaussian, discrete time process. We then look at the transition densities between the observations in our sample of this discrete time process.

However, this transition density is also unknown. So, we estimate it by taking a large number of simulations of the discretized process between the two observations. In doing these simulations, we make use of the fact that the transition density between observations of the Euler approximation is a convolution of Gaussian densities. Our simulation method can be interpreted simply as a numeric method for computing the multiple integrals involved in this convolution.

We are able to control the discretization and the simulation errors asymp-

totically and we show that our estimator converges to the (unknown) true maximum likelihood estimator and that it has the same properties in large samples.

This method has several advantages. First, as a likelihood maximizing estimator, this estimator uses all the information about the transition density of the process instead of just using the information about some arbitrarily chosen moments. This advantage is even more marked over methods that use only unconditional moments and that, therefore, do not use information about the transitions of the process.

Second, the method is computationally feasible on a personal computer. Although, it requires the computation of a large number of trajectories of a finely discretized process between each two consecutive observations, these are simple to compute recursively.

And, finally, our method imposes relatively mild conditions on the diffusion to be estimated. In particular, we do not impose conditions on the rate of mixing of the process in order to show that our estimator is equal asymptotically to the true likelihood estimator. This is particularly important since a great many diffusions that have been used in models of asset prices are not stationary.

This paper is organized as follows. In section 2, we explain the econometric method. Section 3 shows the convergence of our simulated likelihood estimator to the true likelihood estimator. Section 4 shows the asymptotic properties of the true likelihood estimator of our estimator simulated likelihood estimator. In section 5, we illustrate our method, estimating a parametric process for the spot interest rate. Section 6 concludes the paper.

2   Simulated Maximum Likelihood

In this section, we describe the econometric method. For notational simplicity, only univariate diffusion processes are considered here. The results can easily be extended to the multidimensional case.

Consider a complete probability space, $(\Omega, \mathcal{F}, P)$, where a standard Brownian motion, W, is defined. Let Y be defined by the SDE

$$dY(t) = a(Y(t), t; \theta)dt + b(Y(t), t; \theta)dW(t) \tag{1}$$

with initial value $Y(0) \in I\!R$. The functions a and b of Y and t depend on an unknown parameter vector θ. In order to characterize the dynamics of Y, we are interested in estimating this parameter. We make the following assumptions.

Assumption 1 The drift and the diffusion function are sufficiently regular for the existence of a unique strong solution to (1). Sufficient conditions are the usual linear growth and uniform Lipschitz continuity[5].

Assumption 2 Let $\theta \in \Theta \subset I\!R^K$, where Θ is a compact parameter set that contains the true parameter vector θ_0.

Now, suppose the process Y is sampled at $N + 1$ discrete points in time $t_0, t_1, ..., t_N$, and let $\mathbf{Y}_{(N)} \equiv (Y_0, Y_1, ..., Y_N)$ denote this random sample, where Y_n is the observed realization of Y at time t_n for $n = 0, 1, ..., N$.

Given the discretely sampled data and the specification of the process Y, we denote by $P(\mathbf{Y}_{(N)}; \theta)$ the joint distribution of $\mathbf{Y}_{(N)}$ and let $p(\mathbf{Y}_{(N)}; \theta)$ denote the density representation of $P(\mathbf{Y}_{(N)}; \theta)$. When considered as a function of θ, this joint density is obviously the likelihood function[6] of $\mathbf{Y}_{(N)}$, denoted $L(\mathbf{Y}_{(N)}; \theta)$. Since Y is a Markovian process, the joint density p may be written as the following product of conditional densities

$$L(\mathbf{Y}_{(N)}; \theta) = p(Y_0, t_0; \theta) \prod_{n=0}^{N-1} p(Y_{n+1}, t_{n+1} | Y_n, t_n; \theta) \tag{2}$$

For simplicity, we denote $L(\mathbf{Y}_{(N)}; \theta)$ by $L_N(\theta)$.

Assumption 3 In a neighborhood of the true parameter θ^0, $L_N(\theta)$ is twice continuously differentiable in θ. Furthermore, $E[(\partial L_N(\theta)/\partial \theta)(\partial L_N(\theta)/\partial \theta')]$ has full rank and is bounded for all $\theta \in \Theta$.

---

[5] See Karatzas and Shreve (1988).

[6] See Dhrymes (1994, Sections 5.2 and 5.3) for a discussion of the definition of the likelihood function and the problem of estimation within the probability space defined.

This assumption insures the identifiability of θ_0 in the parametric class of likelihood functions considered.

Deriving the likelihood function then reduces to calculating the transition density functions $p(Y_{n+1}, t_{n+1} | Y_n, t_n; \theta)$ and the unconditional density at the initial time[7].

The problem we face is that the likelihood function is in general unknown, since the transition densities (and unconditional density) that constitute it are unknown. We therefore procede to approximate the transition densities between each pair of consecutive observations by the transition densities of a discretization of the diffusion. We take this discretization to have a time interval many times smaller than the interval between observations.

Now, we still have the problem that not even the transition densities of the discretized process between observations are known. We can nevertheless compute them numerically by taking many simulated paths of the discretized process and taking the mean of an appropriate function of these simulated values.

In a simpler way, we propose a method of numerically computing the transition densities of the diffusion that can then be replaced in the likelihood function. For all $n = 0, 1, ..., N - 1$, and given θ, we discretize the process Y, starting from Y_n at time t_n, in order to approximate the density of Y at time t_{n+1}, and evaluate it at the observed realization Y_{n+1}. Then, we can use this value to substitute for $p(Y_{n+1}, t_{n+1} | Y_n, t_n; \theta)$ in the likelihood function.

We first discretize the process Y between times t_n and t_{n+1}[8]. There is an infinity of discrete time processes that approximate Y in this interval. We thus choose a particularly simple one: The Euler approximation[9].

---

[7]See Karlin and Taylor (1981) for a characterization of the unconditional density of a diffusion.

[8]For an excellent reference to the discretization of diffusions see Kloeden and Platen (1992).

[9]There are more efficient schemes than this one, namely the one proposed by Milshtein (1974) and further enhanced by Pardoux and Talay (1985). However, for our purposes the Euler scheme seems appropriate, although any other scheme with Gaussian transitions would do. GARCH approximations of stochastic volatility processes as in Nelson (1990) are of particular interest.

7

Without loss of generality, normalize the length of the intervals $[t_n, t_{n+1}]$ to 1 and split them into M equal subperiods of length $h = 1/M$ and let $\hat{Y}_n^{(M)}$ be the discrete-time process defined, for $m = 0, ..., M - 1$, by[10]

$$
\begin{aligned}
\hat{Y}_{t_n+(m+1)h}^{(M)} &= \hat{Y}_{t_n+mh}^{(M)} + ha\left(\hat{Y}_{t_n+mh}^{(M)}, t_n + mh\right) \\
&+ b\left(\hat{Y}_{t_n+mh}^{(M)}, t_n + mh\right)\Delta_{t_n+(m+1)h}^{(M)}W
\end{aligned}
\tag{3}
$$

with initial value $\hat{Y}_{t_n}^{(M)} = Y_n$, where

$$
\Delta_{t_n+(m+1)h}^{(M)}W = W(t_n + (m + 1)h) - W(t_n + mh)
$$

Note that the processes $\hat{Y}_n^{(M)}$, for all n, are defined on the same probability space as Y.

Now, the one-step transition density of the discrete time process, $\hat{Y}_n^{(M)}$, evaluated at y at time $t_n + (m + 1)h$, given that x was observed at time $t_n + mh$, is normal, that is,

$$
q^{(M)}(y, t_n + (m + 1)h | x, t_n + mh) = \phi(y; x + ha(x, t_n + mh), hb(x, t_n + mh)^2)
$$

where $\phi(.; M, V)$ denotes the normal density with mean M and variance V.

The transition density for several discretization periods can be obtained recursively, for $i + 1 > j$ varying as m, as

$$
\begin{aligned}
q^{(M)}(y, t_n + (i + 1)h | x, t_n + jh) &= \int_{-\infty}^{+\infty} q^{(M)}(y, t_n + (i + 1)h | z, t_n + ih) \\
&\quad q^{(M)}(z, t_n + ih | x, t_n + jh)dz
\end{aligned}
$$

Thus, the conditional density of the discrete time process at the point Y_{n+1} at time t_{n+1} given the observation Y_n at time t_n is

$$
\begin{aligned}
q^{(M)}(Y_{n+1}, t_{n+1} | Y_n, t_n) &= \int_{-\infty}^{+\infty} q^{(M)}(Y_{n+1}, t_{n+1} | z, t_n + (M - 1)h) \\
&\quad q^{(M)}(z, t_n + (M - 1)h | Y_n, t_n)dz
\end{aligned}
\tag{4}
$$

---

[10]We assume that we are given θ, and do not make explicit the dependence on this parameter vector for notational ease.

This conditional density is the convolution of $m - 1$ normal densities and consists of $m - 2$ integrals that cannot in general be computed analytically. In principle, this multiple integral could be computed by traditional methods of numerical integration, but this becomes computationaly infeasible very quickly[11]. We propose computing this integral by Monte Carlo simulation.

A natural approximation to the transition density can be obtained by looking at $q^{(M)}(Y_{n+1}, t_{n+1} | Y_n, t_n)$ as an expectation of a function of the random variable $\hat{Y}_{t_n + (M-1)h}^{(M)}$ - that is, the variable z in (4) - with density $q^{(M)}(z, t_n + (M-1)h | Y_n, t_n)$, which can be easily approximated by simulation.

We thus take a large number P of simulations of the stochastic process $\hat{Y}_n^{(M)}$, using a normal random number generator to produce P independent Gaussian sequences $\{\Delta_{t_n + (m+1)h}^{(M)} W(p)\}_{m=0}^{M-1}$, $p = 1, 2, ..., P$, that correspond to drawings of the original Brownian motion[12]. In this way, P observations of $\hat{Y}_{t_n + (M-1)h}^{(M)}$ are produced, denoted $\hat{Y}_{t_n + (M-1)h}^{(M)}(p)$ or simply $Z_p^{(M)}$, for $p = 1, 2, ..., P$, and an estimate of $q^{(M)}(Y_{n+1}, t_{n+1} | Y_n, t_n)$ can be computed as

$$\hat{q}^{(M,P)}(Y_{n+1}, t_{n+1} | Y_n, t_n) = \frac{1}{P} \sum_{p=1}^{P} \phi(Y_{n+1}; M(Z_p^{(M)}), V(Z_p^{(M)}))$$

where

$$M(Z_p^{(M)}) = Z_p^{(M)} + ha(Z_p^{(M)}, t_n + (M-1)h)$$
$$V(Z_p^{(M)}) = hb(Z_p^{(M)}, t_n + (M-1)h)^2$$

We can compute the unconditional density for the initial observation when the process is stationary (and thus time independent) and ergodic. We can make use of the fact that $p(y) = \lim_{t \to \infty} p(y, t | x, 0)$ and simulate the conditional density over a long enough time interval, for any given initial condition, x, and evaluate it at $y = Y_0$. Alternatively, we can use the following equality obtained from the forward Kolmogorov equation[13]

$$p(y) = \frac{C}{b^2(y)} \exp\{ \int_{-\infty}^{y} \frac{2a(v)}{b^2(v)} dv \}$$

---

[11]See McFadden (1989))

[12]We can apply variance reduction techniques to the Monte Carlo draws in order to increase the precision of this integration method. See Kloeden and Platen (1992) and Newton (1990).

[13]See Karlin and Taylor (1981) or Ait-Sahalia (1995).

where C is just a constant to make the density integrate to one. This integral can in some parametrizations be computed explicitly, or, in other cases, numerically. In any case, the first term of the likelihood function will not effect the value of the function for large enough samples. Of course, when the process is not stationary, it must be assumed that the first point in the sample is deterministic.

Finally, we approximate the true likelihood function by its simulated couterpart

$$\hat{L}_N^{(M,P)}(\theta) = \hat{q}^{(M,P)}(Y_0, t_0; \theta) \prod_{n=0}^{N-1} \hat{q}^{(M,P)}(Y_{n+1}, t_{n+1}|Y_n, t_n; \theta)$$

and the log-likelihood function by

$$\hat{l}_N^{(M,P)}(\theta) = \ln \hat{q}^{(M,P)}(Y_0, t_0; \theta) + \sum_{n=0}^{N-1} \ln \hat{q}^{(M,P)}(Y_{n+1}, t_{n+1}|Y_n, t_n; \theta)$$

The estimator obtained by maximizing $\hat{l}_N^{(M,P)}(\theta)$ is denoted by $\hat{\theta}_N^{(M,P)}$.

When using a numerical optimization algorithm, it is necessary to compute the log-likelihood function for different values of θ in every iteration. Note that when θ changes, the same drawings of the P independent sequences $\{\Delta_{t_n+(m+1)h}^{(M)} W(p)\}_{m=0}^{M-1}$, $p = 1, 2, ..., P$ are used. This allows us to obtain an approximation that is a smooth function of θ and it is necessary to obtain the asymptotic results provided in the next section.

The gradient and Hessian of the simulated log-likelihood function with respect to θ can be computed explicitly. We see that, for each simulated transition density,

$$\frac{d\hat{q}^{(M,P)}(.|.)}{d\theta} = \frac{1}{P} \sum_{p=1}^{P} \left[\frac{\partial \phi(Z_p^{(M)})}{\partial \theta} + \frac{\partial \phi(Z_p^{(M)})}{\partial Z_p^{(M)}} \frac{\partial Z_p^{(M)}}{\partial \theta} \right]$$

where we write $\phi(Z_p^{(M)})$ for $\phi(., M(Z_p^{(M)}), V(Z_p^{(M)}))$ All derivatives can be computed explicitly, except for $\hat{D}_p^{(M)} \equiv \partial Z_p^{(M)}/\partial \theta$. This derivative can be obtained from the Monte Carlo simulation of the discretized process as[14]

$$\hat{D}_{t_n+(m+1)h}^{(M)} = \hat{D}_{t_n+mh}^{(M)} + h(a_\theta(\hat{Y}_{t_n+mh}^{(M)}, t_n + mh) + a_y(\hat{Y}_{t_n+mh}^{(M)}, t_n + mh)\hat{D}_{t_n+mh}^{(M)})$$

---

[14]Subscripts indicate partial derivatives.

$$+(b_\theta(\hat{Y}^{(M)}_{t_n+mh}, t_n + mh) + b_y(\hat{Y}^{(M)}_{t_n+mh}, t_n + mh)\hat{D}^{(M)}_{t_n+mh})\Delta^{(M)}_{t_n+(m+1)h}W$$

with initial value $\hat{D}^{(M)}_{t_n} = 0$. The Hessian can be obtained in a similar way.

In order to show the asymptotic properties of our estimator, we continue as follows. First, we show the convergence of the simulated maximum likelihood estimator to the true likelihood estimator. Then, we give the asymptotics of the true likelihood estimator. This is the subject of the next two sections.

3 Convergence of the Simulated Likelihood Estimator to the True Likelihood Estimator

In this section we show that the simulated likelihood estimator converges to, and is asymptotically normal around, the true likelihood estimator. We first show the convergence of the simulated densities to the true densities as both the number of discretization intervals, M, and the number of simulations, P, go to infinity. Then, we show the convergence of the simulated likelihood function to the true likelihood function, letting the sample size, N, also go to infinity. We conclude the section with the convergence of the simulated likelihood estimator to the true likelihood estimator.

We want to characterize the convergence of $\hat{q}^{(M,P)}(Y_{n+1}, t_{n+1}|Y_n, t_n)$, a triangular array of random variables indexed by M and P, to $p(Y_{n+1}, t_{n+1}|Y_n, t_n)$, a fixed number. We start by giving the following proposition.

Proposition 1 Under assumption 1, as $M \to \infty$,

$$q^{(M)}(Y_{n+1}, t_{n+1}|Y_n, t_n) - p(Y_{n+1}, t_{n+1}|Y_n, t_n) = o\left(M^{1/2}\right).$$

The strong consistency of our estimator of the transition density follows immediately from an application of the strong law of large numbers.

Proposition 2 Under assumption 1, as $M \to \infty$ and $P \to \infty$, almost surely

$$\hat{q}^{(M,P)}(Y_{n+1}, t_{n+1}|Y_n, t_n) - p(Y_{n+1}, t_{n+1}|Y_n, t_n) \to 0.$$

11

Corolllary As $M \to \infty$ and $P \to \infty$, almost surely the simulated likelihood function converges to the true likelihood function.

We now characterize the rate of convergence of the simulated transition density to the true transition density.

Proposition 3 Under assumption 1, as $M \to \infty$ and $P \to \infty$, in such a way that $\sqrt{P}/M \to 1$,

$$\sqrt{P}[\hat{q}^{(M,P)}(Y_{n+1}, t_{n+1}|Y_n, t_n) \quad - \quad p(Y_{n+1}, t_{n+1}|Y_n, t_n)]$$
$$\sim \quad N(0, \mathrm{Var}(\phi(Y(t_n); M(Y_{n+1}), V(Y_{n+1}))))$$

Note that the asymptotic variance is a finite number that can be consistently estimated by its simulated counterpart, $\mathrm{Var}(\phi(Z^{(M)}; M(Y_{n+1}), V(Y_{n+1}))$. The asymptotic ratio of M and P means that we should take the number of simulations to be the square of the number of discretizations. Duffie and Glynn (1983) show that, for this ratio, the number of basic computations necessary to compute the simulated density is proportional to M^3.

Proposition 4 Under assumptions 1 and 3, as $N \to \infty$, $M \to \infty$ and $P \to \infty$, with $\sqrt{P}/M \to 1$,

$$\hat{l}_N^{(M,P)}(\theta) - l_N(\theta) = o(NP^{-1/2}).$$

Since the simulated likelihood function converges to the true likelihood function almost surely for any fixed θ, it follows that the estimator of θ from the simulated likelihood function also converges to the value of the parameters maximizing the true likelihood function. The following result gives the rate of this convergence.

Proposition 5 Under assumptions 1 to 3, as $N \to \infty$, $M \to \infty$ and as $P \to \infty$, with $\sqrt{P}/M \to 1$,

$$\hat{\theta}_N^{(M,P)} - \theta_N = o(N^{1/2}P^{-1/4}).$$

where θ_N is the parameter that maximizes $l_N(\theta)$.

Therefore, the simulated likelihood estimator converges to the true likelihood estimator as long as P goes to infinity at a rate faster than the square of N.

4 Asymptotics

In this section, we characterize the consistency and asymptotic normality of the true likelihood estimator[15] and of our simulated likelihood estimator.

We first introduce some notation. Denote the gradient of the conditional density of observation n by $u_n(\theta)$. Then

$$u_n(\theta) = \frac{d}{d\theta} \ln p(Y_n|Y_{n-1})$$

and thus

$$\frac{d \ln L_N(\theta)}{d\theta} = \sum_{n=1}^{N} u_n(\theta).$$

By construction, $E_\theta[u_n(\theta)|\mathcal{F}_{n-1}] = 0$, implying that $\{d \ln L_N(\theta)/d\theta, \mathcal{F}_n\}$ is a square integrable martingale, given the boundedness of second derivatives in assumption 3. Denote the second derivatives of the transition densities by

$$v_n(\theta) = \frac{d^2}{d\theta^2} \ln p(Y_n|Y_{n-1}) = \frac{du_n(\theta)}{d\theta},$$

It is well known that, almost surely,

$$E_\theta[v_n(\theta)|\mathcal{F}_{n-1}] = -E_\theta[u_n^2(\theta)|\mathcal{F}_{n-1}].$$

In order to simplify notation, we now introduce

$$
\begin{aligned}
J_N(\theta) &= \sum_{n=1}^{N} v_n(\theta) \\
&= \sum_{n=1}^{N} \frac{d^2}{d\theta^2} \ln p(Y_n|Y_{n-1}) \\
I_N(\theta) &= \sum_{n=1}^{N} E_\theta[u_n^2(\theta)|\mathcal{F}_{n-1}] \\
&= \sum_{n=1}^{N} E_\theta \left[\frac{d}{d\theta} \ln p(Y_n|Y_{n-1}) \right]^2,
\end{aligned}
$$

---

[15]We broadly follow Hall and Heyde (1980) in obtaining the asymptotic properties of the true likelihood estimator and specialize the relevant conditions to the case of diffusions.

from what follows that

$$I_N(\theta) = -E_\theta[J_N(\theta)|\mathcal{F}_{n-1}]$$

and therefore $J_N(\theta) + I_N(\theta)$ is a martingale.

We now investigate the consistency of the maximum likelihood estimator. In order to do that, the function $d\ln L_N/d\theta$ is expanded around the true value θ_0 of the parameter, and evaluated at the optimum, θ_N,

$$\begin{aligned}
\frac{d}{d\theta}\ln L_N(\theta_N) &= \sum_{n=1}^{N} u_n(\theta_0) - (\theta_N - \theta_0)I_N(\theta_0) \\
&+ (\theta_N - \theta_0)[J_N(\theta_N^*) + I_N(\theta_0)]
\end{aligned}$$

where $\theta_N^* = \theta_0 + \gamma(\theta_N - \theta_0)$ with $|\gamma| < 1$. Notice that, since θ_N maximizes $\ln L_N$, the left hand side is simply zero. Rearranging the terms of this expansion in order to isolate $(\theta_N - \theta_0)$, it follows that

$$\begin{aligned}
(\theta_N &- \theta_0)[1 - I_N(\theta_0)^{-1}(J_N(\theta_N^*) + I_N(\theta_0))] \\
&= I_N(\theta_0)^{-1}\sum_{n=1}^{N} u_n(\theta_0).
\end{aligned}$$

We make a further assumption.

Assumption 4 $I_N(\theta_0) \to \infty$, in the sense that $\lambda' I_N(\theta_0)\lambda \to \infty$, for any $\lambda \in \mathbb{R}^K$. A sufficient condition is that the gradient of the transition densities be bounded.

We can now state the following proposition, that shows the consistency of the true likelihood estimator.

Proposition 6 Under assumptions 1 to 4, as $N \to \infty$,

$$\theta_N - \theta_0 \to 0$$

Regarding the fluctuations of the estimator around the true value of the parameter, we can see that, as $N \to \infty$,

$$I_N(\theta_0)^{1/2}(\theta_N - \theta_0) \sim I_N(\theta_0)^{-1/2}\sum_{n=1}^{N} u_n(\theta_0),$$

14

Now, the right hand side is a martingale difference array, and we can apply to it a Central Limit Theorem[16].

Assumption 5 The ratio of the gradient of the transition densities to the transition densities converges, or, if divergent, does so at a rate that is slower that the rate of convergence of $I_N(\theta_0)^{-1/2}$ to zero. A sufficient condition is that the transtion densities be strictly positive and their gradient be bounded[17].

Proposition 7 Under assumptions 1 to 5, as $N \to \infty$,

$$I_N(\theta_0)^{1/2}(\theta_N - \theta_0) \sim N(0,1),$$

We are now ready to establish the asymptotic fluctuations of the simulated likelihood estimator, $\hat{\theta}_N^{(M,P)}$, around the true value of the parameter, θ_0.

Write
$$\hat{\theta}_N^{(M,P)} - \theta_0 = (\hat{\theta}_N^{(M,P)} - \theta_N) + (\theta_N - \theta_0) \tag{5}$$

Notice that the first term on the right hand side is shown in proposition 5 to behave asymptotically, as $N \to \infty$, as

$$\hat{\theta}_N^{(M,P)} - \theta_N = o(N^{1/2}P^{-1/4})$$

The second term is shown in proposition 6 to behave as

$$I_N^{1/2}(\theta_0)(\theta_N - \theta_0) \sim N(0,1)$$

Hence, multiplying equation (5) by $I_N^{1/2}(\theta_0)$, gives, as $N \to \infty$,

$$I_N^{1/2}(\theta_0)(\hat{\theta}_N^{(M,P)} - \theta_0) = I_N^{1/2}(\theta_0)(\theta_N - \theta_0) + I_N^{1/2}o(N^{1/2}P^{-1/4})$$

We can show that the second term in the right hand side is asymptotically negligible when P goes to infinity at a rate faster than the fourth power of N and state the following proposition.

---

[16]See Davidson (1994, Section 24.2).

[17]See the comments on the proof of proposition 7.

Proposition 8 Under assumptions 1 to 5, as $M \to \infty$, $P \to \infty$ and $N \to \infty$, with $\sqrt{P}/M \to 1$ and $N/P^{1/4} \to 0$,

$$I_N^{1/2}(\theta_0)(\hat{\theta}_N^{(M,P)} - \theta_0) \sim N(0,1)$$

Finally, note that a consistent estimator of $I_N(\theta_0)$ is $I_N(\hat{\theta}_N^{(M,P)})$.

5 Application to the Short Term Interest Rate

In this section we apply our econometric method to estimate a parametric diffusion for the short term interest rate that has been proposed by Ait-Sahalia (1995). The short rate is parametrized as follows

$$dr(t) = [\alpha_0 + \alpha_1 r(t) + \alpha_2 r^2(t) + \alpha_3/r(t)]dt + [\beta_0 + \beta_1 r(t) + \beta_2 r(t)^{\beta_3}]dW(t)$$

Ait-Sahalia provides conditions on the parameter values to insure existence, stationarity and positivity of the process.

We study this parametrization for two reasons. First, because it encompasses several of the specifications for the short rate process found in the literature, which can therefore be tested. Second, because it allows us to verify the existence of nonlinearities in the drift reported by Ait-Sahalia.

Our data set consists of observations of the same interest rate that has been studied by Ait-Sahalia. We use the seven day eurodollar rate as a proxy for the instantaneous interest rate. Our data were taken from Datastream and consist of Bankers' Trust middle quotes for deposits at the close of the London market. We obtain 514 weekly[18] observations from October 14, 1983 to August 13, 1993. In Figure 1 we show a plot of the time series of the data and Table 1 reports summary statistics.

We estimate the diffusion by simulated likelihood, taking 15 discretization intervals between each pair of observations and 225 simulated paths of each

---

[18]Using weekly data can be seen as a compromise between high frequency and avoiding microstructure problems.

discretized process[19]. We use the numeric optimization method proposed by Berndt, Hall, Hall and Hausman (1974) to maximize the simulated likelihood function. The computation time of the simulated likelihood on a 486DX2 personal computer is of the order of 15 minutes.

The results of the simulated likelihood estimation are presented in Table 2 and the corresponding drift and diffusion functions are plotted in Figures 2 and 3. We obtain results that are similar to the results of Ait-Sahalia, except that our estimated diffusion function is practically linear whereas his looks more like a parabola. Some of our parameter estimates are not significative, namely the estimates for β_0, β_1 and β_2. The estimate of β_3 is very close to one, which gives some support to a linear diffusion function.

We test three constrained models. In all of these, we model the diffusion function as constant elasticity of variance and try different drift functions. Table 3 reports the simulated log-likelihood values for the constrained models as well as for the general model. We use a likelihood ratio test to compare the different nested specifications. We find that all constrained models can be rejected with respect to the general model. However, among the different models with constant elasticity of variance, we find that we cannot reject a quadratic drift with respect to the general drift, and that we cannot reject a linear drift with respect to a quadratic drift. The case of linear drift is nevertheless rejected when compared with the quadratic drift. We conclude that the case for a nonlinear drift is not as strong as reported by Ait-Sahalia[20].

6 Conclusion and Future Work

We develop a new econometric method to estimate parametric diffusions from discretely sampled data. The method consists on the optimization of a numerical approximation to the likelihood function. This approximation is based on a Monte Carlo simulation of a discretization of the continuous

---

[19]We experimented several diferent sizes for the discretization interval and found that, for the interval chosen, the simulated densities were very stable to variations in the number of intervals (and corresponding changes in the number of simulations).

[20]Although for identification of the drift, it is fundamental to have a long sample, and the sample used by Ait-Sahalia covers a longer time span.

time process. We are able to control the rate of this approximation and show that our estimator inherits the asymptotic properties of the true likelihood estimator.

Our method has four important advantages over others that have been proposed in the literature. First, it shares the asymptotic properties of true likelihood estimation. Second, it makes use of all the information in the sample, rather than just the information in some moments. Third, it is computationally feasible. And, fourth, it can be applied to non-stationary processes.

A number of issues remain to be investigated for this econometric method. First, the small sample properties of the method should be examined. Second, other discretization schemes can be used, in particular it would be interesting to use the GARCH approximations to diffusions studied by Nelson (1990). Finally, the method can in principle be extended to cover regime switching models, jump-diffusions and processes with barriers.

Appendix

Proof of Proposition 1 Under regularity conditions for the drift and variance of the process[21], it is known that

$$E\left[|\hat{Y}_{t_{n+1}}^{(M)} - Y(t_{n+1})|^2|\hat{Y}_{t_n}^{(M)} = Y(t_n) = Y_n\right] = O\left(\frac{1}{M}\right).$$

From Chebyshev's inequality, it follows easily that the convergence of the process in mean square to zero at a given rate implies convergence of the transition densities at the square root of that rate. □

Proof of Proposition 2 Recall that

$$\hat{q}^{(M,P)}(Y_{n+1}, t_{n+1}|Y_n, t_n) = \frac{1}{P}\sum_{p=1}^{P}\phi(Y_{n+1}; M(Z_p^{(M)}), V(Z_p^{(M)}))$$

where the elements in the sum are i.i.d. and such that

$$E[\phi(Y_{n+1}; M(Z_p^{(M)}), V(Z_p^{(M)}))] = q^{(M)}(Y_{n+1}, t_{n+1}|Y_n, t_n).$$

Hence the Strong Law of Large Numbers applies and, almost surely,

$$\hat{q}^{(M,P)}(Y_{n+1}, t_{n+1}|Y_n, t_n) - q^{(M)}(Y_{n+1}, t_{n+1}|Y_n, t_n) \to 0$$

as $P \to \infty$. Use of proposition 1 completes the proof. □

Proof of Proposition 3 Write

$$
\begin{aligned}
&\sqrt{P}\ \ [\hat{q}^{(M,P)}(Y_{n+1}, t_{n+1}|Y_n, t_n) - p(Y_{n+1}, t_{n+1}|Y_n, t_n)]\\
&= \sum_{p=1}^{P}\frac{\phi(Z_p^{(M)}) - q^{(M)}(Y_{n+1}, t_{n+1}|Y_n, t_n)}{\sqrt{P}}\\
&+ \sqrt{P}[q^{(M)}(Y_{n+1}, t_{n+1}|Y_n, t_n) - p(Y_{n+1}, t_{n+1}|Y_n, t_n)].
\end{aligned}
$$

Now, proposition 1 ensures that the second term converges pointwise to zero whenever $M \to \infty$.

---

[21]Assumption 1 is sufficient, see Pardoux and Talay (1985).

We can apply the Lindeberg-Feller Central Limit Theorem to the first term in the right hand side of the equality above as in Duffie and Glynn (1993). Notice that, from the local Brownian property of diffusions[22], as $M \rightarrow \infty$,

$$\phi(Y_{n+1}; M(Z_p^{(M)}), V(Z_p^{(M)})) = \phi(Z_p^{(M)}; M(Y_{n+1}), V(Y_{n+1})) + o(1/\sqrt{M})$$

and it follows that

$$\text{Var}[\phi(Y_{n+1}; M(Z_p^{(M)}), V(Z_p^{(M)}))] = \text{Var}[\phi(Z_p^{(M)}; M(Y_{n+1}), V(Y_{n+1}))] + o(1/\sqrt{M})$$

which converges to a finite positive constant. $\square$

Proof of Proposition 4 Denote by $x_n^{(M,P)}$ the difference between the simulated transition and the true one

$$x_n^{(M,P)} \equiv \hat{q}^{(M,P)}(Y_{n+1}, t_{n+1}|Y_n, t_n) - p(Y_{n+1}, t_{n+1}|Y_n, t_n)$$

Proposition 3 implies that $x_n^{(M,P)}$ is of order $P^{-1/2}$ as P tends to infinity. It is then possible to characterize the rate of convergence of the log-likelihood functions as follows.

Taking $x = x_n^{(M,P)}$ and $p = p(Y_{n+1}, t_{n+1}|Y_n, t_n)$, write

$$\begin{aligned}
\ln[\hat{q}^{(M,P)}(Y_{n+1}, t_{n+1}|Y_n, t_n)] &- \ln[p(Y_{n+1}, t_{n+1}|Y_n, t_n)] \\
&= \ln(x + p) - \ln p = \ln(1 + x/p)
\end{aligned}$$

Use of a simple Taylor expansion for fixed p, as $x \rightarrow 0$, gives

$$\ln(1 + x/p) \approx x/p + o(x) = o(P^{-1/2})$$

where the last equality follows from the previous results. Summing this difference of logarithms over all the N elements of the sample, gives the result. $\square$

Proof of Proposition 5 Expanding $l_N(\theta)$ around θ_N and evaluating it at $\hat{\theta}_N^{(M,P)}$ gives

$$l_N(\hat{\theta}_N^{(M,P)}) = l_N(\theta_N) + (\hat{\theta}_N^{(M,P)} - \theta_N)\frac{\partial l_N}{\partial \theta}(\theta_N) + \frac{1}{2}(\hat{\theta}_N^{(M,P)} - \theta_N)^2\frac{\partial^2 l_N}{\partial \theta^2}(\theta_N^*)$$

---

[22]See Amaro de Matos (1995).

20

where θ_N^* is a convex combination of $\hat{\theta}_N^{(M,P)}$ and θ_N. The term corresponding to the first derivative in the expansion is zero since, by construction,

$$\frac{\partial l_N}{\partial \theta}(\theta_N) = 0$$

A similar expansion of $\hat{l}_N^{(M,P)}(\theta)$ around $\hat{\theta}_N^{(M,P)}$ and evaluated at θ_N gives

$$\hat{l}_N^{(M,P)}(\theta_N) = \hat{l}_N^{(M,P)}(\hat{\theta}_N^{(M,P)}) + (\hat{\theta}_N^{(M,P)} - \theta_N)\frac{\partial \hat{l}_N^{(M,P)}}{\partial \theta}(\hat{\theta}_N^{(M,P)}) + \frac{1}{2}(\theta_N - \hat{\theta}_N^{(M,P)})^2 \frac{\partial^2 \hat{l}_N^{(M,P)}}{\partial \theta^2}(\theta_N^{**})$$

where θ_N^{**} is a convex combination of $\hat{\theta}_N^{(M,P)}$ and θ_N. Since $\partial^2 l_N / \partial\theta^2$ is bounded by assumption 3 and $\hat{l}_N^{(M,P)}$ converges to l_N, we can sum both expansions, rearrange the terms and use the result of the previous proposition to obtain the result. $\square$

Proof of Proposition 6 As in Hall and Heyde[23], it suffices that $I_N(\theta_0) \to \infty$ as $N \to \infty$ for the right hand side to go almost surely to zero. Hence, as $N \to \infty$ and almost surely,

$$(\hat{\theta} - \theta_0)[1 - I_N(\theta_0)^{-1}(J_N(\theta^*) + I_N(\theta_0))] \to 0.$$

Now, since there exists θ^{**} between θ^* and θ such that

$$J_N(\theta^*) = J_N(\theta_0) + (\theta^* - \theta_0)\frac{dJ_N(\theta^{**})}{d\theta}$$

the term converging to zero may be decomposed in

$$(\hat{\theta} - \theta_0)[1 - I_N(\theta_0)^{-1}(J_N(\theta_0) + I_N(\theta_0))] + (\hat{\theta} - \theta_0)(\theta^* - \theta_0)I_N(\theta_0)^{-1}\frac{dJ_N(\theta^{**})}{d\theta}$$

If $I_N(\theta_0) \to \infty$ as $N \to \infty$, the second term converges to zero provided that J_N has bounded gradient, which is insured by assumption 3. Regarding the first term, since $J_N + I_N$ is a martingale, the sum converges to zero in the mean. Divided by I_N and summed to 1, the term in square brackets

---

[23]See Section 6.2 and Theorem 2.18.

Table 1
Summary Statistics for Short Term
Interest Rate

	r	Δr
Mean	0.07147	-0.000125
Std. Dev.	0.02178	0.005229
ρ(1 week)	0.97109	-0.420478
ρ(1 month)	0.95945	-0.000626
ρ(3 months)	0.92019	0.104542
ρ(6 months)	0.84208	0.063245

Table 2
Simulated Likelihood
Parameter Estimates

	ln L = 2048.56	
	Estimate	Std.Error
α_0	-6.66E-3	3.28E-3
α_1	4.98E-2	2.20E-2
α_2	-3.065	1.2097
α_3	6.27E-4	3.31E-4
β_0	-1.10E-3	7.20E-4
β_1	-1.61E-3	5.17E-3
β_2	2.87E-2	1.80E-2
β_3	9.06E-1	1.01E-1

Table 3
Simulated Likelihood Values
Different Models

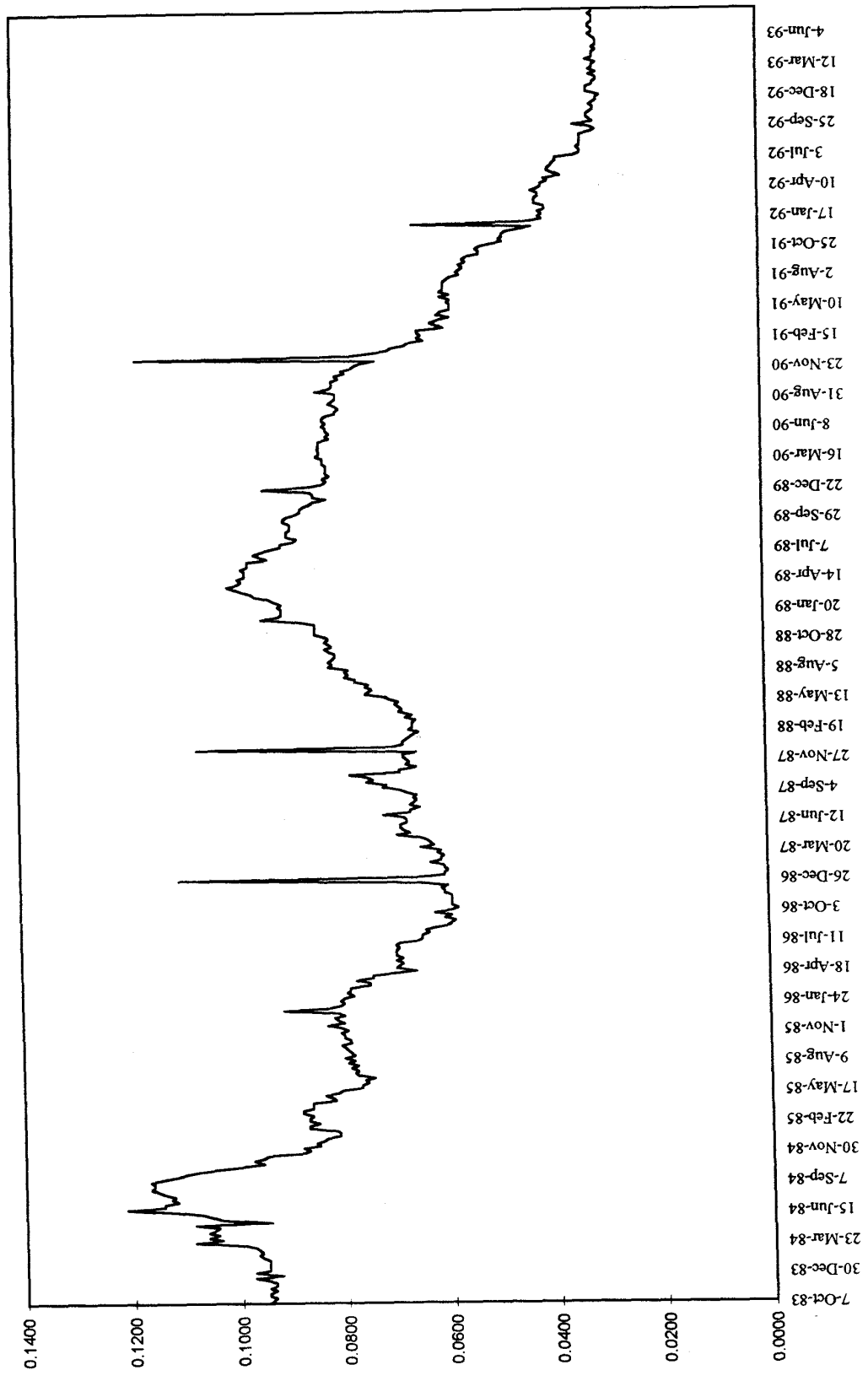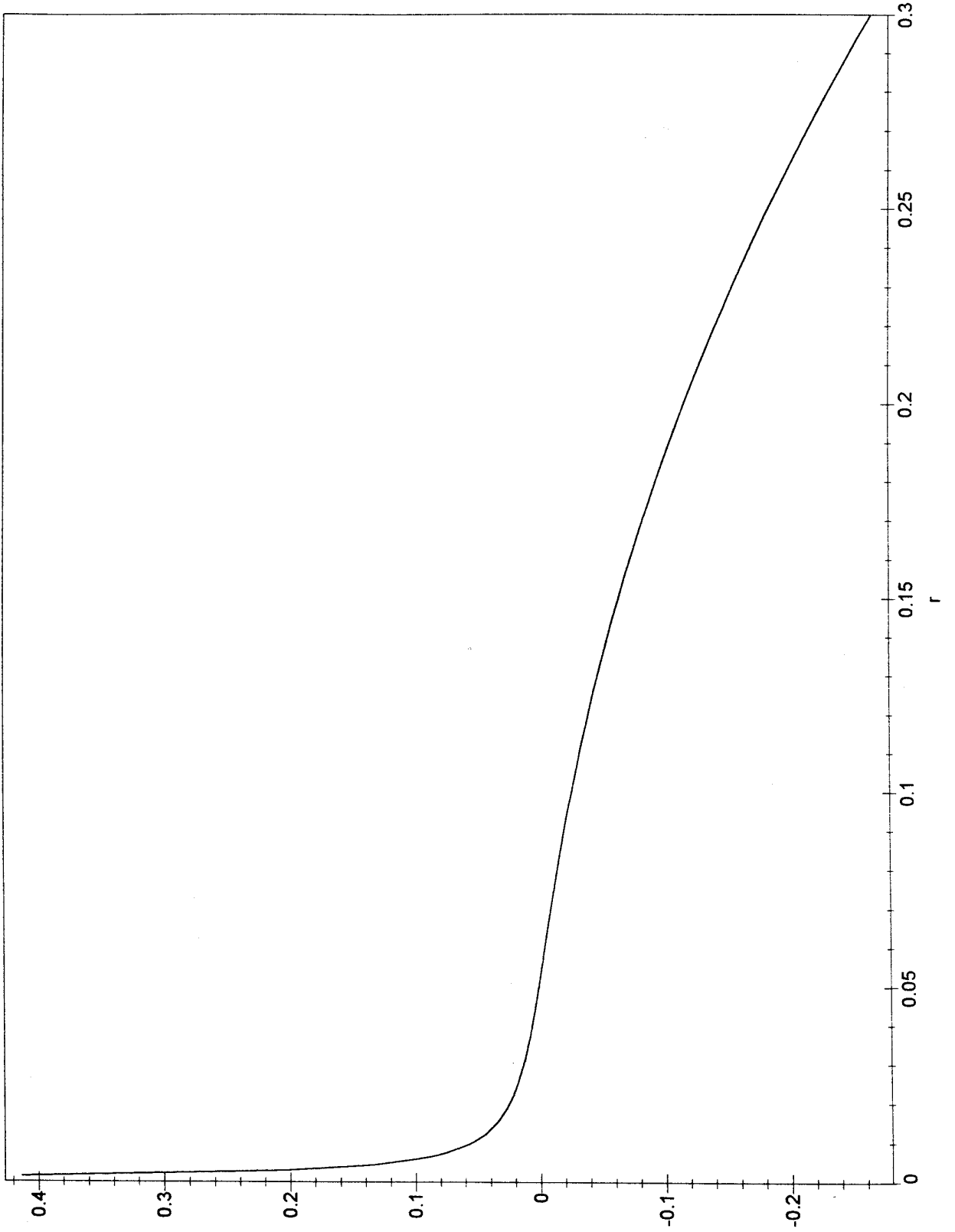| Model	SLV
Linear Drift, CEV Diffusion	2028.98
Quadratic Drift, CEV Diffusion	2032.60
General Drift, CEV Diffusion	2034.43
General Model	2056.59

Figure 1
Seven Day Interest Rate

Figure 2: Drift Function

Figure 3: Diffusion Function