Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

# FITTING DIFFUSION MODELS IN FINANCE

Don L. McLeish and Adam W. Kolkiewicz
University of Waterloo

## ABSTRACT

This paper is concerned with the problem of estimation for stochastic differential equations based on discrete observations when the likelihood formula is unknown. Often in the financial literature the first order discrete-time approximation to the diffusion process is considered adequate for the purpose of simulation, estimation and fitting the model to historical data. We propose methods of estimation based on higher order Ito-Taylor expansions. Different methods of generating optimal estimating functions are considered and a method of quantifying the loss of information due to using lower order approximations is proposed. An important feature of these methods is that an assessment of the goodness of fit to data is possible. These ideas are illustrated using a model which generalizes most of the single factor diffusion models of the short-rate interest rate used in finance.

**Key Words:** Diffusion models, estimating functions, finance.

# 1 Introduction

Many models common in finance take the form of one or more diffusion equations. Such equations are generally described by means of a stochastic differential equation of the form

$$dX_t = a(X_t)dt + \sigma(X_t)dW_t, \qquad 0 \le t \le T, \tag{1.1}$$

where $W_t$ is an ordinary Wiener process, and the *drift* coefficient $a$ and the diffusion coefficient $\sigma$ may depend on unknown parameters. Markov diffusion models have played a pre-eminent role in the theoretical literature on the term structure of interest rates (e.g. see Brennan and Schwartz (1979), Cox, Ingersoll and Ross (1985), and Longstaff, F.A. (1989)). A few of the most common models are listed in Table 1. Others, such as geometric Brownian motion, are special cases.

**Table 1**

| Model | $a(x,\theta)$ | $\sigma(x,\theta)$ |
|---|---|---|
| Vasicek | $\theta_1 + \theta_2 x$ | $\theta_3$ |
| Cox, Ingersoll, Ross | $\theta_1 + \theta_2 x$ | $\theta_3 x^{1/2}$ |
| Brennan, Schwartz | $\theta_1 + \theta_2 x$ | $\theta_3 x$ |
| Black, Karasinski | $\theta_1 x + \theta_2 x \; log(x)$ | $\theta_3 x$ |
| Cox | $\theta_1 x$ | $\theta_2 x^{\theta_3}$ |
| Pearson, Sun | $\theta_1 + \theta_2 x$ | $(\theta_3 + \theta_4 x)^{1/2}$ |
| Constantinides-Ingersoll | 0 | $\theta x^{3/2}$ |

In this paper, we discuss the estimation of parameters $\theta_i$ in models such as those above for discretely sampled data. That is, on the basis of observations on $X_t$ at discrete time points $t_1 < t_2 < \ldots$, we wish to construct reasonably efficient estimators of the parameters.

Let us consider for example a process defined by the following diffusion equation;

$$dX_t = (\alpha + \beta X_t)dt + cX_t^{\gamma} dW_t.$$

This model generalizes all but one of the diffusions used above and is investigated by Chan, Karolyi, Longstarr and Sanders (1992). When $\beta$ is negative, the process is mean reverting in the sense that it tends towards the value $-\alpha/\beta$, its equilibrium mean. In this case, since the diffusion coefficient is 0 at 0 and the drift term positive in a neighbourhood around 0, the process, if initialized at a positive value $X_0$, remains positive with probability 1. This is a simple consequence of Theorem 2, page 149 of Gihman and Skorohod (1972).

Provided that the function below is integrable, this process has equilibrium distribution given by the probability density function

$$\frac{K}{x^{2\gamma}} exp\{-\frac{2\alpha(\gamma-1) + \beta x(2\gamma-1)}{2x^{2\gamma-1}c^2(2\gamma-1)(\gamma-1)}\}$$

and this equilibrium distribution is well-defined, for example, for $\gamma > 1$. Note that although the process is driven by a Brownian motion, the stationary distribution has tails far from Gaussian. Indeed, moments are only finite up to order less than $2\gamma - 1$.

Now consider a naive discrete time approximation to this process. The obvious discrete approximation is

$$X_{t+h}-X_t= a(X_t)h + \sigma(X_t)\epsilon_t,$$

where $\epsilon_t$ is a sequence of independent normal random variables with mean 0 and variance $h$. This simple *Euler* approximation to the process has some undesirable features. For example, while the original process has an equilibrium distribution and remains positive, the discrete approximation may have neither property. Although the qualitative behaviour of the discrete and continuous processes differ, the maximum likelihood estimator of the drift parameters have similar forms. For example, for the continuous time process observed on the interval $[0, T]$, the maximum likelihood estimator of the parameters $\alpha$ and $\beta$ are given by the solutions of the two estimating equations

$$\int_0^T X_t^{-2\gamma} dX_t - \int_0^T X_t^{-2\gamma}(\hat{\alpha} + \hat{\beta} X_t)dt = 0 \tag{1.2}$$

$$\int_0^T X_t^{1-2\gamma} dX_t - \int_0^T X_t^{1-2\gamma}(\hat{\alpha} + \hat{\beta} X_t)dt = 0 \tag{1.3}$$

and the maximum likelihood estimator for the discretely observed process is an analogous function of $X_{nh}$. Generally, under reasonable conditions, the continuous time estimator is consistent as the period of observation approaches infinity i.e. $T \to \infty$. The discrete time estimators are consistent as $T \to \infty$ and $h \to 0$.

The tails of the stationary distribution are large, certainly substantially larger than those of the normal distribution. A diffusion model, although driven by a process with Gaussian tails, can generate a process with tails similar to those, for example, of the stable distributions with index less than two. In order to determine whether diffusion models provide an adequate fit to financial data, we begin with data consisting of the yield of 30 year US bonds, the data obtained daily over the period April 13, 1987 to June 13, 1994. There are a total of 1818 recorded daily observations in this period. We begin by attempting to fit a general diffusion model of the above form (cf. Chan, Karolyi, Longstarr and Sanders (1992))

$$dX_t = (\alpha + \beta X_t)dt + c{X_t}^{\gamma} dW_t.$$

The yields are plotted in Figure 1.

We discuss the estimation of the parameters in the next section. At the moment we simply note that estimated values in this case are $\hat{\alpha} = 3.1827$, $\hat{\beta} = -0.3962$, $\hat{c} = 0.0075$, $\hat{\gamma} = 2.5813$, indicating a tendency for the process to fluctuate around the mean $-\alpha/\beta$ of approximately eight

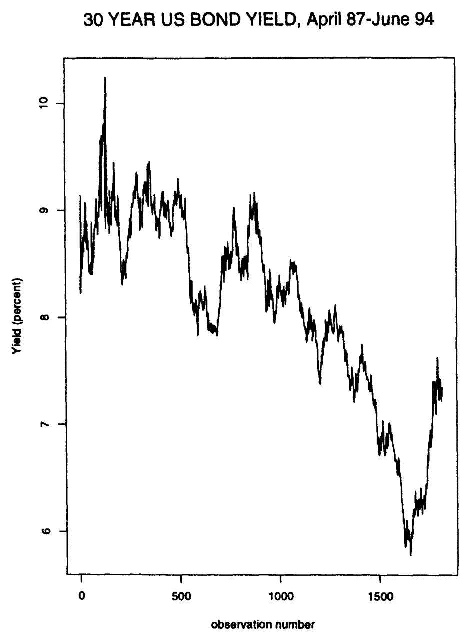


Figure 1:

percent. The graph of the fitted equilibrium probability density function in Figure 2 further illustrates that the right tails are distinctly non-Gaussian.

Does the fitted diffusion adequately explain the tails of the increments of the process? Consider the standardized "Euler residuals", defined by $r_i = (\Delta X_{t_i} - a(X_{t_i})\Delta_{t_i})/\sigma(X_{t_i})$, where $\Delta_{t_i} = t_{i+1} - t_i$ and $\Delta X_{t_i} = X_{t_{i+1}} - X_{t_i}$. Provided that the discretization intervals are sufficiently small, this should be approximately a sequence of independent random normal variables. Figure 3 displays a Normal probability plot of these values against the normal quantiles. This plot shows clear evidence that the residuals are non-normal, indeed have tails more like those of a stable law. In fact if we fit a symmetric stable law to these residuals, we obtain index $\alpha$ =1.69. Almost the same value is obtained if we fit a symmetric stable law to the marginal distribution of the increments $\Delta X_{t_i}$.

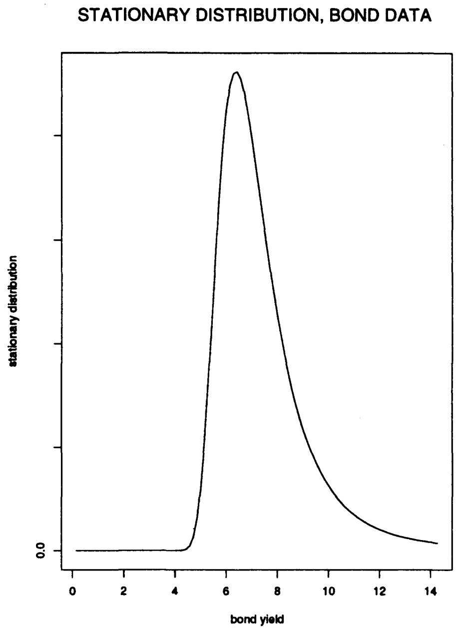


Figure 2:

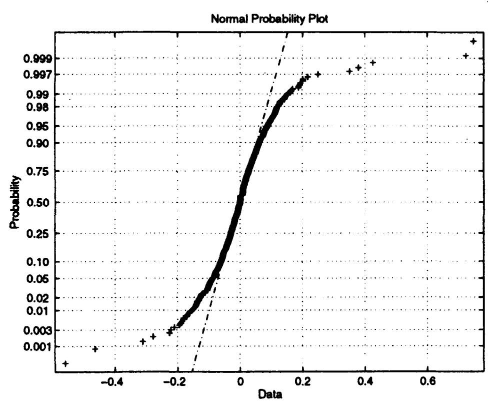


Figure 3:

We seek an explanation of these tails in the Ito-Taylor expansion of the diffusion equation. According to the Ito-Taylor expansion (cf. Kloeden and Platen (1992), page 164), $\Delta X_{t_i}$ can be expressed using a linear combination of the Hermite polynomials in the normalized increments of the Brownian motion. Indeed, with $Z_1 = \Delta W_{t_i}/\sqrt{\Delta_{t_i}}$, we have

$$\Delta X_{t_i} - E(\Delta X_{t_i}) = a_1 Z_1 + a_2(Z_1^2 - 1) + a_3 Z_1(Z_1^2 - 3) + a_4 Z_2 \tag{1.4}$$

where the coefficients $a_i$ are functions of $X_{t_i}$ and $Z_2$ is a standard normal random variable independent of $Z_1$. Notice that the distribution of the increment, centered at its expectation is not exactly normal, and indeed if the coefficients $a_2$, and $a_3$ are reasonably large compared with $a_1$, they increase the weight in the tails of the distribution. Two questions arise immediately from this observation.

- How do we use the representation (1.4) to estimate parameters in a diffusion?
- Does this representation adequately explain the increased weight in the tails of the residuals?

Despite a vast literature, many practical issues related to the problem of estimation and fitting a model from discretely sampled diffusion processes remain unanswered and only few quantitative results regarding consequences of discretization are available. For example, it is common practice to estimate the parameters using low order approximations to the diffusion process, for example the Euler scheme (e.g. Chan, Karolyi, Longstarr, and Sanders (1992)) or the Milstein scheme (e.g. Chesney, Elliott, Madan, and Yang (1993)). This is justified by asymptotic results which apply when the time intervals $t_{i+1} - t_i$ converge to zero, but also by a lack of simple estimating procedures based on higher order approximations. It is not always clear whether the observed discretization is fine enough to justify the use of the lowest order approximations or whether higher order approximations will contribute something significant from the perspective of statistical modeling. In this study we propose several methods of estimation based on different order approximations and also methods which allow for assessing the goodness-of-fit. One advantage of the proposed methods is that it is possible to compare their relative performance.

## 2 Estimation of Parameters

Let us suppose that a process $X_t$ satisfies (1.1), and on the basis of observations at discrete time points $t_1 < t_2 < \dots$, we wish to construct estimators of the parameters. There are several reasonable approaches to this problem.

(1) When the parameter lies in the drift term, we may construct the continuous time maximum likelihood estimators as in (1.2) and (1.3) above and then approximate the integrals by sums.

(2) Again when the parameter lies in the drift term, we may construct the continuous time maximum likelihood estimating functions and then condition these estimating functions on the observed discrete data.

(3) We may base the estimation on the discrete data alone, using the exact or approximate score function for discrete observations. This is equivalent to (1.2) in the case of drift parameters.

Unfortunately, except for few simple examples, only the first approach is generally feasible. Both the second and third approaches require some simplification to the problem.

How should we estimate the parameters if a given sequence of observations, centered at the expected values, has distribution given exactly by equation (1.4)? One possibility is to determine the score function for the distribution, and, using it as an estimating function, obtain the maximum likelihood estimators. Unfortunately, this is a rather difficult task. Alternatively, we may project this score function on some more suitable subspace. Such an approach guarantees the optimal estimating function in the chosen class. For example, the Hermite polynomials $h_i$, given by $h_i(x) = x$, $(x^2-1)$, $(x^3-3x)$ respectively for $i = 1, 2, 3$, provide a reasonable basis for expanding functions of near-normal random variables. These functions have mean zero and variance $i!$ and they are uncorrelated under a normal assumption. Projecting the score function onto these polynomials is equivalent to projecting onto a space spanned by the powers of $\Delta X_{t_i}$. Kessler and Sorensen (1995), for example, choose as basis functions for the space the eigenfunctions of the infinitesimal generator of the Markov process.

## 2.1 The Ito Taylor Expansion.

Higher order Ito–Taylor expansions may be used to approximate score functions by their projections onto a space spanned by polynomials.

Let $\{t_1, \ldots, t_n\}$ be the points at which we observe the diffusion process $\{X_t\}$. In the representation of $X_{t_{i+1}}$

$$X_{t_{i+1}} = X_{t_i} + \int_{t_i}^{t_{i+1}} a(X_s, \theta)ds + \int_{t_i}^{t_{i+1}} \sigma(X_s, \theta)dW_s \tag{2.1}$$

Ito's lemma can be written in terms of two operators on twice differentiable functions $f$:

$$L^o = a\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2},$$

and

$$L^1 = \sigma\frac{\partial}{\partial x}.$$

Then for any twice differentiable function $f$,

$$f(X_{t_{i+1}}) = f(X_{t_i}) + \int_{t_i}^{t_{i+1}} L^0 f(X_s)ds + \int_{t_i}^{t_{i+1}} L^1 f(X_s)dW_s. \qquad (2.2)$$

By substituting in each of the integrands in (2.1) using the above identity and iterating this process we arrive at the Ito-Taylor expansions (e.g. Kloeden and Platen, (1992) page 164). It is easy to observe that terms with nonzero expectation will come only from the first integral and they will be of the form

$$M_{j,i} \stackrel{def}{=} (L^o)^{(j-1)} a(X_{t_i}, \theta) \frac{(\Delta_{t_i})^j}{j!}, \quad j = 1, \ldots, \qquad (2.3)$$

where $(L^o)^{(j)}$ denotes the j-th iteration of the operator $L^o$ with $(L^o)^{(0)}$ being the identity operator. These terms provide successive approximations to the conditional expectations of $\Delta X_{t_i}$ $(= X_{t_{i+1}} - X_{t_i})$:

$$E(\Delta X_{t_i} | X_{t_i}) \approx m_{r,i} \stackrel{def}{=} \sum_{j=1}^{r} M_{j,i}, \quad r = 1, \ldots. \qquad (2.4)$$

The first two approximations are $m_{1,i} = a(X_{t_i}, \theta)\Delta_{t_i}$,which is equal to the first term in the Euler expansion, and

$$m_{2,i} = a(X_{t_i}, \theta)\Delta_{t_i} + \frac{1}{2}[aa' + \frac{1}{2}\sigma^2 a^{(2)}](X_{t_i}) \cdot \Delta_{t_i}^2,$$

which corresponds to the terms with nonzero expectation in the strong Ito–Taylor expansion of order 1.5. In the latter approximation $a'$ and $a^{(2)}$ denote the first and second derivatives of $a$ with respect to $x$, respectively. In general, the difference

$$\Delta X_{t_i} - m_{r,i} \qquad (2.5)$$

will have conditionally on $X_{t_i}$ nonzero expectation of order $\Delta_{t_i}^{r+1}$. Using these differences we shall find estimating equations for $\theta$ by finding moments of approximations to the distribution of (2.5).

The distribution of (2.5) is determined by terms coming from the Ito–Taylor expansions of both integrals in (2.1). By gathering these terms, we have the following approximations to the distribution of (2.5 ):

**(a1)** - terms of orders up to $0_P(\Delta_{t_i}^{\frac{1}{2}})$:

$$\sigma(X_{t_i}, \theta) \int_{t_i}^{t_{i+1}} dW_s = \sigma \Delta W_{t_i},$$

which together with the term $m_1$ corresponds to the Euler expansion.

**(a3)** - terms of orders up to $0_P(\Delta_{t_i}^{\frac{3}{2}})$:

$$\begin{aligned}\sigma\Delta W_{t_i} &+ \frac{L^1\sigma}{2}[(\Delta W_{t_i})^2 - \Delta_{t_i}] + \frac{1}{2}L^1 a(X_{t_i})[(\Delta W_{t_i})^2 - \Delta_{t_i}] \\ &+ \frac{1}{6}(L^1)^{(2)}\sigma(X_{t_i})[(\Delta W_{t_i})^2 - 3\Delta_{t_i}]\Delta W_{t_i} \\ &+ L^o\sigma(X_{t_i})[\Delta W_{t_i}\Delta_{t_i} - \Delta Z_{t_i}],\end{aligned}$$

where $\Delta Z_{t_i}$ is a normally distributed random variable with mean, variance and correlation

$$E(\Delta Z_{t_i}) = 0, \quad E((\Delta Z_{t_i})^2) = \frac{1}{3}\Delta_{t_i}^3, \quad \text{and} \quad E(\Delta W_{t_i}\Delta Z_{t_i}) = \frac{1}{2}(\Delta_{t_i})^2,$$

respectively. These terms together with $m_2$ correspond to the strong Ito-Taylor approximation of order 1.5.

By considering more terms we can obtain more accurate approximations to the transition distribution of the discretized process $X_t$. Since it seems that there is no easy method of finding an explicit form of the density function for higher order approximations one may propose to approximate the score functions by their projections onto the space spanned by polynomial functions. It is interesting to notice that the multiple stochastic integrals that arise in Ito-Taylor approximations lead to the Hermite polynomials:

$$\int_{t_i}^{t_{i+1}}\int_{t_i}^{s_n}\cdots\int_{t_i}^{s_2} dW_{s_1}dW_{s_2}\dots dW_{s_n} = \frac{\Delta_{t_i}{}^{n/2}}{n!}h_n(\frac{\Delta W_{t_i}}{\sqrt{\Delta_{t_i}}}),$$

where $h_n$ is the Hermite polynomial of degree $n$.

As an example of the proposed method, we shall consider the approximation (a3) and henceforth we shall assume that it provides adequate approximation to the distribution of the difference (2.5). We rewrite the approximation (a3) in terms of two standard normally distributed and uncorrelated random variables $Z_1$ and $Z_2$:

$$a_{1,i}Z_1 + a_{2,i}(Z_1^2 - 1) + a_{3,i}(Z_1^2 - 3)Z_1 + a_{4,i}Z_2, \tag{2.6}$$

where in the last expression we used the following notation:

$$a_{1,i} = a_{1,i}(\theta) = \sigma\sqrt{\Delta_{t_i}} + \frac{\Delta_{t_i}{}^{3/2}}{2}[\sigma a' + a\sigma' + \frac{1}{2}\sigma^2\sigma^{(2)}],$$

$$a_{2,i} = a_{2,i}(\theta) = \frac{1}{2}\sigma\sigma'\Delta_{t_i}$$

$$a_{3,i} = a_{3,i}(\theta) = \frac{\Delta_{t_i}{}^{3/2}}{6}\sigma[\sigma\sigma^{(2)} + (\sigma')^2],$$

$$a_{4,i} = a_{4,i}(\theta) = \frac{(\Delta_{t_i})^{3/2}}{2\sqrt{3}}[\sigma a' - a\sigma' - \frac{1}{2}\sigma^2\sigma^{(2)}].$$

Since $Z_1$, $Z_1^2 - 1$, $[Z_1^2 - 3]Z_1$, and $Z_2$ are all uncorrelated, it is relatively simple to find higher moments of the random variable (2.6). Using the general theory of estimating functions we can employ these moments to generate optimal estimating equations.

We shall denote the variance, skewness and kurtosis of the differences $\Delta X_{t_i}$, conditionally on $X_{t_i}$, by $\mu_{2,i}$, $\gamma_{1,i}$, and $3 + \gamma_{2,i}$, respectively. From (2.6), we can find the following explicit forms for these moments:

$$\mu_{2,i} = a_{1,i}^2 + 2a_{2,i}^2 + 6a_{3,i}^2 + a_{4,i}^2,$$

$$\gamma_{1,i} = (\frac{1}{\mu_{2,i}})^{\frac{3}{2}}[6a_{1,i}^2 a_{2,i} + 36a_{1,i}a_{2,i}a_{3,i} + 8a_{2,i}^3 + 108a_{2,i}a_{3,i}^2],$$

$$\begin{aligned}\gamma_{2,i} = {} & \frac{1}{\mu_{2,i}^2}[3a_{1,i}^4 + 60a_{2,i}^4 + 3348a_{3,i}^4 + 1296a_{1,i}a_{3,i}^3 + 24a_{1,i}^3 a_{3,i} + 60a_{1,i}^2 a_{2,i}^2 \\ & +252a_{1,i}^2 a_{3,1}^2 + 576a_{1,i}a_{2,i}^2 a_{3,i} + 2232a_{2,i}^2 a_{3,i}^3 + 3a_{4,i}^4 + 12a_{4,i}^2 a_{3,i}^2 \\ & +6a_{4,i}^2 a_{1,i}^2 + 36a_{4,i}^2 a_{3,i}^2] - 3.\end{aligned}$$

We shall use these moments to project the score functions corresponding to the approximation (2.6) onto the space spanned by the function $\{1, x, x^2\}$.

Now, let us assume that $m_{r,i}$ provides, for some value of $r$, a good approximation to the conditional expectation of $\Delta X_{t_i}$. Then it is easy to observe that the functions:

$$f_{1,i}(\Delta X_{t_i}, \theta) = \Delta X_{t_i} - m_{r,i}(\theta),$$

and

$$f_{2,i}(\Delta X_{t_i}, \theta) = (\Delta X_{t_i} - m_{r,i}(\theta))^2 - \mu_{2,i} - \gamma_{1,i}\mu_{2,i}^{\frac{1}{2}}(\Delta X_{t_i} - m_{r,i}(\theta))$$

are orthogonal estimating functions and that the projection of the score function for estimating the parameter $\theta_j$ is of the form:

$$\frac{1}{\mu_{2,i}}\frac{\partial m_{r,i}}{\partial \theta_j} f_{1,i}(\Delta X_{t_i}, \theta) - \frac{\mu_{2,i}^{\frac{1}{2}}\gamma_{1,i}\frac{\partial}{\partial \theta_j}m_{r,i} - \frac{\partial}{\partial \theta_j}\mu_{2,i}}{\mu_{2,i}^2(\gamma_{2,i} + 2 - \gamma_{1,i}^2)} f_{2,i}(\Delta X_{t_i}, \theta). \quad (2.7)$$

If we decide to estimate the parameters using the Euler scheme, then the score function for estimating the parameter $\theta_j$ can be obtained from the normal distribution of the difference $\Delta x_i - m_{1,i}$ and it is of the form

$$\frac{1}{\mu_{2,i}}\frac{\partial m_{1,i}}{\partial \theta_j}[\Delta X_{t_i} - m_{1,i}] + \frac{\frac{\partial}{\partial \theta_j}\mu_{2,i}}{2\mu_{2,i}^2}[(\Delta X_{t_i} - m_{1,i})^2 - \mu_{2,i}], \quad (2.8)$$

where $\mu_{2,i} = \sigma^2(X_{t_i}, \theta)\Delta_{t_i}$. By taking $\gamma_{1,i} = 0$ and $\gamma_{2,i} = 0$, which correspond to a normal distribution, we can obtain from (2.7) the projection of the score function onto the space spanned by the functions $\{1, x, x^2\}$. In this case, the resulting estimating equations are the likelihood equations given by (2.8). Thus, for small $\Delta_{t_i}$ the optimal estimating function (2.8) will be very close to the true score function. When the time interval $\Delta_{t_i}$ is not small enough then instead of (2.8 ) we should consider estimating functions given by (2.7), which are based on higher order approximations.

Let us observe that the estimating equations (2.7) and (2.8) are of the same form but they have different approximations to the first two moments of $\Delta X_{t_i}$ and different weights on the functions $f_{1,i}$ and $f_{2,i}$. If a higher order expansion (like (2.6)) is a good approximation to the distribution of $\Delta X_{t_i}$, then it is possible to quantify the loss of efficiency due to using a lower order approximation, since in this case one may assume that the two estimating functions differ only by their weights in the representation

$$w_{1,i} f_{1,i}(\Delta X_{t_i}, \theta) + w_{2,i} f_{2,i}(\Delta X_{t_i}, \theta). \tag{2.9}$$

Another situation when it is possible to generate a number of unbiased estimating equations which differ only by their weights on the functions $f_{1,i}$ and $f_{2,i}$ is the case when either of the first two moments of $\Delta X_{t_i}$ is known explicitly. For example, when the drift function $a(x, \theta)$ is a linear function of $x$, then the conditional expectation of $\Delta X_{t_i}$ can be found explicitly and then it may be of interest to compare efficiency of different estimating functions based on $f_{1,i}(\Delta X_{t_i}, \theta)$.

Finally, when the time interval $\Delta t_i$ is not sufficiently small to justify the use of estimating functions based on low order expansions and when explicit forms of the first two moments are unknown then Monte Carlo simulations provide one more method of generating unbiased estimating function of form (2.9) (Bibby and Sorensen, 1995). In the next section we consider a method of comparison of such estimating equations.

## 2.2 Estimating Functions of Higher Degree.

So far we have discussed only quadratic estimating functions but the method of projection can be applied to generate polynomial estimating functions. For example, it is possible using only knowledge of the first three moments of $\Delta X_{t_i}$ to project the score function onto the space spanned by the three components of the vector estimating function

$$f_i(\theta) = \begin{pmatrix} (\Delta X_{t_i} - m_{r,i}(\theta)) \\ (\Delta X_{t_i} - m_{r,i}(\theta))^2 - \mu_{2,i}(\theta) \\ (\Delta X_{t_i} - m_{r,i}(\theta))^3 - \gamma_{1,i}(\theta)\mu_{2,i}^{3/2}(\theta) \end{pmatrix}$$

provided we settle for an approximation to the covariance of these terms. These covariances are used only in determining the weights on the estimating functions and so while the resulting estimating function may be slightly sub-optimal, it will be unbiased. For the following, in order to simplify the notation slightly, we assume we are estimating a single scalar parameter $\theta$ which may be one component of the vector of parameters.

Define

$$b_i = \frac{\partial E_\eta f_i}{\partial \eta}|_{\eta=\theta}$$

$$= \begin{pmatrix} \frac{\partial}{\partial\theta} m_{r,i}(\theta) \\ \frac{\partial}{\partial\theta}\mu_{2,i}(\theta) \\ 3(\mu_{2,i}(\theta))\frac{\partial}{\partial\theta}m_{r,i}(\theta) + \frac{\partial}{\partial\theta}\gamma_{1,i}(\theta)\mu_{2,i}^{3/2}(\theta) + \frac{3}{2}\gamma_{1,i}(\theta)\mu_{2,i}^{1/2}(\theta)\frac{\partial}{\partial\theta}\mu_{2,i}(\theta) \end{pmatrix}$$

The conditional covariance matrix of $f_i$ is obtained by using the normal approximation to the distribution to determine the moments of order $\geq 5$ , since the distribution as $\Delta_{t_i} \to 0$ is normal. This yields

$$\Sigma_f = E(f_i f_i^T/\mathcal{F}_{i-1}) = D\begin{pmatrix} 1 & \gamma_{1,i}(\theta) & 3+\gamma_{2,i}(\theta) \\ \gamma_{1,i}(\theta) & 2+\gamma_{2,i}(\theta) & 0 \\ 3+\gamma_{2,i}(\theta) & 0 & 15 \end{pmatrix} D,$$

where $D$ denotes a diagonal matrix with diagonal elements $(\mu_{2,i}(\theta))^{1/2}, \mu_{2,i}(\theta), (\mu_{2,i}(\theta))^{3/2}$ . In this case the projection of the score function onto the linear space spanned by the three components of $f_i(\theta)$ is given by

$$g_n^*(\theta) = \sum_i b_i^T \Sigma_f^{-1} f(\theta).$$

If for small $\Delta_{t_i}$ we replace both $\gamma_{2,i}(\theta)$ and $\gamma_{1,i}(\theta)$ in $\Sigma_f$ by their asymptotic value 0, we obtain an approximation

$$\Sigma_f^{-1} \approx D^{-1}\begin{pmatrix} 5/2 & 0 & -1/2 \\ 0 & 1/2 & 0 \\ -1/2 & 0 & 1/6 \end{pmatrix} D^{-1}.$$

It follows that

$$g_n^*(\theta) \approx [\frac{1}{\mu_{2,i}}\frac{\partial}{\partial\theta}m_{r,i}(\theta) - \frac{1}{2}\{\frac{\partial}{\partial\theta}\gamma_{1,i}(\theta)\mu_{2,i}^{-1/2} + \frac{3}{2}\gamma_{1,i}\mu_{2,i}^{-3/2}(\theta)\frac{\partial}{\partial\theta}\mu_{2,i}(\theta)\}]f_{1,i}(\theta)$$

$$+\frac{1}{2(\mu_{2,i}(\theta))^2}\frac{\partial}{\partial\theta}\mu_{2,i}(\theta)f_{2,i}(\theta)$$

$$+\frac{1}{6}\{\frac{\partial}{\partial\theta}\gamma_{1,i}(\theta)\mu_{2,i}^{-3/2}(\theta) + \frac{3}{2}\gamma_{1,i}\mu_{2,i}^{-5/2}(\theta)\frac{\partial}{\partial\theta}\mu_{2,i}(\theta)\}f_{3,i}(\theta).$$

Now as $\Delta_{t_i} \to 0$ , $\mu_{2,i}(\theta) \sim \sigma^2(\theta)\Delta_{t_i}$ and $\gamma_{1,i} \to 0$ . Assuming this convergence is sufficiently rapid, the above estimating function is asymptotically equivalent to (2.8). Observe that only the first term is used if the parameter lies only in the drift, and only the second term if it is a diffusion parameter.

The normal approximation is not the only alternative to approximating higher order moments of the diffusion. For example, we could assume that (2.6) holds exactly and obtain moments of orders 5 and 6 using this approximation. An alternative approach uses Ito's lemma combined with the approximations (2.3) and (2.4) to the conditional expectation of $\Delta X_{t_i}$. If we first apply the Ito's lemma to the processes $X_t^2$ , $X_t^3$ , and $X_t^4$ , and then use (2.3) and (2.4), we can approximate, in principle to any degree, the first four conditional moments of $\Delta X_{t_i}$. For example, we have the following approximations to the conditional second moment

$$E[X_{t_{i+1}}^2 - X_{t_i}^2 | X_{t_i}] \approx \sum_{j=1}^{r} (\underline{L}^o)^{(j-1)} \underline{a}(X_{t_i}^2, \theta) \frac{(\Delta_{t_i})^j}{j!}, \quad r = 1, \ldots,$$

where

$$\underline{L}^o = \underline{a}\frac{\partial}{\partial x} + \frac{1}{2}\underline{\sigma}^2 \frac{\partial^2}{\partial x^2},$$

and

$$\begin{aligned} \underline{a}(x,\theta) &= 2a(\sqrt{x})\sqrt{x} + \sigma^2(\sqrt{x}), \\ \underline{\sigma}(x,\theta) &= 2\sigma(\sqrt{x})\sqrt{x}. \end{aligned}$$

Since for any admissible function $h$ we have $\underline{L}^o h|_{x^2} = L^o h(x^2)$, we arrive at

$$E[X_{t_{i+1}}^2 - X_{t_i}^2 | X_{t_i}] \approx \sum_{j=1}^{r} (L^o)^{(j)} X_{t_i}^2 \frac{\Delta_{t_i}^j}{j!}, \quad r = 1, \ldots,$$

Similar approximations can be obtained for higher moments of the process $X_t$:

$$E[X_{t_{i+1}}^k - X_{t_i}^k | X_{t_i}] \approx \sum_{j=1}^{r} (L^o)^{(j)} X_{t_i}^k \frac{\Delta_{t_i}^j}{j!}, \quad r = 1, \ldots,$$

From the differences of moments of $X_t$, we obtain the moments of $\Delta X_{t_i}$. Although computations involved in these approximations are still complex, they are simpler than for the method which uses approximations to the distribution of $\Delta X_{t_i}$. In addition, now calculations can be carried out using symbolic computer languages, like, for example, Maple. A drawback of this approach is that it does not allow for a simple assessment of the goodness-of-fit of a model.

# 3 Relative efficiency for estimators based on different order approximations

In this section we use methods of the general theory of estimating equations to compare estimating functions generated from different approximations to the distribution of the increments $\Delta X_{t_i}$.

Let $\mathcal{G}$ be a class of zero mean, square integrable p-dimensional estimating functions $g_n(X_1, \ldots, X_n; \theta)$ which are almost surely differentiable with respect to the components of $\theta$ and such that $E(\dot{g}_n) = E(\frac{\partial}{\partial \theta_j} g_{n,i})$ and $E(g_n g_n^T)$ are nonsingular. Suppose also that $\{g_n, \mathcal{F}_n\}$ is a martingale whose quadratic characteristic is $\{< g >_n, \mathcal{F}_n\}$.

Let $f_i = f(X_1, \ldots, X_i; \theta)$, $1 \leq i \leq n$, be specified d-dimensional vectors that are martingale differences and suppose that we want to find an optimal estimating function from the class $\mathcal{M} \subseteq \mathcal{G}$ of martingale estimating functions of the form

$$g_n = \sum_{i=1}^{n} w_i^T f_i,$$

with the $w_i$ being matrices which are $\mathcal{F}_{i-1}$ measurable. In the theory of optimal estimating equations two criteria are used: the small sample optimality criterion ($O_F$-optimality) and the asymptotic optimality criterion ($O_A$-optimality) (cf. Godambe and Heyde, 1987). Both optimality criteria are satisfied by the same estimating function

$$g_n^* = \sum_{i=1}^{n} w_i^{*T} f_i, \tag{3.1}$$

with

$$w_i^* = -(E(\dot{f}_i|\mathcal{F}_{i-1}))^T (E(f_i f_i^T / \mathcal{F}_{i-1}))^{-1}.$$

In the one dimensional case ($d = 1$), the two criteria are equivalent to finding an estimating function which minimizes either

$$\frac{[E(\frac{\partial}{\partial \theta} g_n)]^2}{E(g_n^2)}, \tag{3.2}$$

for $O_F$-optimality, or

$$\frac{\sum_{i=1}^{n} E((w_i^T f_i)^2 | \mathcal{F}_{i-1})}{[\sum_{i=1}^{n} w_i^T E(\frac{\partial}{\partial \theta} f_i | \mathcal{F}_{i-1})]^2}, \tag{3.3}$$

for $O_A$-optimality.

The reciprocal $I_{g_n}$ of the quantity (3.3) is called the martingale information in $g_n$. This information occurs as a scale variable in the asymptotic distribution of the estimator obtained as the solution to equation $g_n(\theta) = 0$.

Thus maximizing $I_{g_n}$ leads to asymptotic confidence regions of minimum size. Using this interpretation, it seems reasonable to define the conditional relative information $CRI(g_1, g_2)$ of an estimating function $g_1$ with respect to a second estimating function $g_2$ as the ratio $I_{g_1}/I_{g_2}$.

When $g = g^*$ we can find the martingale information using the formula

$$I_{g^*} =< g^* >_n= \sum_{i=1}^{n} E((w_i^{*T} f_i)^2 | \mathcal{F}_{i-1}), \tag{3.4}$$

however, for a general estimating function we have to use (3.3).

The above concepts can be applied to the problem of estimation described in the previous section. Under the measure $\mathcal{P}$ derived from the 1.5 strong approximation (2.6) and under the assumption that $m_{r,i}$ is an adequate approximation to the conditional expectation of $\Delta X_{t_i}$, the optimal estimating function $g_n^*$, which is in the space spanned by the functions $f_{1,i}$ and $f_{2,i}$, is given by (2.7). We shall denote the optimal coefficients by $w_{1,i}^*$ and $w_{2,i}^*$. Suppose also that we have another estimating function which is of the form

$$g_n = \sum_{i=1}^{n} w_{1,i} h_{1,i} + w_{2,i} h_{2,i} ,$$

where

$$h_{1,i} = \Delta X_{t_i} - m_{r,i}, \quad \text{and} \quad h_{2,i} = (\Delta X_{t_i} - m_{r,i})^2 - \mu_{2,i} ,$$

and we would like to compare the martingale information contained in both functions. Note that in the above representation of $g_n$ the moments $m_{r,1}$ and $\mu_{2,1}$ are the same as in the optimal estimating function $g_n^*$. These may be different from what we are actually using when building an estimating function based on lower order approximation but when comparing the martingale information we have to deal with unbiased estimating equations. For example, the weights in $g_n$ may come from the Euler scheme but $m_{r,i}$ and $\mu_{2,i}$ may be based on a higher order scheme or obtained from simulation.

For the optimal estimating function (2.7) the martingale information can be determined using (3.4)

$$\begin{aligned} I_{g_n^*} &= \sum_{i=1}^{n} [w_{1,i}^{*2} E(f_{1,i}^2 | \mathcal{F}_{i-1}) + w_{2,i}^{*2} E(f_{2,i}^2 | \mathcal{F}_{i-1})] \\ &= \sum_{i=1}^{n} [w_{1,i}^{*2} \mu_{2,i} + w_{2,i}^{*2} \mu_{2,i}^2 (2 + \gamma_{2,i} - \gamma_{1,i}^2)] \\ &= \sum_{i=1}^{n} \left[ \frac{1}{\mu_{2,i}} \left( \frac{\partial m_{r,i}}{\partial \theta_j} \right)^2 + \frac{[\mu_{2,i}^{\frac{1}{2}} \gamma_{1,i} \frac{\partial}{\partial \theta_j} m_{r,i} - \frac{\partial}{\partial \theta_j} \mu_{2,i}]^2}{\mu_{2,i}^2 (2 + \gamma_{2,i} - \gamma_{1,i}^2)} \right], \end{aligned}$$

where we used notation from the previous section. For the second estimating function $g_n$ we have the following formula

$$I_{g_n} = \frac{[\sum_{i=1}^{n} w_{1,i}\frac{\partial}{\partial\theta_j}m_{r,i} + w_{2,i}\frac{\partial}{\partial\theta_j}\mu_{2,i}]^2}{\sum_{i=1}^{n}[w_{1,i}^2\mu_{2,i} + w_{2,i}^2\mu_{2,i}^2(2+\gamma_{2,i}) + 2w_{1,i}w_{2,i}\gamma_{1,i}\mu_{2,i}^{\frac{3}{2}}]}. \quad (3.5)$$

Using these formulae we can find the information contained in estimating functions based on different approximations to the transition distribution and in that way we can more easily assess the merits of using higher order expansions for a particular model-data combination.

For illustration of the effects of discretization on different estimating equations let us consider again the model

$$dX_t = (\alpha + \beta X_t)dt + cX_t^{\gamma}dW_t, \quad (3.6)$$

with the following values of the parameters:

$$\alpha = 3.2, \quad \beta = -0.4, \quad c = 0.01, \quad \text{and } \gamma = 2.5,$$

which are close to the values which are observed in practice (Section 1).

The relative conditional efficiencies of the optimal quadratic estimating functions based on the Euler scheme with respect to the quadratic estimating functions based on the strong Ito-Taylor approximation of order 1.5 are plotted in Figure 4 .

For simulation purposes we assumed that the process was observed at discrete equidistant points $\Delta, 2\Delta, \ldots, n\Delta$, with $n = 260$. Then we compared the conditional information of the two estimating equations at different values of $\Delta$, with $\Delta = 1$ corresponding to daily observations. The reported values are means of the relative information calculated from five different trajectories of the process.

The graphs show that the effect of discretization may be different for different parameters. While for the parameters in the drift term the efficiency of the two methods remains virtually unchanged for all values of $\Delta$, for $c$ and $\gamma$ the changes are quite visible suggesting a slightly higher efficiency of the estimating equation based on the higher order approximation. The largest drop in efficiency is about 5% and it occurs for the biweekly observations. Overall, for the given values of the parameters, the two methods of estimation show very similar performance.

One may argue that in the above comparison the two estimating equations give rise to estimators with almost the same efficiency because in the estimating equation based on the Euler scheme we used the moments from the higher scheme, "borrowing" in this way efficiency from the more accurate approximation. We now consider a method of comparison which allows us to compare estimating equations without this adjustment.

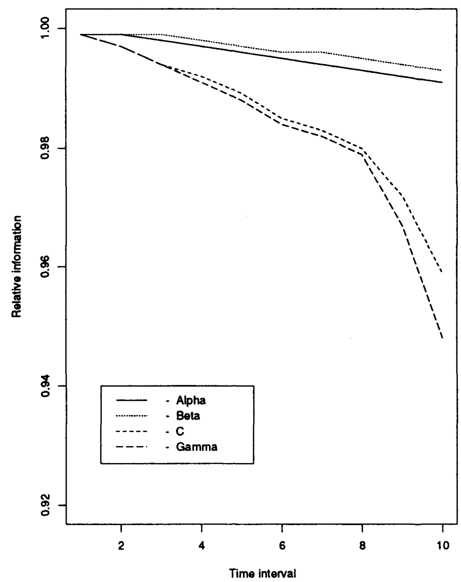


Figure 4:

Suppose that we want to compare the efficiency of the estimating equation (2.7) and (2.8), assuming that $\{\Delta t_i\}$ are not small enough to use the Euler scheme but the strong approximation of order 1.5 provides a good approximation to the conditional distribution of $\Delta X_{t_i}$. This would imply that the estimating equation (2.8) generated from the Euler scheme is not unbiased and therefore the previous methods of comparison cannot be applied directly. It is possible, however, to compare the information contained in each of these two estimating equations, at least up to order $O(\Delta^2)$, if we use the asymptotic results presented by Florens-Zmirou (1989).

Suppose that a parameter $\theta$ is to be estimated from a discrete equidistant observations, $X_\Delta, \ldots, X_{n\Delta}$, of the process $\{X_t\}$ with constant diffusion defined by

$$dX_t = a(X_t, \theta)dt + \sigma dW_t,$$

and we are using estimating equations of the form

$$g_n = \sum_{i=1}^{n} h(X_{(i-1)\Delta}, X_{i\Delta}; \theta).$$

Assume that the process $X_t$ is ergodic and its invariant measure is given by $\mu_\theta$. Furthermore, let $Q_t^\theta = \pi_t^\theta \times \mu_\theta$, where $\pi_t^\theta(dy, x) = P_\theta(X_t \in dy | X_o = x)$ is the transition density of $X_t$. By $\theta_o$ we shall denote the true value of the parameter.

Florens-Zmirou (1989) shows that if $h$ is such that $\int h^2 dQ_\Delta^\theta < \infty$ and $a$ satisfies some regularity conditions then

$$\frac{1}{n}\sum_{i=1}^{n} h(X_{(i-1)\Delta}, X_{i\Delta}; \theta) \to \int h(x, y; \theta) dQ_\Delta^{\theta_o} \quad \text{in} \quad L^2(P_{\theta_o}), \quad \text{as} \quad n \to \infty.$$

Also, if the equation $\int h(x, y; \theta) dQ_\Delta^{\theta_o} = 0$ has an unique solution, $\theta_\Delta$ say, and $\sigma$ and $h$ satisfy some regularity conditions then the estimating equation $g_n = 0$ gives estimates $\hat{\theta}_n$ such that $\hat{\theta}_n \to \theta_\Delta$ in probability under $P_{\theta_o}$, and

$$\sqrt{n}(\hat{\theta}_n - \theta_\Delta) \xrightarrow{\mathcal{D}} N(0, V_\Delta), \tag{3.7}$$

where

$$V_\Delta = \frac{[\int h^2(x, y; \theta_\Delta) dQ_\Delta^{\theta_o} + O(\Delta^4)]}{\int \dot{h}(x, y; \theta_\Delta) dQ_\Delta^{\theta_o}}. \tag{3.8}$$

These results can be generalized to processes with non-constant diffusion term by the well known transformation

$$t(x) = \int_0^x \sigma(y; \theta)^{-1} dy.$$

In view of (3.7) and (3.8), it seems reasonable to define information contained in the estimating function $g_n$, which now may be biased, as

$$I_{g_n}^* = \frac{[E_{\theta_o} \dot{h}(X_0, X_\Delta; \theta_\Delta)]^2}{E_{\theta_o} h^2(X_0, X_\Delta; \theta_\Delta)}.$$

Since we do not know $\theta_o$, it is not possible in practice to calculate $I_{g_n}^*$. We may use, however, the observed information

$$\hat{I}_{g_n}^* = \frac{[\sum_{i=1}^n \dot{h}(X_{(i-1)\Delta}, X_{i\Delta}; \hat{\theta}_\Delta)]^2}{\sum_{i=1}^n h^2(X_{(i-1)\Delta}, X_{i\Delta}; \hat{\theta}_\Delta)}, \tag{3.9}$$

which under some regularity conditions and suitably normalized will converge to $I_{g_n}^*$. The observed information $\hat{I}_{g_n}^*$ is analogous in form to the martingale information $I_{g_n}$ which we used for unbiased estimating equations. The main difference is that the information $\hat{I}_{g_n}^*$ does not involve conditional expectations calculated under the specified model and therefore can be regarded as robust information (for general discussion about robust versions of information see Barndorff-Nielsen and Sorensen, (1994)). In addition, $\hat{I}_{g_n}^*$

retains its meaning even if $\hat{\theta}_\Delta$ does not converge to the true value of the parameter, provided $\Delta$ is small enough.

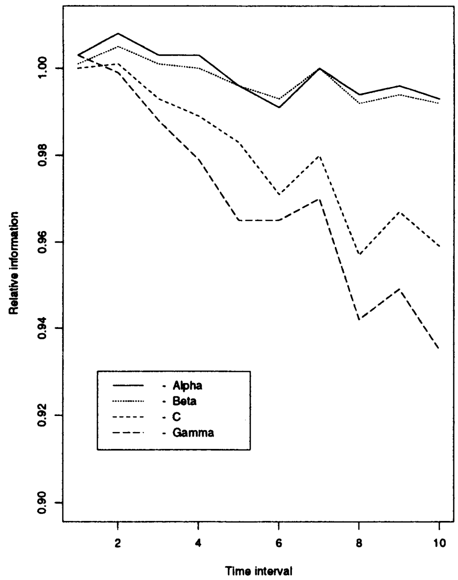


Figure 5:

We repeated the simulation study for the process (3.6) with the same values of the parameters and the same number of observations. This time each of the four parameters was estimated individually using two methods: the optimal quadratic estimating equation derived from the Euler scheme and the optimal quadratic estimating equation based on the order 1.5 strong Taylor approximation. Then, using the expression (3.9), the observed information was calculated for both methods. The procedure was repeated for different values of the time increments $\Delta$, with $\Delta = 1$ corresponding, as before, to the daily observations. Figure 5 shows the averages of the relative observed informations for the two methods, based on five different trajectories of the process.

The relative observed information exhibits a higher variability than in the previous simulation, but, otherwise, the two methods of estimation show very similar performance. It seems that for the given values of the parameters of the process (3.6) and the time intervals $\Delta$ which range from daily up to biweekly observations, there is not much gain in efficiency when instead of using the optimal quadratic estimating equations based on the Euler scheme we use the quadratic estimating equations based on the order 1.5 strong Taylor approximation. Obviously, this may not be true for a different set of the parameters and/or different sampling intervals $\Delta$.

# 4 Bond Data Example and Model Assessment

We now return to the bond data example and write the estimating functions in more explicit form. Here $a(x) = \alpha + \beta x, \quad \sigma(x) = cx^\gamma$. For simplicity, we retain only terms up to order $(\Delta_{t_i})$ in the distribution on the right side of (2.6). These may influence the efficiency of the estimator but not the consistency. In this case we are able to determine the mean exactly since $m(t) = E(X_t)$ satisfies the linear differential equation

$$m'(t) = (\alpha + \beta m(t))\Delta_{t_i}.$$

It follows from the solution of this differential equation that

$$m_{\infty,i} = \frac{1}{\beta}\{(\alpha + \beta X_{t_i})[e^{\beta\Delta_{t_i}} - 1]\}.$$

Also, as in the Milstein approximation,

$$\mu_{2,i} \approx \sigma^2 \Delta_{t_i}\{1 + \frac{1}{2}(\sigma')^2\Delta_{t_i}\} = c^2 X_{t_i}^{2\gamma}\Delta_{t_i}[1 + \frac{1}{2}c^2\gamma^2 X_{t_i}^{2\gamma-2}\Delta_{t_i}].$$

Solving for the root of the estimating functions (2.8) reduces to finding $\alpha, \beta$ minimizing

$$\sum_i w_i(\Delta X_{t_i} - m_{\infty,i})^2, \quad w_i = \frac{1}{\mu_{2,i}}, \tag{4.1}$$

where the weights are held constant while minimizing. For fixed weights, and constant $\Delta t$ this has an explicit solution. Similarly, in principle at least, $c, \ \gamma$ can be found by minimizing the sum of squares

$$\sum 2w_i^2[(\Delta X_{t_i} - m_{\infty,i})^2 - \mu_{2,i}]^2, \tag{4.2}$$

where again the weights are held constant while minimizing. One might substitute an initial consistent estimator of the parameters $c, \gamma$ in the weights, followed by the minimization. Unfortunately, there is little information in this data available for estimating *both* parameters $c, \gamma$ as the contour plot of the sum of squares function illustrated in Figure 6 indicates. There are local minima along a wide curved valley of nearly constant depth. Because of this near unidentifiability, convergence is extremely slow but the values $\gamma = 2.5910$, $\hat{\alpha} = 3.188$, $\hat{\beta} = -.3966$ and $\hat{c} = .0052$ appear to correspond to a local minimum.

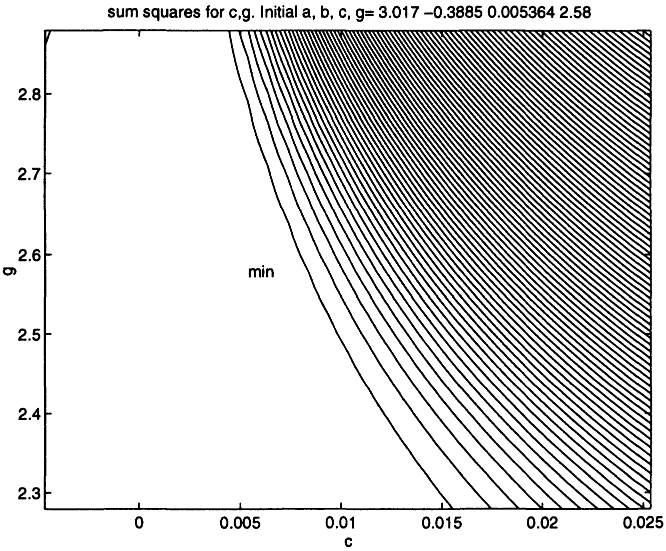


Figure 6:

There remains the question of whether the 1.5 strong order model (1.4) adequately explains the increased weight in the tail observed in Figure 3. Unfortunately, each increment centered at its expectation as in the left side of (1.4) is a function of *two* independent normal random variables $Z_1$ and $Z_2$, either of which might be considered residuals. Since there is not a unique $Z_1$ for each observed increment, we are unable to directly define and analyze normal residuals in the model of (1.4). We propose one possible solution to this problem. Assume for simplicity that $\gamma > 1$. Consider the transformed process $Y_t = X_t^{1-\gamma}$. Then by Ito's lemma, $Y_t$ satisfies a diffusion equation with constant diffusion term

$$dY_t = (\gamma - 1)\{[\frac{\gamma c}{2} Y_t^{-1/(1-\gamma)} - \alpha Y_t^{-\gamma/(1-\gamma)} - \beta Y_t]dt + c dW_t\}$$

$$= \bar{a}(Y_t)dt + c(\gamma - 1)dW_t, \quad say. \tag{4.3}$$

It follows from the representation (2.6) for the process $Y_t$ that

$$a_{1,i} = c(\gamma - 1)\sqrt{\Delta_{t_i}}[1 + \frac{\Delta_{t_i}}{2}\bar{a}'(Y_t)], \quad a_{2,i} = a_{3,i} = 0,$$

$$a_{4,i} = \frac{c(\gamma - 1)(\Delta_{t_i})^{3/2}}{2\sqrt{3}}\bar{a}'(Y_t). \tag{4.4}$$

Now because (2.6) is a *linear* combination of two independent normal random variables, if we divide by the standard deviation, $\sqrt{a_{1,i}^2 + a_{4,i}^2}$, the result is a standard normal variate that can be regarded as a standardized residual. Thus, in this case, the standardized residuals are of the form

$$r_i = \frac{\Delta Y_{t_i} - \bar{a}(Y_{t_i})\Delta_{t_i} - \frac{\Delta_{t_i}{}^2}{2}[\bar{a}(Y_{t_i})\bar{a}'(Y_{t_i}) + \frac{1}{2}(c(\gamma-1))^2\bar{a}^{(2)}(Y_{t_i})]}{c(\gamma-1)\sqrt{\Delta_{t_i}}\sqrt{1 + \frac{\Delta_{t_i}\bar{a}'(Y_t)}{2} + \frac{[\Delta_{t_i}\bar{a}'(Y_t)]^2}{3}}}. \quad (4.5)$$

The plot of these residuals is in Figure 7.

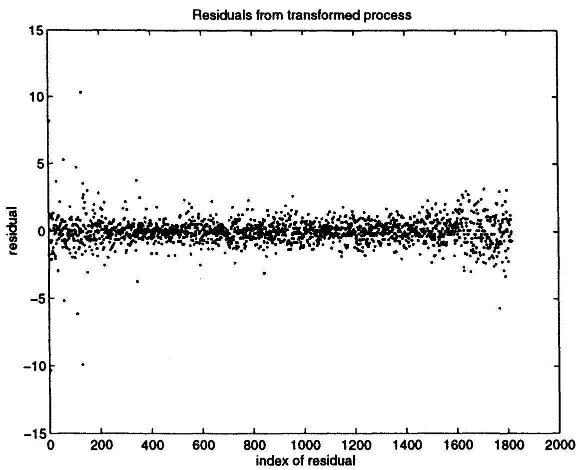


Figure 7:

They seem to indicate a reasonable fit of the model, although there is possible evidence that the diffusion term is not homogeneous in time. We also generate a normal probability plot for these residuals n Figure 8. Note that these have the same basic character as do the Euler residuals, indicating wide tails more consistent with the stable laws, for example, than the Gaussian assumption. This leads us to speculate that the wide-tail phenomenon is not solved by increasing the order of the Ito-Taylor approximation, but requires developing models defined as stochastic integrals with respect to wider-tailed distributions driving the process than Brownian motion. The most obvious of these, while analytically complex, are the stable processes with index less than two.

The transformation, which we use largely so that we are able to define approximately normally distributed residuals from the suggested model, provides alternative estimating functions as well. In general, it is usually possible to transform the original diffusion process so that the diffusion term is constant in $Y_t$, say $\bar{\sigma}$. In this case, the third order Ito-Taylor expansion is particularly simple, and as in this case, it results in a normal distribution since $a_{2,i} = a_{3,i} = 0$. Therefore, assuming the normal approximation to be accurate, the maximum likelihood estimators of the parameters may be obtained by weighted least squares. For example, for a parameter in the drift term $\bar{a}$ only, we minimize

$$\sum_i w_i(\Delta Y_{t_i} - \bar{a}(Y_{t_i})\Delta_{t_i} - \frac{{\Delta_{t_i}}^2}{2}[\bar{a}(Y_{t_i})\bar{a}'(Y_{t_i}) + \frac{1}{2}\bar{\sigma}^2\bar{a}^{(2)}(Y_{t_i})])^2 \tag{4.6}$$

with weights

$$w_i^{-1} = \Delta_{t_i}(1 + \frac{\Delta_{t_i}\bar{a}'(Y_t)}{2} + \frac{[\Delta_{t_i}\bar{a}'(Y_t)]^2}{3}). \tag{4.7}$$

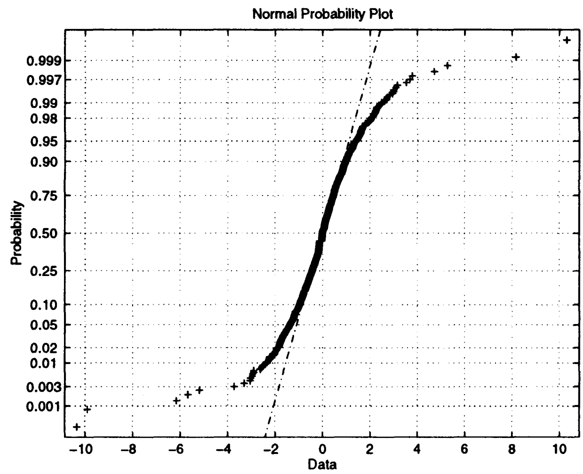


Figure 8:

# 5 Conclusion

It is common practice in finance to use a particular diffusion model, sometimes with several factors, to model a given process and to price derivatives. In many cases, the choice of model is motivated by analytic convenience. Models such as the CIR model have easy solutions and pricing some derivatives is straightforward. We have shown that the Ito-Taylor expansion can be useful for two purposes. The first is calibrating the model or estimating the parameters. This may also be effected by first transforming the model to one which has constant variance term. The order of the Ito-Taylor expansion seems to have limited influence on the efficiency of the estimators when data is collected at discrete time intervals, provided these are not too far apart (e.g. daily data). The second application of the Ito-Taylor expansion, perhaps the more important one, is in assessing the goodness of fit of the data to the model. We speculate that many of the standard diffusion models will tend to fit observed data poorly in the tails of the distribution, and these tails may have considerable influence on the price of derivative products. Alternative models constructed as stochastic integrals with respect to wider tailed alternatives such as the stable laws are likely needed to achieve a satisfactory fit.

## References

Barndorff-Nielsen, O.E. and Sorensen, M. (1994) A review of some aspects of asymptotic likelihood theory for stochastic processes. *Internat. Statist. Rev.* , 61(1), 133-165.

Brennan, M.J. and Schwartz, E.S. (1979) A continuous time approach to the pricing of bounds, *Journal of Banking and Finance*, 3, 133-155.

Bibby, B.M. and Sorensen, M. (1995) Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1(1/2), 17-39.

Chan, K.C., Karolyi, G.A., Longstaff, F.A. and Sanders, A.B. (1992) A n empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance* , Vol. XLVII, 1209-1227.

Chesney, M., Elliott, R.J., Madan, D., and Yang H. (1993) Diffusion coefficient estimation and asset pricing when risk premia and sensitivities are time varying, *Mathematical Finance*, Vol. 3, No. 2, 85-99.

Cox, J.C., Ingersoll J.E., and Ross, S. A. (1985) A theory of the term structure of interest rates, *Econometrica*, 53, 385-406.

Florens-Zmirou, D. (1989) Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 20(4), 547-557.

Gihman, I.I. and Skorohod, A.V. (1972) *Stochastic Differential Equations* Springer-Verlag, Berlin.

Godambe, V.P. and Heyde, C.C. (1987) Quasi-likelihood and optimal estimation. *International Statistical Review*, 55, 3, 231-244.

Kessler, M. and Sorensen, M. (1995) Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Research Report No. 332, Department of Theoretical Statistics, University of Aarhus.*

Kloeden, P.E. and Platen, E. (1992) Numerical Solution of Stochastic Differential Equations. *Springer-Verlag.*

Longstaff, F.A. (1989) A nonlinear general equilibrium model of the term structure of interest rates, *Journal of Financial Economics*, 23, 195-224.