Report n. 115

One dimensional SDE models, low order numerical methods and simulation based estimation: a comparison of alternative estimators

Eugene M. Cleur, Piero Manfredi

Pisa, Marzo 1997

One dimensional SDE models, low order numerical methods and simulation based estimation: a comparison of alternative estimators*

Eugene M. Cleur† Piero Manfredi

March 16, 1998

Abstract

We evaluate the effects of several discretisation schemes on alternative estimators of the drift parameters of stochastic differential equations, namely the continuous time MLE, a so-called naive estimator and an indirect estimator obtained through calibration. Two main results are evidenced: first, the importance of correctly generating data in a simulation based estimation procedure and second, the role of an indirect estimation procedure through calibration as a general strategy to be used every time the conditions of the estimation experiment are not the optimal ones.

Key words: *stochastic differential equation models, simulation of trajectories, estimation of drift parameters, maximum-likelihood estimation, indirect estimation, calibration*

1 Introduction

In this paper a simulation based approach is used to evaluate the effects of several discretisation schemes on alternative estimators of the drift parameters of stochastic differential equations (SDEs), namely the discretised

*The authors very warmly thank Carlo Bianchi, Giorgio Calzolari and Lorenzo Panattoni for invaluable comments on an earlier draft of this paper and the two anonymous referees whose critical observations have led to a great improvement in the presentation. The authors would also like to thank the MURST for financial support through the Fondo Ministeriale 40%.

†address of both authors: Facoltà di Economia, Via Ridolfi 10, 56124 Pisa, Italy. e-mail: cleur@ec.unipi.it and manfredi@ec.unipi.it

continuous time maximum likelihood estimator, a so called naïve estimator and an indirect estimator recently proposed by Gourieroux, Monfort and Renault (1993) (GMR (1993) for short).

We consider, in particular, several first-order SDE models taken from the recent financial literature (namely the well known Vasicek model, the Cox, Ingersoll, Ross (CIR) model and the Brennan - Schwarz model[1]) and the so-called heuristically derived logistic equation which is a stochastic variant of the classical logistic equation of population dynamics. These models were deliberately chosen in view of their common structure; they exhibit, in fact, analogous problems concerning the numerical approximation of their time paths, and they provide a common form for the maximum likelihood estimator. In addition, they all possess an "exact" solution in the trajectory domain or in the density domain or in both. This means that they may be generated "exactly" in some sense.

In fact, a commonly encountered problem in simulation-based approaches to estimation is the need for well generated data. This requirement is fundamental (in some sense it constitutes a condition of well-posedness of our estimation experiments) in that, otherwise the interpretation of our results could be masked by badly generated data. A dangerous consequence of this could be that the failure of an estimate is attributed incorrectly to a lack of nice statistical properties rather than to data generated incorrectly. These considerations emphasise the fact that a simulation based estimation experiment is well posed if and only if data from the underlying model to be estimated are correctly generated

We investigate the effects of several discretisation schemes on alternative estimators of the drift parameters of stochastic differential equations (SDEs), because SDE having exact solutions constitute a very small family. When no exact solution is available we have to resort to approximate numerical techniques. For this reason, the estimators obtained from the exact solutions are used as a benchmark in the evaluation of the performances of the estimators obtained from data generated from the most common low order discrete time approximations such as the Euler, the Milstein and the Taylor 1.5 order schemes. We show that when the discretisation time step is not sufficiently small, the use of a low order approximation to generate the process may lead to an asymptotic bias in the maximum likelihood estimator. This bias depends on the degree of non-linearity of the underlying process which in turn depends on the value of the tuning parameter k (see equation (1) below).

[1]These are the names commonly employed in the financial literature. As it is known from the literature on stochastic processes, the Vasicek model is an Ornstein and Uhlenbeck process with positive long term equilibrium, the CIR model is a Feller process and the Brennan and Schwarz is a geometric brownian motion with positive long term equilibrium.

We hope to obtain, in this way, at least some insight as to how to approach more complex problems, such as higher order and/or nonlinear SDEs, in which the lack of analytical results that may serve as benchmarks oblige us to completely trust in the results obtained by resorting to approximate numerical schemes. An obvious suggestion would be to use higher order schemes in order to increase the accuracy of the approximations.

We show that when the "experimental conditions" are correct, i.e. when the underlying continuous processes are generated correctly, the discretised continuous time maximum likelihood estimator (DMLE) seems to work quite well. A serious bias appears as expected (see Kloeden et al. (1995, 2.13)) only under quite stressed situations such as when the length of the observation period [0,T] is quite small, or when the amplitude of the discretisation step is large. For these situations other types of ML "allied" estimators have been proposed, based on notions such as the "martingale compensator" (Bibby and Soerensen 1995). For such "seriously stressed" situations we consider the performances of an indirect estimator proposed by GMR (1993) in removing the bias in the DMLE.

The indirect estimator is obtained by calibrating on the MLEs of the underlying process and on the parameters of an auxiliary model as, for example, in GMR (1993).

The results obtained by calibrating both on the MLE and the naive estimator seem to suggest that, independently of the nature of a particular indirect estimator used in a specific problem, the calibration procedure, which is basic to the indirect approach, rather than constituting an alternative estimation technique, seems to constitute a fundamental strategy which should be employed in order to improve the quality of our estimates every time the conditions of the experiment are not the optimal ones.

Finally we show that a so-called "naive" estimator, which has often been used due mostly to its simplicity, can give very misleading results and should therefore be resorted to only when other more reliable estimators are not available

2 Estimation of drift parameters

Let us consider the problem of the estimation, from a continuous signal over a prescribed interval [0,T], of the drift parameter Θ (Θ can be a vector; it is assumed that it does not appear among the diffusion parameters[2]) of the SDE:

[2] We will not be concerned in this paper with the problem of estimation of the diffusion coefficients.

$$dX_t = a(X_t, \Theta)dt + b(X_t, \sigma)dW_t \tag{1}$$

Equation (1), under suitable regularity conditions, possesses a well behaved solution and, moreover, admits the so called Continuous Time Maximum Likelihood Estimator (CTMLE, see Liptser and Shiryayev 1981, and Kloeden et al. 1995), which is obtained by maximising the likelihood function:

$$L(\Theta) = \exp\left\{ \int_0^T \frac{a(X_s, \Theta)}{\{b(X_s, \sigma)\}^2} dX_s - \frac{1}{2} \int_0^T \frac{\{a(X_s, \Theta)\}^2}{\{b(X_s, \sigma)\}^2} ds \right\} \tag{2}$$

When (1) is linear in the drift parameters, the likelihood (2) assumes a simple structure in which the various "integral statistics" of the underlying process appear only as coefficients of the relevant variables (i.e. the parameters to be estimated), thereby making the optimisation problem simple[3] and hence permitting us to derive explicit expressions for the CTMLE. In this case not only can the CTMLE be explicitly computed, but it also has nice asymptotic properties (Kloeden et al. 1995 and references therein), due to the fact that the corresponding stochastic process belongs to the exponential family[4]. The performances of such estimators for "short signals", on the other hand, are more complex to study analytically. These difficulties are further amplified when, as often happens, data from diffusion processes are not available under the form of a continuous realization of the underlying stochastic process, but rather under the form of a sample taken from it (i.e. a discrete time series). In such cases, it is in general quite difficult to determine explicit likelihood functions on the "discretization grid". A common estimation strategy would then be to construct approximate estimators through a suitable discretization of the various integral statistics. We call such estimators "discretized" CTMLE (DMLE for short).

The investigation of the formal properties of DMLE is difficult so that a simulation based inference, based on the construction of (discrete) pathwise or strong numerical approximations (Kloeden and Platen 1992) of the trajectories of the underlying SDE, becomes particularly appealing. Once the discrete approximated path is obtained, we could substitute it into the corresponding formula for the DMLE and thereby check the performance of the estimator in a straightforward way (see Kloeden et al. 1995).

We argue that if the true underlying process could be generated "exactly" or approximated to a desired level of accuracy, the problem of evaluating the

[3]These considerations on MLE are no longer true if the drift is non-linear in parameters.

[4]Note that the CTMLE is equivalent to the Minimum Distance estimator (weighted least square) for exponential families.

performance of the "discretized" estimator could then be considered as being correctly posed and this should make it possible for us to analyse correctly the effects of factors such as the method of numerical evaluation of the integral statistics, the maximum amplitude, of the time step used in the discretisation and the length T of the solution interval.

3 The "test models" and the common structure of their CTMLE

The four models considered in this paper are quite similar. They can be encompassed by the following compact equation, by varying the two parameters α, β:

$$dX_t = k(\vartheta - X_t)X_t^{\alpha}dt + \sigma X_t^{\beta}dW_t \qquad (3)$$

where k, ϑ and σ are strictly positive. We identify them in the following table:

Table 1. Values of α and β for specific models of the family (3)

MODEL	α	β
Vasicek	0	0
CIR	0	0.5
Brennan and Schwarz	0	1
Heuristically Derived Logistic	1	1

The formal properties of these models are well known. A broad literature is at present available for what concerns their utilisation as mathematical models in finance (Chan et al. (1992)) and in population dynamics (Gard 1987, 1992).

With regards to the problem of estimation of (3), the common dependency of the drift on its structural parameters gives rise to a common form of the likelihood functions of the four models and hence also to common forms for the CTMLE, providing they exist, of the drift parameters. The only differences present are in the "integral statistics" involved (see Appendix A for details). The CTMLE of the parameters k and θ have the form:

$$\widehat{k} = \frac{I_1 I_5 - I_2 I_3}{I_3 I_4 - I_5^2} \quad \widehat{\theta} = \frac{I_1 I_4 - I_2 I_5}{I_1 I_5 - I_2 I_3} \qquad (4)$$

where the I_j are the corresponding "integral statistics" reported in Table 2.

Table 2. Integral statistics involved in the CTMLE for the four models

MODEL	I_1	I_2	I_3	I_4	I_5
Vasicek	$\int_0^T dX_s$	$\int_0^T X_s dX_s$	$\int_0^T ds$	$\int_0^T X_s^2 ds$	$\int_0^T X_s ds$;
CIR	$\int_0^T \frac{dX_s}{X_s}$	$\int_0^T dX_s$	$\int_0^T \frac{ds}{X_s}$	$\int_0^T X_s ds$	$\int_0^T ds$;
Brennan and Schwarz	$\int_0^T \frac{dX_s}{X_s^2}$	$\int_0^T \frac{dX_s}{X_s}$	$\int_0^T \frac{ds}{X_s^2}$	$\int_0^T ds$	$\int_0^T \frac{ds}{X_s}$
HD Logistic	$\int_0^T \frac{dX_s}{X_s}$	$\int_0^T dX_s$	$\int_0^T ds$	$\int_0^T X_s^2 ds$	$\int_0^T X_s ds$;

For the CIR model, for example, thanks to Ito's formula and a trapezoidal evaluation of the various (nonstochastic) integrals we obtain the following discrete estimates of the various integral statistics:

$$\begin{aligned}
I_1 &= \log \tfrac{X_T}{X_0} + \tfrac{\sigma^2}{2} I_3 \\
I_2 &= X_T - X_0 \\
I_3 &= \tfrac{\Delta}{2} \textstyle\sum_{n=1}^{n_T} \left(X_{n-1}^{-1} + X_n^{-1} \right) \\
I_4 &= \tfrac{\Delta}{2} \textstyle\sum_{n=1}^{n_T} \left(X_{n-1} + X_n \right) \\
I_5 &= T
\end{aligned}$$

where Δ is the discretisation step. By introducing these expressions in (4) we obtain a[5] discrete maximum likelihood estimator (DMLE). In Tables A1 and A2 (see the Appendix) we report the corresponding expressions for all the models considered in this paper.

4 Computational experiments: Discrete Maximum Likelihood Estimates

We first estimate the four models using the DMLE. The given models were considered under parameter constellations guaranteeing the existence of a stationary distribution; this is a useful, but not always necessary assumption (see Kloeden et al. 1995). In all cases we only estimate the two drift parameters k and ϑ by assuming always known the diffusion coefficient σ. This is a simplifying assumption quite common in the specialised literature on the estimation of SDE (see Kloeden et al. 1995, Bibby and Sorensen 1994). It is justified by the theorem of quadratic variation of semi-martingales which

[5] Clearly it is not unique since alternative expressions may be obtained by using different integration rules, such as Simpson's rule, etc.

ensures that, in the presence of highly frequent data, the diffusion coefficient becomes known with probability one. However, in many empirical situations, such as in empirical finance, σ is unknown and must be estimated. This problem has been investigated in the existing literature (see, for instance, Overbeck and Ryden 1997) and by the authors in a forthcoming paper. In our numerical experiments, the estimate of ϑ does not appear to suffer from any bias whatever the conditions of the experiment might be. This is in agreement with the fact that, as expected, the estimation of an "equilibrium" parameter is usually not a problem, at least as long as a sufficiently long series of data is available. For this reason, in what follows, we only report the results related to the estimation of k.

We consider only "equilibrium dynamics", that is, we always set the initial conditions of the processes considered equal to their long term mean ϑ. We do not consider the effects on estimation of transient phases which are "long" as compared to the "equilibrium phase" in which the process has already reached its equilibrium regime.[6]

4.0.1 The CIR model: alternative generators and estimation

We compare the performances of the estimates obtained from data generated by the approximate Euler, Milstein and Taylor 1.5 schemes with the estimates obtained from data generated by the exact conditional chi-squared distribution (CCSD) of the CIR model in order to show to what extent badly generated data can lead to inconsistent results.

The CTMLE has nice asymptotic properties and hence we expect that for sufficiently long (T large) series of well generated data the estimate should be consistent. This result should remain true even for the corresponding discrete-time estimator at least as long as the discretisation step is not large and the numerical evaluation of the given integral statistics is sufficiently accurate. These expected results are fully confirmed by our numerical experiments.

On the other hand, the CTMLE does not necessarily possess good small sample properties. Hence, even if the process is correctly generated but the estimation time span [0,T] is small, the CTMLE could be biased. We must also note that even when T is large the CTMLE will be biased if the process is badly generated. We expect that these results could worsen when we consider their discretized counterparts DMLE. Our experiments fully confirm

[6] The study of the sole transient phase is necessary for instance in processes of physiological growth, where the only relevant parameters are those connected with the dynamics of transition to maturity. Results on some of these effects were presented in (Cleur and Manfredi 1996).

this conjecture as well.

Tables 3-6 report numerical results of a hierarchy of simulation runs for an increasing sequence of values of the tuning parameter k. Our simulations were performed for fixed values of ϑ (ϑ=0.1) and σ(σ=0.06), which are reasonable values for the CIR model, as used in other similar studies (Bianchi et al. 1993, Bianchi and Cleur (1995)). The standard deviations reported in parentheses were calculated from the Monte Carlo estimates obtained from 1000 replications.

Table 3. CIR model: DMLE of k for four generators; T=2000, Δ =0.01 (standard deviations in brackets)

True values of k	Chisquare	Euler	Milstein	Taylor 1.5
0.3	0.3020 (0.0171)	0.3015 (0.0170)	0.3016 (0.0171)	0.3020 (0.0171)
0.8	0.8021 (0.0277)	0.7987 (0.0276)	0.7990 (0.0257)	0.8021 (0.0277)
1.5	1.5018 (0.0378)	1.4903 (0.0373)	1.4908 (0.0373)	1.5019 (0.0378)

Results obtained from 1000 replications

Table 4. CIR model: DMLE of k for four generators; T=100, Δ =0.01 (standard deviations in brackets)

True values of k	Chisquare	Euler	Milstein	Taylor 1.5
0.3	0.3426 (0.0865)	0.3419 (0.0862)	0.3420 (0.0863)	0.3426 (0.0865)
0.8	0.8403 (0.1282)	0.8366 (0.1274)	0.8369 (0.1274)	0.8403 (0.1282)
1.5	1.5387 (0.1719)	1.5264 (0.1699)	1.5269 (0.1699)	1.5387 (0.1720)

Results obtained from 1000 replications

Table 5. CIR model: DMLE of k for four generators; T=5000, Δ =0.1 (standard deviations in brackets)

True values of k	Chisquare	Euler	Milstein	Taylor 1.5
0.3	0.3007 (0.0170)	0.2956 (0.0104)	0.2966 (0.0105)	0.3008 (0.0107)
0.8	0.8009 (0.0175)	0.7673 (0.0164)	0.7700 (0.0164)	0.8019 (0.0175)
1.5	1.5012 (0.0242)	1.3855 (0.0215)	1.3904 (0.0216)	1.5073 (0.0243)

Results obtained from 1000 replications

Table 6. CIR model: DMLE of k for four generators;
T=100, Δ =0.1 (standard deviations in brackets)

True values of k	Chisquare	Euler	Milstein	Taylor 1.5
0.3	0.3395 (0.0856)	0.3330 (0.0835)	0.3341 (0.0835)	0.3396 (0.0857)
0.8	0.8385 (0.1295)	0.8013 (0.1214)	0.8041 (0.1215)	0.8394 (0.1299)
1.5	1.5383 (0.1737)	1.4162 (0.1541)	1.4212 (0.1542)	1.5441 (0.1751)

Results obtained from 1000 replications

When T is large (T=2000) (see Table 3) and the discretisation step is sufficiently small (of the order of magnitude of Δ=0.01), the estimates obtained from the data generated by all four generators are close to the true values and there appears to be no significant bias. When T is small, (T=100) (see Table 4), although we have a reasonably small discretisation step, (Δ=0.01), the estimates from all four generators are heavily biased. These results confirm the previous observation that the DMLE preserves nice asymptotic properties, like all maximum likelihood estimators, but suffers from a significant bias for short series.

However (see Table 5), when the discretisation step is relatively large (Δ=0.1), notable differences set in even though T is maintained large (T=5000). In particular, when k is set to 0.8 and then to 1.5, thereby increasing the convexity in the trajectories of the process, the estimates obtained from the data generated by the CCSD of the process and by the Taylor 1.5 scheme continue to be close to the true values, but those obtained from data generated by the Euler and Milstein schemes exhibit an increasingly significant bias.

The reason for this is simply that as the convexity increases, the approximation of the trajectory by a "linear" method such as the Euler and Milstein schemes provide very "low quality" data to the DMLE. Hence the bad performance of the DMLE for these cases is, basically, a consequence of a badly posed experiment. A heuristic explanation is the following: in our experiments we estimate k by using "regime" data, i. e. the stochastic fluctuation around the long term equilibrium of the process. The approximation of the "true" fluctuation which is provided by a linear scheme, such as Euler and Milstein, is systematically "biased" as compared to both the exact fluctuation and the fluctuation generated by higher order schemes. In particular, when keeping fixed, i. e. keeping under control, the stochastic effects, this bias becomes a monotonically increasing function of both the discretisation step and the degree of convexity of the mean solution curve. Thus, when the mean solution is a convex curve, as in the CIR model, by increasing k and keeping all other factors fixed, we increase the degree of convexity of the mean solution, so making the "linear" approximation more and more inadequate. Clearly, if the approximation is "bad", then to resort to a very large T does not help at all; we would just repeat the same error over a larger scale.

It should also be noted that the inaccuracy in the Euler and Milstein schemes is relatively small when k is small thereby giving the impression to an unaware reader of not being present.

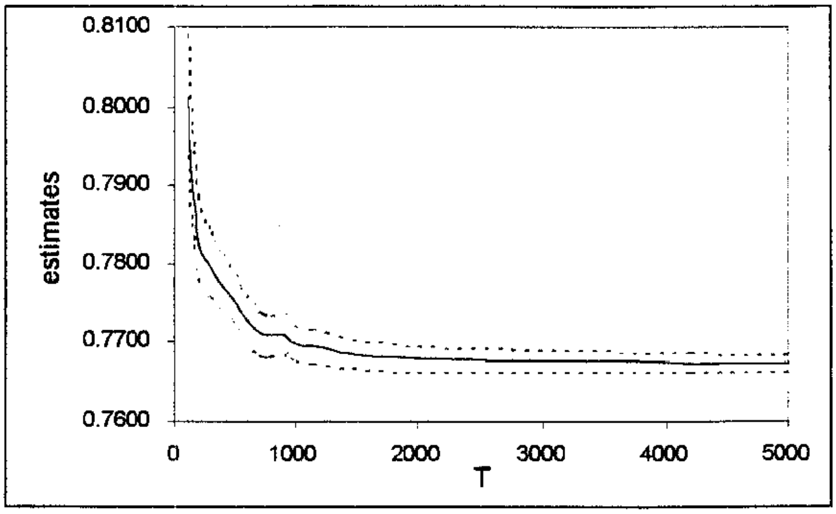
The bias reported above, which is actually a "discretization bias" due to bad data generation, should be distinguished from the notion of "discretization bias" employed for instance in GMR (1993), to define the bias which appears when we employ a "wrong" discrete auxiliary model to naively estimate the underlying continuous stochastic process. By progressively reducing Δ this bias can be reduced, as is to be expected, since a smaller discretisation step obviously leads to a better approximation of the continuous process, thereby improving the quality of the data employed in the estimation.

In recent published literature on simulation based estimation there has been a common practice which consists in the generation of the CIR model, or other models of the family (3), by means of a linear scheme such as the Euler and Milstein schemes, with, at times, a too large discretisation step (such as Δ=0.1) combined with a small T and a small value of the tuning parameter k (see Broze et al. 1995a and 1995b, Bianchi et al. 1994 and 1996). Our results show that such a practice can very easily be misleading if it is used to evaluate numerically the performances of an estimator such as a DMLE. We should point out that the paper by Broze et al. (1995b) and ours could be considered as being complementary to each other. Whereas we have investigated the very concrete problem of a correct generation of the underlying process and to what extent convexity effects may lead to biased estimates, they investigate, in a very formal way, the effects on the estimation of an Ornstein-Uhlenbeck process and a geometric brownian motion due to badly generated data using a linear scheme.

Finally, we note that (see Tab. 4) for a small T (T=100), even if discretisation step is small (such as Δ=0.01), all the generators provide data leading to strongly biased estimates. This result was expected (see for instance Kloeden et al. 1995).

Figure 1 shows a graph of the DMLE of k, corresponding to the true value k=0.8, against T (100<T<5000) obtained from the data generated by means on the Euler scheme (results from Milstein are very similar) using an inadequate discretisation step Δ. The curve practically always underestimates the true value of k. In particular the bias monotonically increases with T.

Figure 1: Cir Model: estimates of k (true value=0.8) from data generated by the Euler scheme (Δ =0.1, 1000 replications) for 100<T<5000. Estimates with continuous line, 2σ confidence band with dashed line.



it is not possible to establish, a priori, rules which will always guarantee that our experiment is well posed. Even in very simple problems, such as those considered here, in order to be sure of generating good data by resorting to linear methods (i.e. Euler and Milstein), we should continuously "tune" the value of the discretization step Δ as k increases. This is clearly very difficult in practice and obviously silly!

On the other hand, our experiments show that, even for a relatively large discretization step, the Taylor 1.5 scheme continues to provide sufficiently good approximations to the trajectories of the underlying SDEs, i.e. close to the values provided by the exact criterion CCSD, even for values of k which are well beyond the restricted range (0, 1.5) considered here. As a consequence, the DMLE estimates calculated from data generated by these two methods continue to be very close to each other and close to the true value of the parameter. For instance, up to a value of k=5.0 (detailed results for this particular case are not presented in this paper) the relative difference in the estimates from the two data was less that 0.003. This is certainly strong evidence in favour of a systematic use of a higher order scheme in simulation based estimation of SDE parameters.

The results obtained for the CIR model carry over, mutatis mutandis, to the remaining models. We report some selected results for the Vasicek model.

4.0.2 The Vasicek model

The Vasicek model, obtained from equation (3) by setting α=β=0, possesses the formal solution:

$$X_t = \vartheta + (X_0 - \vartheta)e^{-kt} + \sigma \int_0^t e^{ks} dW_s (4.1) \tag{5}$$

which may be considered an "exact" generator of the process. As alternative generators of the model we use only the Euler and Taylor 1.5 schemes, since the Euler and Milstein schemes are equivalent for all models with constant diffusion coefficients. The formal solution (5), may be written in the following recurrent form particularly convenient from the simulation point of view:

$$X_{n+1} = e^{-k\Delta} X_n + \vartheta(1 - e^{-k\Delta}) + U_{n+1} \tag{6}$$

where the U_n are iid normally distributed random variables, with zero mean and variance given by:

$$Var(U_n) = \frac{\sigma^2}{2k}\left(1 - e^{-2k\Delta}\right) \tag{7}$$

Tables 7-10 report results for a series of simulations on the Vasicek model which, as can be observed, are very similar to those reported above for the CIR model. In particular, we may note that, provided the data are generated correctly (see Tables 7 and 8), the DMLE preserves nice asymptotic properties, but suffers from a significant bias for short series.

Table 7. Vasicek model: DMLE of k for three generators; T=2000, Δ=0.01 (standard deviations in brackets)

True values of k	Exact solution	Euler /Milstein	Taylor 1.5
0.3	0.3020 (0.0171)	0.3016 (0.0176)	0.3020 (0.0171)
0.8	0.8021 (0.0278)	0.7989 (0.0276)	0.8021 (0.0278)
1.5	1.5040 (0.0363)	1.4926 (0.0358)	1.5031 (0.0363)

Results obtained from n=1000 replications

Table 8. Vasicek model: DMLE of k for three generators; T=100, Δ=0.01 (standard deviations in brackets)

True values of k	Exact solution	Euler /Milstein	Taylor 1.5
0.3	0.3424 (0.0821)	0.3418 (0.0860)	0.3424 (0.0862)
0.8	0.8405 (0.1279)	0.8369 (0.1271)	0.8405 (0.1279)
1.5	1.5388 (0.1716)	1.5269 (0.1696)	1.5388 (0.1716)

Results obtained from n=1000 replications

Table 9. Vasicek model: DMLE of k for three generators; T=20000, Δ=0.1 (standard deviations in brackets)

True values of k	Exact solution	Euler /Milstein	Taylor 1.5
0.3	0.3001 (0.0053)	0.2956 (0.0052)	0.3001 (0.0053)
0.8	0.8002 (0.0089)	0.7682 (0.0084)	0.8010 (0.0090)
1.5	1.5012 (0.0129)	1.3885 (0.0115)	1.5070 (0.0130)

Results obtained from n=1000 replications

Table 10. Vasicek model: DMLE of k for three generators; T=100, Δ=0.1 (standard deviations in brackets)

True values of k	Exact solution	Euler /Milstein	Taylor 1.5
0.3	0.3401 (0.0861)	0.3342 (0.0840)	0.3402 (0.0815)
0.8	0.8389 (0.1301)	0.8034 (0.1221)	0.8389 (0.1305)
1.5	1.5386 (0.1740)	1.4196 (0.1544)	1.5441 (0.1754)

Results obtained from n=1000 replications

5 Computational experiments: Naive Estimates

The so-called naive estimator used in GMR (1993) and examined in some detail in Bianchi et al. (1994 and 1996) for the CIR model, has been very common in statistical practice (see Seber (1989)). This estimator is defined as an OLS estimator of the discrete "auxiliary model" (see GMR 1993) which is nothing other than a reparameterisation of the equation obtained from the Euler scheme approximation of a given SDE. For the CIR model, for example, the "auxiliary model" is given by

$$\frac{Y_{n+1} - Y_n}{\sqrt{Y_n}} = \frac{k\vartheta\Delta}{\sqrt{Y_n}} - k\sqrt{Y_n} + \sigma\epsilon_{n+1} \tag{8}$$

where ϵ_n is a sequence of iid normal random variables with zero mean and variance equal to the amplitude of the discretization step. Model (8) can of course be estimated using OLS.

Clearly, the naive estimator will exhibit good statistical properties only if the data are generated by the Euler scheme, which would be the underlying "true model", and with the same time step. On the other hand, if the process is generated by the corresponding "exact" solution or by a sufficiently accurate higher order scheme (in our experiments a Taylor scheme of order 1.5 satisfies this request), it becomes quite obvious that an "auxiliary model" defined by the Euler scheme cannot be expected to give good estimates unless, as was evidenced above, T is sufficiently large and the discretisation time step is sufficiently small. In studies similar to the one carried out in this paper (see Broze et al (1995b) and Bianchi et al. (1996)) the data were generated assuming that the true underlying process to be generated was not the underlying SDE, but its Euler discretization. Consequently, it should come as no surprise that the naive estimator did not exhibit significant bias even for a relatively large discretisation step.

The Monte Carlo experiments carried out with the DMLE were repeated for the naive estimate of the four models considered here. For brevity, we report only a few of these, for the CIR model, in Tables 11 and 12. In particular, we report the results relative to data generated "exactly" using the non-central chi-square distribution and to data generated from the Euler scheme and suggest that their interpretation be made keeping in mind the results in Tables 3-6..

Table 11. Naïve estimates of the CIR model.
Data are generated from the exact chisquare conditional distribution and from the approximate Euler scheme (standard deviations in brackets)

	T = 100	Δ= 0.01	T = 1000	Δ= 0.01
True values of k	Chisquare	Euler	Chisquare	Euler
0.3	0.3420 (0.0863)	0.3424 (0.0864)	0.3042 (0.0249)	0.3046 (0.0250)
0.8	0.8368 (0.1275)	0.8400 (0.1277)	0.8008 (0.0392)	0.8040 (0.0393)
1.5	1.5269 (0.1702)	1.5380 (0.1709)	1.4918 (0.0532)	1.5030 (0.0534)

Results obtained from 1000 replications

Table 12. Naïve estimates of the CIR model.
Data are generated from the exact chisquare conditional distribution and from the approximate Euler scheme (standard deviations in brackets)

	T=100	Δ=0.1	T=1000	Δ=0.1
True values of k	Chisquare	Euler	Chisquare	Euler
0.3	0.3340 (0.0847)	0.3383 (0.08508)	0.2990 (0.02379)	0.3035 (0.02392)
0.8	0.8046 (0.1253)	0.8355 (0.1272)	0.7724 (0.03824)	0.8036 (0.03890)
1.5	1.4266 (0.1621)	1.5326 (0.1672)	1.3968 (0.05084)	1.5039 (0.05246)

Results obtained from 1000 replications

The naive estimates in Table 11 are very close to their DMLE counterparts in Tables 3 and 4. This is evidently due to the fact that for a small discretisation step, Δ=0.01, the Euler scheme, which is also the model used for obtaining the naive estimate, gives a reasonably good approximation to the underlying process. Hence we would expect the estimates obtained using the DMLE and the naive estimator to be very similar; this is exactly what we find in comparing Tables 3, 4 with Table 11.

We note from Tables 11 and 12 that, as expected, for large T the naive estimator does not suffer from a significant bias only when the data are generated by the Euler scheme, i.e. when there is perfect agreement between the data generator and the estimated model. On the other hand, for small T or when the data are generated by the "exact" solution of the underlying process, the naive estimator exhibits a large bias. The value of 0.8046 in Table 12 obtained from data generated from the "exact" solution could be very misleading and underlies the need for a consideration of a variety of parameter values and data lengths in establishing, via simulation, the properties of statistical estimators. Under such circumstances, it is our belief that the naive estimator should never be used for estimating SDE.

6 Calibration as a general strategy?

As seen before (see Table 4), even if the data are well generated, but T is small, the DMLE is seriously biased. This bias is totally unavoidable but can, in many situations, be corrected by resorting to indirect estimation methods (see GMR (1993)). In Table 13 we present the results relating to the DMLE and Indirect estimate of the CIR model when T and the discretisation step are small (T=100, Δ=0.01). H defines the number of simulated series used in the calibration (for details see GMR (93)). The series in this experiment were generated by using the Taylor 1.5 scheme[7] (results from the CCSD are almost identical and hence not reported here for brevity).

[7] We have done so deliberately in order to further evaluate the performance of the Taylor 1.5 scheme which so far appears to have provided very good results.

Table 13. CIR model: Indirect estimation of k:
a)by calibrating the DMLE (second and third columns);
b) by calibrating the naïve OLS estimator (fourth and fifth columns).
T=100, Δ=0.01, H=4 (standard deviations in brackets)

True values of k	DMLE	DMLE Indirect est.	Naïve OLS	Naïve OLS Indirect
0.3	0.3468 (0.0894)	0.3013 (0.0990)	0.3414 (0.0904)	0.3011 (0.1003)
0.8	0.8488 (0.1546)	0.8049 (0.1670)	0.8373 (0.1383)	0.8021 (0.1514)
1.5	1.5453 (0.1865)	1.5012 (0.1996)	1.5301 (0.1851)	1.5040 (0.2021)

Results obtained from n=500 replications

As we may see, both the DMLE and the naive estimator give very similar results before calibration and both successfully correct for the bias after calibration. This suggests that calibration (i.e. indirect estimation) can be used to correct for bias and hence is a good strategy which may be applied in practice whenever the experimental conditions are "stressed". In addition, Table 13 suggests that for short, but highly frequent data, a calibration on the simplest available estimate could be sufficient to correct for the bias. This conclusion is supported by our comments above on the results in Table 3, 4, and 11.

Completely analogous results for the indirect estimate were obtained for all the four models considered. We report, in Table 14, the results obtained for the Vasicek model in the special case where T=100 and Δ=0.01, i.e. when the DMLE suffers from a serious bias and hence is comparable with Table 13.

Table 14 Vasicek model: indirect estimation of k:
a)by calibrating the DMLE (second and third columns);
b)by calibrating the naïve OLS estimator (fourth and fifth columns).
T=100, Δ=0.01, H=4 (standard deviations in brackets)

True values of k	DMLE	DMLE Indirect est.	Naïve OLS	Naïve OLS Indirect
0.3	0.3436 (0.0889)	0.3025 (0.0994)	0.3434 (0.0889)	0.3033 (0.0997)
0.8	0.8401 (0.1382)	0.8004 (0.1501)	0.8374 (0.1378)	0.8019 (0.1515)
1.5	1.5400 (0.1856)	1.5012 (0.1990)	1.5303 (0.1846)	1.5039 (0.2021)

Results obtained from n=500 replications

7 Conclusions

This paper has shown the importance of correctly generating data in a simulation-based estimation procedure. In the absence of "exact" solutions, for the generation of the sample paths of SDEs we should always resort to higher order approximate schemes. The importance of using higher order schemes is fully evidenced in our experiments, based on the comparison between a Taylor scheme of strong order 1.5 , which in many reasonable cases appears to work quite well (its results are comparable to those obtained from data generated "exactly") and the lower order Euler and Milstein schemes which do not, in many cases, provide adequate approximations to the underlying processes.

Our results suggest that, whenever we have a fairly long approximation of the time path of a given SDE observed on a sufficiently fine grid of time points, the Discretised Maximum Likelihood Estimator appears to provide nice estimates of the drift parameters. The results also indicate that the naive estimator, which has been very often used in the literature (see for example Seber (1989, chapter 7) and references therein), should not be used in any circumstances.

The DMLE, which has good large sample properties, is seriously biased for small samples. In this case, i.e. when we have short series of even well generated data, we suggest that an indirect simulation-based estimation procedure should be used.

References

[1] Bianchi C., R. Cesari and L. Panattoni, (1994), Alternative estimators of a diffusion model of the term structure of interest rates. A Monte Carlo comparison, Temi di discussione della Banca d'Italia

[2] Bianchi C. and E. Cleur (1996), Indirect Estimation of Stochastic Differential Equation Models: some Computational Experiments, Computational Economics, 9, 257-274

[3] Bibby B.M.and M. Soerensen (1994), Martingale estimation functions for discretely observed diffusion processes, technical report n. 277, Department of Theoretical Statistics, Univ. of Aarhus

[4] Broze L., O. Scaillet and J.-M. Zakoian (1995a), Testing for continuous-time models of the short term interest rate, Journal of Empirical Finance, 2, 199-223.

[5] Broze L., O. Scaillet and J.-M. Zakoian (1995b), Quasi Indirect Inference for Diffusion Processes,Discussion Paper n. 9095, CORE, Louvain.

[6] Chan K.G., G.A. Karolyi , F. A. Longstaff and A. B. Sanders (1992), An empirical comparison of alternative models of the short term interest rate, Journal of Finance, 47, 1209-1227

[7] Cleur E.M.and P. Manfredi (1996), Non-linear Stochastic Differential Equation Population Models: Numerical Methods and Estimation Procedures, paper presented at the Third International Meeting on Quantitative Methods for Applied Sciences, June 12-14 1996, Siena.

[8] Gard T. (1987), Stochastic differential equations, Marcel Dekker, New York.

[9] Gard T. (1992), Stochastic models for toxicant-stressed populations, Bulletin of Mathematical Biology, Vol. 54, 5, 827-837

[10] Gouriéroux C., A. Monfort and E. Renault (1993), Indirect Estimation, J. App. Econometrics, Vol. 8, S85-S118

[11] Johnson N. and T. Kotz (1992), Continuous Univariate Distributions, Wiley, New York.

[12] Kloeden P.E. and E. Platen (1992), Numerical Solution of SDE, Springer Verlag, New York.

[13] Kloeden P.E., E. Platen and H. Schurz (1994), Numerical Solution of SDE through Computer Experiments, Springer, New York.

[14] Kloeden P.E., E. Platen , H. Schurz and M. Sorensen (1995), On the effects of Discretisation on Estimators of Drift Parameters for Diffusion Processes, to appear in Advances in Applied Probability.

[15] Kuchler U. and M. Sorensen (1989), Exponential Families of Stoc. Processes: An Unifying Semimartingales Approach, Int. Stat. Review, 123–144

[16] Liptser L.S. and A. N. Shiryayev (1981), Statistics of random Processes, Springer Verlag, New York.

[17] Overbeck L. and T. Ryden (1997), Estimation in the Cox-Ingersoll-Ross Mode, Econometric Theory, 13, 430-461.

[18] Seber G.A.F. (1989), Nonlinear regressions, Wiley, New York

8 Appendix A

8.1 MLE of drift parameters in continuous time.

In the main text we considered the Continuous Time Maximum Likelihood Estimation of the one-dimensional SDE $dX_t = a(X_t, \Theta)dt + b(X_t, \sigma)dW_t$ for the restricted but important family

$$dX_t = k(\vartheta - X_t)X_t^{\alpha}dt + \sigma X_t^{\beta}dW_t \tag{9}$$

where the parameters α and β are defined as in Table 1 of the main text. We now briefly sketch the calculations involved in deriving the CTMLE.

The likelihood function for model (9) is given by (for a more detailed discussion of the existence problem see Liptser and Shiryayev (1981) and Kloeden et al. (1995)):

$$L(k,\vartheta) = \exp\left\{\int_0^T \frac{k(\vartheta - X_s)X_s^{\alpha}}{\left(\sigma X_s^{\beta}\right)^2}dX_s - \frac{1}{2}\int_0^T \frac{\left(k(\vartheta - X_s)X_s^{\alpha}\right)^2}{\left(\sigma X_s^{\beta}\right)^2}ds\right\}$$

or:

$$\log L(k,\vartheta) = \sigma^{-2}\left(\int_0^T \frac{k(\vartheta - X_s)X_s^{\alpha}}{X_s^{2\beta-\alpha}}dX_s - \frac{1}{2}\int_0^T \frac{k^2(\vartheta - X_s)^2}{X_s^{2(\beta-\alpha)}}ds\right) \tag{10}$$

The general form (10) may appear somewhat complicated, but it has the advantage of being highly general and encompasses all the cases considered in this paper. The likelihoods corresponding to the four models considered are easily found by just setting the parameters α and β to the values in Table 1. Simple calculations show that the log-likelihood reduces to

$$\sigma^2 \log L(k,\vartheta) = I_1 k\vartheta - I_2 k - \frac{1}{2}\left[I_3\vartheta^2 + I_4 - 2I_5\vartheta\right]k^2 \tag{11}$$

where the I_j are the "integral statistics"[8] resulting from the explicit development of (10) and are given by

$$I_1 = I_1(\alpha,\beta) = \int_0^T X_s^{\alpha-2\beta}dX_s\ ; \qquad I_2 = I_2(\alpha,\beta) = \int_0^T X_s^{1+\alpha-2\beta}dX_s;$$
$$I_3 = I_3(\alpha,\beta) = \int_0^T X_s^{-2(\beta-\alpha)}ds; \qquad I_4 = I_4(\alpha,\beta) = \int_0^T X_s^{2-2(\beta-\alpha)}ds\ ;$$
$$I_5 = I_5(\alpha,\beta) = \int_0^T X_s^{1-2(\beta-\alpha)}ds. \tag{12}$$

[8]This terminology is uncommon but its meaning should be perfectly clear.

We note that the first two integral statistics, $I_1$and I_2 , are stochastic while the remaining are not.

We may obtain the integral statistics for all the models considered in this paper through a straightforward substitution of the values reported in Table1; for instance, by putting $\alpha = \beta = 1$[9], we get the following integral statistics for the logistic equation

$$\begin{array}{lll} I_1(1,1) = \int_0^T X_s^{-1} dX_s \ ; & I_2(1,1) = \int_0^T dX_s; & \\ I_3(1,1) = \int_0^T ds; & I_4(1,1) = \int_0^T X_s^2 ds \ ; & I_5(1,1) = \int_0^T X_s^1 ds. \end{array} \tag{13}$$

Noting that the drift of the family of SDEs considered here is linear in the parameters k and $z = k\vartheta$, we may reparameterize (11) to obtain

$$L^*(k,z) = \sigma^2 \log L(k,z) = I_1 z - I_2 k - \frac{1}{2}\left[I_3 z^2 + I_4 k^2 - 2 I_5 k z\right]. \tag{14}$$

The first order conditions

$$\begin{array}{l} \frac{\partial L*(z,k)}{\partial k} = -I_2 - I_4 k + I_5 z = 0 \\ \frac{\partial L*(z,k)}{\partial z} = I_1 - I_3 z + I_5 k = 0 \end{array} \tag{15}$$

lead to the following unique solution

$$\hat{k} = \frac{I_1 I_5 - I_2 I_3}{I_3 I_4 - I_5^2} \ ; \qquad \hat{z} = \frac{I_1 I_4 - I_2 I_5}{I_3 I_4 - I_5^2} \ ; \qquad \hat{\vartheta} = \frac{I_1 I_4 - I_2 I_5}{I_1 I_5 - I_2 I_3} \tag{16}$$

It is straightforward to prove that the second order conditions for a maximum are satisfied and hence the expressions in (16) define the unique form of the CTMLE for all the models of the family (9).

8.2 MLE of drift parameters in discrete time

The empirical use of the CTMLEs is made possible by a conversion to their discrete counterparts namely the Discrete (time) Maximum Likelihood Estimates. This is obtained via suitable discretisations of the integral statistics, I_j, some of which, as we have already observed, are stochastic and some are not.

[9]Notice that this approach enables us to derive the CTMLE for a much greater family of models than the one defined in Table 2.

The non-stochastic integrals, I_3, I_4 and I_5 may be evaluated by using an appropriate discretisation rule; we have always applied the trapezoidal rule. The expressions thus obtained are summarised in the following Table:

Table A1. Trapezoidal evaluation of the non-stochastic integral statistics.

	I_3	I_4	I_5
$Vasicek$	T	$\Delta\sum_{n=1}^{n_T}\frac{X_{n-1}^2+X_n^2}{2}$	$\Delta\sum_{n=1}^{n_T}\frac{X_{n-1}+X_n}{2}$
CIR	$\Delta\sum_{n=1}^{n_T}\frac{X_{n-1}^{-1}+X_n^{-1}}{2}$	$\Delta\sum_{n=1}^{n_T}\frac{X_{n-1}+X_n}{2}$	T
BS	$\Delta\sum_{n=1}^{n_T}\frac{X_{n-1}^{-2}+X_n^{-2}}{2}$	T	$\Delta\sum_{n=1}^{n_T}\frac{X_{n-1}^{-1}+X_n^{-1}}{2}$
$Logistic$	T	$\Delta\sum_{n=1}^{n_T}\frac{X_{n-1}^2+X_n^2}{2}$	$\Delta\sum_{n=1}^{n_T}\frac{X_{n-1}+X_n}{2}$

The stochastic integrals, on the other hand, can be reduced to a non-stochastic form by systematically applying Ito's formula (see Kloeden et al. 1995). Remember that the stochastic integrals appearing in (12) are

$$I_1 = I_1(\alpha,\beta) = \int_0^T X_s^{\alpha-2\beta} dX_s \quad \text{and} \quad I_2 = I_2(\alpha,\beta) = \int_0^T X_s^{1+\alpha-2\beta} dX_s. \quad (17)$$

Straightforward calculations lead to the following Table:

Table A2. Ito's formula evaluation of the stochastic integral statistics

	I_1	I_2
$Vasicek$	$X_T - X_0$	$\frac{X_T^2-X_0^2-\sigma^2}{2}$
CIR	$\log\frac{X_T}{X_0}+\frac{\sigma^2}{2}I_3$	$X_T - X_0$
BS	$\sigma^2 I_5 - (\frac{1}{X_T}-\frac{1}{X_0})$	$\log\frac{X_T}{X_0}+\frac{\sigma^2}{2}T$
$Logistic$	$\log\frac{X_T}{X_0}+\frac{\sigma^2}{2}T$	$X_T - X_0$

9 Appendix B: Numerical schemes

We briefly report the time-discrete schemes used in this paper for the numerical solution of SDEs. We employed the approximate Euler, Milstein and Taylor 1.5 schemes (see Kloeden and Platen (1992), chapter 10, and Kloeden et al. (1994), chapter 4). In addition, in order to check on the accuracy of the methods considered, we employed a number of alternative generation techniques; these ranged from the exact formal solution, when available (as in the case of the Ornstein - Uhlenbeck, Brennan - Schwartz and logistic models), to the exact conditional probability density function of the process (as in the case of the CIR model).

Let us consider the SDEs

$$dX_t = a(X_t)dt + b(X_t)dW_t \tag{18}$$

where the functions $a(.)$ and $b(.)$ are assumed to be differentiable as many times as we need. Of course (18) must define a "Stochastic Cauchy's problem", i.e.it has to be completed by prescribed initial conditions. The Euler, Milstein and Taylor 1.5 numerical approximation schemes all belong to the class of "Taylor type" numerical schemes, or "strong Taylor approximations". They provide "strong", ie pathwise, approximations (let us denote them by Y_n) to the trajectories of the underlying SDE (Kloeden and Platen 1992) over a prescribed discretization grid with time step Δ_n.

9.0.1 The Euler scheme

This is the simplest strong Taylor approximation and it usually has an order of strong convergence of 0.5. For equations such as (18) above, it is given by

$$Y_{n+1} = Y_n + a(Y_n)\Delta_n + b(Y_n)\Delta W_n \tag{19}$$

In (19) Y_{n+1} is the approximated value of the underlying true trajectory at the $(n+1)$-th point of the discretization grid (ie: Y_{n+1} is the estimate provided by the method for X_{n+1}), Δ_n is the discretisation time step, ΔW_n is a sequence of I.I.D normal random variables with zero mean and variance given by the amplitude of the corresponding time step. For instance, in the case of the CIR model, where: $a(X) = k(\vartheta - X)$, $b(X) = \sigma\sqrt{X}$ we have

$$Y_{n+1} = Y_n + k(\vartheta - X_n)\Delta_n + \sigma\sqrt{X_n}\Delta W_n$$

9.0.2 Milstein scheme

The Milstein scheme is defined by

$$Y_{n+1} = Y_n + a(Y_n)\Delta_n + b(Y_n)\Delta W_n + \frac{1}{2}b(Y_n)b'(Y_n)\left[(\Delta W_n)^2 - \Delta_n\right]$$

This scheme has, under suitable regularity conditions, the order 1.0 of strong convergence.

9.0.3 Taylor 1.5 scheme

Following Kloeden et al. (1994, page 146) the Taylor 1.5 scheme, i.e. the approximation which converges strongly with order 1.5, is given by

$$Y_{n+1} = M_n + RT_n \qquad (20)$$

where M_n is the corresponding Milstein scheme and RT_n is given by the following expression:

$$RT_n = a'b\Delta Z_n + \frac{1}{2}\left[aa' + \frac{1}{2}b^2a''\right]\Delta_n^2 + \left[ab' + \frac{1}{2}b^2b''\right][\Delta W_n \Delta_n - \Delta Z_n] + \frac{1}{2}b\left[bb'' + (b')^2\right]\cdot\left[\frac{1}{3}(\Delta W_n)^2 - \Delta_n\right]\cdot \Delta W_n$$

where the various functions are evaluated at the point Y_n. The Taylor 1.5 scheme needs the generation of a pair of dependent random variables ΔW_n and ΔZ_n (for details see Kloeden et al. 1994, 146)

9.0.4 CIR model: exact conditional distribution

The CIR model does not have a closed form formal solution in the trajectory domain but has a closed form solution for its conditional distribution. This conditional distribution $p(X(s)/X(t))$ is a stationary non-central chi-square distribution (for its generation on the computer see, for ex., Johnson and Kotz (1992)) with $2q$ degrees of freedom and noncentrality parameter u, where

$$u(t) = 2c(t)X(t)e^{-k(s-t)}; \quad c(t) = \frac{2k}{\sigma^2(1-e^{-k(s-t)})}; \quad q = \frac{2k\vartheta}{\sigma^2}$$