Supporting information to: Estimation in discretely observed diffusions killed at a threshold

Enrico Bibbona
University of Torino, Torino, Italy
enrico.bibbona@unito.it

Susanne Ditlevsen
University of Copenhagen, Copenhagen, Denmark.
susanne@math.ku.dk

June 14, 2012

A Numerical details

In this Section we provide some details about the numerical procedure used in the simulation studies of the main paper, both to simulate the sample and to compute the estimated values. We warn the reader that many references are made to formulae and Sections of the main document. We worked in the R environment (R Development Core Team, 2011). Most of our routines have been designed modifying and adapting functions that were implemented in the `sde` package which is thoroughly documented in Iacus (2008).

A.1 Simulations

The simulation of diffusion processes up to their FPT through a barrier b requires some care. If at a given time the process was at level $x_n < b$, and we generate the next point and find $x_{n+1} < b$, we cannot assure that the underlying continuous process did not cross the barrier between the two points. If we stop the simulation only when $x_{n+1} \geq b$, we significantly overestimate the FPT. To solve this problem two competitive methods were proposed (Giraudo & Sacerdote, 1999; Baldi and Caramellino, 2002). For each couple of simulated points x_n and x_{n+1} (if both are below b), the probability p of the process crossing the threshold between the two points is evaluated and a corresponding Bernoulli random variable is generated: if you get 1 a crossing occurred and x_n is the last point of the path, while 0 means no crossing and the simulation continues. The first method is slightly more accurate when the discretization step gets larger, the second much faster to compute. We choose the second. To avoid the influence of using the same approximation both in the simulation scheme and in the estimation method, we simulated with a smaller discretization step w.r.t. the one considered for estimation. To assess the accuracy of the simulation we compare the mean FPT estimated from the simulations with the one prescribed by the theory. In particular, for the WD we calculate analytically the probability that the passage occurs between two steps of the discretization and the quantity

$$\mathbb{E}(N) = \sum_{n=1}^{\infty} n\Delta \cdot \mathbb{P}((n-1)\Delta < T \leq n\Delta),$$

which is a discretized version of the mean FPT. A comparison between $\mathbb{E}(N)$ and its sample values derived from the simulations shows good agreement as reported in Table 6, whose second row is already reported in Table 1.

A.2 Different implementation of formula (10)

Another approximation of the distribution of the FPT is the following,

$$G^b_\theta(\Delta|x) = 1 - \int_{-\infty}^{S} f^b_\theta(X_\Delta = y \,|\, X_0 = x)\, dy. \tag{21}$$

In most cases the numerical evaluation of this integral is much slower than using (10) without providing better performance. Nevertheless, there might be occasions where this alternative turns out to be useful. In particular, the possibility of calculating one of the integrals in (10) or (21) analytically would drastically speed up the algorithm.

In particular, for the OU process we can approximate $f^b_\theta(y, \Delta|x)$ in (21) by expression (12) for the Wiener process, which is the first order approximation in Δ of (15), and we can calculate the integral analytically. If this expression replaces (10), the algorithm becomes significantly faster but less precise, especially if the discretization step is not extremely small. Numerical evaluation of (10) in the parameter settings used here is in any case reasonably fast so we suggest its adoption.

A.3 Minimization algorithm

To minimize numerically the negative log-likelihood function we used the standard Nelder-Mead algorithm provided by the R function `optim`, cf. the R manual (R Development Core Team, 2011) and references quoted therein. There are restrictions on the admissible values for some parameters: In the SR model

Table 6: Comparison between theoretical mean first passage step and sample averages.

	CASE 1	CASE 2	CASE 3	CASE 4
$\mathbb{E}(N)$	33.83	33.83	100.50	98.99
avg(N)	33.72	34.28	100.73	101.61

$\mu \leq \sigma^2/2$, and σ, and β have to be positive in the SR and in the OU model. The likelihood function is evaluated as `NA` (missing value) if the minimizer tries to calculate it for parameters out of this range, and the minimum is calculated just among the admissible values. The effect of this constraint can be seen in Figures 1 and 4 both in the densities and in the Q-Q plots referring to the estimation from a single trajectory. As soon as the estimates get better the effect is lost. Further care is required due to the fact that, when the minimizer tries to compute the likelihood function at some region of the parameter space where it returns missing values (`NA`) or infinite values (for example when the constraints just mentioned are not satisfied, but also when numerical approximation is not good enough), the algorithm should not halt, but go on calculating the likelihood at another value. The box-constrained algorithm, which is denoted "L-BFGS-B" in R, turned out not to be feasible as it halts if the likelihood function returns `NA`s of infinite values. Especially for the SR model the transition density (even in absence of a barrier) might be problematic to evaluate when the parameters provided by the minimizer are not close to the true ones. In this case we need to be sure that the function returns `NA` when it is not evaluated with satisfactory precision. For the transition density of the SR model the R functions `dchisq` and `pchisq` turned out to be the best choice among the different possibilities discussed in (Iacus, 2008, Section 3.1.3). Nealder-Mead minimizers require reasonable initial values. These values are provided by estimators that can be calculated explicitly. We used the standard choices suggested in the literature in the absence of a barrier. The initial estimators for the WD model are

$$\hat{\mu} = \frac{X_N}{(N-1)\,\Delta}; \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^N (X_i - X_{i-1} - \hat{\mu}\,\Delta)^2}{(N-1)\,\Delta}.$$

The initial estimators for the OU model are

$$\hat{\beta} = -\frac{1}{\Delta}\log\left(\frac{\sum_{i=1}^N (X_i - \bar{X})(X_{i-1} - \bar{X})}{\sum_{i=0}^N (X_i - \bar{X})^2}\right); \quad \hat{\mu} = \hat{\beta}\bar{X}; \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^N (X_i - X_{i-1})^2}{(N-1)\,\Delta}.$$

The initial estimators for the SR model are

$$\hat{\beta} = -\frac{1}{\Delta}\log\left(\frac{N\sum_{i=1}^N \frac{X_i}{X_{i-1}} - \sum_{i=1}^N X_i \sum_{i=1}^N \frac{1}{X_{i-1}}}{N^2 - \sum_{i=1}^N X_{i-1}\sum_{i=1}^N \frac{1}{X_{i-1}}}\right);$$

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^N X_i + \frac{\mathrm{e}^{-\hat{\beta}\Delta}(X_N - x_0)}{N\hat{\beta}(1 - \mathrm{e}^{-\hat{\beta}\Delta})};$$

$$\hat{\sigma}^2 = \frac{2\hat{\beta}\sum_{i=1}^N \frac{1}{X_{i-1}}\left(X_i - \mathrm{e}^{-\hat{\beta}\Delta}X_{i-1} - \frac{\hat{\mu}}{\hat{\beta}}(1 - \mathrm{e}^{-\hat{\beta}\Delta})\right)^2}{(1 - \mathrm{e}^{-\hat{\beta}\Delta})\sum_{i=1}^N \frac{1}{X_{i-1}}\left[\frac{\hat{\mu}}{\hat{\beta}}(1 - \mathrm{e}^{-\hat{\beta}\Delta}) + 2\mathrm{e}^{-\hat{\beta}\Delta}X_{i-1}\right]}.$$

References

Baldi P. & Caramellino L. (2002). Asymptotics of hitting probabilities for general one-dimensional pinned diffusions. *Ann. Appl. Probab. 12*(3), 1071–1095.

Giraudo M. T. & Sacerdote L. (1999). An improved technique for the simulation of first passage times for diffusion processes. *Comm. Statist. Simulation Comput. 28*(4), 1135–1163.

Iacus, S. M. (2008). *Simulation and inference for stochastic differential equations*. Springer Series in Statistics. New York: Springer. With R examples.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.