Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Econometrics 134 (2006) 1–68
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Asymptotic properties of Monte Carlo estimators of diffusion processes

Jérôme Detemple[a,b], René Garcia[c,b,d,*], Marcel Rindisbacher[b,e]

[a]*Boston University School of Management, USA*
[b]*CIRANO, Canada*
[c]*Economics Department, Université de Montréal, Canada*
[d]*CIREQ, Canada*
[e]*Rotman School of Management, University of Toronto, Canada*

Available online 31 August 2005

Abstract

This paper studies the limit distributions of Monte Carlo estimators of diffusion processes. We examine two types of estimators based on the Euler scheme, one applied to the original processes, the other to a Doss transformation of the processes. We show that the transformation increases the speed of convergence of the Euler scheme. We also study estimators of conditional expectations of diffusions. After characterizing expected approximation errors, we construct second-order bias-corrected estimators. We also derive new convergence results for the Mihlstein scheme. Illustrations of the results are provided in the context of simulation-based estimation of diffusion processes.

JEL classification: C13; C15

Keywords: Monte Carlo estimators; Diffusion processes; Simulation-based estimation; Doss transformation

*Corresponding author. Tel.: +1 514 343 5960; fax: +1 514 343 5831.
E-mail address: rene.garcia@umontreal.ca (R. Garcia).

0304-4076/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2005.06.028

1. Introduction

Standard applications in finance, such as portfolio allocation, asset pricing and risk management, rely on models of prices and state variables described by diffusion processes. For practical implementations of these financial models estimates of the drift and volatility coefficients of the underlying processes are needed. Precise statistical methods are essential for that purpose. The importance of accurate methods is further enhanced by the potentially large impact of parameter uncertainty and estimation errors on economic decisions.[1]

While maximum likelihood is the estimation method of choice, it is, in general, infeasible as explicit formulas for transition densities are often unknown. Inference can, therefore, only proceed with the use of an auxiliary numerical procedure designed to approximate the transition density. An example is the simulated maximum likelihood estimation (SMLE) method, suggested by Pedersen (1995a,b).[2] This method approximates the transition density by applying a Euler discretization to the diffusion process. It splits the time between observations into subintervals and evaluates the transition of the process between any two observations by integrating out convolutions of subdensities (generally Gaussian) using a Monte Carlo procedure. Consistency of the estimator is established when the numbers of subintervals and Monte Carlo replications go to infinity, while a particular ratio of the two converges to zero.[3] Any feasible estimator, however, involves finite numbers of subintervals and replications and will therefore be biased and inefficient. Moreover, there are trade-offs between the number of subintervals and that of replications. Increasing the former will decrease the bias but will boost the variance of the estimator. Increasing the latter will reduce the variance, but will also, if the estimator is biased, diminish the probability that a confidence interval of a given size covers the true value of the parameter. The main contribution of our paper is to characterize the asymptotic properties of the errors associated with efficient approximation procedures for the estimation of diffusion processes. Our results enable us to construct (second-order) bias-corrected estimators that are asymptotically equivalent to estimators obtained by sampling from the true distribution of the processes.

In a recent paper, Durham and Gallant (2002) stress that SMLE can become computationally expensive, if the goal is to achieve a reasonable degree of accuracy, and propose various ways to reduce this cost. Their approach uses bias- and variance-reduction schemes to limit both the number of subintervals and that of replications. They investigate, by experimentation, the trade-offs between these two control parameters for various discretization and simulation schemes. Our goal is to supplement experimentation with an explicit characterization of the asymptotic error properties of Monte Carlo estimators for general diffusion processes.

[1] For optimal portfolio choice, see, for example, Bawa et al. (1979) and Barberis (2000).

[2] See also Santa-Clara (1995) and Brandt and Santa-Clara (2002).

[3] In Brandt and Santa-Clara (2002), the ratio $S^{1/2}/M \to 0$, with S the number of replications and M the number of subintervals. A discussion of this ratio is provided in Section 5.1.

As an illustration, we apply our convergence results to simulated methods of moments.[4] These include the simulated method of moments (Duffie and Singleton, 1993), indirect inference procedures (Smith, 1990; Gouriéroux et al., 1993) and efficient method of moment procedures (EMM) (Gallant and Tauchen, 1996).[5] Recently, Broze et al. (1998) studied the asymptotic bias associated with indirect inference. This procedure relies on the simulation of diffusion processes at a step that is finer than the observation step to correct the discretization bias. They emphasize that the use of any fixed time interval to perform the simulation implies an asymptotic bias.[6] Although their theoretical asymptotic results support the choice of an arbitrarily small discretization step, their numerical experiments show a trade-off between the length of the simulated data MT (where T is the length of the observed sample and M the number of trajectories) and the size $h = T/N$ of the discretization step. Our results provide an analytical characterization of this trade-off. We show, in particular, that the second-order bias is a function of both M and N and provide a procedure to correct simulation-based method-of-moment estimators for it.

To state the problem in general terms, note that auxiliary numerical procedures for SMM estimators involve the calculation of conditional expectations of functions of the solutions of stochastic differential equations. The use of a Monte Carlo procedure for that purpose induces two types of approximation errors. The first error is due to the numerical solution of a stochastic differential equation based on a time discretization scheme. The second error is the Monte Carlo error due to the approximation of the expectation in the moment condition, by an average over independent replications. Suppose that one seeks to compute the conditional expectation $f(t,x) = \mathbf{E}_t[g(X_T)]$, where X_T is the terminal value of the solution of the stochastic differential equation (SDE)

$$\mathrm{d}X_v = A(X_v)\,\mathrm{d}v + B(X_v)\,\mathrm{d}W_v, \quad X_t = x. \tag{1}$$

To approximate the terminal value X_T, of the solution of (1), several discretization schemes can be used.[7] The most popular, perhaps because of its ease of implementation, is the Euler scheme. This iterative procedure evaluates the drift and volatility functions at the value $X(t_n)$ at time t_n in order to infer the value $X(t_{n+1})$ at t_{n+1}, and proceeds in this manner until $t_N = T$. The second approximation is in the computation of the conditional expectation that is performed by averaging over

[4]The methodology developed here could be applied to other simulation-based estimation methods such as SMLE. In those applications the steps needed to characterize the bias will typically be more involved than for the simulated method of moments.

[5]Other numerical approaches have been proposed to estimate a diffusion process when the transition density is not known in closed form. Let us mention the numerical resolution of the Kolmogorov forward equation (Lo, 1988), the estimation of the infinitesimal generator based on a truncated set of eigenfunctions (Kessler and Sørensen, 1999; Hansen and Scheinkman, 1995), orthogonal series expansions (Ait-Sahalia, 2002), and Markov chain Monte Carlo (MCMC) Bayesian techniques (Eraker, 2001; Chib et al., 2001). Except for Ait-Sahalia (2002), all other methods require that the sampling interval go to zero to obtain convergence and involve the two types of approximation errors mentioned above.

[6]Because the simulation is based on a fixed time step, and therefore not performed using the true probability density function, they rename the method *quasi-indirect inference*.

[7]A detailed analysis of discretization schemes available can be found in Kloeden and Platen (1997).

a finite sample of approximated terminal values $X(T)$. Justification for this averaging rests on the law of large numbers. The combination of these two operations, labelled MCE (Monte Carlo with Euler discretization), produces an estimate of $f(t, x)$ that involves the two types of errors mentioned above. Understanding the trade-off between these errors requires the asymptotic error distribution.

In an insightful paper, Duffie and Glynn (1995) highlighted the trade-off between the discretization error and the Monte Carlo averaging error, and showed the existence of an efficient choice of discretization steps and Monte Carlo replications. For this efficient MCE scheme, they also characterized the asymptotic distribution of the approximation error and found it to be non-centered. A consequence is that the efficient procedure has a second-order bias. In this paper we characterize the second-order bias as the expected value of a known random variable. This random variable can be simulated along with the diffusion and used to design a new approximation that corrects for second-order bias. The bias corrected estimate is asymptotically equivalent to a Monte Carlo procedure that samples directly from the true distribution of the terminal point of the diffusion.[8]

In the first part of this paper (Sections 2–4), we study the asymptotic distributions of errors associated with discretization schemes for general diffusion processes and of Monte Carlo estimators of conditional expectations of diffusions. Error properties for approximations of solutions of SDEs (the first type of error) have been studied before. For the Euler discretization scheme the asymptotic error distribution was found by Kurtz and Protter (1991a) and Jacod and Protter (1998). We extend their results by proposing a change of variables, commonly referred to as a Doss transformation (see Doss, 1977; Detemple et al., 2005) that reduces the diffusion coefficient of the SDE to unity. This transformation has enjoyed recent popularity in financial econometrics (see, for instance, Ait-Sahalia, 2002; Durham and Gallant, 2002) and has been used for the computation of optimal portfolios in dynamic asset allocation models (Detemple et al., 2003). We show that a Doss transformation of the SDE can improve the speed of convergence of the discretization scheme as the martingale part of the transformed SDE can be approximated without error. The asymptotic law of the estimate of the Doss transformed state variable is derived and found to be non-centered. This can be contrasted with the simple Euler scheme applied to the original (non-transformed) SDE that produces an error whose asymptotic law is centered.

One of the numerical schemes used to reduce bias is the Milshtein second-order scheme (Milshtein, 1995). We characterize its asymptotic error distribution and show that it does not dominate the Euler scheme with transformation, in terms of convergence behavior. This highlights the benefits of the transformation. Our results for the weak limit of the solution of SDEs based on the Milshtein scheme and the second-order bias of estimators of conditional expectations are new and therefore of

[8]Our method is simpler than the one in Talay and Tubaro (1990) as it does not require solving a PDE in addition to calculating an expectation. In addition, it leads to the construction of computationally feasible second-order bias corrected approximation schemes.

independent interest. Moreover, they enlighten some of the experimentation results of Durham and Gallant (2002).

In the second part of the paper (Section 5), we apply these results to the estimation of the parameters of diffusions by simulated methods of moments. We first provide a general setup for parameter estimation and then develop an explicit simulated extended quasi-maximum likelihood estimator based on the estimator presented by Wefelmeyer (1996). We illustrate the method with L-CEV and CIR processes that are widely used in the literature to model the short rate of interest.

As mentioned above, our theoretical convergence results enable us to design bias-corrected estimators that have the same asymptotic distributions as the infeasible estimators based on the unknown true transition densities. We illustrate the improved performance of these estimators in simulation experiments similar to those performed by Durham and Gallant (2002) for the CIR model. Their simulations show that the Doss transformation reduces the bias of the parameter estimates. Our asymptotic convergence results establish that this transformation indeed increases the speed of convergence, hence provides a theoretical explanation for their finding.

The paper is organized as follows. In Section 2 we study the asymptotic error distributions of approximations of solutions of SDEs and provide numerical illustrations for standard processes in finance. Section 3 describes the asymptotic laws of estimators of conditional expectations and characterizes expected approximation errors, second-order discretization biases and bias-corrected estimators. New asymptotic convergence results for the Milshtein scheme are presented in Section 4. Section 5 provides an application to a simulation-based method of moments. Conclusions are formulated in Section 6. All proofs are collected in Appendix A. Appendix B contains expressions needed to characterize the second-order bias-corrected estimators.

2. Asymptotic laws of estimators of solutions of SDEs

Continuous-time financial models are often based on multivariate diffusions with general drift and diffusion functions. The solution of the financial application, be it asset pricing, portfolio allocation or risk management, relies on the simulation of discretized versions of these stochastic differential equations (SDEs). The Euler scheme is most often used for this purpose. This discretization involves an approximation error. In this section we study the asymptotic error distribution of Euler approximations of solutions of SDEs. We also study the error distribution associated with a Doss transformation of the state variables. This change of variables is useful for numerical efficiency, as shown by the dynamic portfolio application in Detemple et al. (2003). It has also been used by Ait-Sahalia (2002) and Durham and Gallant (2002) as a variance reduction device. The importance of obtaining an explicit expression for the asymptotic distribution of the approximation error cannot be underestimated. As we will see, it is not possible to approximate the distribution of this error in a finite simulation experiment using a simulated benchmark for the true value X_T, no matter how finely the process is sampled.

Convergence results for the Euler scheme are reviewed in Section 2.1. Their application to Doss-transformed processes is studied in Section 2.2. Numerical illustrations are provided in Section 2.3.[9]

2.1. Euler approximation without transformation

To set the stage for the convergence results with the transformation and the Milshtein scheme, we recall known results for the standard Euler scheme. These results are also essential for finding an explicit expression for the second-order bias.

Consider the $d \times 1$ random vector X_T given by the terminal value of the solution of the SDE

$$\mathrm{d}X_v = A(X_v)\,\mathrm{d}v + \sum_{j=1}^{d} B_j(X_v)\,\mathrm{d}W_v^j, \tag{2}$$

where A and B_j are $d \times 1$ vectors such that $A \in \mathscr{C}^3(\mathbb{R}^d)$, $B_j \in \mathscr{C}^3(\mathbb{R}^d)$ and A, B_j are at most of linear growth.[10] The Euler approximation of (2) is

$$X_T^N = X_0 + \sum_{n=0}^{N-1} A(X_{nh}^N)h + \sum_{n=0}^{N-1}\sum_{j=1}^{d} B_j(X_{nh}^N)\Delta W_{nh}^j, \tag{3}$$

where $h = T/N$ and $\Delta W_{nh}^j = W_{(n+1)h}^j - W_{nh}^j$.[11]

Kurtz and Protter (1991a) and Jacod and Protter (1998) deduce the asymptotic behavior of the error associated with the Euler approximation of X_T (see Jacod and Protter, 1998, Theorem 3.2, p. 276).

Theorem 1. *The approximation error $X_T^N - X_T$ converges weakly at the rate $1/\sqrt{N}$ (i.e. $\sqrt{N}(X_T^N - X_T) \Rightarrow U_T^X$).[12] The asymptotic error is*

$$U_T^X = -\frac{1}{\sqrt{2}}\,\Omega_T \int_0^T \Omega_v^{-1} \sum_{l,j=1}^{d} [\partial B_j B_l](X_v)\,\mathrm{d}Z_v^{l,j} \tag{4}$$

with $[Z^{l,j}]_{l,j\in\{1,\dots,d\}}$ a $d^2 \times 1$ standard Brownian motion independent of W, ∂B_j a $d \times d$ matrix of derivatives of B_j with respect to X and

$$\Omega_v = \mathscr{E}^R\left(\int_0^{\cdot} \partial A(X_s)\,\mathrm{d}s + \sum_{j=1}^{d}\int_0^{\cdot} \partial B_j(X_s)\,\mathrm{d}W_s^j\right)_v. \tag{5}$$

[9]To simplify the presentation we assume homogeneous dynamics of the process to be simulated.

[10]The space $\mathscr{C}^k(\mathbb{R}^d)$ denotes the space of k times continuously differentiable, $\mathbb{R}^d$-valued functions.

[11]To simplify the notation we restrict the error analysis to equidistant discretization schemes.

[12]Let S be a metric space and $\mathscr{S}$ its Borel sets. A sequence of random variables X^N is said to converge weakly to a random variable X whenever, with $\mathbf{P}_{X^N} \equiv \mathbf{P} \circ (X^N)^{-1}$ and $\mathbf{P}_X \equiv \mathbf{P} \circ X^{-1}$, we have $\int_S f(s)\,\mathrm{d}\mathbf{P}_{X^N}(s) \to \int_S f(s)\,\mathrm{d}\mathbf{P}_X(s)$ for all continuous and bounded functions f on S (see Billingsley, 1968).

In this last expression ∂A is the $d \times d$ matrix of derivatives of the vector A with respect to the elements of X and $\mathscr{E}^R(\cdot)$ is the right stochastic exponential.[13]

Theorem 1 says that the error converges in law at the rate $1/\sqrt{N}$ to the random variable U_T^X, as the number of discretization points N becomes large. The asymptotic error U_T^X depends on the coefficients of the SDE and their derivatives. Surprisingly, it also depends on new Brownian motions ($[Z^{l,j}]_{l,j\in\{1,\dots,d\}}$), which are orthogonal to the original ones (W). These appear because the stochastic integral of a time-dependent function with respect to a Brownian motion is imperfectly correlated with the terminal value of the Brownian motion. A second Brownian motion is then needed to describe the law of the integral. The result in Theorem 1 follows because the limit law of X_T^N depends on stochastic integrals of this sort.

To illustrate the result consider the simple case of a geometric Brownian motion

$$\mathrm{d}X_v = aX_v\,\mathrm{d}v + bX_v\,\mathrm{d}W_v. \tag{6}$$

The asymptotic error is $U_T^X = -(b/\sqrt{2})X_T Z_T = Z_{(b^2/2)X_T^2}$, where X_T is log-normally distributed and Z_T is normal. The error distribution is a mixture of normals where the mixing distribution is the square of the geometric Brownian motion X_T.

The scaling property of the Brownian motions $Z^{l,j}$ shows that the asymptotic distribution, in the univariate case, is always a mixture of normals. But the law of the mixing distribution is not always explicit. For instance, in the case of a CIR process

$$\mathrm{d}X_v = \kappa(\bar{X} - X_v)\,\mathrm{d}v + \sigma\sqrt{X_v}\,\mathrm{d}W_v, \tag{7}$$

one finds that

$$U_T^X = -\frac{\sigma^2}{2\sqrt{2}}\Omega_T \int_0^T \Omega_v^{-1}\,\mathrm{d}Z_v = Z_{\frac{\sigma^4}{8}\Omega_T^2 \int_0^T \Omega_v^{-2}\,\mathrm{d}v}.$$

The law of the mixing random variable $(\sigma^4/8)\Omega_T^2 \int_0^T \Omega_v^{-2}\,\mathrm{d}v$ is unknown as Ω satisfies the equation $\mathrm{d}\Omega_v = (-\kappa\,\mathrm{d}v + ((\sigma/2)/\sqrt{X_v})\,\mathrm{d}W_v)\Omega_v$ where $\Omega_0 = 1$, whose solution depends on the path of X. One must then resort to numerical procedures to examine the asymptotic error distribution. Illustrations of this are provided in Figs.1 and 2 for the CIR process and the constant elasticity of variance process with linear drift (L-CEV).

The dependence of the asymptotic distribution on an independent Brownian motion, that does not exist on the original probability space, shows that it is not possible to approximate the distribution of the approximation error in a finite simulation experiment using a simulated benchmark for the true value X_T. This underscores the importance of an explicit formula for the asymptotic error. In the absence of an exact expression for U_T^X, error analysis using a simulated benchmark $X_T^{N^\star}$ with $N^\star$ large (i.e. analysis of $\sqrt{N}(X_T^N - X_T^{N^\star})$) will always depend just on the original Brownian motions W^j and not on the independent Brownian motions $Z^{l,j}$ that appear in the random limit.

[13]For a $d \times d$ semimartingale M, the right stochastic exponential $Z_v = \mathscr{E}^R(M)_v$ is the unique solution of the $d \times d$ matrix SDE $\mathrm{d}Z_v = \mathrm{d}M_v Z_v$ with $Z_0 = I_d$.

Theorem 1 shows that the standard Euler procedure converges at the rate of $1/\sqrt{N}$. Next, we introduce a change of variables that simplifies the volatility of the underlying SDE, thereby leading to an approximation of the true value X_T with an improved rate of convergence, equal to $1/N$. Durham and Gallant (2002) have already remarked in their experiments that this transformation improves the performance of their simulation and importance sampling schemes. They suggest that the improvement might be related to the fact that the transformed process is "closer" to a Gaussian process. We formalize this intuition and illustrate it with examples in the next sub-sections.

2.2. Euler approximation with Doss transformation

Let us first introduce the Doss transformation.[14] Consider the transformed volatility coefficient, $B^{\hat{B}}(x) \equiv B(x)\hat{B}^{-1}$, where $\hat{B}$ is an arbitrary matrix of constants. Suppose that the rotated volatility coefficient $B^{\hat{B}}$ satisfies the rank condition,

$$rank(B^{\hat{B}}) = d, \quad \text{a.e.}, \tag{8}$$

and the commutativity condition,

$$\partial B_j^{\hat{B}} B_i^{\hat{B}} = \partial B_i^{\hat{B}} B_j^{\hat{B}} \quad \text{for all } i,j = 1,\ldots,d. \tag{9}$$

Then, there exists a function $G^{\hat{B}} : \mathbb{R}^d \to \mathbb{R}^d$ solution of the total ODE

$$\partial_z G^{\hat{B}}(z) = B^{\hat{B}}(G^{\hat{B}}(z)), \quad G^{\hat{B}}(0) = 0, \tag{10}$$

such that $X_t = G^{\hat{B}}(\hat{X}_t)$, where

$$\mathrm{d}\hat{X}_v = \hat{A}(\hat{X}_v)\,\mathrm{d}v + \sum_{j=1}^{d} \hat{B}_j\,\mathrm{d}W_v^j \quad \text{with } G^{\hat{B}}(\hat{X}_0) = X_0, \tag{11}$$

and

$$\hat{A}(x) \equiv B^{\hat{B}}(x)^{-1}\mathscr{A}_t G^{\hat{B}}(x), \tag{12}$$

where the operator $\mathscr{A}$ is defined by

$$\mathscr{A} G^{\hat{B}} \equiv A(G^{\hat{B}}) - \frac{1}{2}\sum_{j=1}^{d} \partial B_j^{\hat{B}}(G^{\hat{B}}). \tag{13}$$

Note that in the case where B is commutative, one can choose $\hat{B} = I_d$. In this instance the transformed state variables $\hat{X}_j$ have unit volatility coefficients.

The Euler approximation of the transformed state variables satisfying (11) is

$$\hat{X}_T^N = \hat{X}_0 + \sum_{n=0}^{N-1} \hat{A}(\hat{X}_{nh}^N)h + \sum_{n=0}^{N-1}\sum_{j=1}^{d} \hat{B}_j \Delta W_{nh}^j.$$

The error distribution of this approximation of the d-vector $\hat{X}_T$ is given next.

[14]See Doss (1977) and Detemple et al. (2005) for details.

Theorem 2. *Suppose that the rank and commutativity conditions* (8) *and* (9) *are satisfied. The approximation error* $\hat{X}_T^N - \hat{X}_T$ *converges weakly at the rate* $1/N$ (*i.e.* $\mathrm{N}(\hat{X}_T^N - \hat{X}_T) \Rightarrow U_T^{\hat{X}}$). *The asymptotic error is*

$$U_T^{\hat{X}} = -\hat{\Omega}_T \int_0^T \hat{\Omega}_v^{-1} \partial \hat{A}(\hat{X}_v) \times \left(\frac{1}{2} \mathrm{d}\hat{X}_v + \frac{1}{\sqrt{12}} \sum_{j=1}^{d} \hat{B}_j \, \mathrm{d}Z_v^j + \frac{1}{2} \sum_{j,k,l=1}^{d} \partial_{l,k} \hat{A}(\hat{X}_v) \hat{B}_{k,j} \hat{B}_{l,j} \, \mathrm{d}v \right) \tag{14}$$

with $[Z^j]_{j \in \{1,\ldots,d\}}$ *a* $d \times 1$ *standard Brownian motion independent of* W *and* $Z^{h,j}$, $\partial \hat{A}(\hat{X}_v) = [\partial_1 \hat{A}(\hat{X}_v), \ldots, \partial_d \hat{A}(\hat{X}_v)]$ *the* $d \times d$ *matrix with columns given by the derivatives of the vector* $\hat{A}(\hat{X}_v)$, *and* $\partial_{l,k} \hat{A}(\hat{X}_s)$ *the* $d \times 1$ *vector of cross derivatives of* $\hat{A}(\hat{X}_v)$ *with respect to arguments* l, k. *The* $d \times d$ *matrix* $\hat{\Omega}_v$ *is*

$$\hat{\Omega}_v = \mathscr{E}^R \left(\int_0^{\cdot} \partial \hat{A}(\hat{X}_s) \, \mathrm{d}s \right)_v . \tag{15}$$

Theorem 2 shows that the speed of convergence increases after application of the transformation. It also shows that the limiting random variable is different and, in particular, involves exponentials of a bounded total variation process instead of a stochastic integral. But, in contrast to the limit without transformation U_T^X, the asymptotic error $U_T^{\hat{X}}$ does not have a zero mean.

The simplest example is the Ornstein–Uhlenbeck process

$$\mathrm{d}X_v = \kappa(\bar{X} - X_v) \, \mathrm{d}v + \sigma \, \mathrm{d}W_v. \tag{16}$$

Its asymptotic error $U_T^{\hat{X}}$ is the sum of two normal random variables

$$\begin{aligned} U_T^{\hat{X}} &= \frac{\kappa}{2} \mathrm{e}^{-\kappa T} \int_0^T \mathrm{e}^{\kappa v} \, \mathrm{d}X_v + \frac{\kappa\sigma}{\sqrt{12}} \mathrm{e}^{-\kappa T} \int_0^T \mathrm{e}^{\kappa v} \, \mathrm{d}Z_v \\ &\equiv \frac{\kappa^2}{2} (\bar{X} - X_0) \mathrm{e}^{-\kappa T} T + W_{\alpha(T)} + Z_{\beta(T)}, \end{aligned} \tag{17}$$

where

$$\alpha(T) \equiv \frac{\sigma^2 \kappa (\mathrm{e}^{2\kappa T} - 1 + 2\kappa T(1 - \kappa T))}{16 \mathrm{e}^{2\kappa T}} \quad \text{and} \quad \beta(T) \equiv \frac{\kappa\sigma^2}{24} (1 - \mathrm{e}^{-2\kappa T}).$$

If $X_0 \neq \bar{X}$ the asymptotic law is clearly non-centered.

The result in Theorem 2 can now be used to construct an approximation of $X_T = G(\hat{X}_T)$ with an improved speed of convergence.

Corollary 1. *Under the conditions of Theorem* 2, $N(G(\hat{X}_T^N) - X_T) \Rightarrow B(X_T) U_T^{\hat{X}}$.

The convergence rate $1/N$ attained by $G(\hat{X}_T^N)$ is the same as the convergence rate of the Euler scheme applied to an ordinary differential equation. This is the best rate that can be attained with an Euler scheme.

2.3. A numerical example

For concreteness we illustrate Theorems 1 and 2 with a mean reverting, constant elasticity of variance process

$$\mathrm{d}X_v = \kappa(\overline{X} - X_v)\,\mathrm{d}v + \sigma X_v^{\gamma}\,\mathrm{d}W_v. \tag{18}$$

This specification is often used to model the short rate in term structure models, as in Chan et al. (1992). For $\gamma = 0.5$ we obtain a CIR process; otherwise it is a standard L-CEV process.

For a precise characterization of the asymptotic error we need to identify the expressions in Ω_T and U_T^X. Straightforward computations give $\partial B(x) = \sigma\gamma x^{\gamma-1}$ and $\partial A(x) = -\kappa$. The SDE for the transformed process is[15]

$$\mathrm{d}\hat{X}_v = \left[\left[\frac{\partial A}{B} - \frac{1}{2}\,\partial B\right] \circ G\right](\hat{X}_v) + \mathrm{d}W_v, \tag{19}$$

where $G(x) = (\sigma(1-\gamma)x)^{1/(1-\gamma)}$. Similarly, expressions for $\partial\hat{A}$ and $\partial^2\hat{A}$ that appear in $\hat{\Omega}_T$ and $U_T^{\hat{X}}$, are

$$\partial\hat{A}(x) = \left[\left[\partial A - \frac{A\partial B}{B} - \frac{1}{2}\partial^2 BB\right] \circ G\right](x)$$

and

$$\partial^2\hat{A}(x) = \left[\left[\partial^2 AB - \partial A\,\partial B - A\partial^2 B + \frac{A(\partial B)^2}{B} - \frac{1}{2}(\partial^3 B\,B + \partial^2 B\partial\,B)B\right] \circ G\right](x) \tag{20}$$

with $\partial^2 B(x) = \sigma\gamma(\gamma-1)x^{\gamma-2}$, $\partial^3 B(x) = \sigma\gamma(\gamma-1)(\gamma-2)x^{\gamma-3}$ and $\partial^2 A(x) = 0$.

For the L-CEV process we adopt the parameter values in Chan et al. (1992): $\kappa = 0.0171$, $\overline{X} = 0.1138$, $\sigma = -0.0655$, $\gamma = 0.9997$. Parameter values for the CIR process are taken from Broze et al. (1998): $\kappa = 0.0305$, $\overline{X} = 0.0791$, $\sigma = -0.0219$.

Figs. 1 and 2 graph the asymptotic error distributions for CIR and L-CEV. The graphs plot the empirical distribution functions based on $M = 50\,000$ replications and use $T = 1$ and $N = 365$. Observe that the empirical distributions using the transformation are non-centered and that the errors, with the transformation, are considerably smaller.

Comparing Figs. 3 and 4 for the CIR process and Figs. 5 and 6 for the L-CEV process reveals the increased speed of convergence with the transformation. In those experiments the benchmark "true" value is computed without transformation and taking $N = 2^{14}$.[16] Approximation errors, relative to this benchmark, are then computed using $N = 2^x$ with $x = 2, \ldots, 9$. Distribution functions are again based on $M = 50\,000$ replications.

The simulation of the discretized version of the process of interest is usually the first step of a Monte Carlo procedure for an econometric technique such as the

[15] Given two functions f and g, the notation $\circ$ reads $[f \circ g](x) \equiv f(g(x))$.

[16] We take this shortcut to illustrate the relative speed of convergence.

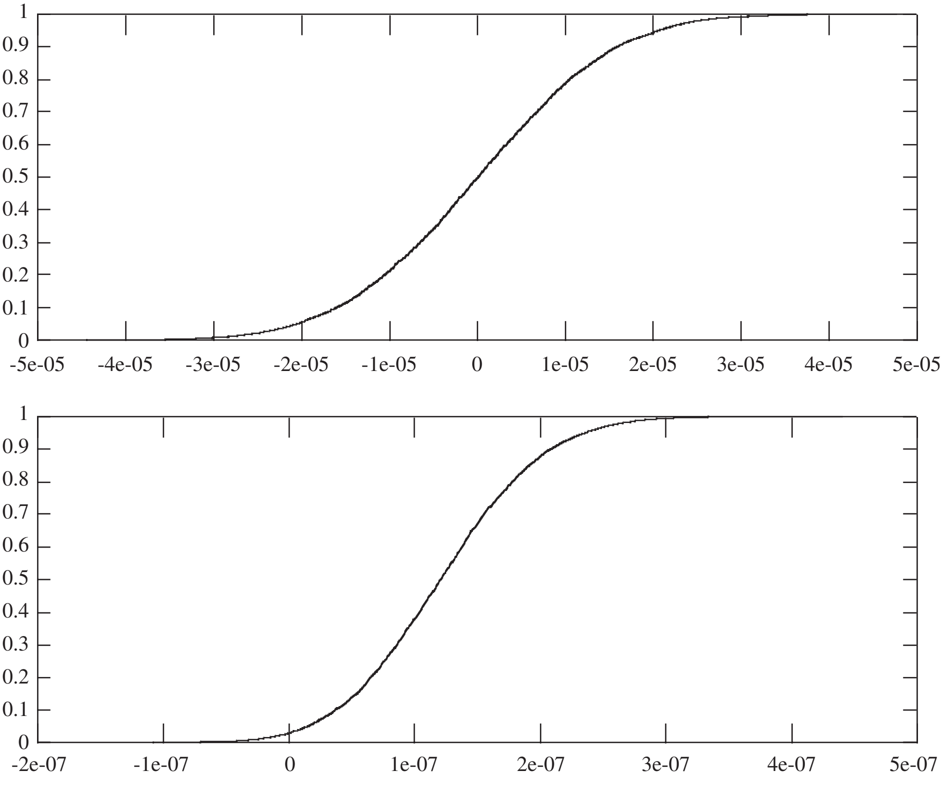

Fig. 1. Asymptotic error distribution function of CIR process, approximated with a Euler scheme without Doss transformation ($\mathbf{P}(U_1^X \leqslant x)$ upper graph), and with Doss transformation ($\mathbf{P}(U_1^{G(\hat{X})} \leqslant x)$ lower graph).

simulated method of moments. The next step is to generate a large number of discretized trajectories to compute an average designed to approximate the moment condition. Common wisdom suggests that the precision of this estimator can be improved by discretizing the process as finely as possible and taking a very large number of independent replications. In practice, however, one faces a limited budget of computation time. Duffie and Glynn (1995) propose a computationally efficient scheme which optimizes the gains achieved by reducing the length of the discretization step and by increasing the number of simulations of the sample path of the discretized process. For this efficient MCE scheme, they also characterized the asymptotic distribution of the approximation error and found it to be non-centered. The efficient procedure has therefore a second-order bias. In the next section, we extend their results in several directions. We start with a characterization of the second-order bias as the expected value of a known random variable. As this random variable can be simulated, along with the diffusion, a new approximation that corrects for second-order bias can be designed. We also provide equivalent results for the Doss-transformed process introduced in the previous section.

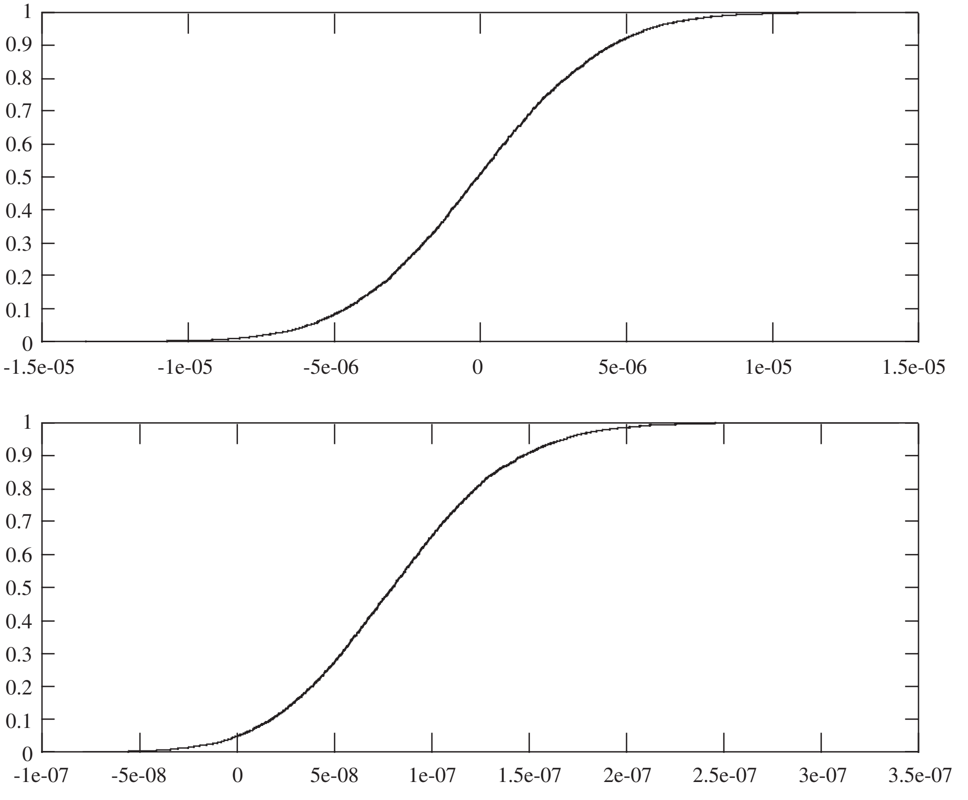

Fig. 2. Asymptotic error distribution function of a CEV process, approximated with a Euler scheme without Doss transformation ($\mathbf{P}(U_1^X \leqslant x)$ upper graph), and with Doss transformation ($\mathbf{P}(U_1^{G(\hat{X})} \leqslant x)$ lower graph).

3. Asymptotic laws of estimators of conditional expectations

We now derive the asymptotic error of the estimate of the conditional expectation of a function of the terminal value of an SDE, X_T. When the distribution of X_T is unknown an estimator of the expected value is obtained by sampling independent replications of the numerical solution of the SDE and averaging over the sampled values. The approximation error of this scheme has two components (Duffie and Glynn, 1995). The first is the error due to the discretization of the SDE. The second is the error in the approximation of the conditional expectation by a sample average.

Section 3.1 presents our central result, namely the asymptotic error distributions associated with estimators of conditional expectations. Auxiliary results concerning the error component associated with the discretization scheme are described in Section 3.2. The second-order biases of these estimators are discussed in Section 3.3, and bias correction is performed in Section 3.4.

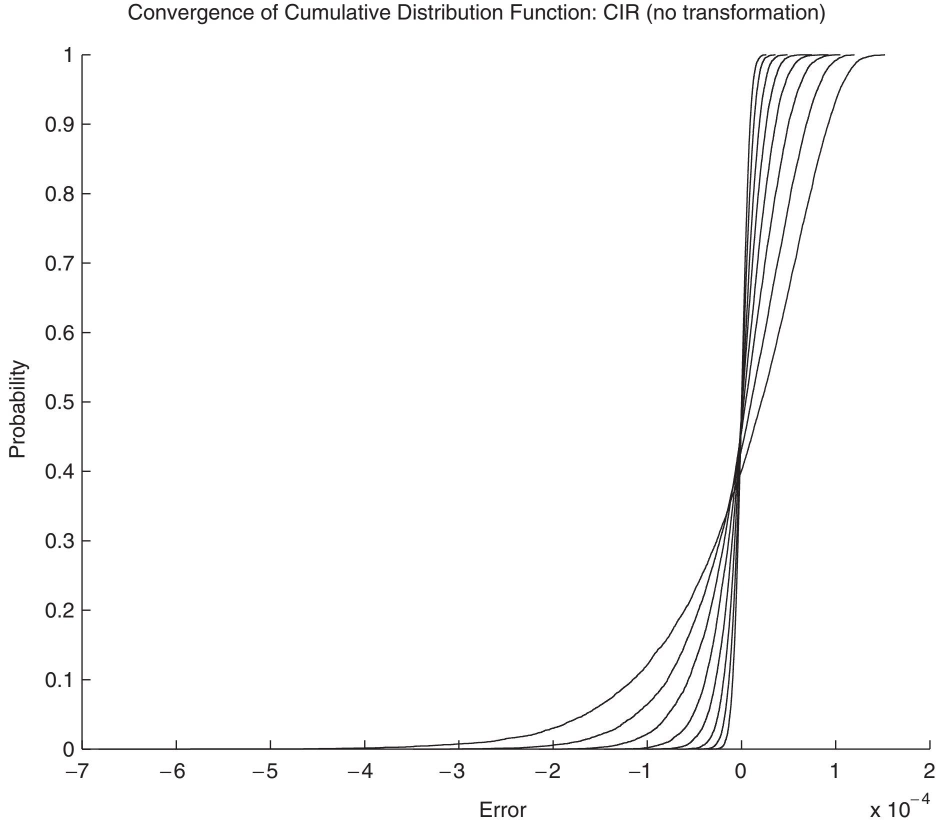

Fig. 3. Speed of convergence of distribution function of the approximation error $X_1^N - X_1$ for a CIR process approximated with a Euler scheme for $N = 2^x$ and $x = 2, \ldots, 9$.

3.1. Asymptotic error distributions

Suppose that we wish to calculate $\mathbf{E}[g(X_T)|\mathscr{F}_0] = \mathbf{E}_0[g(X_T)] = \mathbf{E}_0[\hat{g}(\hat{X}_T)]$ where X solves (2) or (11). The estimators without and with transformation are

$$g^{N,M} \equiv \frac{1}{M} \sum_{i=1}^{M} g(X_T^{i,N}), \tag{21}$$

and

$$\hat{g}^{N,M} \equiv \frac{1}{M} \sum_{i=1}^{M} \hat{g}(\hat{X}_T^{i,N}). \tag{22}$$

These estimators of the conditional expectation rely on the law of large numbers and draw independent replications $X_T^{i,N}$(resp. $\hat{X}_T^{i,N}$) of the terminal points

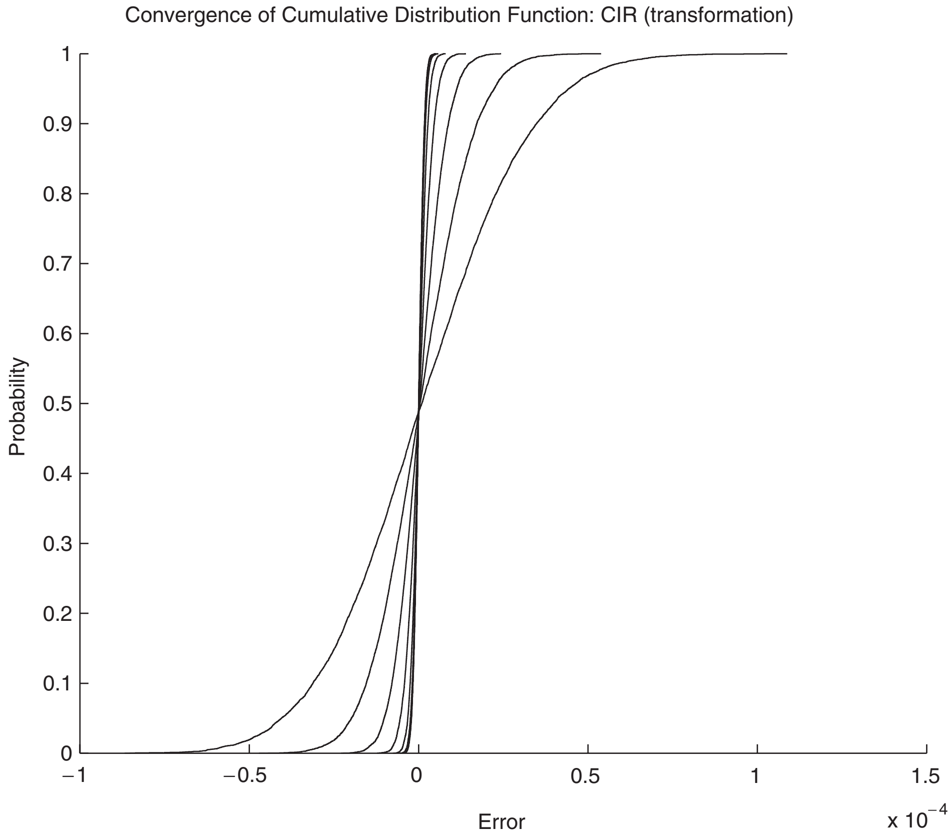

Fig. 4. Speed of convergence of the distribution function of the approximation error $G(\hat{X}_1^N) - X_1$ for a Doss-transformed CIR process approximated with a Euler scheme for $N = 2^x$ and $x = 2, \ldots, 9$.

X_T^N (resp. $\hat{X}_T^N$) of the Euler discretized diffusion without (resp. with) Doss transformation. Our next theorem describes their asymptotic laws.

Theorem 3. *Let* $\hat{g} \in \mathscr{C}^1(\mathbb{R}^d)$, $g \in \mathscr{C}^3(\mathbb{R}^d)$, *and suppose that* $g(X_T) \in \mathbb{D}^{1,2}$.[17] *Also suppose that the assumptions of Theorems* 4 *and* 5 *hold. For the schemes without and with transformation, we have, as* $M \to \infty$,

$$\sqrt{M}(g^{N_M,M} - \mathbf{E}_0[g(X_T)]) \Rightarrow \varepsilon_{\frac{1}{2}} K_T(X_0) + L_T(X_0), \tag{23}$$

$$\sqrt{M}(\hat{g}^{N_M,M} - \mathbf{E}_0[g(X_T)]) \Rightarrow \varepsilon_{\frac{1}{2}} \hat{K}_T(X_0) + L_T(X_0), \tag{24}$$

[17]The space $\mathbb{D}^{1,2}$ is defined as the domain of the Malliavin derivative operator (see Nualart (1995) for an exact definition and more on Malliavin calculus). A brief introduction to Malliavin calculus also appears in Appendix D of Detemple et al. (2003).

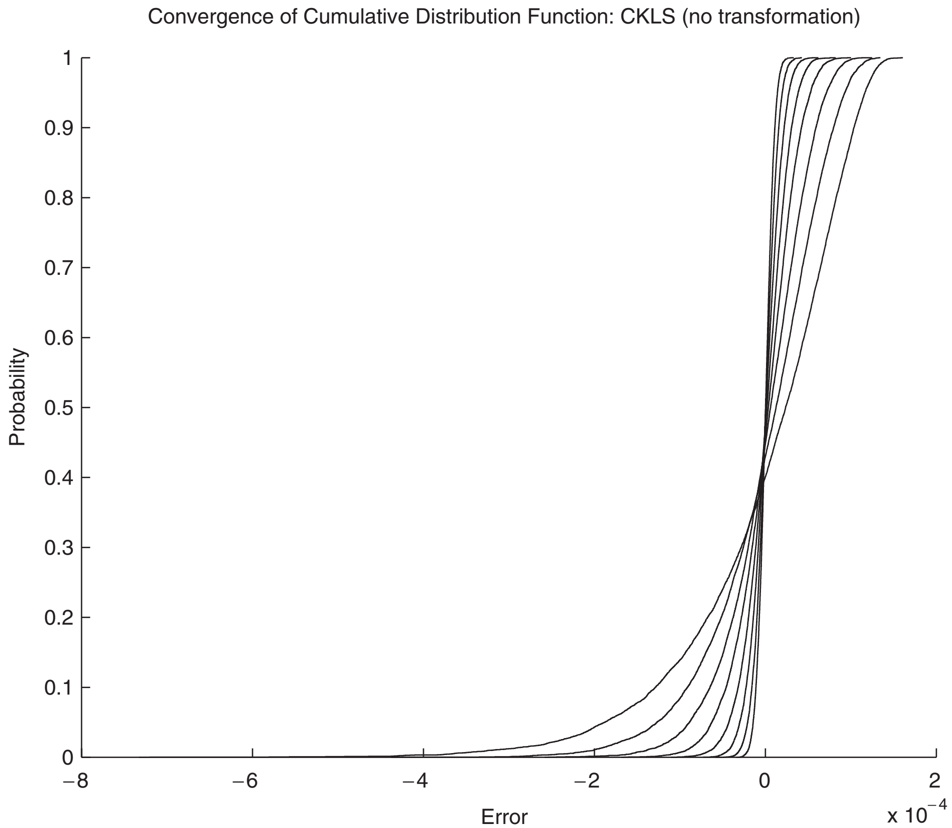

Fig. 5. Speed of convergence of the distribution function of the approximation error $X_1^N - X_1$ for a CEV process approximated with a Euler scheme for $N = 2^x$ and $x = 2, \ldots, 9$.

where $\lim_{M\to\infty} N_M = +\infty$ *and* $\varepsilon = \lim_{M\to\infty} \sqrt{M}/N_M$, *and* $L_T(X_0)$ *is the terminal value of a centered Gaussian martingale with* (*deterministic*) *quadratic variation and conditional variance given by*

$$[L, L]_T = \int_0^T \mathbf{E}_0[N_v(N_v)'] \, \mathrm{d}v \equiv \mathbf{VAR}[g(X_T)|\mathscr{F}_0], \tag{25}$$

$$N_v = \mathbf{E}_v[\partial g(X_T) \mathscr{D}_v X_T]. \tag{26}$$

In these expressions $\mathscr{D}_s X_T$ *is the Malliavin derivative of* X_T. *The deterministic functions* K *and* $\hat{K}$ *are defined in Theorems* 4 *and* 5 *below.*

The random variable $\mathscr{D}_s X_T$ captures the impact of an innovation in the Brownian motion W at time s on the state variable X at time T. In essence this derivative measures the persistence of a shock in the state variable. It is similar to an impulse

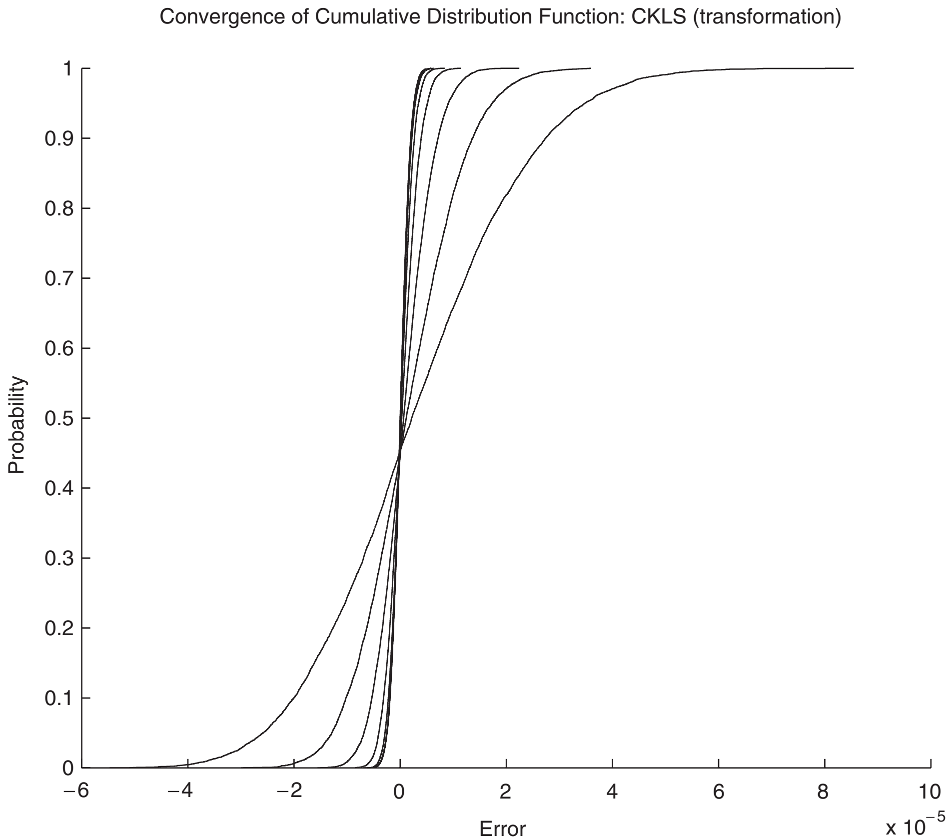

Fig. 6. Speed of convergence of the distribution function of the approximation error $G(\hat{X}_T^N) - X_T$ for a Doss-transformed CEV process approximated with a Euler scheme for $N = 2^x$ and $x = 2, \ldots, 9$.

response function that quantifies the sensitivity of the variable X_T to an uncertainty shock at the prior time s.

The theorem shows that the asymptotic laws of the estimators have two parts. The first, K, corresponds to the discretization bias; the second, L, results from the Monte Carlo estimation of the expectation. Note that L would not vanish, even if samples where taken from the law of X_T. This is because the conditional expectation cannot be calculated in closed form.

The theorem also shows that the estimators converge at the same rate. This follows from the fact that the convergence rate of the expected approximation error, described in Theorems 4 and 5 in the next section, is the same. As the rate $1/\sqrt{M}$ is obtained from a central limit theorem, an additional conclusion is that higher-order schemes would fail to improve the convergence speed. They would just reduce the factor ε in the limits and therefore the second-order bias.

3.2. Expected approximation errors

We now provide auxiliary results concerning the error component associated with the discretization scheme. Let e_T^N ($\hat{e}_T^N$) be the expected approximation error for the scheme without (with) transformation. By definition

$$e_T^N \equiv \mathbf{E}_0[g(X_T^N)] - \mathbf{E}_0[g(X_T)], \tag{27}$$

$$\hat{e}_T^N \equiv \mathbf{E}_0[\hat{g}(\hat{X}_T^N)] - \mathbf{E}_0[g(X_T)], \tag{28}$$

where $\hat{g} \equiv g \circ F$ with F the inverse of G, $g(X_T^N)$ is an approximation of $g(X_T)$ based on the Euler discretization of X, and $\hat{g}(\hat{X}_T^N)$ is an approximation of $g(X_T)$ based on the Euler discretization of the transformed state variables $\hat{X}$.

Next we study the convergence properties of these errors. These results are relevant to the extent that the limits of Ne_T^N and $N\hat{e}_T^N$ determine the means of the asymptotic error distributions of the corresponding efficient estimators of conditional expectations. As we will show, these limits characterize the second-order approximation bias of efficient Monte Carlo estimators of diffusions.

3.2.1. Euler scheme on the original state variables

Our first result describes the convergence of the expected approximation error e_T^N, in (27). Define the random variables

$$\begin{aligned} V_{1,T} \equiv & -\Omega_T \int_0^T \Omega_s^{-1} \left(\partial A(X_s)\,\mathrm{d}X_s + \sum_{j=1}^d [\partial B_j A](X_s)\,\mathrm{d}W_s^j \right. \\ & \left. - \sum_{i,j=1}^d [\partial B_j\, \partial B_j\, B_i](X_s)\,\mathrm{d}W_s^i \right) \\ & + \Omega_T \int_0^T \Omega_s^{-1} \left[\sum_{j=1}^d [\partial B_j\, \partial B_j\, A] + \sum_{j,k,l=1}^d [\partial_k(\partial_l A B_{k,j})] \right] (X_s)\,\mathrm{d}s \\ & + \Omega_T \int_0^T \Omega_s^{-1} \sum_{i,j=1}^d ([\partial[\partial B_j\, \partial B_j\, B_i] B_i - \partial B_i\, \partial B_j\, \partial B_j\, B_i](X_s))\,\mathrm{d}s, \end{aligned} \tag{29}$$

$$V_{2,T} \equiv -\int_0^T \sum_{i,j=1}^d v_{i,j}(s,T)\,\mathrm{d}s, \tag{30}$$

where ∂A, ∂B_j are $d \times d$ matrices of Jacobians, Ω is defined in Theorem 1 and $v_{i,j}(s,T)$ is defined in (137)–(141). With this notation we have:

Theorem 4. *Suppose that $A, B_j \in \mathscr{C}^2(\mathbb{R}^d)$. Let $g \in \mathscr{C}^3(\mathbb{R}^d)$ be such that*

$$\lim_{r\to\infty} \limsup_N \mathbf{E}_0[\mathbf{1}_{\{|N(g(X_T^N)-g(X_T))|>r\}} N|g(X_T^N) - g(X_T)|] = 0 \tag{31}$$

(**P**-*a.s.*). *Then*,

$$Ne_T^N \to \tfrac{1}{2}K_T(X_0) \equiv \tfrac{1}{2}\mathbf{E}_0[\partial g(X_T)V_{1,T} + V_{2,T}], \tag{32}$$

where $V_{1,T}$, $V_{2,T}$ *are given by* (29) *and* (30), *and* e_T^N *is defined in* (27).

Theorem 4 provides a probabilistic characterization of the asymptotic expected error. The expressions in (32) depend on random variables V_1 and V_2 that are determined in closed form by the derivatives of the drift and volatility coefficients of the SDE. They can therefore be simulated along with the state variables to derive bias-corrected estimators (see Section 3.4). This characterization of the asymptotic expected error is easier to evaluate than the expressions presented in Talay and Tubaro (1990) and Bally and Talay (1996a,b). These authors show that the asymptotic expected error for the Euler scheme can be written as the expectation of a function of a random variable, where the function solves a PDE. In contrast, our characterization is fully probabilistic as it does not require the resolution of a PDE, and therefore, does not suffer from the curse of dimensionality affecting numerical solutions of PDEs. As a consequence, it remains applicable for multivariate diffusions. Even in the univariate case, using Monte Carlo simulation in combination with the solution of a PDE is computationally costly. This may explain why the theoretical results of Talay and Tubaro (1990) and Bally and Talay (1996a,b) have seen limited use in applications.

Implementation, in numerous applications, requires the computation of conditional expectations of path-dependent functionals of diffusion processes, such as the Riemann integral $\int_0^T g(X_s)\,\mathrm{d}s$ (see the example in Section 5.2). A convergent estimator for this integral, based on the Euler scheme, is $\sum_{n=0}^{N-1} g(X_{nh})h$, where $h = T/N$, and X^N is the solution of the Euler-discretized SDE starting at X_0. Theorem 4 can be used to deduce the asymptotic expected approximation error in these cases.

Corollary 2. *Under the assumptions of Theorem* 4, *we have*, *as* $N \to \infty$,

$$N\mathbf{E}_0\left[\sum_{n=0}^{N-1} g(X_{nh}^N)h - \int_0^T g(X_s)\,\mathrm{d}s\right] \to \frac{1}{2}K_{1T}(X_0)$$

P–*a.s.*, *with*

$$K_{1T}(X_0) \equiv \int_0^T K_s(X_0)\,\mathrm{d}s + K_{2T}(X_0), \tag{33}$$

where K *is defined in Theorem* 4, *and*

$$K_{2T}(X_0) \equiv -\mathbf{E}_0\left[\int_0^T \left(\partial g(X_s)\left(A(X_s) + \sum_{j=1}^d [(\partial B_j)B_j](X_s)\right) + \sum_{j=1}^d [B_j'\,\partial^2 g\, B_j](X_s)\right)\mathrm{d}s\right]. \tag{34}$$

The expected approximation error has two terms. The first term, $\int_0^T K_s(X_0)\,\mathrm{d}s$, captures the cumulated expected approximation error $X^N - X$ in the Riemann integral over the interval $[0, T]$. The second term, K_{2T}, emerges because the continuous Riemann integral, $\int_0^T g(X_s)\,\mathrm{d}s$, is approximated by a discrete sum, $\sum_{n=0}^{N-1} g(X_{nh}^N)h$.

3.2.2. Euler scheme on the transformed state variables

To derive the expected approximation error for the estimator with transformation define the random variable

$$\hat{V}_T = -\hat{\Omega}_T \int_0^T \hat{\Omega}_v^{-1} \partial \hat{A}(\hat{X}_v) \left(\mathrm{d}\hat{X}_v + \sum_{j,k,l=1}^{d} \partial_{l,k} \hat{A}(\hat{X}_v) \hat{B}_{k,j} \hat{B}_{l,j} \, \mathrm{d}v \right) \tag{35}$$

with $\hat{\Omega}_T = \mathscr{E}^R(\int_0^v \partial \hat{A}(\hat{X}_s)\,\mathrm{d}s)$. We obtain

Theorem 5. *Suppose that $\hat{A} \in \mathscr{C}^1(\mathbb{R}^d)$, and that the conditions of Theorem 2 hold. For $\hat{g} \in \mathscr{C}^1(\mathbb{R}^d)$ such that*

$$\lim_{r\to\infty} \limsup_{N} \mathbf{E}_0[\mathbf{1}_{\{|N(\hat{g}(\hat{X}_T^N) - \hat{g}(\hat{X}_T))| > r\}} N|\hat{g}(\hat{X}_T^N) - \hat{g}(\hat{X}_T)|] = 0 \tag{36}$$

P-*a.s. we have*, **P**-*a.s., as $N \to \infty$,*

$$N\hat{e}_T^N \to \tfrac{1}{2}\hat{K}_T(X_0) \equiv \tfrac{1}{2}\mathbf{E}_0[\partial \hat{g}(\hat{X}_T)\hat{V}_T], \tag{37}$$

where $\hat{V}_T$ is defined in (35), *and $\hat{e}_T^N$ is defined in* (28).

A comparison of (32) with (37) suggests that it will be difficult, in general, to establish the dominance of one method over the other on the basis of the asymptotic expected error. Indeed, the formulas reveal that both methods converge at the same speed $1/N$ and that the second-order biases, while different $(K_T(X_0) \neq \hat{K}_T(X_0))$, do not appear to be ordered in a systematic manner. To compare the two methods one may want to use additional criteria, such as the computational cost.

The difference in the convergence rates to the limit errors (Theorems 1 and 2) and those to the limit expected errors (Theorems 4 and 5) can be explained as follows. The speed of convergence for the limit error is determined by the martingale part of the error expansion that converges more slowly $(1/\sqrt{N})$ than the bounded variation part $(1/N)$. The transformation eliminates the martingale part of the error. Taking the expectation also eliminates the martingale part of the error. It follows immediately that the expected errors will converge at the same rate. To see that the expectation eliminates the martingale part of the error in the absence of a transformation it suffices to note that this term converges weakly to a stochastic integral whose expectation is null.[18]

Let us now consider estimators of Riemann integrals. The counterpart of Corollary 2, for the scheme based on the transformed state variables, is

[18]The martingale part of the error converges to the product of a random variable and a stochastic integral with independent integrator.

Corollary 3. *Under the assumptions of Theorem* 5, *we have, as* $N \to \infty$,

$$N\mathbf{E}_0\left[\sum_{n=0}^{N-1} \hat{g}(\hat{X}^N_{nh})h - \int_0^T \hat{g}(\hat{X}_s)\,\mathrm{d}s\right] \to \frac{1}{2}\hat{K}_{1T}(X_0)$$

$\mathbf{P}$*–a.s., with*

$$\hat{K}_{1T}(X_0) \equiv \int_0^T \hat{K}_s(X_0)\,\mathrm{d}s + \hat{K}_{2T}(X_0), \tag{38}$$

where $\hat{K}$ *is defined in Theorem* 5, *and*

$$\hat{K}_{2T}(X_0) \equiv -\mathbf{E}_0\left[\int_0^T \left(\partial\hat{g}(\hat{X}_s)\hat{A}(\hat{X}_s) + \sum_{j=1}^{d} \hat{B}_j'\partial^2\hat{g}(\hat{X}_s)\hat{B}_j\right)\mathrm{d}s\right]. \tag{39}$$

The two terms in this decomposition have the same interpretation as those of Corollary 2. Note, in particular, that the expression for $\hat{K}_2$ follows immediately from K_2 and the deterministic nature of the volatility coefficient $\hat{B}$ (i.e. $\partial\hat{B}_j = 0$).

3.3. Second-order discretization biases

Theorem 3 shows that the two procedures have asymptotic second-order discretization biases that are, respectively, given by $\frac{\varepsilon}{2}K$ and $\frac{\varepsilon}{2}\hat{K}$. It follows that any confidence interval, based solely on the Gaussian process L, will suffer from a size distortion. More precisely, when $M \to \infty$ we have

$$\mathbf{P}\left(\mathbf{E}_0[g(X_t)] \in C_g^{N_M,M}(\alpha)\right) \to \Psi(\delta) \tag{40}$$

with

$$C_g^{N_M,M}(\alpha) \equiv \left[\sqrt{M}g^{N_M,M} + \Phi^{-1}(\alpha/2)\frac{\sigma^{N_M,M}}{\sqrt{M}}, \sqrt{M}g^{N_M,M} + \Phi^{-1}(\alpha/2)\frac{\sigma^{N_M,M}}{\sqrt{M}}\right]$$

and

$$\Psi(\delta) \equiv \Phi(\Phi^{-1}(1-\alpha/2) - \delta) - \Phi(\Phi^{-1}(\alpha/2) - \delta), \tag{41}$$

where Φ denotes the cumulative Gaussian distribution, $\delta = \varepsilon K_T(X_0)/2\sqrt{\mathbf{VAR}[L_T|\mathscr{F}_0]}$ and where $(\sigma^{N,M})^2 = \mathbf{VAR}^{N,M}[L_T|\mathscr{F}_0]$ is a convergent estimator of the variance. A confidence interval of nominal size α, based on L, will cover the true value $\mathbf{E}_0[g(X_T)]$ only with probability $\Psi(\delta)$ and not $1-\alpha$. For the method with transformation the coverage probability of the true value is $\Psi(\hat{\delta})$ with $\hat{\delta} = \varepsilon\hat{K}_T(X_0)/2\sqrt{\mathbf{VAR}[L_T|\mathscr{F}_0]}$.

The degree of size distortion can be measured by $s(z) \equiv 1 - \alpha - \Psi(z)$, where $z \in \{\delta, \hat{\delta}\}$. As $\Psi(z)$ is strictly monotone with $\Psi(0) = 1-\alpha$ we conclude that these confidence intervals have the requested nominal size if and only if there is no second-order bias, i.e. $\delta = 0$. Clearly, an increase in the second-order bias reduces the real coverage probability. Likewise, a decrease in the asymptotic variance or an increase in ε will increase the size distortion.

A benefit of formulas (32), (37) for the second-order biases $K_T(X_0)$ and $\hat{K}_T(X_0)$, is that they can be computed by simulation. Theorems 4 and 5 can then be used to develop approximation schemes that correct for second-order bias. Likewise, asymptotically valid confidence intervals can easily be implemented. Furthermore, bias correction is feasible even when the number of state variables is large. In contrast bias correction based on the solutions of PDEs quickly becomes infeasible when the number of state variables increases.[19]

3.4. Bias-corrected estimators

Bias-corrected estimators are asymptotically equivalent to Monte Carlo estimators obtained by sampling from the true distribution of X_T. As a result they do not suffer from the size distortion problem described above. Bias-corrected estimators of conditional expectations can be constructed as

$$g_{\mathrm{c}}^{N,M} = \frac{1}{M} \sum_{i=1}^{M} [g(X_{Nh}^{i,N}) + \tfrac{1}{2} \partial g(X_{Nh}^{i,N}) C_{1,Nh}^{i,N} + \tfrac{1}{2} C_{2,Nh}^{i,N}], \tag{42}$$

$$\hat{g}_{\mathrm{c}}^{N,M} = \frac{1}{M} \sum_{i=1}^{M} \left[\hat{g}(\hat{X}_{Nh}^{i,N}) + \frac{1}{2} \partial \hat{g}(\hat{X}_{Nh}^{i,N}) \hat{C}_{Nh}^{i,N} \right], \tag{43}$$

where $C_{1,Nh}^{i,N}, C_{2,Nh}^{i,N}, \hat{C}_{Nh}^{i,N}$, for $n = 0, \ldots, N-1$, are defined in Appendix B and $\hat{g}$ is implicitly defined at the beginning of Section 3.2.2. Our next result shows that $g_{\mathrm{c}}^{N,M}$ and $\hat{g}_{\mathrm{c}}^{N,M}$ are bias-corrected estimators.

Theorem 6. *Suppose that the conditions of Theorems* 4 *and* 5 *hold. Then*,

$$\sqrt{M}(g_{\mathrm{c}}^{N_M,M} - \mathbf{E}_0[g(X_T)]) \Rightarrow L_T(X_0) \tag{44}$$

and

$$\sqrt{M}(g_{\mathrm{c}}^{N_M,M} - \hat{g}_{\mathrm{c}}^{N_M,M}) \Rightarrow 0 \tag{45}$$

as $M \to \infty$, *with* $\lim_{M\to\infty} N_M = +\infty$.

Theorem 6 shows that $g_{\mathrm{c}}^{N,M}$ and $\hat{g}_{\mathrm{c}}^{N,M}$ correct for the second-order bias and are asymptotically equivalent.[20] This means that a bias correction eliminates all the potential benefits of the transformation. However, one should bear in mind that the equivalence is asymptotic and that bias-corrected estimators may perform differently in finite samples.

[19] Duffie and Glynn (1995) propose the Richardson–Romberg type of estimator $(2/4M)\sum_{i=1}^{4M} g(X_T^{i,2N}) - (1/M)\sum_{i=1}^{M} g(X_T^{i,N})$ to eliminate the second-order bias asymptotically. To calculate this estimator one must quadruple the number of replications and double the number of discretization points. In our method the second-order bias can be simulated along with the state variables, which is computationally cheaper than a Richardson–Romberg approximation scheme.

[20] Two estimators are called equivalent if they share the same asymptotic error distribution.

Corresponding bias-corrected estimators for conditional expectations of Riemann integrals

$$f(X_0) = \mathbf{E}_0\left[\int_0^T g(X_s)\,\mathrm{d}s\right] \tag{46}$$

are

$$f_{\mathrm{c}}^{N,M} = \frac{1}{M}\sum_{i=1}^{M}\left(\sum_{n=0}^{N-1}\left[g(X_{nh}^{i,N}) + \frac{1}{2}\,\partial g(X_{nh}^{i,N})C_{1,nh}^{i,N} + \frac{1}{2}\,C_{2,nh}^{i,N}\right]h + \frac{1}{2}\,C_{3,Nh}^{i,N}\right), \tag{47}$$

$$\hat{f}_{\mathrm{c}}^{N,M} = \frac{1}{M}\sum_{i=1}^{M}\left(\sum_{n=0}^{N-1}\left[\hat{g}(\hat{X}_{nh}^{i,N}) + \frac{1}{2}\,\partial\hat{g}(\hat{X}_{nh}^{i,N})\hat{C}_{nh}^{i,N}\right]h + \frac{1}{2}\,\hat{C}_{1,Nh}^{i,N}\right). \tag{48}$$

These estimators are sums over (42) and (43). Each involves an additional term (C_3 and $\hat{C}_1$), that corrects for the bias due to the approximation of a continuous integral by a discrete sum. Appendix B provides exact expressions for these additional bias correction terms. The next result parallels Theorem 6.

Corollary 4. *Suppose that the conditions of Theorem* 4 *and* 5 *hold. Then,* $\sqrt{M}(f_{\mathrm{c}}^{N_M,M} - f(X_0)) \Rightarrow L_T^f(X_0)$, *where* $L_T^f(X_0)$ *is a centered Gaussian martingale with* (*deterministic*) *quadratic variation and conditional variance given by*

$$[L^f, L^f]_T = \int_0^T \mathbf{E}_0[N_v^f(N_v^f)']\,\mathrm{d}v \equiv \mathbf{VAR}\left[\int_0^T g(X_s)\,\mathrm{d}s\middle|\mathscr{F}_0\right], \tag{49}$$

$$N_v^f = \mathbf{E}_v\left[\int_v^T \partial g(X_s)\mathscr{D}_v X_s\,\mathrm{d}s\right]. \tag{50}$$

Furthermore,

$$\sqrt{M}(f_{\mathrm{c}}^{N_M,M} - \hat{f}_{\mathrm{c}}^{N_M,M}) \Rightarrow 0 \tag{51}$$

as $M \to \infty$ *with* $\lim_{M\to\infty} N_M = \infty$.

4. A comparison with Milshtein's second-order approximation

While Euler schemes for SDEs are appealing from a computational point of view, they might be judged insufficiently accurate. Second-order schemes such as Milshtein's scheme (see Milshtein, 1984, 1995; Talay, 1984, 1986, 1996) have in fact been proposed to provide improved approximations. In this section, we extend our analysis to the Milshtein second-order approximation.

The Milshtein approximation of X_T in (2) is

$$\tilde{X}_T^N = X_0 + \sum_{n=0}^{N-1}\left(A(\tilde{X}_{nh}^N)h + \sum_{j=1}^{d} B_j(\tilde{X}_{nh}^N)\Delta W_{nh}^j + \sum_{j,l=1}^{d}[\partial B_l\, B_j](\tilde{X}_{nh}^N)\Delta F^{l,j}\right), \tag{52}$$

where

$$\Delta F^{l,j} \equiv \int_{nh}^{(n+1)h} \int_{nh}^{s} \mathrm{d}W_v^l \,\mathrm{d}W_s^j, \tag{53}$$

with $h = T/N$ and $\Delta W_{nh}^j = W_{(n+1)h}^j - W_{nh}^j$. This scheme is obtained using a stochastic Taylor expansion for the diffusion coefficient.[21]

The Milshtein scheme is often difficult to implement for multivariate diffusions. The increments $\Delta F^{l,j}$, defined in (53) cannot, in general, be written in a form that is easy to simulate. In fact, for multivariate diffusions without commutative noise ($\partial B_l\, B_k \neq \partial B_k B_l$ for some $k \neq l$), the increment $\Delta F^{l,j}$ can only be simulated by using a further discretization of the intervals $[nh, (n+1)h]$ (see Gaines and Lyons, 1997). This entails a substantial increase in computational cost. The comparison to a Euler scheme with a number of discretization points equal to the total number of points required to implement the Milshtein scheme is unclear.[22]

For commutative noise (i.e. $\partial B_l\, B_k = \partial B_k\, B_l$), the last term in (52) simplifies to[23]

$$\sum_{j,l=1}^{d} [\partial B_l\, B_j](\tilde{X}_{nh}^N)\Delta F^{l,j} = \frac{1}{2}\left(\sum_{j=1}^{d} [\partial B_j\, B_j](\tilde{X}_{nh}^N)((\Delta W_{nh}^j)^2 - h)\right) + \frac{1}{2}\left(\sum_{\substack{j,l=1\\ j\neq l}}^{d} [\partial B_l\, B_j](\tilde{X}_{nh}^N)\Delta W_{nh}^j\, \Delta W_{nh}^l\right). \tag{54}$$

The representation of $\tilde{X}_T^N$ as a functional of the Brownian increments ΔW_{nh}^k for $k = l,j$ follows. Note, in particular, that every univariate diffusion has commutative noise and can be implemented on the basis of (54).

Our next result describes the asymptotic distribution of the approximation error associated with (52). It will enable us to find an explicit expression for the Monte Carlo estimator of a conditional expectation based on the Milshtein scheme.

Theorem 7. *The approximation error $\tilde{X}_T^N - X_T$ converges weakly at the rate $1/N$ (i.e. $N(\tilde{X}_T^N - X_T) \Rightarrow \tilde{U}_T^X$). The asymptotic error is*

$$\tilde{U}_T^X = -\frac{1}{2}\,\Omega_T \int_0^T \Omega_s^{-1}\left(\partial A(X_s)\,\mathrm{d}X_s - \sum_{j=1}^{d} [(\partial B_j)(\partial A)B_j](X_s)\,\mathrm{d}s\right)$$

[21] For details on the stochastic Taylor expansion and the derivation of this result see Kloeden and Platen (1997) or Milshtein (1995).

[22] The commutative noise condition necessary for this representation is the same condition that is needed for the Doss transformation. When commutativity fails the transformation cannot be used (see Remarks 3.1–3.4 in Detemple et al., 2005 for further details) and the implementation of the Milshtein scheme requires the simulation of the iterated Wiener integrals $\Delta F^{l,j}$. In those cases implementation based on a standard Euler scheme without transformation is considerably less demanding.

[23] See Milshtein (1995) or Gaines and Lyons (1997).

$$
\begin{aligned}
&-\frac{1}{2}\,\Omega_T \int_0^T \Omega_s^{-1} \sum_{j=1}^{d} [(\partial[(\partial A)B_j])B_j - (\partial B_j)(\partial B_j)A](X_s)\,\mathrm{d}s \\
&-\frac{1}{2}\,\Omega_T \int_0^T \Omega_s^{-1} \sum_{j=1}^{d} [(\partial B_j)A](X_s)\,\mathrm{d}W_s^j \\
&-\frac{1}{\sqrt{12}}\,\Omega_T \int_0^T \Omega_s^{-1} \sum_{j=1}^{d} [(\partial A)B_j - (\partial B_j)A](X_s)\,\mathrm{d}Z_s^j \\
&-\frac{1}{\sqrt{6}}\,\Omega_T \int_0^T \Omega_s^{-1} \sum_{i,l,j=1}^{d} [\partial B_i\,\partial B_l\,B_j](X_s)\,\mathrm{d}\tilde{Z}_s^{l,j,i},
\end{aligned}
$$

where the process $((Z^j)_{j\in\{1,\dots,d\}}, (\tilde{Z}^{l,j,i})_{i,l,j=1,\dots,d})$ *is a* $d + d^3 \times 1$ *standard Brownian motion independent of* W. *The process* Ω_T *is given in Theorem* 1.

Theorem 7 shows that the speed of convergence increases when one uses the stochastic Taylor expansion of the diffusion term. This expansion eliminates the error component of order $1/\sqrt{N}$ in the martingale part of the Euler approximation. It follows that the error for the Milshtein scheme converges at the same speed as the error for the Euler scheme with transformation: both schemes improve on the standard Euler approach. Note also that, in the special case of a constant volatility coefficient, estimators with and without transformation are identical, and asymptotic error distributions for Milshtein and Euler with transformation are the same (i.e. $\tilde{U}_T^X = U_T^{\hat{X}}$). In this instance, all the schemes have the same asymptotic distribution.

The asymptotic error distribution of estimators of conditional expectations is described next.

Theorem 8. *Suppose that the assumptions of Theorem* 9 *below hold. Let* $g \in \mathscr{C}^1(\mathbb{R}^d)$ *and suppose that* $g(X_T) \in \mathbb{D}^{1,2}$. *For the Milshtein scheme, we have, as* $M \to \infty$,

$$
\sqrt{M}\left(\frac{1}{M}\sum_{i=1}^{M} g(\tilde{X}_T^{i,N_M}) - \mathbf{E}_0[g(X_T)]\right) \Rightarrow \varepsilon\frac{1}{2}\,\tilde{K}_T(X_0) + L_T(X_0), \tag{55}
$$

where $\lim_{M\to\infty} N_M = +\infty$ *and* $\varepsilon = \lim_{M\to\infty} \sqrt{M}/N_M$. *The random variable* $L_T(X_0)$ *is the terminal value of a centered Gaussian martingale with (deterministic) quadratic variation and conditional variance defined in Theorem* 3. *The deterministic function* $\tilde{K}_T$ *is defined in Theorem* 9.

For estimates of conditional expectations the rate of convergence of the Milshtein scheme is identical to the rates of the Euler schemes with and without transformation. The three schemes differ only in their second-order biases. The second-order bias $\tilde{K}$ can be found explicitly, as shown in Theorem 9 below. Its size relative to the second-order biases of Euler schemes depends on the coefficients of the underlying processes. A global ordering of the three schemes in terms of second-order asymptotic properties is not readily apparent.

As for the estimator with transformation, the expected approximation error

$$\tilde{e}_t^N \equiv \mathbf{E}_0[g(\tilde{X}_T^N) - g(X_T)], \tag{56}$$

can be directly deduced from Theorem 7 under an additional uniform integrability condition. In order to do this, define the random variable

$$\begin{aligned}\tilde{V}_T = & -\Omega_T \int_0^T \Omega_s^{-1}\left(\partial A(X_s)\,\mathrm{d}X_s - \sum_{j=1}^d [(\partial B_j)(\partial A)B_j](X_s)\,\mathrm{d}s\right) \\ & -\Omega_T \int_0^T \Omega_s^{-1} \sum_{j=1}^d [(\partial[(\partial A)B_j])B_j - (\partial B_j)(\partial B_j)A](X_s)\,\mathrm{d}s \\ & -\Omega_T \int_0^T \Omega_s^{-1} \sum_{j=1}^d [(\partial B_j)A](X_s)\,\mathrm{d}W_s^j. \end{aligned} \tag{57}$$

The equivalent of Theorems 4 and 5, is

Theorem 9. *For $g \in \mathscr{C}^1(\mathbb{R}^d)$ such that*

$$\lim_{r\to\infty} \limsup_N \mathbf{E}_0[\mathbf{1}_{\{|N(g(\tilde{X}_T^N) - g(X_T))| > r\}} N|g(\tilde{X}_T^N) - g(X_T)|] = 0 \tag{58}$$

we have, $\mathbf{P}$-*a.s.*, *as* $N \to \infty$,

$$N\tilde{e}_T^N \to \tfrac{1}{2}\tilde{K}_T(X_0) \equiv \tfrac{1}{2}\mathbf{E}_0[\partial g(X_T)\tilde{V}_T], \tag{59}$$

where $\tilde{V}_T$ *is defined in* (57) *and* $\tilde{e}_T^N$ *is defined in* (56).

Our next result extends Theorem 9 to Riemann integrals. The result parallels Corollary 2.

Corollary 5. *Under the assumptions of Theorem* 9, *we have*, *as* $N \to \infty$,

$$N\mathbf{E}_0\left[\sum_{n=0}^{N-1} g(\tilde{X}_{nh}^N)h - \int_0^T g(X_s)\,\mathrm{d}s\right] \to \frac{1}{2}\tilde{K}_{1T}(X_0)$$

$\mathbf{P}$–*a.s.*, *with*

$$\tilde{K}_{1T}(X_0) \equiv \int_0^T \tilde{K}_s(X_0)\,\mathrm{d}s + \tilde{K}_{2T}(X_0), \tag{60}$$

where $\tilde{K}$ *is given in Theorem* 9, *and where* $\tilde{K}_{2T} = K_{2T}$ *is given in Corollary* 2.

The asymptotic expected errors, and hence the second-order bias, for estimators of Riemann integrals based on the Milshtein approximation, have the same structures as the corresponding quantities for Euler schemes in Corollaries 2 and 3. The first component captures the cumulative expected error of the Milshtein approximation of the state variables in the Riemann integral. The second component is the expected error due to the approximation of the continuous integral by a discrete sum.

Bias-corrected estimators based on the Milshtein scheme are asymptotically equivalent to Monte Carlo estimators obtained by sampling from the true distribution of X_T. As a result they do not suffer from the size distortion problem described earlier. Bias-corrected estimators of conditional expectations can be constructed as

$$\tilde{g}_c^{N,M} = \frac{1}{M}\sum_{i=1}^{M}\left[g(\tilde{X}_{Nh}^{i,N}) + \frac{1}{2}\,\partial g(\tilde{X}_{Nh}^{i,N})\tilde{C}_{Nh}^{i,N}\right], \tag{61}$$

where $\tilde{C}_{Nh}^{i,N}$ is a convergent approximation of $-\tilde{V}_T$, defined in Appendix B. Our next result shows that $\tilde{g}_c^{N,M}$ is a bias-corrected estimator of the conditional expectation based on the Milshtein scheme.

Theorem 10. *Suppose that the conditions of Theorems* 4, 5 *and* 9 *hold. Then*,

$$\sqrt{M}(\tilde{g}_c^{N_M,M} - \mathbf{E}_0[g(X_T)]) \Rightarrow L_T(X_0), \tag{62}$$

$$\sqrt{M}(g_c^{N_M,M} - \hat{g}_c^{N_M,M}) \Rightarrow 0 \tag{63}$$

and

$$\sqrt{M}(g_c^{N_M,M} - \tilde{g}_c^{N_M,M}) \Rightarrow 0 \tag{64}$$

as $M \to \infty$, *with* $\lim_{M\to\infty} N_M = \infty$.

It is important to note that bias-corrected Milshtein does not improve over bias-corrected Euler, even without transformation. Indeed, the theorem shows that all estimators (Milshtein and Euler) are asymptotically equivalent after second-order bias corrections. For non-commutative noise, our explicit expressions for the second-order biases can be used to develop estimators based on the Euler scheme that are computationally superior to the Milshtein scheme because they do not require further subdivisions of discretization intervals to approximate the increments $\Delta F^{l,j}$.

Second-order bias-corrected estimators for conditional expectations of Riemann integrals based on the Milshtein scheme (as in (47)–(48)) are

$$\tilde{f}_c^{N,M} = \frac{1}{M}\sum_{i=1}^{M}\left(\sum_{n=0}^{N-1}\left[g(\tilde{X}_{nh}^{i,N}) + \frac{1}{2}\,\partial g(\tilde{X}_{nh}^{i,N})\tilde{C}_{nh}^{i,N}\right]h + \frac{1}{2}\,\tilde{C}_{1,Nh}^{i,N}\right). \tag{65}$$

As for Euler-based estimators, they involve a new term $\tilde{C}_1$ that corrects the second-order bias due to the approximation of a continuous integral by a discrete sum. An explicit expression can be found in Appendix B. In parallel with Corollary 4, we have:

Corollary 6. *Suppose that the conditions of Theorems* 4 *and* 9 *hold. Then*

$$\sqrt{M}(f_c^{N_M,M} - \tilde{f}_c^{N_M,M}) \Rightarrow 0 \tag{66}$$

as $M \to \infty$, *with* $\lim_{M\to\infty} N_M = \infty$.

Corollary 6 combined with Corollary 4 enable us to conclude that bias-corrected estimators based on the Euler scheme $f_{c}^{N,M}$, the Euler scheme applied to the transformed variables $\hat{f}_{c}^{N,M}$, and the Milshtein scheme $\tilde{f}_{c}^{N,M}$, are asymptotically equivalent.

5. Simulation-based inference for diffusions

Standard dynamic models in finance involve diffusion processes with unknown coefficients. Estimation and statistical inference are therefore essential for implementation purposes. As mentioned in the introduction, the absence of explicit formulas for transition densities of general diffusions implies that maximum likelihood inference can only proceed with the use of an auxiliary numerical procedure designed to approximate the transition density.[24] Our characterizations of asymptotic errors apply directly to these procedures and can be used to find efficient experimental designs for simulation-based inference methods for diffusions.

Section 5.1 sets up a general framework for the estimation of the parameters of a diffusion process using simulation-based methods of moments. A simulation-based version of extended quasi maximum likelihood estimation is presented in Section 5.2. Illustrations of the results are provided in Section 5.3.

5.1. A simulated method-of-moment approach for diffusions

We briefly sketch a general setup for parameter estimation of diffusion processes. Suppose that we observe an equally spaced discrete-time sample $\{Y_0, \dots, Y_L\}$, where $Y_l \equiv Y_{t_l}$ and $\Delta \equiv t_{l+1} - t_l$, of the following diffusion process, supposed to be univariate for notational convenience,

$$\mathrm{d}Y_t = A(Y_t;\theta)\,\mathrm{d}t + B(Y_t;\theta)\,\mathrm{d}W_t \quad \text{and} \quad Y_{t_0} \text{ given.} \tag{67}$$

Assume that coefficients depend on an unknown parameter vector $\theta \in \Psi \subset \mathbb{R}^p$. The transition law of the Markov chain $\{Y_0, \dots, Y_L\}$ is assumed to be absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}^d$,

$$\mathbf{P}^{\theta_0}(Y_{t_{l+1}} \in \mathrm{d}z | \mathscr{F}_{t_l}) \equiv p_{\Delta}^{\theta_0}(Y_l, z)\lambda(\mathrm{d}z), \tag{68}$$

[24] Numerical approximations can be avoided if we let the observation interval decrease to zero. For this type of fill-in asymptotics, efficient estimators have been obtained by Dacunha-Castelle and Florens-Zmirou (1986), Florens-Zmirou (1989, 1993), Genon-Catalot (1990), Genon-Catalot and Jacod (1993) and Bibby and Sørensen, 1995. This type of asymptotics seems less appropriate for non-experimental setups like those encountered in financial econometrics.

where $p_{\Delta}^{\theta_0}(Y_l, z)$ is the transition density. Also suppose that the Markov chain $\{Y_0, \ldots, Y_L\}$ is geometrically ergodic.[25] This last assumption guarantees the existence of a stationary density, denoted $\bar{p}^{\theta_0}(y)$.[26]

We seek to estimate the parameter θ using the conditional constraints:

$$\int_{\mathbb{R}^d} h_{j,\Delta}(Y_l, z; \theta) p_{\Delta}^{\theta}(Y_l, z)\lambda(\mathrm{d}z) = 0 \quad \text{for all } j = 1, \ldots, q. \tag{70}$$

The structure of (70) implies that $h_{\Delta}(Y_l, z; \theta_0) \equiv (h_{j,\Delta}(Y_l, z; \theta_0))_{j=1,\ldots,q}$ is a vector of martingale increments. This suggests estimating θ_0 by solutions $\hat{\theta}_{\Delta}^{L}$ of martingale estimating equations

$$\frac{1}{L}\sum_{l=0}^{L-1} g_{\Delta}(Y_l, Y_{l+1}; \theta) = 0, \tag{71}$$

where $g_{\Delta}(Y_l, Y_{l+1}; \theta) \equiv J_{\Delta}(Y_l)' h_{\Delta}(Y_l, Y_{l+1}; \theta)$ and $J_{\Delta}(Y_l)$ is a $q \times p$ matrix of adapted weights.

In contrast to the estimator above, GMM estimators can be written as M estimators,

$$\tilde{\theta}_{\Delta}^{L} = \arg\min_{\theta} \left\| S_L^{1/2}\left(\frac{1}{L}\sum_{l=0}^{L-1} g_{\Delta}(Y_l, Y_{l+1}; \theta)\right)\right\|^2, \tag{72}$$

where the weight $J_{\Delta}(Y_l)$ is now interpreted as the optimal instrument. The matrix S_L is a weighting matrix that controls the efficiency of the GMM estimator.

It is known from Hansen (1985); Hansen et al. (1988); Chamberlain (1987, 1992); Newey (1990) and Wefelmeyer (1996) that the efficient estimator $\tilde{\theta}_{\Delta}^{L}$ is obtained for instruments given by

$$J_{\Delta}(y) = \Psi_{\Delta}(y)^{-1}\Gamma_{\Delta}(y),$$

$$\Psi_{\Delta}(y) \equiv \int_{\mathbb{R}} h_{\Delta}(y, z; \theta_0)(h_{\Delta}(y, z; \theta_0))' p_{\Delta}^{\theta_0}(y, z)\lambda(\mathrm{d}z),$$

[25]This assumption can be avoided by introducing a stochastic normalization for limit theorems, i.e. a random sequence c_L such that $c_L(\hat{\theta}_{\Delta}^{L} - \theta_0) \Rightarrow Z$ where Z is a standard normal variate and $c_L \to \infty$. Random normalizations arise naturally in the definitions of locally asymptotically mixed normal (LAMN) and locally asymptotically Brownian functional (LABF) estimators. See Basawa and Scott (1982) for more on this. As shown by Heyde (1992) confidence intervals based on a stochastic normalization of the estimator are, in general, shorter than those based on a deterministic normalization.

[26]A Markov chain is geometrically ergodic if there exists a constant $r > 1$ such that

$$\sum_{k=1}^{\infty} r^k \left(\left\| \int_B p_{\Delta}^{\theta_0}(Y_l, z)\lambda(\mathrm{d}z) - \int_B \bar{p}^{\theta_0}(z)\lambda(\mathrm{d}z) \right\|_V \right)^k < \infty \quad \text{for all } B \in \mathbb{R}, \tag{69}$$

where $\|\cdot\|_V$ is the variational norm (see Meyn and Tweedie, 1993). See Duffie and Singleton (1993) for an application of this in a discrete time model. General convergence results for Markov chains appear in Meyn and Tweedie (1993).

$$\Gamma_\Delta(y) \equiv \int_{\mathbb{R}} (\partial_\theta h_\Delta(y,z;\theta_0))' p_\Delta^{\theta_0}(y,z)\lambda(\mathrm{d}z).$$

Our next result shows that the optimal weight S_L does not play any role if optimal instruments are used. This follows from the first-order asymptotic equivalence of $\tilde{\theta}_\Delta^L$ and $\hat{\theta}_\Delta^L$.

Theorem 11. *In an ergodic model, the efficient estimator $\hat{\theta}_\Delta^L$ has the asymptotic error distribution*

$$\sqrt{L}(\hat{\theta}_\Delta^L - \theta_0) \Rightarrow \Sigma_\Delta^{-1/2} Z \tag{73}$$

as $L \to \infty$, where

$$\Sigma_\Delta \equiv \left(\int_{\mathbb{R}} \Gamma_\Delta(y)' \Psi_\Delta(y)^{-1} \Gamma_\Delta(y) \bar{p}^{\theta_0}(y)\lambda(\mathrm{d}y) \right), \tag{74}$$

and $Z \sim N(0, I_p)$ is a standard normal random variable vector. Furthermore,

$$\sqrt{L}(\hat{\theta}_\Delta^L - \tilde{\theta}_\Delta^L) \Rightarrow 0, \tag{75}$$

as $L \to \infty$.

A similar result for d-order Markov chains can be found in Müller and Wefelmeyer (2002) (see also references therein). They show that the martingale estimating function framework covers GMM, generalized quasi-likelihood, extended quasi-likelihood, and quasi-likelihood estimations. Our result simply expresses the fact that the GMM estimator with optimal instruments is just identified and consequently can be obtained as the solution of (71).

If the transition density $p_\Delta^\theta(y,z)$ is known, the estimation of the unknown parameters based on (70) is equivalent to the estimation of unknown parameters in a Markovian ergodic time series model.[27] It is of particular interest to note that the discrete nature of observations has no impact on the inference procedure.

Unfortunately, the transition density $p_\Delta^\theta(y,z)$ is usually unknown or the optimal weights $J(y)$ cannot be calculated in closed form. The corresponding optimal estimating function cannot therefore be used as it stands for estimating θ, i.e. the efficient estimator is infeasible. Similarly, for moment-based constraints, indirect inference or EMM estimators, the functions $h_{j,\Delta}(Y_l, Y_{l+1};\theta)$ that identify the parameters through (70), are not known in explicit form and are obtained by Monte Carlo simulation.[28]

[27] The particular choice $h_{j,\Delta}(y,z;\theta) = \partial_{\theta_j} p_\Delta^\theta(y,z)$ leads to a maximum likelihood estimator which attains the Cramér–Rao lower bound $(\int_{\mathbb{R}^d} \Psi_\Delta(y)\bar{p}^{\theta_0}(y)\lambda(\mathrm{d}y))^{-1}$.

[28] See Gallant and Tauchen (2002) for a survey. The moment-based estimating functions h_j are often of the form $h_{j,\Delta}(Y_l, Y_{l+1};\theta) = f_j(Y_l, Y_{l+1}) - \int_{\mathbb{R}^d} f_j(Y_l, z) p_\Delta^\theta(Y_l, z)\lambda(\mathrm{d}z)$. The functions h_j cannot be obtained in closed form, if the transition density is unknown or the Riemann integral cannot be solved explicitly, as it is often the case in applications. Similarly, for indirect inference or EMM the functions h_j are associated with the score function of an auxiliary model, where the parameters of the auxiliary model are replaced by their estimates based on the sample of data and where the structural processes are simulated for given values of the structural parameters. Estimates of the latter are the values that set the moment conditions as close to zero as possible based on a suitable GMM metric.

In a diffusion setting, simulations and numerical solution techniques for SDEs can be used to approximate the optimal weights J and/or the functions h_j in the constraints that identify the parameters.[29] For simulation-based estimators computations can be performed by first discretizing time and using the Euler or the Milshtein schemes to approximate the solution of the stochastic differential equation, and then replicating this approximation M times to generate M independent realizations of the terminal point $\{Y_{l+1}^{i,N}(Y_l) : i = 1, \ldots, M\}$. The integral in the expression for h_j and/or J can then be replaced by an empirical mean over independent replications given the initial value.

If the transition density is unknown, the estimator $\hat{\theta}^{L,M,N}$ is obtained from a simulation-based martingale estimating function, i.e. as the unique root of

$$\frac{1}{L}\sum_{l=0}^{L-1} g_{\Delta}^{M,N}(Y_l, Y_{l+1}; \hat{\theta}^{L,M,N}) \equiv \frac{1}{L}\sum_{l=0}^{L-1} J_{\Delta}^{M,N}(Y_l) h_{\Delta}^{M,N}(Y_l, Y_{l+1}; \hat{\theta}^{L,M,N}) = 0. \tag{76}$$

The results of this paper can be used to study the asymptotic properties of the normalized error $\sqrt{M}(g_{\Delta}^{M,N} - g_{\Delta})$ resulting from this procedure. In particular, we now show that an approximation of the true estimating function based on M independent Monte Carlo replications with N discretization points introduces a second-order bias. As a result simulation-based estimators have higher-order asymptotic properties that differ from those of their infeasible optimal counterparts. The second-order bias has important implications for the construction of asymptotic confidence intervals and for parametric hypothesis testing.

To describe the second-order bias of the simulation-based estimator, introduce the expression,

$$\kappa_{j,\Delta}(\theta_0) \equiv \int_{\mathbb{R}\times\mathbb{R}} K_{j,\Delta}(y; y, z, \theta_0) p_{\Delta}(y, z; \theta_0) \lambda(\mathrm{d}z) \bar{p}^{\theta_0}(y) \lambda(\mathrm{d}y) \quad \text{for } j = 1, \ldots, p, \tag{77}$$

where $K_{j,\Delta}(y; y, z, \theta_0)$ is defined in Theorem 4 (the notation emphasizes that the function being averaged over depends, for simulation-based estimating functions, on the observations Y_l and Y_{l+1}. As in Theorem 4, the first argument corresponds to the starting point of the simulations Y_l, i.e., the state for which the conditional expectation is calculated).[30]

Our next result shows that the simulation procedure introduces an additional second-order bias.

Theorem 12. *Assume that the conditions of Theorem* 4, *for the Euler scheme without transformation, are satisfied. Then, for geometrically ergodic diffusions observed over*

[29] Alternative kernel-based estimation methods for conditional expectations converge at a slower rate ($L^{-2/(d+4)}$) that depends on the dimension of the diffusion. Therefore, for multivariate diffusions, a large sample is necessary to get reasonably accurate estimates of the weights and/or the functions h_j. Simulation-based estimators are particularly interesting in those cases.

[30] For MCE estimators these observations do not affect the convergence properties.

time intervals of equal fixed length Δ, we have, as $L \to \infty$,

$$\sqrt{L}(\hat{\theta}_{\Delta}^{L,M_L,N_L} - \hat{\theta}_{\Delta}^{L}) \Rightarrow -\varepsilon_1\varepsilon_2\Sigma_{\Delta}(\theta_0)^{-1}\kappa_{\Delta}(\theta_0), \quad (78)$$

where $\varepsilon_1 = \lim_{L\to\infty}\sqrt{L}/\sqrt{M_L}$ *and* $\varepsilon_2 = \lim_{L\to\infty}\sqrt{M_L}/N_L$, *with* $\lim_{L\to\infty} M_L = \lim_{L\to\infty} N_L = \infty$.

Corresponding estimators for the Euler scheme with Doss transformation (under the assumptions of Theorem 5) and the Milshtein scheme (under the assumptions of Theorem 9), are obtained if $K_{j,\Delta}(\cdot, y, z; \theta_0)$ in (77) is replaced by $\hat{K}_{j,\Delta}(\cdot, y, z; \theta_0)$ of Theorem 5 and $\tilde{K}_{j,\Delta}(\cdot, y, z; \theta_0)$ of Theorem 9, both adjusted for the dependence of h_j and J on the observations $(Y_l, Y_{l+1}) = (y, z)$.

Theorem 12 shows that the efficient simulated estimator has a second-order bias that depends on the discretization scheme chosen. The numerical scheme used to solve the SDE becomes irrelevant only if $\varepsilon_1\varepsilon_2\kappa_{\Delta}(\theta_0) = 0$. The conditions stating that ε_1 and ε_2 are finite, which are conditions linking the sample size L, the number of replications M_L and the number of discretization points N_L, are therefore necessary. If these conditions fail, the growth of L has to be restricted, and the resulting estimators become inefficient. Ignoring the second-order bias leads to size-distorted statistical tests.

The size of the second-order bias depends on the choice of the discretization scheme only through the functions K, $\hat{K}$ and $\tilde{K}$ that characterize the expected approximation errors of the different methods. This follows because the speed of convergence of the expected error is the same for all discretization schemes. We also see that the second-order bias is inversely related to the asymptotic variance of the estimator. Variance reduction, without second-order bias correction, will therefore increase the size distortion of asymptotic confidence intervals and hypothesis tests.

It is also important to note that the results of Theorem 12 do not directly cover simulated maximum likelihood estimators obtained from the numerical integration of the Chapman–Kolmogorov equation (see Pedersen, 1995a,b; Brandt and Santa-Clara, 2002). For an estimator of this type, the Monte Carlo integration of the true transition density kernel is based on a Gaussian kernel approximation with bandwidth $1/N$ determined by the number of discretization points N. The resulting score function depends on the time discretization parameter N not just through the simulated discretized trajectories but also through its functional form. As described by Milshtein et al. (2004), this estimator of the transition density can be interpreted as a kernel estimator based on simulated i.i.d. data. It is well known that such an estimator will converge at a rate slower than $M^{-1/2}$. In fact the rate for the score is $M^{-2/(d+4)}$, a quantity that depends on the dimension of the diffusion. Optimality for a simulated maximum likelihood estimator requires that $\sqrt{L}/M_L^{2/(d+4)} \to \varepsilon_1$ and $M_L^{2/(d+4)}/N_L \to \varepsilon_2$. It follows that efficient simulated maximum likelihood estimators based on the numerical integration of the Chapman–Kolmogorov equation are less efficient than maximum likelihood estimators, even in the univariate case.

Theorem 12 shows that estimators with $\varepsilon_1\varepsilon_2 = 0$ are inefficient. If L, M_L, N_L are chosen such that $\varepsilon_1\varepsilon_2 = 0$, L has to grow at a slower rate than if $0 < \varepsilon_1\varepsilon_2 < \infty$. It follows that the asymptotic variance and therefore the lengths of confidence intervals are larger than the variance and length of the asymptotic confidence interval of the efficient estimator, as L is too small compared to the efficient estimator. The efficient estimator requires quadrupling the sample length, quadrupling the number of simulations and doubling the number of discretization points, in order to cut the asymptotic variance and therefore the length of the confidence interval in half.

Corresponding optimal designs for the simulated maximum likelihood estimators of Pedersen (1995a,b) and Brandt and Santa-Clara (2002), depend in addition on the dimension of the diffusion. These authors assume $\varepsilon_2 = 0$. This assumption, unfortunately, is not sufficient. To preclude an exploding second-order bias it must also be the case that $\sqrt{L}/\sqrt{M_L}$ does not diverge faster than $\sqrt{M_L}/N_L$ vanishes. The rate of divergence of M_L can therefore not be chosen independent of L.

For estimators with a fixed number of discretization points and a fixed number of Monte Carlo simulations the second-order bias always explodes when the sample size becomes large. Asymptotic confidence intervals will then cover the true values with null probability. Hypothesis tests become invalid as their effective size can become considerably smaller than the nominal significance level of the test.

The results in Theorems 4, 5 and 9 describing the second-order biases for the Euler schemes with and without Doss transformation and for the Milshtein scheme, enable us to construct second-order bias-corrected estimators that are both efficient and have asymptotic confidence intervals that do not suffer from size distortion. It is important to note that asymptotic efficiency (i.e. estimators with the shortest asymptotic confidence intervals) can only be achieved by estimators that correct for second-order bias. Non-corrected estimators require a refinement of the time discretization in the simulation of SDEs as the sample size increases. As a consequence, the computational cost of efficient estimators without second-order bias corrections explodes. Second-order bias-corrected estimators, alone, have the same asymptotic distributions as estimators obtained from the infeasible optimal estimating function.[31]

5.2. *Simulated extended quasi-maximum likelihood estimator*

We now present a simulation-based version of the extended quasi-maximum likelihood estimator (EQMLE) introduced for Markovian semi-martingales by Wefelmeyer (1996). This estimator is a special case of a moment-based estimator. It is based on parameter restrictions

$$a(Y_l;\theta) \equiv \mathbf{E}\left[\int_{t_l}^{t_{l+1}} A(Y_s;\theta)\,\mathrm{d}s \,\middle|\, Y_l\right], \tag{79}$$

[31]This does not imply that simulation-based inference methods without second-order bias correction are inferior to estimators based on analytic orthogonal series expansions of the likelihood, such as those proposed by Ait-Sahalia (2002). Explosions of the second-order biases associated with the efficient estimators of this type can only be avoided by increasing the dimensionality of the orthogonal basis as the sample increases. Second-order biases for such estimators are unknown.

$$b(Y_l;\theta) \equiv \mathbf{E}\left[\int_{t_l}^{t_{l+1}} B(Y_s;\theta)B(Y_s;\theta)'\,\mathrm{d}s \middle| Y_l\right] \tag{80}$$

implied by the drift and volatility functions of the diffusion process. The moment conditions associated with

$$h_{1,\Delta}(Y_l, Y_{l+1};\theta) \equiv Y_{l+1} - Y_l - a(Y_l;\theta), \tag{81}$$

$$h_{2,\Delta}(Y_l, Y_{l+1};\theta) \equiv (Y_{l+1} - Y_l - a(Y_l;\theta))^2 - b(Y_l,\theta), \tag{82}$$

together with the optimal instruments $J_\Delta(Y_l)$ are then used to obtain an estimating function, as in the previous section, based on

$$g_i(Y_l, Y_{l+1};\theta) = J_\Delta(Y_l)h_i(Y_l, Y_{l+1};\theta) \quad \text{for } i = 1, 2. \tag{83}$$

Such an estimator is labelled *extended quasi-maximum likelihood estimator* (EQMLE). This estimator is the most efficient in the class of estimating functions based on h_i and is asymptotically equivalent to the GMM estimator based on the optimal instruments (see Theorem 11).

The EQMLE is infeasible if both the drift and volatility functions (a, b) are not available in closed form. This is the case for general diffusions for which the transition density is unknown or the expectation cannot be obtained in closed form. In this case calculation of the drift and volatility functions can proceed by simulation. The resulting estimator, which uses the same optimal weights as the EQMLE, is called the *simulated extended quasi-maximum likelihood estimator* (SEQMLE). The asymptotic error distribution of this estimator can be studied within our theoretical framework. We also design estimators that correct for second-order bias. They are called *simulated*, *bias-corrected*, *extended quasi-maximum likelihood estimator* (SBCEQMLE). The next section presents results for both estimators for CIR and L-CEV processes.

5.3. An application of the simulated quasi-maximum likelihood estimator to CIR and L-CEV processes

We perform a Monte Carlo study to analyze the properties of the simulation-based estimators associated with various discretization and transformation schemes introduced in the previous sections. The candidate methods are the Euler discretization scheme, with and without the Doss transformation, and the Milshtein discretization scheme. The Monte Carlo setup is inspired by the study of Chan et al. (1992), also used by Ait-Sahalia (1999). We simulate 500 series of 306 monthly observations and compute our estimators, with and without bias correction (SEQMLE and SBCEQMLE) for the three methods just mentioned. For comparison purposes, we also report the results obtained with maximum likelihood estimators based on the Euler approximation (ML-Euler) and on the approximations of the transition densities proposed by Ait-Sahalia (1999) (ML-YAS) with Hermite polynomials of order one ($J = 1$) and two ($J = 2$). The series in our setup are short and the models considered are non-linear. The estimation task is therefore

Table 1
RMSE and Average Bias of SEQMLE and SBCEQMLE estimators for CIR process

CIR: $\mathrm{d}Y_t = \kappa(\bar{Y} - Y_t)\,\mathrm{d}t + \sigma\sqrt{Y_t}\,\mathrm{d}W_t$
True values: $\kappa = 0.5$, $\bar{Y} = 0.06$, $\sigma = 0.072$

Method	$\bar{Y}$	κ	σ
RMSE (500 trials)			
MLE-Euler	0.0073	0.2778	0.0032
MLE-YAS ($J = 1$)	0.0073	0.2945	0.0029
MLE-YAS ($J = 2$)	0.0073	0.3011	0.0029
SEQMLE-Euler	0.0108	0.2740	0.0047
SEQMLE-Doss	0.0108	0.2740	0.0047
SEQMLE-Milshtein	0.0108	0.2740	0.0047
SBCEQMLE-Euler	0.0109	0.2867	0.0048
SBCEQMLE-Doss	0.0109	0.2866	0.0048
SBCEQMLE-Milshtein	0.0109	0.2867	0.0048
Average bias (500 trials)			
ML-Euler	−0.0002	0.1342	−0.0016
MLE-YAS ($J = 1$)	−0.0002	0.1502	0.0002
MLE-YAS ($J = 2$)	−0.0002	0.1544	0.0002
SEQMLE-Euler	0.0043	0.0962	−0.0039
SEQMLE-Doss	0.0043	0.0962	−0.0039
SEQMLE-Milshtein	0.0043	0.0962	−0.0039
SBCEQMLE-Euler	0.0044	0.1085	−0.0040
SBCEQMLE-Doss	0.0044	0.1084	−0.0040
SBCEQMLE-Milshtein	0.0044	0.1085	−0.0040

difficult and standard errors of estimators will generally be high, especially for the L-CEV process.

In the simulation-based methods, the discretization parameter $N = 2^4$ and the number of replications $M = 500$ per observation. For the optimization procedure, we used quasi-Newton methods for computing the SEQMLE and SBCEQMLE, but a simplex method for the maximum likelihood estimators. The choice of the simplex method was motivated by the difficulties we experienced by using gradient-based optimization methods for the estimators based on the Hermite polynomials approximations (ML-YAS). The calculation of numerical derivatives using finite differences appeared numerically unstable and optimization in a Monte Carlo context with gradient-based methods proved simply infeasible. We made sure that the simplex method provided reasonable results by comparing the estimation results with the actual series of short-term interest rates used in Ait-Sahalia (1996). We reproduced the results reported in Table VI of Aït-Sahalia, and verified that the simplex method and a gradient-based method were giving very similar results in terms of likelihood value and parameter estimates.

Table 1 displays the results for the CIR process. Algorithms for the various estimators were started with the same initial values. These are obtained based on the

Table 2
RMSE and Average Bias of SEQMLE and SBCEQMLE estimators for L-CEV process

L-CEV: $dY_t = \kappa(\bar{Y} - Y_t)\,dt + \sigma \min\{Y_t, Y^\star\}^\gamma \,dW_t$
True values: $\kappa = 0.5$, $\bar{Y} = 0.06$, $\sigma = 1.2$, $\gamma = 1.5$ $Y^\star = 10^{10}$
($Y^\star$ guarantees that $\int_0^\cdot \min\{Y_s, Y^\star\}^\gamma \,dW_s$ remains a martingale for $\gamma > 1$)

Method	$\bar{Y}$	κ	σ	γ
RMSE (500 trials)				
ML-Euler	0.0131	0.3124	0.5828	0.1905
MLE-YAS ($J = 1$)	0.0133	0.3232	0.6815	0.1807
SEQMLE-Euler	0.0433	0.3130	0.6173	0.1823
SEQMLE-Doss	0.0438	0.2965	0.6292	0.1822
SEQMLE-Milshtein	0.0432	0.3133	0.6169	0.1822
SBCEQMLE-Euler	0.0418	0.3221	0.5905	0.1699
SBCEQMLE-Doss	0.0460	0.3067	0.5958	0.1765
SBCEQMLE-Milshtein	0.0417	0.3222	0.5925	0.1744
Average bias (500 trials)				
ML-Euler	0.0013	0.1336	−0.0679	−0.0573
MLE-YAS ($J = 1$)	0.0015	0.1448	0.1014	−0.0149
SEQMLE-Euler	0.0243	−0.0357	0.0264	−0.0071
SEQMLE-Doss	0.0249	−0.0495	0.0320	−0.0071
SEQMLE-Milshtein	0.0241	−0.0346	0.0263	−0.0071
SBCEQMLE-Euler	0.0243	−0.0279	0.0151	0.0046
SBCEQMLE-Doss	0.0260	−0.0411	0.0284	0.0077
SBCEQMLE-Milshtein	0.0242	−0.0274	0.0165	0.0047

first-order autocorrelation coefficient for the speed of mean reversion, the unconditional sample mean for the long-term mean and the square-root of the unconditional sample variance divided by the sample mean for the volatility parameter. The results differ across parameters. For the long-term mean, the results are better for the MLE methods both in terms of RMSE and average bias. For the mean reversion parameter, the results of all methods are of the same order of magnitude for the RMSE, but both simulation-based methods, with and without bias correction, reduce the bias with respect to the MLE methods. Finally, for the volatility parameter, the results appear to be slightly better for the MLE methods, both in terms of RMSE and average bias. For all parameters, the RMSE and the average bias obtained with the bias correction methods (SBCEQMLE) are, if anything slightly larger than without bias correction. As discussed below, the effects of bias corrections are more evident for the L-CEV process.

Table 2 provides the estimation results for the L-CEV process.[32] Initial values for all estimators have been obtained by regressing observed increments on a constant

[32]The second-order approximation is not included because of severe numerical difficulties in maximizing this function. Note that Ait-Sahalia (1999) does not report results with the second-order approximation for estimating the parameters of the CEV model with interest rate data in Table VI.

and the level of the process for the parameters of the drift function, and by regressing the logarithm of the observed increments of the quadratic variation on a constant and the level of the process for the diffusion function. In terms of RMSE, the order of magnitude is similar for MLE and simulation-based methods except for the long-term mean and the volatility parameter γ. For the long-term mean, the MLE methods perform better, while for γ the RMSE is smallest for the simulation-based methods with bias correction (SBCEQMLE). However, for the average bias, simulation-based methods improve results by a significant margin with respect to the MLE methods, with the exception again of the long-term mean. The bias correction also appears sizable for κ, σ and γ with respect to the simulation-based methods without correction. Contrary to the results for the CIR process, the second-order bias-corrected estimates dominate in general estimates without bias correction in terms of average bias. If we compare the results between the various simulation-based methods, there are no differences for the CIR process, while the results appear to be better for Euler and Mihlstein than for Doss. In Durham and Gallant (2002), the simulation results were showing that the second-order bias was smallest if the Doss transformation was used.

6. Conclusion

Error analysis for Monte Carlo estimators of numerical solutions of SDEs is a difficult task. The present paper introduced key elements to perform this analysis. Our investigation produced explicit formulas for error distributions that can be used to construct asymptotically valid confidence intervals and to assess the asymptotic errors of Euler schemes, with and without Doss transformation, and of the Milshtein scheme. Numerical experiments showed that the error distributions differ significantly depending on whether the transformation is applied or not. Bias corrections were found to be important in the case of non-linear state variable processes, such as the L-CEV process. This underscores the importance of the error statistics obtained, as processes of relevance in financial economics often have a non-linear structure.

The characterizations obtained here permit a rigorous analysis of the asymptotic error properties of a given estimator. They are therefore crucial for the design of efficient Monte Carlo estimators of diffusion processes. Feasible efficient simulated estimating function estimators were shown to require second-order bias corrections. Our explicit expressions for the second-order biases can be used to derive such estimators.

In principle, error distributions, second-order biases and bias-corrected conditional estimators can also be derived when the conditional expectations to be estimated depend on unknown state variables. This situation is often encountered in financial applications such as stochastic volatility models. The convergence results of simulation schemes presented could serve as a starting point for the study of simulation-based estimators for such latent variable models.

Acknowledgements

The paper was presented at the CIRANO conference on Mathematical Finance and Econometrics (June 2001), the CIRANO conference on Monte Carlo Methods in Finance (March 2002), the Bachelier Society Meetings in Crete (June 2002), and the Econometric Society North American Winter Meeting in San Diego (January 2004). Funding from MITACS is gratefully acknowledged. René Garcia also thanks Hydro-Québec and the Bank of Canada for financial support.

Appendix A. Proofs

We start with a series of auxiliary lemmas that are used to prove the main results of the paper. These lemmas give the weak limits of the components in the error expansions of the estimators (102) and (103).

Let $[Nt]$ be the integer part of Nt and define the time change $\eta_t^N = [Nt]/N$ if $Nt \notin \mathbb{N}$ and $\eta_t^N = t - 1/N$ otherwise. Note that $\lim_{N\to\infty} \eta_t^N = t$. We have

Lemma 1. *The following weak convergence results hold*:

$$V_t^{1,N} \equiv N \int_0^t (s - \eta_s^N)\, \mathrm{d}s \Rightarrow \frac{1}{2}\, t \equiv V_t^1, \tag{84}$$

$$V_t^{2,i,N} \equiv N \int_0^t (W_s^i - W^i \circ \eta_s^N)\, \mathrm{d}s \Rightarrow \frac{1}{2}\, W_t^i + \frac{1}{\sqrt{12}}\, Z_t^i \equiv V_t^{2,i}, \tag{85}$$

$$V_t^{3,i,N} \equiv N \int_0^t (s - \eta_s^N)\, \mathrm{d}W_s^i \Rightarrow \frac{1}{2}\, W_t^i - \frac{1}{\sqrt{12}}\, Z_t^i \equiv V_t^{3,i}, \tag{86}$$

$$V_t^{4,i,j,N} \equiv \sqrt{N} \int_0^t (W_s^i - W^i \circ \eta_s^N)\, \mathrm{d}W_s^j \Rightarrow \frac{1}{\sqrt{2}}\, Z_t^{i,j} \equiv V_t^{4,i,j} \tag{87}$$

as $N \to \infty$, *where* $((W^i)_{i\in\{1,\dots,d\}}, (Z^i)_{i\in\{1,\dots,d\}}, (Z^{i,j})_{i,j\in\{1,\dots,d\}})$ *is a* $(2d + d^2)$*-dimensional standard Brownian motion.*

Proof. As $\eta_t^N = [Nt]/N$ for $Nt \notin \mathbb{N}$ we have $\int_0^t (s - \eta_s^N)\, \mathrm{d}s = (1/N^2) \int_0^{Nt} (s - [s])\, \mathrm{d}s$. Using $N \int_0^t (s - \eta_s^N)\, \mathrm{d}s = (1/N) \sum_{k=1}^{[Nt]} \int_{[k-1,k[} (s - [s])\, \mathrm{d}s + (1/N) \int_{[Nt]}^{Nt} (s - [s])\, \mathrm{d}s$ we then obtain

$$N \int_0^t (s - \eta_s^N)\, \mathrm{d}s = \frac{1}{2} \frac{[Nt]}{N} + \frac{1}{2} \frac{1}{N} (Nt - [Nt])^2 \to \frac{1}{2}\, t \tag{88}$$

as $N \to \infty$. The limit on the right-hand side uses $[s] = k - 1$ for $s \in [k-1, k[$ and the bound $Nt - [Nt] \leqslant 1$.

Similarly, it can be shown, using the scaling property of Brownian motion that

$$N\int_0^t (W_s^i - W^i\circ\eta_s^N)\,\mathrm{d}s = \frac{1}{\sqrt{N}}\sum_{k=1}^{[Nt]}\int_{[k-1,k[}(W_s^i - W_{[s]}^i)\,\mathrm{d}s + \frac{1}{\sqrt{N}}\int_{[Nt]}^{Nt}(W_s^i - W_{[s]}^i)\,\mathrm{d}s. \tag{89}$$

Itô's lemma then implies $\int_{[k-1,k[}(W_s^i - W_{[s]}^i)\,\mathrm{d}s = \int_{k-1}^k (k-s)\,\mathrm{d}W_s^i$ so that $N\int_0^t(W_s^i - W^i\circ\eta_s^N)\,\mathrm{d}s = (1/\sqrt{N})\sum_{k=1}^{[Nt]}\int_{[k-1,k[}(k-s)\,\mathrm{d}W_s^i + (1/\sqrt{N})\int_{[Nt]}^{Nt}(W_s^i - W_{[s]}^i)\,\mathrm{d}s$. Note that the sequence of i.i.d. random variables $\int_{[k-1,k[}(k-s)\,\mathrm{d}W_s^i$ has variance $\frac{1}{3}$ and covariance $\frac{1}{2}1_{\{i=j\}}$ with the Brownian motion W^j. Donsker's functional central limit theorem (see Kallenberg, 1997, Theorem 12.9, p. 225) then gives

$$\frac{1}{\sqrt{N}}\sum_{k=1}^{[Nt]}\int_{[k-1,k[}(k-s)\,\mathrm{d}W_s^i \Rightarrow \frac{1}{2}\,W_t^i + \frac{1}{\sqrt{12}}\,Z_t^i, \tag{90}$$

where Z^i is a standard Brownian motion independent of W^j for all $j\in\{1,\dots,d\}$. This establishes (85) because of the continuity of the pathwise integral with respect to the Lebesgue measure,

$$\mathbf{P}-\lim_{N\to\infty}\frac{1}{\sqrt{N}}\int_{[Nt]}^{Nt}(W_s^i - W_{[s]}^i)\,\mathrm{d}s = 0. \tag{91}$$

The same type of argument shows that

$$N\int_0^t (s-\eta_s^N)\,\mathrm{d}W_s^i = \frac{1}{\sqrt{N}}\sum_{k=1}^{[Nt]}\int_{[k-1,k[}(s-[s])\,\mathrm{d}W_s^i + \frac{1}{\sqrt{N}}\int_{[Nt]}^{Nt}(s-[s])\,\mathrm{d}W_s^i. \tag{92}$$

Donsker's functional central limit theorem implies that the first part converges to a Brownian motion. The second part converges to zero, in probability, by the continuity of the Wiener integral,

$$\mathbf{P}-\lim_{N\to\infty}\frac{1}{\sqrt{N}}\int_{[Nt]}^{Nt}(s-[s])\,\mathrm{d}W_s^i = 0. \tag{93}$$

As the sequence of i.i.d. random variables $\int_{[k-1,k[}(s-[s])\,\mathrm{d}W_s^i$ has variance $\frac{1}{3}$ and covariance $\frac{1}{2}1_{\{i=j\}}$ with W^j as well as covariance $\frac{1}{6}1_{\{i=j\}}$ with $\int_{[k-1,k[}(k-s)\,\mathrm{d}W_s^j$ we have

$$\frac{1}{\sqrt{N}}\sum_{k=1}^{[Nt]}\int_{[k-1,k[}(s-[s])\,\mathrm{d}W_s^i \Rightarrow \frac{1}{2}\,W_t^i - \frac{1}{\sqrt{12}}\,Z_t^i, \tag{94}$$

which establishes (86). It remains to show (87). Again, by the scaling property of Brownian motion

$$\sqrt{N}\int_0^t (W_s^j - W^j \circ \eta_s^N)\,\mathrm{d}W_s^i = \frac{1}{\sqrt{N}}\sum_{k=1}^{[Nt]}\int_{[k-1,k[}(W_s^j - W_{[s]}^j)\,\mathrm{d}W_s^i + \frac{1}{\sqrt{N}}\int_{[Nt]}^{Nt}(W_s^j - W_{[s]}^j)\,\mathrm{d}W_s^i. \tag{95}$$

As the sequence of i.i.d random variables $\int_{[k-1,k[}(W_s^j - W_{[s]}^j)\,\mathrm{d}W_s^i$ has variance $\frac{1}{2}$ and is independent of W^j, $\int_{[k-1,k[}(W_s^i - W_{[s]}^i)\,\mathrm{d}s$ as well as of $\int_{[k-1,k[}(s-[s])\,\mathrm{d}W_s^i$ we can appeal to Donsker's invariance principle to conclude that

$$\frac{1}{\sqrt{N}}\sum_{k=1}^{[Nt]}\int_{[k-1,k[}(W_s^j - W_{[s]}^j)\,\mathrm{d}W_s^i \Rightarrow \frac{1}{\sqrt{2}}\,Z_t^{i,j}. \tag{96}$$

The continuity of the Itô integral implies that $\mathbf{P}-\lim_{N\to\infty}(1/\sqrt{N})\int_{[Nt]}^{Nt}(W_s^j - W_{[s]}^j)\,\mathrm{d}W_s^i = 0$. This establishes (87). □

In the sequel, we need to assess the convergence of stochastic integrals with respect to the processes $V^N \equiv [V^{1,N}, V^{2,N}, V^{3,N}, V^{4,N}]$ defined in Lemma 1. Results of Duffie and Protter (1992), that provide sufficient conditions for the weak convergence of stochastic integrals, i.e. goodness, prove useful in that regard.

Definition 1. A process Y^N is good if $(X^N, Y^N) \Rightarrow (X, Y)$ implies that $(X^N, Y^N, \int_0^{\cdot} X_s^N\,\mathrm{d}Y_s^N) \Rightarrow (X, Y, \int_0^{\cdot} X_s\,\mathrm{d}Y_s)$.

Lemma 2. *The semimartingales* $(V^{1,N}, V^{3,N}, V^{4,N})$ *are good.*

Proof. From condition A in Duffie and Protter (1992) it follows that $V^{1,N}$ is good if $\sup_N N\int_0^t (s-\eta_s^N)\,\mathrm{d}s < \infty$ for all $t \in [0,T]$. Similarly, $V^{3,N}$ and $V^{4,N}$ are good if $\sup_N \mathbf{VAR}[V_t^{3,i,N}] < \infty$ for all $i = 1,\ldots,d$ and $\sup_N \mathbf{VAR}[V_t^{4,i,j,N}] < \infty$ for all $i,j = 1,\ldots,d$ and $t \in [0,T]$. But,

$$\mathbf{VAR}[V_t^{3,i,N}] = N^2\int_0^t (s-\eta_s^N)^2\,\mathrm{d}s, \tag{97}$$

$$\mathbf{VAR}[V_t^{4,i,j,N}] = N\mathbf{E}\left[\int_0^t (W_s^i - W_{\eta_s^N}^i)^2\,\mathrm{d}s\mathbf{1}_{i=j}\right] = N\int_0^t (s-\eta_s^N)\,\mathrm{d}s\mathbf{1}_{i=j}. \tag{98}$$

It is therefore sufficient to show that $N^{\varepsilon}\int_0^t (s-\eta_s^N)^{\varepsilon}\,\mathrm{d}s < \infty$ for $\varepsilon = 1, 2$ and $t \in [0,T]$. The bound $1 > (t-[t])^{\varepsilon} \geq 0$ implies

$$N^{\varepsilon}\int_0^t (s-\eta_s^N)^{\varepsilon}\,\mathrm{d}s = \frac{1}{N}\int_0^{Nt}(s-[s])^{\varepsilon}\,\mathrm{d}s < \frac{1}{N}\int_0^{Nt}\mathrm{d}s = t. \tag{99}$$

We conclude that $V^{1,N}, V^{3,N}, V^{4,N}$ are good. □

The proof of our next lemma shows that the total variation of the semi-martingale $V^{2,N}$ is $\mathbf{O_P}(\sqrt{N})$.[33] As a result $V^{2,N}$ cannot be good.

Lemma 3. *The semi-martingale $V^{2,N}$ is not good.*

Proof. We will show that $N^{-1/2}N\int_0^t |W_s - W_{\eta_s^N}|\,\mathrm{d}s = \mathbf{O_P}(1)$ so that the total variation of $V^{2,N}$ satisfies $\int_0^t |\mathrm{d}V_s^{2,N}| = \mathbf{O_P}(\sqrt{N})$. Note that

$$N^{1/2}\int_0^t |W_s - W_{\eta_s^N}|\,\mathrm{d}s = \frac{1}{N}\int_0^{Nt} |W_s - W_{[s]}|\,\mathrm{d}s$$
$$= \frac{1}{N}\sum_{k=0}^{[Nt]-1}\int_k^{k+1} |W_s - W_k|\,\mathrm{d}s + \frac{1}{N}\int_{[Nt]}^{Nt} |W_s - W_{[Nt]}|\,\mathrm{d}s. \tag{100}$$

By the continuity of the integral the second term $(1/N)\int_{[Nt]}^{Nt} |W_s - W_{[Nt]}|\,\mathrm{d}s = \mathbf{o_P}(1)$. For the first term, note that, by Jensen's inequality,

$$\frac{1}{N^2}\sum_{k=0}^{[Nt]-1}\mathbf{E}\left[\left(\int_k^{k+1} |W_s - W_k|\,\mathrm{d}s\right)^2\right] \leqslant \frac{1}{N^2}\sum_{k=0}^{[Nt]-1}\int_k^{k+1}(s-k)\,\mathrm{d}s$$
$$= \frac{1}{N}\frac{1}{2}\frac{[Nt]}{N} \to 0, \tag{101}$$

as $[Nt]/N \to t$ as $N \to \infty$. By (100), (101) and invoking Chebyshev's weak law of large numbers, we conclude[34]

$$\mathbf{P} - \lim_{N\to\infty} N^{-1/2}N\int_0^t |W_s - W_{\eta_s^N}|\,\mathrm{d}s$$
$$= \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{[Nt]-1}\mathbf{E}\left[\int_k^{k+1} |W_s - W_k|\,\mathrm{d}s\right]$$
$$= \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{[Nt]-1}\mathbf{E}[|W_1|]\int_0^1 \sqrt{s}\,\mathrm{d}s = t\sqrt{\frac{2}{\pi}}\left(\frac{2}{3}\right),$$

where the last equality uses $\mathbf{E}[|W_1|] = \sqrt{2/\pi}$ and $\int_0^1 \sqrt{s}\,\mathrm{d}s = \frac{2}{3}$ as well as $([Nt]/N) \to t$ as $N \to \infty$. This establishes $\int_0^t |\mathrm{d}V_s^{2,N}| = \mathbf{O_P}(\sqrt{N})$. It follows that $\sup_N \int_0^t |\mathrm{d}V_s^{2,N}| = \sup_N N\int_0^t |W_s - W_{\eta_s^N}|\,\mathrm{d}s = \infty$. Theorem 3.2 of Jacod and Protter (1998) then shows that $V_t^{2,N}$ cannot be good. □

Define the errors $U_t^{X^N} = [U_t^{X_1^N}, \ldots, U_t^{X_d^N}]$ where $U_t^{X_l^N} = X_{l,t}^N - X_{l,t}$, and $\overline{U}_t^{X^N} = [\overline{U}_t^{X_l^N}, \ldots, \overline{U}_t^{X_l^N}]$ where $\overline{U}_t^{X_l^N} = X_{l,t}^N - X_{l,\eta_t^N}^N$. To prove Theorem 1 we use the mean

[33] The random variable X^N is $\mathbf{O_P}(N)$ if $\mathbf{P} - \lim_{N\to\infty}(1/N)X^N = K \neq 0$ for some random variable K. The random variable X^N is $\mathbf{o_P}(N)$ if $\mathbf{P} - \lim_{N\to\infty}(1/N)X^N = 0$. If $\mathbf{P} - \lim_{N\to\infty} X^N = K \neq 0$ (resp. $\mathbf{P} - \lim_{N\to\infty} X^N = 0$) we say that X^N is $\mathbf{O_P}(1)$ (resp. $\mathbf{o_P}(1)$).

[34] Chebyshev's weak law of large numbers states that if $(1/N)\sum_{i=1}^N Z^i$ is such that $(1/N^2)\sum_{i=1}^N \mathbf{E}[(Z^i)^2] \to 0$ then $\mathbf{P} - \lim_{N\to\infty}((1/N)\sum_{i=1}^N (Z^i - \mathbf{E}[Z^i])) = 0$.

value theorem to write

$$
\begin{aligned}
A(X^N_{\eta^N_t}) - A(X_t) &= A(X^N_t) - A(X_t) - (A(X^N_t) - A(X^N_{\eta^N_t}))\\
&= \sum_{l=1}^{d} (\partial_l A(X_t + \lambda_{1,l} U_t^{X^N_l} e_l) U_t^{X^N_l} - \partial_l A(X^N_t + \lambda_{3,l} \overline{U}_t^{X^N_l} e_l) \overline{U}_t^{X^N_l}),
\end{aligned}
$$

where $\lambda_{\cdot,l} \in]0,1[$ for all $l = 1, \ldots, d$, and $e'_l = [0, \ldots, 0, 1, 0, \ldots, 0]$ is the lth unit vector of dimension d. A similar expression holds for $B(X^N_{\eta^N_t}) - B(X_t)$. The expansion of the error $U_t^{X^N}$ of the Euler continuous approximation can be decomposed as

$$
\begin{aligned}
U_T^{X^N} = & \int_0^T \sum_{l=1}^{d} \partial_l A(X_s + \lambda_{1,l} e_l U_s^{X^N_l}) U_s^{X^N_l} \, \mathrm{d}s\\
& + \int_0^T \sum_{l=1}^{d} \sum_{j=1}^{d} \partial_l B_j(X_s + \lambda_{2,l} e_l U_s^{X^N_l}) U_s^{X^N_l} \, \mathrm{d}W_s^j\\
& - \int_0^T \sum_{l=1}^{d} \partial_l A(X^N_s + \lambda_{3,l} e_l \overline{U}_s^{X^N_l}) \overline{U}_s^{X^N_l} \, \mathrm{d}s\\
& - \int_0^T \sum_{l=1}^{d} \sum_{j=1}^{d} \partial_l B_j(X^N_s + \lambda_{4,l} e_l \overline{U}_s^{X^N_l}) \overline{U}_s^{X^N_l} \, \mathrm{d}W_s^j.
\end{aligned}
$$

As $\overline{U}_s^{X^N_l} = A_l(X^N_{\eta^N_s})(s - \eta^N_s) + \sum_{i=1}^{d} B_{l,i}(X^N_{\eta^N_s})(W_s^i - W^i_{\eta^N_s})$ we have

$$
\begin{aligned}
U_T^{X^N} = & \int_0^T \sum_{l=1}^{d} \partial_l A(X_s + \lambda_{1,l} e_l U_s^{X^N_l}) U_s^{X^N_l} \, \mathrm{d}s\\
& + \int_0^T \sum_{l=1}^{d} \sum_{j=1}^{d} \partial_l B_j(X_s + \lambda_{2,l} e_l U_s^{X^N_l}) U_s^{X^N_l} \, \mathrm{d}W_s^j\\
& - \frac{1}{N} \int_0^T \sum_{l=1}^{d} \partial_l A(X^N_s + \lambda_{3,l} e_l \overline{U}_s^{X^N_l}) A_l(X^N_{\eta^N_s}) \, \mathrm{d}V_s^{1,N}\\
& - \frac{1}{N} \int_0^T \sum_{l=1}^{d} \partial_l A(X^N_s + \lambda_{3,l} e_l \overline{U}_s^{X^N_l}) \sum_{j=1}^{d} B_{l,j}(X^N_{\eta^N_s}) \, \mathrm{d}V_s^{2,j,N}\\
& - \frac{1}{N} \int_0^T \sum_{l=1}^{d} \sum_{j=1}^{d} \partial_l B_j(X^N_s + \lambda_{4,l} e_l \overline{U}_s^{X^N_l}) A_l(X^N_{\eta^N_s}) \, \mathrm{d}V_s^{3,j,N}\\
& - \frac{1}{\sqrt{N}} \int_0^T \sum_{l=1}^{d} \sum_{j=1}^{d} \partial_l B_j(X^N_s + \lambda_{4,l} e_l \overline{U}_s^{X^N_l}) \sum_{i=1}^{d} B_{l,i}(X^N_{\eta^N_s}) \, \mathrm{d}V_s^{4,i,j,N}. \quad (102)
\end{aligned}
$$

Lemmas 2 and 3 imply that the integral with respect to $V^{2,N}$ is the only term whose limit is difficult to find. Our next result shows that its limit must account for a "Wong-Zakai correction term" (Wong and Zakai, 1964).

Lemma 4. *For a sequence of good semimartingales* α^N *such that* $(\alpha^N, V^{2,N}) \Rightarrow (\alpha, V^2)$ *we have* $\int_0^T \alpha_s^N \, \mathrm{d}V_s^{2,j,N} \Rightarrow \int_0^T \alpha_s \, \mathrm{d}V_s^{2,j} + [\alpha, V^{2,j}]_T$, *for* $j = 1, \ldots, d$.

Proof. The integration by parts formula combined with the fact that $V^{2,j,N}$ is of bounded total variation for fixed N, and that α^N is good gives

$$\begin{aligned}\int_0^t \alpha_s^N \, \mathrm{d}V_s^{2,j,N} &= \alpha_t^N V_t^{2,j,N} - \int_0^t V_s^{2,j,N} \, \mathrm{d}\alpha_s^N - [V^{2,j,N}, \alpha^N]_t \\ &= \alpha_t^N V_t^{2,j,N} - \int_0^t V_s^{2,j,N} \, \mathrm{d}\alpha_s^N \\ &\Rightarrow \alpha_t V_t^{2,j} - \int_0^t V_s^{2,j} \, \mathrm{d}\alpha_s \\ &= \int_0^t \alpha_s \, \mathrm{d}V_s^{2,j} + [\alpha, V^{2,j}]_t.\end{aligned}$$

The second equality follows from $[V^{2,j,N}, \alpha^N] = 0$, a consequence of the fact that $V^{2,j,N}$ is of bounded total variation for fixed N. The last equality is an application of the integration by parts formula. □

The asymptotic error expansion for X^N in (102) involves an integral with respect to the bounded variation process $V^{2,j,N}$ that is not good (by Lemma 3). The limit of this integral is therefore not the integral of the weak limit of the integrand with respect to the weak limit of the integrator. Lemma 4 shows how the limit can be constructed if the integrand is good. To apply this result to the integral with respect to $V^{2,N}$ in (102) it remains to show that the integrand is good. For this we need the following lemma.

Lemma 5. *The semimartingale* X^N *is good.*

Proof. By Theorem 2.5 of Kurtz and Protter (1991b) goodness is equivalent to uniform tightness, defined as follows,

Definition 2 (*Jakubowski et al., 1989*). A sequence of semimartingales X^N is uniformly tight if and only if for each t and any simple predictable H^N such that $|H^N| \leqslant 1$ (**P**-a.s.), the set $\{\int_0^t H_{s-}^N \, \mathrm{d}X_s^N, \ N \geqslant 1\}$, is stochastically bounded (uniformly in N).[35,36]

To prove Lemma 5 it suffices to show that X^N is uniformly tight. The proof is by contradiction and uses arguments in Kurtz and Protter (1996).

To simplify notation set $Y_t' = [t, W_t']$ and $f = [A, B_1, \ldots, B_d]$, and write $X_T^N = X_0 + \sum_{j=0}^d \int_0^T f_j(X_{\eta_s^N}^N) \, \mathrm{d}Y_s^j$. As $f \in \mathscr{C}^2$ (by assumption), $\mathbf{P} - \lim_{N\to\infty} X^N = X$ (by Theorem 3.1 of Jacod and Protter (1998)) and $\eta_t^N \to t$ as $N \to \infty$, we can apply the continuous mapping Theorem to conclude $(X^N, f(X_{\eta^N}^N), Y) \Rightarrow (X, f(X), Y)$.

[35] H is simple predictable if there exists a sequence of stopping times $0 = \tau_1 \leqslant \cdot \leqslant \tau_{n+1} = T$, and $\mathscr{F}_{\tau_i}$ measurable finite random variables $(H_i)_{0 \leqslant i \leqslant n}$ such that $H_t = H_0 \mathbf{1}_{\{0\}}(t) + \sum_{i=1}^n H_i \mathbf{1}_{]\tau_i, \tau_{i+1}]}(t)$.

[36] A set of random variables $\{X^N, \ N \geqslant 1\}$ is uniformly stochastically bounded if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $\sup_N \mathbf{P}(|X^N| > \delta) < \varepsilon$.

Suppose now that X^N is not uniformly tight. By Definition 2 it follows that there exist a time t and simple predictable processes H^N with $|H^N|\leqslant 1$ (**P**-a.s.), for which the set $\{\int_0^t H^N_{s-}\,\mathrm{d}X^N_s,\ N\geqslant 1\}$ is not stochastically bounded, i.e. there exists an ε for which there is no $\delta>0$ such that $\sup_N \mathbf{P}(|\int_0^t H^N_{s-}\,\mathrm{d}X_s|>\delta)<\varepsilon$. For this sequence $|H^N|\leqslant 1$ (**P**-a.s.), and this ε, we have $\sup_{N^0}\inf_{N>N^0}\mathbf{P}(|\int_0^t H^N_{s-}\,\mathrm{d}X^N_s|>\delta)\geqslant\varepsilon$ for any $\delta>0$. Thus, there exists a sequence of positive constants $c^N\to\infty$, such that

$$\begin{aligned}
\varepsilon &\leqslant \liminf_{N\to\infty}\mathbf{P}\left(\left|\int_0^t H^N_{s-}\,\mathrm{d}X^N_s\right|>c^N\right)\\
&\leqslant \liminf_{N\to\infty}\left(\mathbf{P}\left(\int_0^t H^N_{s-}\,\mathrm{d}X^N_s>c_N\right)+\mathbf{P}\left(-\int_0^t H^N_{s-}\,\mathrm{d}X^N_s>c_N\right)\right)\\
&= \liminf_{N\to\infty}\left(\mathbf{P}\left(\int_0^t \frac{H^N_{s-}}{c^N}\,\mathrm{d}X^N_s>1\right)+\mathbf{P}\left(-\int_0^t \frac{H^N_{s-}}{c^N}\,\mathrm{d}X^N_s>1\right)\right)\\
&= \liminf_{N\to\infty}\left(\mathbf{P}\left(\sum_{j=0}^d\int_0^t \frac{H^N_{j,s-}}{c^N}f_j(X^N_{\eta^N_s})\,\mathrm{d}Y^j_s>1\right)\right.\\
&\qquad\left.+\mathbf{P}\left(-\sum_{j=0}^d\int_0^t \frac{H^N_{j,s-}}{c^N}f_j(X^N_{\eta^N_s})\,\mathrm{d}Y^j_s>1\right)\right)\\
&= 0.
\end{aligned}$$

The null limit on the last line uses goodness of Y and $(H^N_{j,}f_j(X^N_{\eta^N_\cdot})/c^N,\ Y^j)_{j=0,\dots,d}\Rightarrow(0,Y)$. The limit for the integrand follows from $|H^N_j|\leqslant 1$, $f_j(X^N_{\eta^N_\cdot})\Rightarrow f_j(X)$, and $c^N\to\infty$ (so that $H^N_jf_j(X^N\eta^N_\cdot)/c^N\Rightarrow 0$ uniformly in N). Goodness of Y then implies $\sum_{j=0}^d\int_0^t(H^N_{j,s}/c^N)f_j(X^N_{\eta^N_s})\,\mathrm{d}Y^j_s\Rightarrow 0$.

From this contradiction we conclude that X^N is uniformly tight and hence good. □

The weak limit results in Lemmas 1–5 are sufficient to prove Theorem 1.

Proof of Theorem 1 (*Kurtz and Protter, 1991a*). It follows from the weak limits in Lemmas 1–5 that lines 2, 3 and 4 of the approximation error $U^{X^N}_T$ in (102) converge weakly to zero. The remaining terms (lines 1 and 5) converge weakly to a linear SDE whose solution corresponds to the expression for U^X_T in the theorem. □

If we apply the Doss transformation and sample Gaussian increments $W^j_{t+(n+1)h}-W^j_{t+nh}$ then the error $U^{\hat{X}^N}=\hat{X}^N-\hat{X}$ of the Euler approximation with transformation can be expanded as

$$\begin{aligned}
U^{\hat{X}^N}_T &= \int_0^T\sum_{l=1}^d \partial_l\hat{A}(\hat{X}_s+\lambda_{1,l}e_lU^{\hat{X}^N_l}_s)U^{\hat{X}^N_l}_s\,\mathrm{d}s\\
&\qquad-\int_0^T\sum_{l=1}^d \partial_l\hat{A}(\hat{X}^N_s+\lambda_{3,l}e_l\overline{U}^{\hat{X}^N_l}_s)\overline{U}^{\hat{X}^N_l}_s\,\mathrm{d}s,
\end{aligned}$$

where $\lambda_{\cdot,l} \in]0,1[$ for all $l = 1,\ldots,d$, $e_l = [0,\ldots,1,\ldots,0]'$ is the lth unit vector and $\overline{U}_s^{\hat{X}_l^N} = \hat{X}_{l,s}^N - \hat{X}_{l,\eta_s^N}^N$. As $\overline{U}_s^{\hat{X}_l^N} = \hat{A}_l(\hat{X}_{\eta_s^N})(s-\eta_s^N) + \sum_{j=1}^d \hat{B}_{l,j}(W_s^j - W_{\eta_s^N}^j)$ we obtain

$$
\begin{aligned}
U_T^{\hat{X}^N} = & \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s + \lambda_{1,l} e_l U_s^{\hat{X}_l^N}) U_s^{\hat{X}_l^N}\, \mathrm{d}s \\
& - \frac{1}{N}\int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s^N + \lambda_{3,l} e_l \overline{U}_s^{\hat{X}_l^N}) \hat{A}_l(\hat{X}_{\eta_s^N})\, \mathrm{d}V_s^{1,N} \\
& - \frac{1}{N}\int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s^N + \lambda_{3,l} e_l \overline{U}_s^{\hat{X}_l^N}) \sum_{j=1}^d \hat{B}_{l,j}\, \mathrm{d}V_s^{2,j,N}.
\end{aligned} \tag{103}
$$

As for the error expansion without transformation (102) we have integrals with respect to $\hat{X}^N$. To find the limit using integration by parts we need to establish goodness of $\hat{X}^N$. Our next lemma verifies this property.

Lemma 6. *The semimartingale $\hat{X}^N$ is good.*

Proof. The proof parallels the proof of Lemma 5. □

Proof of Theorem 2. As in the proof of Theorem 1 integrals with respect to $V^{2,j,N}$ need special treatment. Let $\hat{\alpha}_s^{l,j,N} \equiv \partial_l \hat{A}(\hat{X}_s^N + \lambda_{3,l} e_l \bar{U}_s^{\hat{X}_l^N})\hat{B}_{l,j}$ and note that $\hat{\alpha}^{l,j,N}$ is good by the continuity of $\partial_l \hat{A}$. An application of Lemma 4 then shows that

$$
\int_0^T \hat{\alpha}_s^{l,j,N}\, \mathrm{d}V_s^{2,j,N} \Rightarrow \int_0^T \partial_l \hat{A}(\hat{X}_s)\hat{B}_{l,j}\, \mathrm{d}V_s^{2,j} + \frac{1}{2}\int_0^T \sum_{k=1}^d \partial_{l,k}\hat{A}(\hat{X}_s)\hat{B}_{k,j}\hat{B}_{l,j}\, \mathrm{d}s. \tag{104}
$$

As $\lim_{N\to\infty} \eta_s^N = s$ and $\mathbf{P} - \lim_{N\to\infty} \overline{U}^{\hat{X}_l^N} = \mathbf{P} - \lim_{N\to\infty} U^{\hat{X}_l^N} = 0$ and as $V^{1,N}$ is good by Lemma 2, it follows that $NU^{\hat{X}^N} \Rightarrow U^{\hat{X}}$ where

$$
\begin{aligned}
U_T^{\hat{X}} = & \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s) U_s^{\hat{X}_l}\, \mathrm{d}s - \frac{1}{2}\int_0^T \sum_{l=1}^d [\partial_l \hat{A}(\hat{X}_s)\hat{A}_l(\hat{X}_s) \\
& + \sum_{j,k=1}^d \partial_{l,k}\hat{A}(\hat{X}_s)\hat{B}_{k,j}\hat{B}_{l,j}]\, \mathrm{d}s \\
& - \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s) \sum_{j=1}^d \hat{B}_{l,j}\left(\frac{1}{2}\, \mathrm{d}W_s^j + \frac{1}{\sqrt{12}}\, \mathrm{d}Z_s^j\right).
\end{aligned}
$$

This SDE is linear and its solution corresponds to the result announced. □

Proof of Corollary 1. The result follows from $N(G(\hat{X}_T^N) - X_T) = \partial G(\hat{X}_T)U_T^{\hat{X}^N} + \mathbf{o_P}(1)$, where $\mathbf{o_P}(1)$ denotes terms that vanish in probability (see footnote 33), and $\partial G(z) = B(G(z))$. □

Proof of Theorem 3. The proofs are the same for the cases with and without transformation. The approximation error can be written as

$$\frac{1}{\sqrt{M}}\left(\sum_{i=1}^{M} g(X_T^{i,N}) - \mathbf{E}_0[g(X_T)]\right) = \frac{1}{\sqrt{M}}\sum_{i=1}^{M}(g(X_T^{i,N}) - g(X_T^i)) + \frac{1}{\sqrt{M}}\sum_{i=1}^{M}(g(X_T^i) - \mathbf{E}_0[g(X_T)]). \tag{105}$$

The Lindeberg central limit theorem for i.i.d. random variables gives

$$\frac{1}{\sqrt{M}}\sum_{i=1}^{M}(g(X_T^i) - \mathbf{E}_0[g(X_T)]) \Rightarrow \sqrt{\mathbf{VAR}[g(X_T)|\mathscr{F}_0]}Z, \tag{106}$$

where $Z \sim \mathrm{N}(0,1)$. The Clark–Ocone formula $g(X_T) - \mathbf{E}_0[g(X_T)] = \int_0^T \mathbf{E}_s[\mathscr{D}_s g(X_T)]\,\mathrm{d}W_s$ combined with the chain rule of Malliavin calculus (see Nualart, 1995) enables us to write $\mathbf{VAR}[g(X_T)|\mathscr{F}_0] = \int_0^T \mathbf{E}_0[\|\mathbf{E}_s[\partial g(X_T)\mathscr{D}_s X_T]\|^2]\,\mathrm{d}s$. This establishes that $\sqrt{\mathbf{VAR}[g(X_T)|\mathscr{F}_0]}Z = L_T(X_0)$ in distribution. It remains to find the weak limit of $(1/\sqrt{M})\sum_{i=1}^{M}(g(X_T^{i,N}) - g(X_T^i))$.

If we introduce two sequences of numbers, N_M and $\varepsilon^M = \sqrt{M}/N_M$ such that $\lim_{M\to\infty}\varepsilon^M = \varepsilon < \infty$ and $\lim_{M\to\infty} N_M = \infty$, we can invoke Kolmogorov's strong law of large numbers, to conclude that

$$\lim_{M\to\infty}\varepsilon^M N_M\left(\frac{1}{M}\sum_{i=1}^{M}(g(X_T^{i,N_M}) - g(X_T^i))\right) = \varepsilon \lim_{M\to\infty} N_M \mathbf{E}_0[g(X_T^{N_M}) - g(X_T)], \quad \mathbf{P}\text{-a.s.} \tag{107}$$

Theorem 4 on the expected approximation error (see below) then gives

$$\frac{1}{\sqrt{M}}\sum_{i=1}^{M}(g(X_T^{i,N_M}) - g(X_T^i)) \to \frac{\varepsilon}{2}K_T(X_0) \tag{108}$$

as $M \to \infty$. The corresponding result for the estimator with transformation is obtained by invoking Theorem 5 instead Theorem 4. This establishes the claim for both estimators. □

We now turn to the proof of the expected approximation error. The error expansion (102) shows that U^{X^N} satisfies a linear SDE. The solution is

$$\begin{aligned} N(\Omega_T^N)^{-1}U_T^{X^N} = &-\int_0^T (\Omega_s^N)^{-1}\,\mathrm{d}I_{1,s}^N - \int_0^T (\Omega_s^N)^{-1}\,\mathrm{d}I_{2,s}^N \\ &-\int_0^T (\Omega_s^N)^{-1}\,\mathrm{d}(I_{3,s}^N - [R^N, I_3^N]_s) \\ &-\sqrt{N}\int_0^T (\Omega_s^N)^{-1}\,\mathrm{d}(I_{4,s}^N - [R^N, I_4^N]_s), \end{aligned} \tag{109}$$

where $\Omega_T^N = \mathscr{E}^R(R^N)_T$ with $R_T^N = [R_{1,T}^N, \ldots, R_{d,T}^N]$ such that

$$R_{i,T}^N = \int_0^T \partial_i A(X_s + \lambda_{1,i} e_i U_s^{X_i^N})\,\mathrm{d}s + \int_0^T \sum_{j=1}^d \partial_i B_j(X_s + \lambda_{2,i} e_i U_s^{X_i^N})\,\mathrm{d}W_s^j \tag{110}$$

for $i = 1, \ldots, d$, and

$$I_{1,T}^N \equiv \int_0^T \sum_{l=1}^d \partial_l A(X_s^N + \lambda_{3,l} e_l \overline{U}_s^{X_l^N}) A_l(X_{\eta_s^N}^N)\,\mathrm{d}V_s^{1,N},$$

$$I_{2,T}^N \equiv \int_0^T \sum_{l=1}^d \partial_l A(X_s^N + \lambda_{3,l} e_l \overline{U}_s^{X_l^N}) \sum_{j=1}^d B_{l,j}(X_{\eta_s^N}^N)\,\mathrm{d}V_s^{2,j,N},$$

$$I_{3,T}^N \equiv \int_0^T \sum_{l=1}^d \sum_{j=1}^d \partial_l B_j(X_s^N + \lambda_{4,l} e_l \overline{U}_s^{X_l^N}) A_l(X_{\eta_s^N}^N)\,\mathrm{d}V_s^{3,j,N},$$

$$I_{4,T}^N \equiv \int_0^T \sum_{i,j=1}^d \left(\sum_{l=1}^d \partial_l B_j(X_s^N + \lambda_{4,l} e_l \overline{U}_s^{X_l^N}) B_{l,i}(X_{\eta_s^N}^N) \right) \mathrm{d}V_s^{4,i,j,N}.$$

In order to find the limit of the expectation of $NU_T^{X^N}$ we need the following lemma.

Lemma 7. *Define* $J_{4,T}^N \equiv \int_0^T (\Omega_s^N)^{-1}\,\mathrm{d}(I_{4,s}^N - [R^N, I_4^N]_s)$. *For any* $\mathscr{F}_T$*-measurable square integrable* $d \times 1$ *random vector* H_T, *and* $\mathscr{F}_T$*-measurable sequence of* $d \times 1$ *random vectors such that* $H_T^N \Rightarrow H_T$ *and*

$$\lim_{r\to\infty} \limsup_N \mathbf{E}_0[|\sqrt{N}(H_T^N)' J_{4,T}^N| \mathbf{1}_{\{|\sqrt{N} H'_T J_{4,T}^N| > r\}}] = 0 \tag{111}$$

(*i.e.* $(H_T^N)' J_{4,T}^N$ *is asymptotically uniformly integrable*) *the cross moment*

$$\sqrt{N}\mathbf{E}_0[(H_T^N)' J_{4,T}^N] \to \tfrac{1}{2}\mathbf{E}_0[U_{2,T}], \tag{112}$$

where

$$U_{2,T} = \sum_{i,j,k=1}^d \left[[h^{k,j} \beta^{k,i,j}, W^i]_T - H_T^k \left(\int_0^T \delta_s^{k,i,j}\,\mathrm{d}W_s^i + [\delta^{k,i,j}, W^i]_T \right) \right] \tag{113}$$

with $\beta_t^{k,i,j}, \delta_t^{k,i,j}$ *the kth elements of the* $d \times 1$ *vectors* $\beta_t^{i,j}, \delta_t^{i,j}$ *given by*

$$\delta_t^{i,j} = \Omega_t^{-1} \left[\partial B_j \sum_{l=1}^d \partial_l B_j\, B_{l,i} \right](X_t) \quad \text{for } i,j = 1, \ldots, d, \tag{114}$$

$$\beta_t^{i,j} = \Omega_t^{-1} \left[\sum_{l=1}^d \partial_l B_j B_{l,i} \right](X_t) \quad \text{for } i,j = 1, \ldots, d \tag{115}$$

and with $\{h_t^{k,j} : j = 1, \ldots, d\}$ *the integrands in the martingale representation of* H_T^k,

$$H_T^k - \mathbf{E}_0[H_T^k] = \sum_{j=1}^{d} \int_0^T h_s^{k,j} \, \mathrm{d}W_s^j \quad \text{for } k = 1, \ldots, d. \tag{116}$$

Proof. To establish this limit, recall that X^N is good and that the coefficients are continuous by assumption. It follows that R^N is good as well. Consider now the limit of $\sqrt{N}\mathbf{E}_0[(H_T^N)' \int_0^T (\Omega_s^N)^{-1} d[R^N, I_4^N]_s]$. Given that

$$\sqrt{N} \int_0^T (\Omega_s^N)^{-1} \, \mathrm{d}[R^N, I_4^N]_s = \sum_{i,j=1}^{d} \int_0^T \delta_s^{i,j} \, \mathrm{d}V_s^{2,i,N} + \mathbf{o_P}(1), \tag{117}$$

where $\delta^{i,j}$ is the fixed random variable (114) we can apply Lemma 4 to obtain the weak limit

$$\sqrt{N} \int_0^T (\Omega_s^N)^{-1} \, \mathrm{d}[R^N, I_4^N]_s \Rightarrow \sum_{i,j=1}^{d} \int_0^T \delta_s^{i,j} \, \mathrm{d}V_s^{2,i} + \sum_{i,j=1}^{d} [\delta^{i,j}, V^{2,i}]_T. \tag{118}$$

Weak convergence, along with the uniform integrability of $\sqrt{N}(H_T^N)' \int_0^T (\Omega_s^N)^{-1} \, \mathrm{d}[R^N, I_4^N]_s$ in assumption (111), implies convergence in means. We obtain,

$$\begin{aligned}
&\lim_{N\to\infty} \sqrt{N}\mathbf{E}_0\left[(H_T^N)' \int_0^T (\Omega_s^N)^{-1} \, \mathrm{d}[R^N, I_4^N]_s\right] \\
&= \mathbf{E}_0\left[H_T'\left(\sum_{i,j=1}^{d} \int_0^T \delta_s^{i,j} \, \mathrm{d}V_s^{2,i} + \sum_{i,j=1}^{d} [\delta^{i,j}, V^{2,i}]_T\right)\right] \\
&= \frac{1}{2} \sum_{i,j=1}^{d} \mathbf{E}_0\left[H_T'\left(\int_0^T \delta_s^{i,j} \, \mathrm{d}W_s^i + [\delta^{i,j}, W^i]_T\right)\right],
\end{aligned}$$

where the last equality uses the independence of $(\delta_\cdot^{i,j}, H_T)$ from $Z_\cdot^i$ in $V^{2,i}$.

Next, consider the limit of $\sqrt{N}\mathbf{E}_0[(H_T^N)' \int_0^T (\Omega_s^N)^{-1} \, \mathrm{d}I_{4,s}^N]$ where H_T is an arbitrary square integrable vector of $\mathscr{F}_T$-measurable random variables. By the Martingale Representation Theorem we can, for each $k = 1, \ldots, d$, write $H_T^{k,N} - \mathbf{E}_0[H_T^{k,N}] = \sum_{l=1}^{d} \int_0^T h_s^{k,l,N} \, \mathrm{d}W_s^l$. As $\sqrt{N} \, \mathrm{d}[V^{4,i,j,N}, W^l]_t = \mathbf{1}_{\{j=l\}} \, \mathrm{d}V_t^{2,i,N}$ we have

$$\sum_{l=1}^{d} \left[\int_0^\cdot h_s^{k,l,N} \, \mathrm{d}W_s^l, \sqrt{N} \int_0^\cdot (\Omega_s^N)^{-1} \, \mathrm{d}I_{4,s}^N\right]_T = \sum_{i,j=1}^{d} \int_0^T h_s^{k,j,N} \beta_s^{i,j} \, \mathrm{d}V_s^{2,i,N} + \mathbf{o_P}(1) \tag{119}$$

$$\Rightarrow \sum_{i,j=1}^{d} \int_0^T h_s^{k,j} \beta_s^{i,j} \, \mathrm{d}V_s^{2,i} + [h^{k,j}\beta^{i,j}, V^{2,i}]_T, \tag{120}$$

where the limit in the second line follows from Lemma 4 and goodness of the sequence $h^{k,j,N}$.

Combining the results above gives, with $h_0^{k,N} \equiv \sqrt{N}\mathbf{E}_0[H_T^{k,N} \int_0^T (\Omega_s^N)^{-1}\,\mathrm{d}I_{4,s}^N]$,

$$\begin{aligned}
\lim_{N\to\infty} h_0^{k,N} &= \lim_{N\to\infty} \sqrt{N}\mathbf{E}_0\left[\left(\mathbf{E}_0[H_T^{k,N}] + \sum_{l=1}^d \int_0^T h_s^{k,l,N}\,\mathrm{d}W_s^l\right)\left(\int_0^T (\Omega_s^N)^{-1}\,\mathrm{d}I_{4,s}^N\right)\right] \\
&= \lim_{N\to\infty} \sqrt{N}\sum_{l=1}^d \mathbf{E}_0\left[\left[\int_0^\cdot h_s^{k,l,N}\,\mathrm{d}W_s^l, \int_0^\cdot (\Omega_s^N)^{-1}\,\mathrm{d}I_{4,s}^N\right]_T\right] \\
&= \lim_{N\to\infty} \sum_{i,j=1}^d \mathbf{E}_0\left[\int_0^T h_s^{k,j,N}\beta_s^{i,j}\,\mathrm{d}V_s^{2,i,N}\right] \\
&= \sum_{i,j=1}^d \mathbf{E}_0\left[\int_0^T h_s^{k,j}\beta_s^{i,j}\,\mathrm{d}V_s^{2,i} + [h^{k,j}\beta^{i,j}, V^{2,i}]_T\right] \\
&= \frac{1}{2}\sum_{i,j=1}^d \mathbf{E}_0[[h^{k,j}\beta^{i,j}, W^i]_T].
\end{aligned}$$

The second line in this derivation uses $\mathbf{E}_0[\int_0^T (\Omega_s^N)^{-1}\,\mathrm{d}I_{4,s}^N] = 0$. The third and fourth lines follow from (119), (120) and the uniform integrability assumption (111). The last line uses $\mathbf{E}_0[\int_0^T h_s^{k,j}\beta_s^{i,j}\,\mathrm{d}V_s^{2,i}] = 0$ for all $i,j,k = 1,\ldots,d$. $\square$

We now proceed with the proof of Theorem 4.

Proof of Theorem 4. By Lemma 7 we know that

$$\sqrt{N}\mathbf{E}_0\left[(H_T^N)'\int_0^T (\Omega_s^N)^{-1}\mathrm{d}(I_{4,s}^N - [R^N, I_4^N]_s)\right] \to \frac{1}{2}\,\mathbf{E}_0[U_{2,T}], \tag{121}$$

with $U_{2,T}$ as in (113), for any $\mathscr{F}_T$-measurable sequence of random vectors $H_T^N \Rightarrow H_T$.

Given that other terms in (109) are of the same order and that each term is asymptotically uniformly integrable by assumption, we can find the expected approximation error from the weak limits of these terms.

As $\mathbf{P} - \lim_{N\to\infty} U^{X^N} = 0$, and

$$\begin{aligned}
\mathrm{d}[R^N, I_3^N]_s &= \sum_{j,k=1}^d \partial B_k(X_s)\gamma_{3,s}^{j,N}\,\mathrm{d}[W^k, V^{3,j,N}]_s \\
&= \sum_{j,k=1}^d \partial B_k(X_s)\gamma_{3,s}^{j,N}\,\mathrm{d}V_s^{1,N}\mathbf{1}_{\{k=j\}} \\
&= \sum_{j=1}^d \partial B_j(X_s)\gamma_{3,s}^{j,N}\,\mathrm{d}V_s^{1,N}
\end{aligned}$$

with

$$\gamma_{3,s}^{j,N} \equiv \sum_{l=1}^d \partial_l B_j(X_s + \lambda_{4,l} e_l \overline{U}_s^{X_l^N}) A_l(X_{\eta_s^N}^N), \tag{122}$$

we get

$$[R^N, I_3^N]_T \Rightarrow \frac{1}{2}\int_0^T \sum_{j=1}^d [\partial B_j \partial B_j A](X_s)\,\mathrm{d}s. \tag{123}$$

For the remaining terms we have $(I_1^N, I_2^N, I_3^N) \Rightarrow (I_1, I_2, I_3)$ where

$$I_{1,T} = \frac{1}{2}\int_0^T \partial A(X_s)A(X_s)\,\mathrm{d}s,$$

$$I_{2,T} = \int_0^T \partial A(X_s)\sum_{j=1}^d B_j(X_s)\left(\frac{1}{2}\,\mathrm{d}W_s^j + \frac{1}{\sqrt{12}}\,\mathrm{d}Z_s^j\right) + \frac{1}{2}\int_0^T \sum_{j,k,l=1}^d [\partial_k(\partial_l A B_{l,j})B_{k,j}](X_s)\,\mathrm{d}s,$$

$$I_{3,T} = \int_0^T \sum_{j=1}^d \partial B_j(X_s)A(X_s)\left(\frac{1}{2}\,\mathrm{d}W_s^j - \frac{1}{\sqrt{12}}\,\mathrm{d}Z_s^j\right).$$

For completeness we establish the limit of $I_{2,T}^N$. The other two limits are obtained along the same lines. Applying Lemma 4 with $\alpha_s^N = \sum_{l,j=1}^d \partial_l A(X_s^N + \lambda_{3,l} e_l \overline{U}_s^{X^N l})B_{l,j}(X_{\eta_s^N}^N)$, where the sequence α^N is good, gives

$$\int_0^T \alpha_s^N\,\mathrm{d}V_s^{2,j,N} \Rightarrow \int_0^T \alpha_s\,\mathrm{d}V_s^{2,j} + [\alpha, V^{2,j}]_T, \tag{124}$$

where $\alpha_s = \sum_{l,j=1}^d [\partial_l A B_{l,j}](X_s)$. Given that $V_t^{2,j} = \frac{1}{2}W_t^j + \frac{1}{\sqrt{12}}Z_t^j$ Ito's lemma implies

$$[\alpha, V^{2,j}]_T = \frac{1}{2}\int_0^T \sum_{l,j=1}^d [\partial[\partial_l A B_{l,j}]](X_s)\,\mathrm{d}[X, W^j]_s = \frac{1}{2}\int_0^T \sum_{l,j,k=1}^d [\partial_k[\partial_l A B_{l,j}]B_{k,j}](X_s)\,\mathrm{d}s, \tag{125}$$

where the second equality follows from $\mathrm{d}[X, W^j]_s = B_j(X_s)\,\mathrm{d}s$. This explains the limit of I_2^N.

As $g \in \mathscr{C}^3(\mathbb{R}^d)$ and as X^N converges, the mean value theorem shows that

$$N(g(X_T^N) - g(X_T)) = \partial g(X_T + diag[\lambda]U_T^{X^N})NU_T^{X^N} = \partial g(X_T + diag[\lambda]U_T^{X^N})\Omega_T^N(N(\Omega_T^N)^{-1}U_T^{X^N})$$

for some vector $\lambda = [\lambda_1, \ldots, \lambda_d]$ with $\lambda_i \in]0,1[, i = 1, \ldots d$, where $diag[\lambda]$ is the diagonal matrix with λ on the diagonal and zeros elsewhere, and where $N(\Omega_T^N)^{-1}U_T^{X^N}$ is defined in (109).

Now define $H_T^k \equiv \partial g(X_T)\Omega_T^k$ for $k = 1, \ldots, d$ where Ω^k is the kth column of the matrix process Ω. As $N(g(X_T^N) - g(X_T))$ is uniformly integrable by assumption (31) we can apply Lemma 7 to conclude that the expected approximation error converges to

$$N\mathbf{E}_0[g(X_T^N) - g(X_T)] \rightarrow \mathbf{E}_0[\partial g(X_T)U_{1,T} - U_{2,T}], \tag{126}$$

where $(\Omega_T)^{-1}U_{1,T}$ is the limit of the first 3 terms in (109), i.e.

$$(\Omega_T)^{-1}U_{1,T} = -\int_0^T (\Omega_s)^{-1}\,\mathrm{d}I_{1,s} - \int_0^T (\Omega_s)^{-1}\,\mathrm{d}I_{2,s} - \int_0^T (\Omega_s)^{-1}\,\mathrm{d}I_{3,s} + \frac{1}{2}\int_0^T \sum_{j=1}^d [\partial B_j\,\partial B_j\,A](X_s)\,\mathrm{d}s] \tag{127}$$

and $U_{2,T}$ is defined in Lemma 7 with $H_T^k = \partial g(X_T)\Omega_T^k$.

Let us now simplify $U_{2,T}$. Note that (113) can be written as $U_{2,T} \equiv \sum_{i,j,k=1}^d (U_{2,1,T}^{k,i,j} - H_T^k U_{2,2,T}^{k,i,j})$ where $U_{2,1,T}^{k,i,j} \equiv [h^{k,j}\beta^{k,i,j}, W^i]_T$ and $U_{2,2,T}^{k,i,j} \equiv \int_0^T \delta_s^{k,i,j}\,\mathrm{d}W_s^i + [\delta^{k,i,j}, W^i]_T$.

To simplify the second term, $U_{2,2,T}^{k,i,j}$, recall the $d\times 1$ process $\delta_t^{i,j} = \Omega_t^{-1}[\partial B_j \sum_{l=1}^d \partial_l B_j\, B_{l,i}](X_t)$. Ito's lemma gives

$$\mathrm{d}\delta_t^{i,j} = \Omega_t^{-1}\,\mathrm{d}\left(\partial B_j \sum_{l=1}^d \partial_l B_j\, B_{l,i}\right) + (\mathrm{d}(\Omega_t^{-1}))\left(\partial B_j \sum_{l=1}^d \partial_l B_j\, B_{l,i}\right) + \mathrm{d}\left[\Omega^{-1}, \partial B_j \sum_{l=1}^d \partial_l B_j\, B_{l,i}\right]. \tag{128}$$

From $\mathrm{d}\Omega_t = [\partial A(X_t)\,\mathrm{d}t + \sum_{j=1}^d \partial B_j(X_t)\,\mathrm{d}W_t^j]\Omega_t$ we deduce

$$\begin{aligned}\mathrm{d}\Omega_t^{-1} &= -\Omega_t^{-1}(\mathrm{d}\Omega_t)\Omega_t^{-1} - \Omega_t^{-1}\,\mathrm{d}[\Omega,\Omega^{-1}]_t \\ &= -\Omega_t^{-1}\left[\partial A(X_t)\,\mathrm{d}t + \sum_{j=1}^d \partial B_j(X_t)\,\mathrm{d}W_t^j\right] + \Omega_t^{-1}\sum_{j=1}^d [\partial B_j\,\partial B_j](X_t)\,\mathrm{d}t\end{aligned} \tag{129}$$

and

$$\begin{aligned}\mathrm{d}[\delta^{i,j}, W^i]_t &= \Omega_t^{-1}\,\mathrm{d}\left[\partial B_j \sum_{l=1}^d \partial_l B_j\, B_{l,i}, W^i\right]_t + \mathrm{d}[\Omega^{-1}, W^i]_t\left(\partial B_j \sum_{l=1}^d \partial_l B_j\, B_{l,i}\right) \\ &= \Omega_t^{-1}\partial\left[\partial B_j \sum_{l=1}^d \partial_l B_j\, B_{l,i}\right]\mathrm{d}[X, W^i]_t \\ &\quad - \Omega^{-1}\sum_j \partial B_j\,\mathrm{d}[W^j, W^i]_t\left(\partial B_j \sum_{l=1}^d \partial_l B_j\, B_{l,i}\right) \\ &= \Omega_t^{-1}\partial\left[\partial B_j \sum_{l=1}^d \partial_l B_j\, B_{l,i}\right]B_i\,\mathrm{d}t - \Omega^{-1}\partial B_i\left(\partial B_j \sum_{l=1}^d \partial_l B_j\, B_{l,i}\right)\mathrm{d}t.\end{aligned}$$

Given that $\sum_{l=1}^d \partial_l B_j\, B_{l,i} = \partial B_j\, B_i$ we can write

$$\sum_{i,j=1}^d [\delta^{i,j}, W^i]_T = \int_0^T \sum_{i,j=1}^d \mathrm{d}[\delta^{i,j}, W^i]_t = \int_0^T \Omega_t^{-1}\eta(X_t)\,\mathrm{d}t, \tag{130}$$

with

$$\eta \equiv \sum_{i,j=1}^{d} (\partial[\partial B_j \, \partial B_j \, B_i] B_i - \partial B_i \, \partial B_j \, \partial B_j \, B_i). \tag{131}$$

Let us now simplify the first term, $U^{k,i,j}_{2,1,T} \equiv [h^{k,j} \beta^{k,i,j}, W^i]_T$. The Clark–Ocone formula (see Nualart, 1995) and the chain rule of Malliavin calculus give

$$\begin{aligned} U^{k,i,j}_{2,1,T} \equiv [h^{k,j} \beta^{k,i,j}, W^i]_T &= \int_0^T E_s[\mathscr{D}_{i,s}(h^{k,j}_s \beta^{k,i,j}_s)] \, \mathrm{d}s \\ &= \int_0^T E_s[(\mathscr{D}_{i,s} h^{k,j}_s) \beta^{k,i,j}_s + h^{k,j}_s (\mathscr{D}_{i,s} \beta^{k,i,j}_s)] \, \mathrm{d}s. \end{aligned}$$

By definition, $h^{k,j}$ is the process in the representation of $H^k_T \equiv \partial g(X_T) \Omega^k_T$. The Clark–Ocone formula gives $h^{k,j}_t = \mathbf{E}_t[\mathscr{D}_{j,t}(\sum_{l=1}^d \partial_l g(X_T) \Omega^{l,k}_T)]$ and the rules of Malliavin calculus establish that

$$\mathscr{D}_{j,t}\left(\sum_{l=1}^{d} \partial_l g(X_T) \Omega^{l,k}_T\right) = \sum_{l,n=1}^{d} \partial^2_{l,n} g(X_T) (\mathscr{D}_{j,t} X_{n,T}) \Omega^{l,k}_T + \sum_{l=1}^{d} \partial_l g(X_T) \mathscr{D}_{j,t} \Omega^{l,k}_T, \tag{132}$$

$$\begin{aligned} \mathscr{D}^2_{i,j,t}\left(\sum_{l=1}^{d} \partial_l g(X_T) \Omega^{l,k}_T\right) = &\sum_{l,m,n=1}^{d} \partial^3_{l,m,n} g(X_T) (\mathscr{D}_{i,t} X_{m,T})(\mathscr{D}_{j,t} X_{n,T}) \Omega^{l,k}_T \\ &+ \sum_{l,n=1}^{d} \partial^2_{l,n} g(X_T) (\mathscr{D}^2_{i,j,t} X_{n,T}) \Omega^{l,k}_T \\ &+ \sum_{l,n=1}^{d} \partial^2_{l,n} g(X_T) (\mathscr{D}_{j,t} X_{n,T}) \mathscr{D}_{i,t} \Omega^{l,k}_T \\ &+ \sum_{l,n=1}^{d} \partial^2_{l,n} g(X_T) (\mathscr{D}_{i,t} X_{n,T}) \mathscr{D}_{j,t} \Omega^{l,k}_T \\ &+ \sum_{l=1}^{d} \partial_l g(X_T) \mathscr{D}^2_{i,j,t} \Omega^{l,k}_T, \end{aligned}$$

where

$$\mathscr{D}_{i,t} X_{m,T} = \sum_{h=1}^{d} \Omega^{m,h}_{t,T} B_{h,i}(X_t), \tag{133}$$

$$\mathscr{D}^2_{i,j,t} X_{m,T} = \sum_{h=1}^{d} (\mathscr{D}_{i,t} \Omega^{m,h}_{t,T}) B_{h,j}(X_t) + \sum_{h=1}^{d} \Omega^{m,h}_{t,T} [(\partial B_{h,j}) B_i](X_t) \tag{134}$$

and where, from

$$\mathrm{d}\Omega^{l,k}_{t,v} = \sum_{h=1}^{d} \left[\partial_h A_l(X_v) \, \mathrm{d}v + \sum_{j=1}^{d} \partial_h B_{l,j}(X_v) \, \mathrm{d}W^j_v\right] \Omega^{h,k}_{t,v}, \quad \Omega^{l,k}_{t,t} = \mathbf{1}_{\{l=k\}}, \tag{135}$$

we deduce that $\Gamma^{l,k}_{1,i}(t,T) \equiv \mathscr{D}_{i,t}\Omega^{l,k}_{t,T}$ and $\Gamma^{l,k}_{2,i,j}(t,T) = \mathscr{D}^2_{i,j,t}\Omega^{l,k}_{t,T}$ solve the linear SDEs,

$$\begin{aligned} \mathrm{d}\Gamma^{l,k}_{1,i}(t,v) = & \sum_{h=1}^{d}\left(\partial_h A_l(X_v)\,\mathrm{d}v + \sum_{j=1}^{d}\partial_h B_{l,j}(X_v)\,\mathrm{d}W^j_v\right)\Gamma^{h,k}_{1,i}(t,v) \\ & + \sum_{h,m=1}^{d}\left(\partial^2_{h,m}A_l(X_v)\,\mathrm{d}v + \sum_{j=1}^{d}\partial^2_{h,m}B_{l,j}(X_v)\,\mathrm{d}W^j_v\right)\Omega^{h,k}_v(\mathscr{D}_{i,t}X_{m,v}) \end{aligned}$$

with $\Gamma^{l,k}_{1,i}(t,t) = \sum_{h=1}^{d}\partial_h B_{l,i}(X_t)\Omega^{h,k}_t$ for $i,l,k = 1,\ldots,d$, and

$$\begin{aligned} \mathrm{d}\Gamma^{l,k}_{2,i,j}(t,v) = & \sum_{h=1}^{d}\left(\partial_h A_l(X_v)\,\mathrm{d}v + \sum_{p=1}^{d}\partial_h B_{l,p}(X_v)\,\mathrm{d}W^p_v\right)\Gamma^{h,k}_{2,i,j}(t,v) \\ & + \sum_{h,m=1}^{d}\left(\partial^2_{h,m}A_l(X_v)\,\mathrm{d}v + \sum_{p=1}^{d}\partial^2_{h,m}B_{l,p}(X_v)\,\mathrm{d}W^p_v\right)(\mathscr{D}_{i,t}X_{m,v})\Gamma^{h,k}_{1,j}(t,v) \\ & + \sum_{h,m=1}^{d}\left(\partial^2_{h,m}A_l(X_v)\,\mathrm{d}v + \sum_{p=1}^{d}\partial^2_{h,m}B_{l,p}(X_v)\,\mathrm{d}W^p_v\right)\Omega^{h,k}_v\mathscr{D}^2_{i,j,t}X_{m,v} \\ & + \sum_{h,m,n=1}^{d}\left(\partial^3_{h,m,n}A_l(X_v)\,\mathrm{d}v + \sum_{p=1}^{d}\partial^3_{h,m,n}B_{l,p}(X_v)\,\mathrm{d}W^p_v\right) \\ & \times\Omega^{h,k}_v(\mathscr{D}_{j,t}X_{m,v})\mathscr{D}_{i,t}X_{n,v}, \end{aligned}$$

with $\Gamma^{l,k}_{2,i,j}(t,t)=\sum_{h,n=1}^{d}\partial^2_{h,n}B_{l,j}(X_t)B_{n,i}(X_t)\Omega^{h,k}_t + \sum_{h=1}^{d}\partial_h B_{l,j}(X_t)\Gamma^{h,k}_i(t,t)$ for $i,j,l,k = 1,\ldots,d$.

From the definition $\beta^{i,j}_t = \Omega^{-1}_t[(\partial B_j)B_i](X_t)$ we also deduce that

$$\mathscr{D}_{i,t}\beta^{i,j}_t = \mathscr{D}_{i,t}(\Omega^{-1}_t[(\partial B_j)B_i](X_t)) = \Omega^{-1}_t[(\partial[(\partial B_j)B_i] - \partial B_i(\partial B_j))B_i](X_t). \tag{136}$$

Setting $v_{i,j}(t,T) \equiv \mathrm{E}_t[(\mathscr{D}_{i,t}h^{k,j}_t)\beta^{k,i,j}_t + h^{k,j}_t(\mathscr{D}_{i,t}\beta^{k,i,j}_t)]$ and using the law of iterated expectations gives

$$\begin{aligned} v_{i,j}(t,T) = & \sum_{k=1}^{d}(\Psi_{k,i,j}(t,T)[(\partial B_j)B_i](X_t) + \Phi_{k,i,j}(t,T)\Omega^{-1}_t \\ & \times[(\partial[(\partial B_j)B_i] - \partial B_i(\partial B_j))B_i](X_t)), \end{aligned} \tag{137}$$

where

$$\Phi_{k,i,j}(t,T) = \sum_{l,n=1}^{d}\partial^2_{l,n}g(X_T)\left(\sum_{h=1}^{d}\Omega^{n,h}_{t,T}B_{h,j}(X_t)\right)\Omega^{l,k}_T + \sum_{l=1}^{d}\partial_l g(X_T)\Gamma^{l,k}_{1,j}(t,T), \tag{138}$$

$$\Psi_{k,i,j}(t,T) = \sum_{l,m,n=1}^{d} \partial^3_{l,m,n} g(X_T)\left(\sum_{h=1}^{d} \Omega^{m,h}_{t,T} B_{h,i}(X_t)\right)\left(\sum_{h=1}^{d} \Omega^{n,h}_{t,T} B_{h,j}(X_t)\right)\Omega^{l,k}_T$$
$$+ \sum_{l,m=1}^{d} \partial^2_{l,m} g(X_T)\left(\sum_{h=1}^{d} \Gamma^{m,h}_{1,i}(t,T) B_{h,j}(X_t) + \sum_{h=1}^{d} \Omega^{m,h}_T [(\partial B_{h,j})B_i](X_t)\right)\Omega^{l,k}_T$$
$$+ \sum_{l,m=1}^{d} \partial^2_{l,m} g(X_T)\left(\sum_{h=1}^{d} \Omega^{m,h}_{t,T} B_{h,j}(X_t)\right)\Gamma^{l,k}_{1,i}(t,T) \tag{139}$$

and where $\Omega^{l,k}_{t,v}$ solves (135) while $\Gamma^{l,k}_{1,i}(t,v)$ and $\Gamma^{l,k}_{2,i,j}(t,v)$ solve the linear SDEs

$$\mathrm{d}\Gamma^{l,k}_{1,i}(t,v) = \sum_{h=1}^{d}\left(\partial_h A_l(X_v)\,\mathrm{d}v + \sum_{j=1}^{d} \partial_h B_{l,j}(X_v)\,\mathrm{d}W^j_v\right)\Gamma^{h,k}_{1,i}(t,v)$$
$$+ \sum_{h,m=1}^{d}\left(\partial^2_{h,m} A_l(X_v)\,\mathrm{d}v + \sum_{j=1}^{d} \partial^2_{h,m} B_{l,j}(X_v)\,\mathrm{d}W^j_v\right)$$
$$\times \Omega^{h,k}_v \left(\sum_{n=1}^{d} \Omega^{m,n}_{t,v} B_{n,i}(X_t)\right) \tag{140}$$

with $\Gamma^{l,k}_{1,i}(t,t) = \sum_{h=1}^{d} \partial_h B_{l,i}(X_t)\Omega^{h,k}_t$, and

$$\mathrm{d}\Gamma^{l,k}_{2,i,j}(t,v) = \sum_{h=1}^{d}\left(\partial_h A_l(X_v)\,\mathrm{d}v + \sum_{p=1}^{d} \partial_h B_{l,p}(X_v)\,\mathrm{d}W^p_v\right)\Gamma^{h,k}_{2,i,j}(t,v)$$
$$+ \sum_{h,m=1}^{d}\left(\partial^2_{h,m} A_l(X_v)\,\mathrm{d}v + \sum_{p=1}^{d} \partial^2_{h,m} B_{l,p}(X_v)\,\mathrm{d}W^p_v\right)$$
$$\times \left(\sum_{n=1}^{d} \Omega^{m,n}_{t,v} B_{n,i}(X_t)\right)\Gamma^{h,k}_{1,j}(t,v)$$
$$+ \sum_{h,m=1}^{d}\left(\partial^2_{h,m} A_l(X_v)\,\mathrm{d}v + \sum_{p=1}^{d} \partial^2_{h,m} B_{l,p}(X_v)\,\mathrm{d}W^p_v\right)\Omega^{h,k}_v$$
$$\times \sum_{h=1}^{d}\left(\Gamma^{m,h}_{1,i}(t,v) B_{h,j}(X_t) + \Omega^{m,h}_{t,v}[(\partial B_{h,j})B_i](X_t)\right)$$
$$+ \sum_{h,m,n=1}^{d}\left(\partial^3_{h,m,n} A_l(X_v)\,\mathrm{d}v + \sum_{p=1}^{d} \partial^3_{h,m,n} B_{l,p}(X_v)\,\mathrm{d}W^p_v\right)\Omega^{h,k}_v$$
$$\times \left(\sum_{p=1}^{d} \Omega^{m,p}_{t,v} B_{p,j}(X_t)\right)\left(\sum_{q=1}^{d} \Omega^{n,q}_{t,v} B_{q,i}(X_t)\right) \tag{141}$$

with $\Gamma^{l,k}_{2,i,j}(t,t) = \sum_{h,n=1}^{d} \partial^2_{h,n} B_{l,j}(X_t) B_{n,i}(X_t)\Omega^{h,k}_t + \sum_{h=1}^{d} \partial_h B_{l,j}(X_t)\Gamma^{h,k}_{1,i}(t,t)$.

Substituting the definitions $U_{2,T} \equiv \sum_{i,j,k=1}^{d}(U^{k,i,j}_{2,1,T} - H^k_T U^{k,i,j}_{2,2,T})$ and $H^k_T \equiv \partial g(X_T)\Omega^k_T$ in (126), using the results above and eliminating expectations of stochastic integrals with respect to Z^j (which is independent from $\mathscr{F}_T$), enables us to rewrite the limit (126) as

$$N\mathbf{E}_0[g(X_T^N)-g(X_T)] \to \mathbf{E}_0\left[\partial g(X_T)U_{1,T} - \sum_{i,j,k=1}^{d}(U_{2,1,T}^{k,i,j} - H_T^k U_{2,2,T}^{k,i,j})\right]$$
$$= \mathbf{E}_0\left[\partial g(X_T)\left(U_{1,T} + \sum_{i,j,k=1}^{d}\Omega_T^k U_{2,2,T}^{k,i,j}\right) - \sum_{i,j,k=1}^{d} U_{2,1,T}^{k,i,j}\right]$$
$$= \frac{1}{2}\mathbf{E}_0[\partial g(X_T)V_{1,T} + V_{2,T}] \tag{142}$$

with $V_{1,T}, V_{2,T}$ as defined in (29) and (30). This establishes the result. □

Proof of Corollary 2. The estimator $\sum_{n=0}^{N-1} g(X_{nh}^N)h$ based on the Euler discrete approximation X_{nh}^N with $h = T/N$, is asymptotically equivalent to the estimator $\int_0^T g(X_{\eta_s^N}^N)\,\mathrm{d}s$ based on the Euler continuous approximation $X_{\eta_s^N}^N$ (Jacod and Protter, 1998, Lemma 1). To find the asymptotic limit of the expected approximation error of the Riemann integral, we can then study the approximation error of the continuous Euler approximation. This error has two parts:

$$\int_0^T (g(X_{\eta_s^N}^N)\,\mathrm{d}s - g(X_s))\,\mathrm{d}s = H_{1T}^N(X_0) + H_{2T}^N(X_0), \tag{143}$$

where

$$H_{1T}^N \equiv \int_0^T (g(X_s^N) - g(X_s))\,\mathrm{d}s,$$

$$H_{2T}^N \equiv -\int_0^T (g(X_s^N) - g(X_{\eta_s^N}^N))\,\mathrm{d}s.$$

By Theorem 4 $N\mathbf{E}_0[g(X_s^N) - g(X_s)] \to \frac{1}{2}K_s(X_0)$, $\mathbf{P}_0$-a.s., so that

$$N\mathbf{E}_0[H_{1T}^N] \to \frac{1}{2}\int_0^T K_s(X_0)\,\mathrm{d}s, \tag{144}$$

$\mathbf{P}_0$-a.s. It remains to study the second component, H_{2T}^N. As shown in the proof of Theorem 1, we get, by the mean value theorem,

$$-NH_{2T}^N = \int_0^T \partial g(X_s)\left(A(X_{\eta_s^N}^N)\,\mathrm{d}V_s^{1,N} + \sum_{j=1}^{d} B_j(X_{\eta_s^N}^N)\,\mathrm{d}V_s^{2,j,N}\right) + \mathbf{o_P}(1). \tag{145}$$

The result now follows using the same arguments as those invoked in the proof of Theorem 4 to find the expected value of limits involving the processes $V^{1,N}$ and $V^{2,N}$. In particular, with $\alpha_{1s}^N \equiv \partial g(X_s)A(X_{\eta_s^N}^N)$ and $\alpha_{2s}^{j,N} \equiv \partial g(X_s)B_j(X_{\eta_s^N}^N)$, we use

$$\mathbf{E}_0\left[\int_0^T \alpha_{1s}^N\,\mathrm{d}V_s^{1,N}\right] \to \frac{1}{2}\mathbf{E}_0\left[\int_0^T \alpha_{1s}\,\mathrm{d}s\right], \tag{146}$$

$$\mathbf{E}_0\left[\int_0^T \alpha_{2s}^{j,N}\,\mathrm{d}V_s^{2,j,N}\right] \to \mathbf{E}_0\left[\int_0^T (\alpha_{2s}^j\,\mathrm{d}V_s^{2,j} + \mathrm{d}[\alpha_2^j, V^{2,j}]_s)\right], \tag{147}$$

$\mathbf{P}_0$-a.s., where $\alpha_{1s} = [(\partial g)A](X_s)$ and $\alpha^j_{2s} = [(\partial g)B_j](X_s)$. The expression for $K_{2T}(X_0)$ follows because the expectation of the integral relative to the martingale $V^{2,j}$ is null and because $\mathrm{d}[\alpha^j_2, V^{2,j}]_s = \frac{1}{2}\mathrm{d}[\alpha^j_2, W^j]_s$. Calculating $\mathrm{d}[\alpha^j_2, W^j]_s$ yields the expression for $K_{2T}(X_0)$. □

We now turn to the expected error associated with the Doss transformation.

Proof of Theorem 5. Note that all the terms of the error expansion (103) are of order $1/N$ and that the limit distribution is non-centered. If the terms in the error expansion are uniformly integrable, the result follows by taking the expectation of the solution of the linear SDE for the approximation error

$$NU_T^{\hat{X}_t^N} = -\hat{\Omega}_T^N \left(\int_0^T (\hat{\Omega}_s^N)^{-1} \, \mathrm{d}\hat{I}_{1,s}^N + \int_0^T (\hat{\Omega}_s^N)^{-1} \, \mathrm{d}\hat{I}_{2,s}^N \right), \tag{148}$$

where $\hat{\Omega}_T^N = \mathscr{E}(\hat{R}^N)_T$ with $\hat{R}_T^N = [\hat{R}_{1,T}^N, \ldots, \hat{R}_{d,T}^N]$ and $\hat{R}_{i,T}^N = \partial_i \hat{A}(\hat{X}_s + \lambda_{1,i} e_i U_s^{\hat{X}^N})$ for $i = 1, \ldots, d$ and where

$$\hat{I}_{1,T}^N \equiv \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s^N + \lambda_{3,l} e_l \overline{U}_s^{\hat{X}_l^N}) \hat{A}_l(\hat{X}_{\eta_s^N}^N) \, \mathrm{d}V_s^{1,N},$$

$$\hat{I}_{2,T}^N \equiv \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s^N + \lambda_{3,l} e_l \overline{U}_s^{\hat{X}_l^N}) \sum_{j=1}^d \hat{B}_{l,j}(\hat{X}_{\eta_s^N}^N) \, \mathrm{d}V_s^{2,j,N}.$$

As $(\hat{\Omega}_t^N, \hat{I}_{1,t}^N, \hat{I}_{2,t}^N, \hat{X}_t^N) \Rightarrow (\hat{\Omega}_t, \hat{I}_{1,t}, \hat{I}_{2,t}, \hat{X}_t)$ with $\hat{\Omega}_t, \hat{X}_t$ as defined in Theorem 2 and

$$\hat{I}_{1,T} \equiv \frac{1}{2} \int_0^T [(\partial \hat{A}) \hat{A}](\hat{X}_s) \, \mathrm{d}s$$

$$\begin{aligned} \hat{I}_{2,T} \equiv & \int_0^T \left[\partial A \sum_{j=1}^d \hat{B}_j \right] (\hat{X}_s) \left(\frac{1}{2} \, \mathrm{d}W_s^j + \frac{1}{\sqrt{12}} \, \mathrm{d}Z_s^j \right) \\ & + \frac{1}{2} \int_0^T \sum_{j,k,l=1}^d \partial_{l,k} \hat{A}(\hat{X}_s) \hat{B}_{k,j} \hat{B}_{l,j} \, \mathrm{d}s \end{aligned}$$

(the expression for $\hat{I}_{2,T}$ follows from (104)), the result follows using the same arguments as in the proof of Theorem 2. □

Proof of Corollary 3. We proceed as in the proof of Corollary 2. From Theorem 5 it follows that

$$N\mathbf{E}_0 \left[\int_0^T (\hat{g}(\hat{X}_s^N) - \hat{g}(\hat{X}_s)) \, \mathrm{d}s \right] \to \frac{1}{2} \, \mathbf{E}_0 \left[\int_0^T \partial \hat{g}(\hat{X}_s) \hat{V}_s \, \mathrm{d}s \right]. \tag{149}$$

The result for the second part of the expected approximation error, $N\mathbf{E}_0[\int_0^T (\hat{g}(\hat{X}_s^N) - \hat{g}(\hat{X}_{\eta_s^N}^N)) \, \mathrm{d}s]$, follows using the same arguments as in the proof of Corollary 2 for $K_{2T}(X_0)$. Simply replace the drift coefficient A by $\hat{A}$ and the volatility coefficient B by $\hat{B}$. With $\hat{B}_j$ deterministic we get $\partial \hat{B}_j = 0$ and the expression announced follows. □

Proof of Theorem 6. Consider the Euler discrete approximations (172)–(173), in Appendix B, of the processes $NC_{j,\cdot}^N$ for $j = 1,2$. The Euler continuous approximations are

$$\begin{aligned}
\mathrm{d}(NC_{1,s}^N) = & \left[\partial A(X_{\eta_s^N}^N)h + \sum_{j=1}^d \partial B_j(X_{\eta_s^N}^N)\,\mathrm{d}W_s^j\right](NC_{1,s}^N) \\
& - \left(\sum_{i,j=1}^d [\partial[\partial B_j\,\partial B_j\,B_i]B_i](X_{\eta_s^N}^N)\right)\mathrm{d}s \\
& + \left(\left[(\partial A)A + \sum_{j=1}^d \partial B_j\,\partial A\,B_j + \sum_{j,k,l=1}^d \partial_k(\partial_l A B_{l,j})B_{k,j}\right](X_{\eta_s^N}^N)\right)\mathrm{d}s \\
& + \sum_{j=1}^d [(\partial A)B_j + (\partial B_j)A](X_{\eta_s^N}^N)\,\mathrm{d}W_s^j - \sum_{i,j=1}^d (\partial B_j)(\partial B_j)B_i](X_{\eta_s^N}^N)\,\mathrm{d}W_s^i,
\end{aligned}$$

$$\mathrm{d}(NC_{2,s}^N) = \sum_{i,j=1}^d v_{i,j}^N(\eta_s^N, T)\,\mathrm{d}s$$

with $C_{1,0}^N = C_{2,0}^N = 0$. Thus, $N\mathbf{E}_0[\partial g(X_{Nh}^N)C_{1,Nh}^N + C_{2,Nh}^N] \to -K_T(X_0)$.

We can use the arguments in the proofs of Theorems 4 and 5 to show that, for N_M and $\varepsilon^M = \sqrt{M}/N_M$ such that $\lim_{M\to\infty}\varepsilon^M = \varepsilon < \infty$ and $\lim_{M\to\infty} N_M = \infty$, we have

$$\begin{aligned}
& \lim_{M\to\infty} \frac{1}{2}\frac{1}{\sqrt{M}}\sum_{i=1}^M \left(\partial g(X_{N_M h}^{i,N_M})C_{1,N_M h}^{i,N_M} + C_{2,N_M h}^{i,N_M}\right) \\
& = \frac{1}{2}\,\varepsilon \lim_{N_M\to\infty} N_M \mathbf{E}_0\left[\partial g(X_T^{N_M})C_{1,T}^{N_M} + C_{2,T}^{N_M}\right] \\
& = -\frac{1}{2}\,\varepsilon K_T(X_0).
\end{aligned}$$

Defining

$$g_{\mathrm{c}}^{N,M} \equiv \frac{1}{M}\sum_{i=1}^M \left[g(X_{Nh}^{i,N}) + \frac{1}{2}(\partial g(X_{Nh}^{i,N})C_{1,Nh}^{i,N} + C_{2,Nh}^{i,N})\right], \tag{150}$$

it then follows that

$$\frac{1}{2}\left(\frac{1}{\sqrt{M}}\sum_{i=1}^M (\partial g(X_{N_M h}^{i,N_M})C_{1,N_M h}^{i,N_M} + C_{2,N_M h}^{i,N_M})\right) \tag{151}$$

corrects the asymptotic second-order bias for the estimator without transformation when $M \to \infty$.

The proof for the estimator with transformation follows the same steps. In this case the average over independent replications of the random variables $\frac{1}{2}\partial\hat{g}(\hat{X}_T^N)\hat{C}_{nh}^N$ approximates the negative of the second-order bias with transformation.

The asymptotic equivalence of the bias-corrected estimators with and without transformation is a consequence of the fact that they have the same asymptotic distribution. □

Proof of Corollary 4. The proof follows from the arguments used to establish Theorem 3 and from Corollaries 2 and 3. □

To prove convergence results for the Milshtein scheme, we need the following lemmas.

Lemma 8. *The following weak convergence results hold*

$$V_t^{5,l,j,N} \equiv N^{3/2} \int_0^t \int_{\eta_v^N}^v \int_{\eta_v^N}^u \mathrm{d}W_r^l \, \mathrm{d}W_u^j \, \mathrm{d}v \Rightarrow \frac{\sqrt{2}}{6} Z_t^{l,j} + \frac{1}{6} \tilde{Z}^{l,j} \equiv V_t^{5,l,j}, \tag{152}$$

$$V_t^{6,l,j,i,N} \equiv N \int_0^t \int_{\eta_v^N}^v \int_{\eta_v^N}^u \mathrm{d}W_r^l \, \mathrm{d}W_u^j \, \mathrm{d}W_v^i \Rightarrow \frac{1}{\sqrt{6}} \tilde{Z}_t^{l,j,i} \equiv V_t^{6,l,j,i} \tag{153}$$

as $N \to \infty$, *where* $((Z^{l,j})_{l,j \in \{1,\ldots,d\}}, (\tilde{Z}^{l,j})_{l,j \in \{1,\ldots,d\}}, (\tilde{Z}^{l,j,i})_{l,j,i \in \{1,\ldots,d\}})$ *is a* $(2d^2 + d^3)$-*dimensional standard Brownian motion and where* $(Z^{l,j})_{l,j \in \{1,\ldots,d\}}$ *is the Brownian motion in the limit* $V^{4,l,j}$ *in Lemma* 1.

Proof. Using the scaling property of Brownian motion we obtain

$$\begin{aligned} V_t^{5,l,j,N} &= \frac{1}{\sqrt{N}} \int_0^{Nt} \int_{[v]}^v \int_{[v]}^u \mathrm{d}W_r^l \, \mathrm{d}W_u^j \, \mathrm{d}v \\ &= \frac{1}{\sqrt{N}} \sum_{k=1}^{[Nt]} \int_{[k-1,k[} \int_{k-1}^v \int_{k-1}^u \mathrm{d}W_r^l \, \mathrm{d}W_u^j \, \mathrm{d}v \\ &\quad + \frac{1}{\sqrt{N}} \int_{[Nt]}^{Nt} \int_{[Nt]}^v \int_{[Nt]}^u \mathrm{d}W_r^l \, \mathrm{d}W_u^j \, \mathrm{d}v. \end{aligned}$$

Integration by parts gives $\int_{[k-1,k[} \int_{k-1}^v \int_{k-1}^u \mathrm{d}W_r^l \, \mathrm{d}W_u^j \, \mathrm{d}v = \int_{[k-1,k[} (k-v) \int_{k-1}^v \mathrm{d}W_r^l \, \mathrm{d}W_v^j$. The sequence of i.i.d. random variables $\int_{[k-1,k[} (k-v) \int_{k-1}^v \mathrm{d}W_r^l \, \mathrm{d}W_v^j$ is independent of W^j, has mean zero, variance $1/12$ and covariance with $\int_{[k-1,k[} \int_{k-1}^v \mathrm{d}W_r^m \, \mathrm{d}W_v^n$ of $(1/6)\mathbf{1}_{\{l=m,j=n\}}$. Donsker's functional central limit Theorem (see Kallenberg, 1997, Theorem 2.9, p. 225) then gives

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{[Nt]} \int_{[k-1,k[} (k-v) \int_{k-1}^v \mathrm{d}W_r^l \, \mathrm{d}W_v^j \Rightarrow \frac{\sqrt{2}}{6} Z_t^{l,j} + \frac{1}{6} \tilde{Z}_t^{l,j}. \tag{154}$$

This result and the continuity of the Riemann integral $\mathbf{P} - \lim_{N \to \infty} (1/\sqrt{N}) \int_{[Nt]}^{Nt} \int_{[Nt]}^v \int_{[Nt]}^u \mathrm{d}W_r^l \, \mathrm{d}W_u^j \, \mathrm{d}v = 0$, establishes (152).

Similarly, using the scaling property of Brownian motion we obtain

$$\begin{aligned} V_t^{6,l,j,i,N} &= \frac{1}{\sqrt{N}} \int_0^{Nt} \int_{[v]}^{v} \int_{[v]}^{u} \mathrm{d}W_r^l \,\mathrm{d}W_u^j \,\mathrm{d}W_v^i \\ &= \frac{1}{\sqrt{N}} \sum_{k=1}^{[Nt]} \int_{[k-1,k[} \int_{k-1}^{v} \int_{k-1}^{u} \mathrm{d}W_r^l \,\mathrm{d}W_u^j \,\mathrm{d}W_v^i \\ &\quad + \frac{1}{\sqrt{N}} \int_{[Nt]}^{Nt} \int_{[Nt]}^{v} \int_{[Nt]}^{u} \mathrm{d}W_r^l \,\mathrm{d}W_u^j \,\mathrm{d}W_v^i. \end{aligned}$$

As the sequence of i.i.d random variables $\int_{[k-1,k[} \int_{k-1}^{v} \int_{k-1}^{u} \mathrm{d}W_r^l \,\mathrm{d}W_u^j \,\mathrm{d}W_v^i$ is independent of W^j and $\int_{[k-1,k[} \int_{k-1}^{v} \mathrm{d}W_r^m \,\mathrm{d}W_v^n$ for all $m,n = 1,\ldots,d$, and has variance $1/6$, Donsker's invariance principle establishes

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{[Nt]} \int_{[k-1,k[} \int_{k-1}^{v} \int_{k-1}^{u} \mathrm{d}W_r^l \,\mathrm{d}W_u^j \,\mathrm{d}W_v^i \Rightarrow \frac{1}{\sqrt{6}} \tilde{Z}_t^{l,j,i}. \tag{155}$$

As by the continuity of the stochastic integral $\mathbf{P} - \lim_{N\to\infty}(1/\sqrt{N}) \int_{[Nt]}^{Nt} \int_{[Nt]}^{v} \int_{[Nt]}^{u} \mathrm{d}W_r^l \,\mathrm{d}W_u^j \,\mathrm{d}W_v^i = 0$, this proves (153). □

Lemma 9. *The semimartingale* $V^{6,l,j,i,N}$ *is good.*

Proof. As for Lemma 2 it is sufficient to show that $\sup_N \mathbf{VAR}[V_t^{6,l,j,i,N}] < \infty$. Using

$$\begin{aligned} \mathbf{VAR}[V_t^{6,l,j,i,N}] &= N^{-1}\mathbf{E}\left[\left(\int_0^{Nt} \int_{[v]}^{v} \int_{[v]}^{u} \mathrm{d}W_r^l \,\mathrm{d}W_u^j \,\mathrm{d}W_v^i\right)^2\right] \\ &= \frac{1}{N}\int_0^{Nt} \int_{[v]}^{v} (u - [v])\,\mathrm{d}u\,\mathrm{d}v = \frac{1}{2}\frac{1}{N}\int_0^{Nt} (v - [v])^2\,\mathrm{d}v \\ &= \frac{1}{2}\frac{1}{N}\sum_{k=1}^{[Nt]} \int_{[k-1,k[} (v-(k-1))^2\,\mathrm{d}v + \frac{1}{2}\frac{1}{N}\int_{[Nt]}^{Nt} (v - [Nt])^2\,\mathrm{d}v \\ &= \frac{1}{6}\left(\frac{[Nt]}{N} + \frac{(Nt-[Nt])^3}{N}\right) < \frac{1}{6}\,t, \end{aligned}$$

we conclude that $V^{6,l,j,i,N}$ is good. □

Lemma 10. *The semimartingale* $V^{5,l,j,N}$ *is not good. For any sequence of good semimartingales* α^N *such that* $(\alpha^N, V^{5,l,j,N}) \Rightarrow (\alpha, V^{5,l,j})$ *we have* $\int_0^T \alpha_s^N \,\mathrm{d}V_s^{5,l,j,N} \Rightarrow \int_0^T \alpha_s \,\mathrm{d}V_s^{5,l,j} + [\alpha, V^{5,l,j}]_T$, *for* $l,j = 1,\ldots,d$.

Proof. Let α^N be an arbitrary sequence of good semimartingales such that $(\alpha^N, V^{5,l,j,N}) \Rightarrow (\alpha, V^{5,l,j})$. Integration by parts gives

$$\begin{aligned}\int_0^T \alpha_s^N \, \mathrm{d}V_s^{5,l,j,N} &= \alpha_T^N V_T^{5,l,j,N} - \int_0^T V_s^{5,l,j,N} \, \mathrm{d}\alpha_s^N \\ &\Rightarrow \alpha_T V_T^{5,l,j} - \int_0^T V_s^{5,l,j} \, \mathrm{d}\alpha_s \\ &= \int_0^T \alpha_s \, \mathrm{d}V_s^{5,l,j} + [\alpha, V^{5,l,j}]_T. \end{aligned} \tag{156}$$

The semimartingale $V^{5,l,j,N}$ is good if and only if $[\alpha, V^{5,l,j}]_T = 0$. To show that goodness fails it suffices to find an α^N such that $[\alpha, V^{5,l,j}]_T \neq 0$. Taking $\alpha^N = V^{4,l,j,N}$ we obtain $[V^{4,l,j}, V^{5,l,j}]_T = \frac{1}{6}T$. The conclusion follows. The limit is given by (156). □

The expansion of $\tilde{U}_T^{\tilde{X}^N} \equiv \tilde{X}_T^N - X_T$ for the approximation error of the Milshtein scheme is obtained along the same lines as (102) for the Euler scheme. We get

$$\begin{aligned}\tilde{U}_T^{\tilde{X}^N} = &\int_0^T \sum_{l=1}^d \partial_l A(X_s + \lambda_{1,l} e_l \tilde{U}_s^{\tilde{X}_l^N}) \tilde{U}_s^{\tilde{X}_l^N} \, \mathrm{d}s \\ &+ \int_0^T \sum_{l=1}^d \sum_{j=1}^d \partial_l B_j(X_s + \lambda_{2,l} e_l \tilde{U}_s^{\tilde{X}_l^N}) \tilde{U}_s^{\tilde{X}_l^N} \, \mathrm{d}W_s^j \\ &- \frac{1}{N} \int_0^T \sum_{l=1}^d \partial_l A(\tilde{X}_s^N + \lambda_{3,l} e_l \overline{U}_s^{\tilde{X}_l^N}) A_l(\tilde{X}_{\eta_s^N}^N) \, \mathrm{d}V_s^{1,N} \\ &- \frac{1}{N} \int_0^T \sum_{l=1}^d \partial_l A(\tilde{X}_s^N + \lambda_{3,l} e_l \overline{U}_s^{\tilde{X}_l^N}) \sum_{j=1}^d B_{l,j}(\tilde{X}_{\eta_s^N}^N) \, \mathrm{d}V_s^{2,j,N} \\ &- \frac{1}{N} \int_0^T \sum_{l,j=1}^d \partial_l B_j(\tilde{X}_s^N + \lambda_{4,l} e_l \overline{U}_s^{\tilde{X}_l^N}) A_l(\tilde{X}_{\eta_s^N}^N) \, \mathrm{d}V_s^{3,j,N} \\ &- \frac{1}{N} \int_0^T \sum_{i,k,l,j=1}^d \partial_k B_i(\tilde{X}_s^N + \lambda_{4,k} e_k \overline{U}_s^{\tilde{X}_k^N}) [\partial B_l B_j]_k(\tilde{X}_{\eta_s^N}^N) \, \mathrm{d}V_s^{6,l,j,i,N} \\ &- \frac{1}{N^{3/2}} \int_0^T \sum_{k,l,j=1}^d \partial_k A(\tilde{X}_s^N + \lambda_{3,k} e_k \overline{U}_s^{\tilde{X}_k^N}) [\partial B_l \, B_j]_k(\tilde{X}_{\eta_s^N}^N) \, \mathrm{d}V_s^{5,l,j,N}, \end{aligned} \tag{157}$$

where $\tilde{U}_s^{\tilde{X}_l^N} = \tilde{X}_{l,s}^N - X_{l,s}$ and $\overline{U}_s^{\tilde{X}_l^N} \equiv \tilde{X}_{l,s}^N - \tilde{X}_{l,\eta_s^N}$. Comparing (157) with (102), shows that the additional term in the Milshtein scheme eliminates integrals with respect to $V^{4,i,j,N}$ in the error expansion of the Euler scheme. At the same time it induces two additional integrals, with respect to $V^{5,l,j,N}$ and $V^{6,l,j,i,N}$ (lines 5 and 6 of (157)), in the approximation errors $A(\tilde{X}_{\eta_s^N}^N) - A(\tilde{X}_s^N)$ and $B_j(\tilde{X}_{\eta_s^N}^N) - B_j(\tilde{X}_s^N)$.

Proof of Theorem 7. Lemmas 9 and 10 imply that the term in line 6 of (157) is of higher order than those in lines 2–5. Lemmas 1–5 and 8–10 imply that lines 1–5 of

(157) converge weakly to a linear SDE whose solution is the expression for $U_T^{\tilde{X}}$ in Theorem 7. □

Proof of Theorem 8. Note that the last two terms of $\tilde{U}_T^X$ are both products of two independent random variables one of which has null expectation. Thus, using

$$\mathbf{E}\left[\Omega_T \int_0^T \Omega_s^{-1}[(\partial A)B_j - (\partial B_j)A](X_s)\,\mathrm{d}Z_s^j\right] = \mathbf{E}\left[\Omega_T \int_0^T \Omega_s^{-1}[(\partial B_i)(\partial B_l)B_j](X_s)\,\mathrm{d}\tilde{Z}_s^{l,j,i}\right] = 0, \tag{158}$$

for $l, j, i = 1, \ldots, d$ leads to the result announced. □

Proof of Theorem 9. The proof is the same as for the Euler scheme with transformation. □

Proof of Corollary 5. We proceed as in the proof of Corollary 2. Decompose the approximation error in two parts, $\tilde{H}_{1T}^N$ and $\tilde{H}_{2T}^N$. These are the equivalents of H_{1T}^N and H_{2T}^N in Corollary 2, obtained by replacing the Euler scheme by the Milshtein scheme. The limit of $N\mathbf{E}_0[\tilde{H}_{1T}^N]$, namely $\frac{1}{2}\int_0^T \tilde{K}_s(X_0)\,\mathrm{d}s$, is found using the same arguments as for the Euler scheme.

The second component $\tilde{H}_{2T}^N$ can be written, using the mean value theorem, as

$$-N\tilde{H}_{2T}^N = \int_0^T \partial g(X_s)\left(A(\tilde{X}_{\eta_s^N}^N)\,\mathrm{d}V_s^{1,N} + \sum_{j=1}^d B_j(\tilde{X}_{\eta_s^N}^N)\,\mathrm{d}V_s^{2,N} + \frac{1}{\sqrt{N}}\sum_{l,j=1}^d [(\partial B_l)B_j](X_s)\,\mathrm{d}V_s^{5,l,j,N}\right) + \mathbf{o_P}(1). \tag{159}$$

The limits for the integrals with respect to $V^{1,N}$ and $V^{2,j,N}$ follow from the arguments in the proof of Corollary 2. For the integral with respect to $V_s^{5,l,j,N}$ note that Lemmas 8 and 10 imply

$$\mathbf{E}\left[\frac{1}{\sqrt{N}}\sum_{l,j=1}^d [(\partial B_l)B_j](X_s)\,\mathrm{d}V_s^{5,l,j,N}\right] \to 0, \tag{160}$$

as $N \to \infty$. This establishes the result announced. □

Proof of Theorem 10. The proof parallels the proof for the Euler scheme with transformation. The average over independent replications of $\frac{1}{2}\partial g(\tilde{X}_T^N)\tilde{C}_{nh}^N$ approximates the negative of the second-order bias with transformation. The bias-corrected estimators with and without transformation are asymptotically equivalent because they share the same asymptotic distribution. □

Proof of Corollary 6. The proof is identical to the proof of Corollary 4. □

Proof of Theorem 11. From (70) we get

$$-\frac{1}{L}\sum_{l=0}^{L-1} g_\Delta(Y_l, Y_{l+1};\theta_0) = \frac{1}{L}\sum_{l=0}^{L-1} J_\Delta(Y_l)'\partial_\theta h_\Delta(Y_l, Y_{l+1};\theta_0)(\hat{\theta}_\Delta^L - \theta_0) + \mathbf{o}_\mathbf{P}(1). \tag{161}$$

By assumption, Y is an ergodic Markov process. We then get, from the central limit theorem for ergodic processes,

$$\sqrt{L}(\hat{\theta}_\Delta^L - \theta_0) \Rightarrow \Sigma_\Delta^{-1/2} Z \tag{162}$$

where $Z \sim N(0, I_p)$, and $\Sigma_\Delta(\theta_0) \equiv (A_\Delta(\theta_0)^{-1})' B_\Delta(\theta_0) A_\Delta(\theta_0)^{-1}$, with

$$A_\Delta(\theta_0)' \equiv \int_{\mathbb{R}^d} J_\Delta(y)'\Gamma_\Delta(y)\bar{p}^{\theta_0}(y)\lambda(\mathrm{d}y),$$

$$B_\Delta(\theta_0) \equiv \int_{\mathbb{R}^d} J_\Delta(y)'\Psi_\Delta(y)J_\Delta(y)\bar{p}^{\theta_0}(y)\lambda(\mathrm{d}y),$$

$$\Gamma_\Delta(y) \equiv \int_{\mathbb{R}^d} \partial_\theta h(y, z;\theta_0)p_\Delta^{\theta_0}(y, z)\lambda(\mathrm{d}z),$$

$$\Psi_\Delta(y) \equiv \int_{\mathbb{R}^d} h(y, z;\theta_0)(h(y, z;\theta_0))'p_\Delta^{\theta_0}(y, z)\lambda(\mathrm{d}z).$$

Setting $J_\Delta(y)' = \Gamma_\Delta(y)'\Psi_\Delta(y)^{-1}$ and using the Cauchy–Schwartz inequality gives the variance lower bound

$$\Sigma_\Delta(\theta_0) = \int_{\mathbb{R}^d} \Gamma_\Delta(y)'\Psi_\Delta(y)^{-1}\Gamma_\Delta(y)\bar{p}^{\theta_0}(y)\lambda(\mathrm{d}y). \tag{163}$$

Similarly, if $\mathbf{P} - \lim_{L\to\infty} S_L = S$, we can use (72) to show that

$$\begin{aligned}
&-\left(\frac{1}{L}\sum_{l=1}^{L}(S^{1/2}g_\Delta(Y_l, Y_{l+1};\theta_0))\right)'\left(\frac{1}{L}\sum_{l=1}^{L}\partial_\theta(S^{1/2}g_\Delta(Y_l, Y_{l+1};\tilde{\theta}_\Delta^L))\right)\\
&\quad = \left(\frac{1}{L}\sum_{l=1}^{L}\partial_\theta(S^{1/2}g_\Delta(Y_l, Y_{l+1};\theta_0))\right)'\left(\frac{1}{L}\sum_{l=1}^{L}\partial_\theta(S^{1/2}g_\Delta(Y_l, Y_{l+1};\tilde{\theta}_\Delta^L))\right)\\
&\qquad \times(\tilde{\theta}_\Delta^L - \theta_0) + \mathbf{o}_\mathbf{P}(1).
\end{aligned}$$

It now follows that

$$\sqrt{L}(\tilde{\theta}_\Delta^L - \theta_0) \Rightarrow \Xi_\Delta(\theta_0)^{-1/2} Z \tag{164}$$

where $\Xi_\Delta(\theta_0) = (C_\Delta(\theta_0)^{-1})' D_\Delta(\theta_0)^{-1} C_\Delta(\theta_0)$, with

$$C_\Delta(\theta_0)' = \int_{\mathbb{R}^d}(S^{1/2}J_\Delta(y))'\Gamma_\Delta(y)\bar{p}^{\theta_0}(y)\lambda(\mathrm{d}y), \tag{165}$$

$$D_{\Delta}(\theta_0) = \int_{\mathbb{R}^d} (S^{1/2} J_{\Delta}(y))' \Psi_{\Delta}(y) (S^{1/2} J_{\Delta}(y)) \bar{p}^{\theta_0}(y) \lambda(\mathrm{d}y). \tag{166}$$

This is of the same form as the variance Σ_{Δ} of $\sqrt{L}(\hat{\theta}_{\Delta}^{L} - \theta_0)$ above for which $J_{\Delta}(y)$ is the optimal weight (by the Cauchy–Schwartz inequality). We conclude that the efficient GMM must have $S_L = I_p$. □

Proof of Theorem 12. The mean value theorem and $\mathbf{P} - \lim_{L\to\infty} \left(\hat{\theta}_{\Delta}^{L,M_L,N_L} - \hat{\theta}_{\Delta}^{L} \right) = 0$ imply the existence of $\theta_{\star}^{L}$ such that $\mathbf{P} - \lim_{L\to\infty} \theta_{\star}^{L} = \theta_0$ and

$$\begin{aligned} &- \Gamma_{\Delta}^{L,M_L,N_L}(Y_l, Y_{l+1}; \theta_{\star}^{L}) \sqrt{L}(\hat{\theta}_{\Delta}^{L,M_L,N_L} - \hat{\theta}_{\Delta}^{L}) \\ &\quad = \frac{\sqrt{L}}{\sqrt{M_L}} \left(\frac{1}{L} \sum_{l=0}^{L-1} U^{g^{M_L,N_L}}(Y_l, Y_{l+1}, \hat{\theta}_{\Delta}^{L,M_L,N_L}, \hat{\theta}_{\Delta}^{L}) \right), \end{aligned} \tag{167}$$

where

$$\begin{aligned} &U^{g^{M_L,N_L}}\left(Y_l, Y_{l+1}, \hat{\theta}_{\Delta}^{L,M_L,N_L}, \hat{\theta}_{\Delta}^{L} \right) \\ &\quad \equiv \sqrt{M_L} \left(g_{\Delta}^{M_L,N_L}(Y_l, Y_{l+1}; \hat{\theta}^{L,M_L,N_L}) - g_{\Delta}(Y_l, Y_{l+1}; \hat{\theta}^{L}) \right), \end{aligned} \tag{168}$$

and

$$\Gamma_{\Delta}^{L,M_L,N_L}(Y_l, Y_{l+1}, \theta_{\star}^{L}) \equiv \frac{1}{L} \sum_{l=0}^{L-1} \partial_{\theta} g_{\Delta}^{M_L,N_L}(Y_l, Y_{l+1}; \theta_{\star}^{L}). \tag{169}$$

By the law of large numbers for ergodic Markov chains

$$\mathbf{P} - \lim_{L\to\infty} \Gamma_{\Delta}^{L,M_L,N_L}(Y_l, Y_{l+1}, \theta_{\star}^{L}) = \Sigma_{\Delta}(\theta_0). \tag{170}$$

Similarly, using Theorems 4, 5, or 9, adjusted for the dependence on the observations Y_l, Y_{l+1}, and the fact that $\mathbf{P} - \lim_{L\to\infty} (\hat{\theta}_{\Delta}^{L,M_L,N_L} - \hat{\theta}_{\Delta}^{L}) = 0$, we conclude that

$$\mathbf{P} - \lim_{L\to\infty} \frac{1}{L} \sum_{l=0}^{L-1} U^{g^{M_L,N_L}}(Y_l, Y_{l+1}, \hat{\theta}_{\Delta}^{L,M_L,N_L}, \hat{\theta}_{\Delta}^{L}) = \varepsilon_2 \kappa_{\Delta}(\theta_0) \tag{171}$$

if $\lim_{L\to\infty} \sqrt{M_L}/N_L = \varepsilon_2$, where $\kappa_{\Delta}(\theta_0)$ is defined in (77). The result announced follows if $\lim_{L\to\infty} \sqrt{L/M_L} = \varepsilon_1 < \infty$. □

Appendix B. Correcting for second-order bias

This appendix provides formulas for the bias-corrected estimators of conditional expectations. In the expressions that follow we use the notation $h = \Delta t$ for the time increment. The terms in the bias-corrected estimator without transformation, (42), are

$$
\begin{aligned}
C^N_{1,(n+1)h} = C^N_{1,nh} &+ \left[\partial A(X^N_{nh})h + \sum_{j=1}^{d} \partial B_j(X^N_{nh})\Delta W^j_{nh}\right] C^N_{1,nh} \\
&+ \left(\left[(\partial A)A + \sum_{j=1}^{d} \partial B_j\,\partial AB_j + \sum_{j,k,l=1}^{d} \partial_k(\partial_l AB_{l,j})B_{k,j}\right](X^N_{nh})\right)\frac{h}{N} \\
&- \left(\sum_{j,k=1}^{d} [\partial[\partial B_j\,\partial B_j\, B_k]B_k](X^N_{nh})\right)\frac{h}{N} \\
&+ \sum_{j=1}^{d} [(\partial A)B_j + (\partial B_j)A](X^N_{nh})\,\frac{\Delta W^j_{nh}}{N} \\
&- \sum_{j,k=1}^{d} [(\partial B_j)(\partial B_j)B_k](X^N_{nh})\,\frac{\Delta W^k_{nh}}{N},
\end{aligned} \tag{172}
$$

$$
C^{i,N}_{2,(n+1)h} = C^N_{2,nh} + \left(\sum_{l,j=1}^{d} v^N_{l,j}(nh, Nh)\right)\frac{h}{N}, \tag{173}
$$

where

$$
\begin{aligned}
v_{i,j}(nh, Nh) = &\sum_{k=1}^{d} \Psi^N_{k,i,j}(nh, Nh)[(\partial B_j)B_i](X^N_{nh}) \\
&+ \sum_{k=1}^{d} \Phi^N_{k,i,j}(nh, Nh)(\Omega^N_{nh})^{-1}[(\partial[(\partial B_j)B_i] - \partial B_i(\partial B_j))B_i](X^N_{nh})
\end{aligned}
$$

with

$$
\begin{aligned}
\Phi^N_{k,i,j}(nh, Nh) = &\sum_{l,n=1}^{d} \partial^2_{l,h} g(X_{nN})\left(\sum_{h=1}^{d} \Omega^{n,h,N}_{nh,Nh} B_{h,j}(X^N_{nh})\right)\Omega^{l,k,N}_{Nh} \\
&+ \sum_{l=1}^{d} \partial_l g(X^N_{Nh})\Gamma^{l,k}_{1,j}(nh, Nh),
\end{aligned} \tag{174}
$$

$$
\begin{aligned}
\Psi^N_{k,i,j}(nh, Nh) = &\sum_{l,m,n=1}^{d} \partial^3_{l,m,n} g(X^N_{Nh})\left(\sum_{h=1}^{d} \Omega^{m,h,N}_{nh,Nh} B_{h,i}(X^N_{nh})\right) \\
&\times\left(\sum_{h=1}^{d} \Omega^{n,h,N}_{nh,Nh} B_{h,j}(X^N_{nh})\right)\Omega^{l,k,N}_{Nh} + \sum_{l,m=1}^{d} \partial^2_{l,m} g(X^N_{Nh}) \\
&\times\left(\sum_{h=1}^{d} \Gamma^{m,h,N}_{1,i}(nh, Nh)B_{h,j}(X^N_{nh}) + \sum_{h=1}^{d} \Omega^{m,h,N}_{Nh}[(\partial B_{h,j})B_i](X^N_{nh})\right)\Omega^{l,k,N}_{Nh} \\
&+ \sum_{l,m=1}^{d} \partial^2_{l,h} g(X^N_{Nh})\left(\sum_{h=1}^{d} \Omega^{m,h,N}_{nh,Nh} B_{h,j}(X^N_{nh})\right)\Gamma^{l,k,N}_{1,i}(nh, Nh),
\end{aligned} \tag{175}
$$

and $\Gamma^{l,k,N}_{1,i}(nh,Nh) = \sum_{p=1}^{d} \partial_p B_{l,i}(X^N_{nh})\Omega^{p,k,N}_{nh} + \sum_{s=nh}^{Nh} \Delta_h \Gamma^{l,k,N}_{1,i}(nh,sh)$ where

$$\begin{aligned}\Delta_h \Gamma^{l,k,N}_{1,i}(nh,sh) = & \sum_{p=1}^{d}\left(\partial_p A_l(X^N_{sh})h + \sum_{j=1}^{d} \partial_p B_{l,j}(X_{sh})\Delta_h W^j_{sh}\right)\Gamma^{p,k,N}_{1,i}(nh,sh) \\ & + \sum_{p,m=1}^{d}\left(\partial^2_{p,m} A_l(X^N_{sh})h + \sum_{j=1}^{d} \partial^2_{p,m} B_{l,j}(X^N_{sh})\Delta_h W^j_{sh}\right)\Omega^{p,k,N}_{sh} \\ & \times\left(\sum_{q=1}^{d} \Omega^{m,q,N}_{nh,sh} B_{q,i}(X^N_{sh})\right), \qquad (176)\end{aligned}$$

and with $\Gamma^{l,k,N}_{2,i,j}(nh,Nh) = \sum_{p,r=1}^{d} \partial^2_{p,r} B_{l,j}(X^N_{nh}) B_{r,i}(X^N_{nh})\Omega^{p,k,N}_{nh} + \sum_{p=1}^{d} \partial_p B_{l,j}(X^N_{nh})\Gamma^{p,k,N}_{1,i}$ $(nh,nh) + \sum_{s=nh}^{Nh} \Delta_h \Gamma^{l,k,N}_{2,i,j}(nh,sh)$ where

$$\begin{aligned}\Delta_h \Gamma^{l,k,N}_{2,i,j}(nh,sh) = & \sum_{p=1}^{d}\left(\partial_p A_l(X^N_{sh})h + \sum_{q=1}^{d} \partial_p B_{l,q}(X^N_{sh})\Delta_h W^q_{sh}\right)\Gamma^{p,k,N}_{2,i,j}(nh,sh) \\ & + \sum_{p,m=1}^{d}\left(\partial^2_{p,m} A_l(X^N_{sh})h + \sum_{q=1}^{d} \partial^2_{p,m} B_{l,q}(X^N_{sh})\Delta_h W^q_{sh}\right) \\ & \times\left(\sum_{r=1}^{d} \Omega^{m,r,N}_{nh,sh} B_{r,i}(X_{nh})\right)\Gamma^{p,k,N}_{1,j}(nh,sh) \\ & + \sum_{p,m=1}^{d}\left(\partial^2_{p,m} A_l(X^N_{sh})h + \sum_{q=1}^{d} \partial^2_{p,m} B_{l,q}(X^N_{sh})\Delta_h W^q_{sh}\right)\Omega^{p,k,N}_{sh} \\ & \times \sum_{p=1}^{d}\left(\Gamma^{m,p,N}_{1,i}(nh,sh)B_{p,j}(X_{nh}) + \Omega^{m,p,N}_{nh,sh}[(\partial B_{p,j})B_i](X^N_{nh})\right) \\ & + \sum_{p,m,r=1}^{d}\left(\partial^3_{p,m,r} A_l(X^N_{sh})h + \sum_{q=1}^{d} \partial^3_{p,m,r} B_{l,q}(X_{sh})\Delta_h W^q_{sh}\right)\Omega^{p,k,N}_{sh} \\ & \times\left(\sum_{q=1}^{d} \Omega^{m,q,N}_{nh,sh} B_{q,j}(X^N_{nh})\right)\left(\sum_{q=1}^{d} \Omega^{r,q,N}_{nh,sh} B_{q,i}(X^N_{nh})\right), \qquad (177)\end{aligned}$$

$$\Omega^N_{(n+1)h} = \Omega^N_{nh} + \left(\partial A(X^N_{nh})h + \sum_{j=1}^{N} \partial B_j(X^N_{nh})(\Delta W^j_{nh})\right)\Omega^N_{nh},$$

$$X^N_{(n+1)h} = X^N_{nh} + A(X^N_{nh})h + \sum_{j=1}^{d} B_j(X^N_{nh})(\Delta W^j_{nh})$$

with $X_0^N = X_0$, $\Omega_0^N = I_d$ and $C_{1,0}^N = C_{2,0}^N = 0$. For the bias-corrected estimator with the transformation, (43), we have

$$\hat{C}_{(n+1)h}^N = \hat{C}_{nh}^N + [\partial\hat{A}(\hat{X}_{nh}^N)h]\hat{C}_{nh}^N + \left(\left[(\partial\hat{A})\hat{A} + \sum_{j,k,l=1}^{d} \partial_{l,k}\hat{A}\hat{B}_{k,j}\hat{B}_{l,j}\right](\hat{X}_{nh}^N)\right)\frac{h}{N} + \sum_{j=1}^{d} \partial\hat{A}(\hat{X}_{nh}^N)\hat{B}_j \frac{\Delta W_{nh}^j}{N},$$

$$\hat{X}_{(n+1)h}^N = \hat{X}_{nh}^N + \hat{A}(\hat{X}_{nh}^N)h + \sum_{j=1}^{d} \hat{B}_j(\Delta W_{nh}^j) \tag{178}$$

with $\hat{X}_0^N = X_0$ and $\hat{C}_0 = 0$, when the transformation is applied.

Additional correction terms for the estimators of conditional expectations of Riemann integrals are, for the Euler scheme,

$$C_{3,(n+1)h}^N = C_{3,nh}^N + \left(\left[(\partial g)\left(A + \sum_{j=1}^{d}(\partial B_j)B_j\right)\right](X_{nh}^N) + \sum_{j=1}^{d}[B_j'(\partial^2 g)B_j](X_{nh}^N)\right)\frac{h}{N}, \tag{179}$$

and for the Euler scheme with transformation,

$$\hat{C}_{1,(n+1)h}^N = \hat{C}_{1,nh}^N + \left([(\partial\hat{g})(\hat{A})](\hat{X}_{nh}^N) + \sum_{j=1}^{d}\hat{B}_j'(\partial^2\hat{g}(\hat{X}_{nh}^N))\hat{B}_j\right)\frac{h}{N}. \tag{180}$$

For the Milshtein scheme the process $\tilde{C}_{(n+1)h}$ in the bias-corrected estimator (61) is

$$\tilde{C}_{(n+1)h}^N = \tilde{C}_{nh}^N + \left[\partial A(\tilde{X}_{nh}^N)h + \sum_{j=1}^{d}\partial B_j(\tilde{X}_{nh}^N)\Delta W_{nh}^j\right]\tilde{C}_{nh}^N + \partial A(\tilde{X}_{nh}^N)\frac{\Delta_h\tilde{X}_{nh}^N}{N} + \sum_{j=1}^{d}[(\partial[(\partial A)B_j])B_j](\tilde{X}_{nh}^N)\frac{h}{N} + \sum_{j=1}^{d}[(\partial B_j)A](\tilde{X}_{nh}^N)\frac{\Delta_h W_{nh}^j}{N}, \tag{181}$$

$$\tilde{X}_{(n+1)h}^N = \tilde{X}_{nh}^N + A(\tilde{X}_{nh}^N)h + \sum_{j=1}^{d} B_j(\tilde{X}_{nh}^N)(\Delta W_{nh}^j) + \sum_{l,j=1}^{d}[(\partial B_l)B_j](\tilde{X}_{nh}^N)\,\Delta F^{l,j}, \tag{182}$$

where $\Delta F^{l,j}$ is defined in (53). The additional correction term for second-order bias-corrected estimators of Riemann integrals based on the Milshtein scheme is

$$\tilde{C}_{1,(n+1)h}^N = \tilde{C}_{1,nh}^N + \left(\left[(\partial g)\left(A + \sum_{j=1}^{d}(\partial B_j)B_j\right)\right](\tilde{X}_{nh}^N) + \sum_{j=1}^{d}[B_j'(\partial^2 g)B_j](\tilde{X}_{nh}^N)\right)\frac{h}{N}. \tag{183}$$

An alternative is to use the Euler scheme to approximate X in the recursion for $\tilde{C}$. As both $\tilde{X}^N_\cdot$ and $X^N_\cdot$ converge to X. the resulting approximation is first-order asymptotically equivalent.

References

Ait-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9, 385–426.

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

Ait-Sahalia, Y., 2002. Maximum likelihood estimation of discretely observed diffusions: a closed form approximation approach. Econometrica 70, 223–262.

Bally, V., Talay, D., 1996a. The law of the Euler scheme for stochastic differential equations (I): convergence rate of the distribution function. Probability Theory and Related Fields 104, 43–60.

Bally, V., Talay, D., 1996b. The law of the Euler scheme for stochastic differential equations (II): convergence rate of the density. Monte Carlo Methods and its Applications 2, 93–128.

Barberis, N., 2000. Investing for the long run when returns are predictable. Journal of Finance 55, 225–264.

Basawa, I.V., Scott, D.J., 1982. Asymptotic optimal inference for non-ergodic models. Lecture Notes in Statistics, vol. 17. Springer, New York.

Bawa, V.S., Brown, S.J., Klein, R.W., 1979. Estimation Risk and Optimal Portfolio Choice. North-Holland, New York.

Bibby, B.M., Sørensen, M., 1995. Martingale estimating functions for discretely observed diffusion processes. Bernoulli 1, 17–39.

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

Brandt, M., Santa-Clara, P., 2002. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. Journal of Financial Economics 63, 161–210.

Broze, L., Scaillet, O., Zakoian, J.-M., 1998. Quasi-indirect inference for diffusion processes. Econometric Theory 14, 161–186.

Chamberlain, G., 1987. Asymptotic efficiency in estimation with conditional moment restrictions. Journal of Econometrics 34, 305–334.

Chamberlain, G., 1992. Efficiency bounds for semiparametric regression. Econometrica 60, 567–596.

Chan, K.C., Karolyi, A., Longstaff, F.A., Saunders, A., 1992. A comparison of alternative models of the short-term interest rates. Journal of Finance 47, 1209–1227.

Chib, S., Elerian, O., Shephard, N., 2001. Likelihood inference for discretely observed non-linear diffusions. Econometrica 69, 959–993.

Dacunha-Castelle, D., Florens-Zmirou, D., 1986. Estimation of the coefficients of a diffusion from discrete observations. Stochastics 19, 263–284.

Detemple, J.B., Garcia, R., Rindisbacher, M., 2003. A Monte Carlo method for optimal portfolios. Journal of Finance 58, 401–446.

Detemple, J.B., Garcia, R., Rindisbacher, M., 2005. Representation formulas for Malliavin derivatives of diffusion processes. Finance and Stochastics 9, 349–369.

Doss, H., 1977. Liens entre équations différentielles stochastiques et ordinaires. Annales de l'Institut H. Poincaré 13, 99–125.

Duffie, D., Glynn, P., 1995. Efficient Monte Carlo simulation of security prices. Annals of Applied Probability 5, 897–905.

Duffie, D., Protter, P., 1992. From discrete to continuous finance: weak convergence of the financial gain process. Mathematical Finance 2, 1–15.

Duffie, D., Singleton, K., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61, 929–952.

Durham, G.B., Gallant, A.R., 2002. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. Journal of Business and Economic Statistics 20, 297–316.

Eraker, B., 2001. MCMC analysis of diffusion models with application to finance. Journal of Business and Economic Statistics 19, 177–191.

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

Florens-Zmirou, D., 1993. On estimating the diffusion coefficient from discrete observations. Journal of Applied Probability 30, 790–804.

Gaines, J.G., Lyons, T.J., 1997. Variable step size control in the numerical solution of stochastic differential equations. SIAM Journal of Applied Mathematics 57, 1455–1484.

Gallant, A.R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657–681.

Gallant, A.R., Tauchen, G., 2002. Simulated score methods and indirect inference for continuous-time models. Handbook of Financial Econometrics, forthcoming.

Genon-Catalot, V., 1990. Maximum contrast estimation for diffusion processes from discrete observations. Statistics 21, 99–116.

Genon-Catalot, V., Jacod, J., 1993. On the estimation of the diffusion coefficient for multidimensional diffusion processes. Annales de l'I.H.P., Probabilité et Statistiques 29, 119–151.

Gouriéroux, C., Montfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, 85–118.

Hansen, L., 1985. A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. Journal of Econometrics 30, 203–238.

Hansen, L.P., Scheinkman, J.A., 1995. Back to the future: generating moment implications for continuous-time Markov processes. Econometrica 63, 767–804.

Hansen, L.P., Heaton, J.C., Ogaki, M., 1988. Efficiency bounds implied by multi-period conditional moment restrictions. Journal of the American Statistical Association 83, 863–871.

Heyde, C.C., 1992. On best asymptotic confidence intervals for parameters of stochastic processes. Annals of Statistics 20, 603–607.

Jacod, J., Protter, P., 1998. Asymptotic error distributions for the Euler method for stochastic differential equations. Annals of Probability 26, 267–307.

Jakubowski, A., Mémin, J., Pagès, G., 1989. Convergence en loi des suites d'intégrals stochastiques sur l'espace $\mathbf{D}^1$ de Skorohod. Probability Theory and Related Fields 81, 111–137.

Kallenberg, O., 1997. Foundations of Modern Probability Theory. Springer, New York.

Kessler, M., Sørensen, M., 1999. Estimating equations based on eigenfunctions for a discretely observed diffusion process. Bernoulli 5 (2), 299–314.

Kloeden, P., Platen, E., 1997. Numerical Solution of Stochastic Differential Equations. Springer, Berlin.

Kurtz, T.G., Protter, P., 1991a. Wong–Zakai corrections, random evolutions and numerical schemes for SDE's. In: Stochastic Analysis. Academic Press, New York, pp. 331–346.

Kurtz, T.G., Protter, P., 1991b. Weak limit theorems of stochastic integrals and stochastic differential equations. Annals of Probability 19, 1035–1070.

Kurtz, T.G., Protter, P., 1996. Weak convergence and stochastic integrals and differential equations, 1995 CIME School in Probability. Lecture Notes in Mathematics, vol. 1627. Springer, Berlin, pp. 1–41.

Lo, A., 1988. Maximum likelihood estimation of generalized Ito processes with discretely-sampled data. Econometric Theory 4, 231–247.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Communications and Control Engineering Series. Springer, Berlin, New York.

Milshtein, G.N., 1984. Weak approximation of solutions of systems of stochastic differential equation. Theory of Probability and its Application 4, 750–768.

Milshtein, G.N., 1995. Numerical Integration of Stochastic Differential Equations. Kluwer Academic Publishers, New York.

Milshtein, G.N., Schoenmakers, J., Spokoiny, V., 2004. Transition density estimation for stochastic differential equations via forward-reverse representations. Bernoulli 10, 281–312.

Müller, U.U., Wefelmeyer, W., 2002. Autoregression, estimating functions, and optimality criteria. In Advances in Statistics, Combinatorics and Related Areas, Selected Papers from the SCRA2001-FIM VIII Proceedings of the Wollongong Conference. World Scientific, Singapore.

Newey, W.K., 1990. Semiparametric efficiency bounds. Journal of Applied Econometrics 5, 90–135.

Nualart, D., 1995. The Malliavin Calculus and Related Topics. Springer, New York.
Pedersen, A.R., 1995a. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scandinavian Journal of Statistics 22, 55–71.
Pedersen, A.R., 1995b. Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. Bernoulli 1, 257–279.
Santa-Clara, P., 1995. Simulated likelihood estimation of diffusions with and application to the short term interest rate. Ph.D. Dissertation, INSEAD.
Smith, A.A., Jr. 1990. Three Essays on the Solution and estimation of dynamic macroeconomic models. Ph.D. Thesis, Duke University.
Talay, D., 1984. Efficient numerical schemes for the approximation of expectations of functionals of S.D.E. In: Korezioglu, H., Maziotto, G., Szpirglas, J. (Eds.), Filtering and Control of Random Processes, Lecture Notes in Control and Information Sciences, vol. 61, Springer, New York, pp. 294–313.
Talay, D., 1986. Discrétisation d'une équation différentielle stochastique et calcul approché d'éspérance de fonctionelles de la solution. Mathematical Modelling and Numerical Analysis 0, 141–179.
Talay, D., 1996. Probabilistic numerical methods for partial differential equations: elements of analysis. CIME School in Probability, Lecture Notes in Mathematics, vol. 1627. Springer, Berlin, pp. 149–196.
Talay, D., Tubaro, L., 1990. Expansion of the global error for numerical schemes solving stochastic differential equations. Stochastic Analysis and its Application 8, 483–509.
Wefelmeyer, W., 1996. Quasi-likelihood models and optimal inference. Annals of Statistics 24, 405–422.
Wong, E., Zakai, M., 1964. On the convergence of ordinary integrals to stochastic integrals. Annals of Mathematical Statistics 36, 1560–1564.