

Asymptotic Properties of Monte Carlo Estimators of Diffusion Processes.

by

Jérôme Detemple, René Garcia and Marcel Rindisbacher.*

This draft: March 30, 2003.

Detemple is from Boston University School of Management and CIRANO, Garcia is from the Economics Department, Université de Montréal, CIRANO and CIREQ and Rindisbacher is from the Rotman School of Management, University of Toronto and CIRANO. 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). This work is funded by MITACS. Address for correspondence: René Garcia, Département de Sciences Économiques et CIREQ, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, Québec, H3C 3J7, Canada.

Abstract

This paper studies the limit distributions of Monte Carlo estimators of diffusion processes. Two types of estimators are examined. The first one is based on the Euler scheme applied to the original processes; the second applies the Euler scheme to a variance-stabilizing transformation of the processes. We show that the transformation increases the speed of convergence of the Euler scheme. The limit distribution of this estimator is derived in explicit form and is found to be non-centered. We also study estimators of conditional expectations of diffusions with known initial state. Expected approximation errors are characterized and used to construct second-order bias corrected estimators. Such bias correction eliminates the size distortion of asymptotic confidence intervals and allows to examine the relative efficiency of estimators. Finally, we derive the limit distributions of Monte Carlo estimators of conditional expectations with unknown initial state. The variance-stabilizing transformation is again found to increase the speed of convergence. For comparison we also study the Milshstein scheme. We derive new convergence results for this scheme and show that it does not improve on the convergence properties of the Euler scheme with transformation. Our results are illustrated in the context of a dynamic portfolio choice problem and of simulated-based estimation of diffusion processes.

1 Introduction

In many financial applications such as portfolio allocation, asset pricing and risk management, we need 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)

$$dX_v = A(X_v)dv + B(X_v)dW_v; \quad X_t = x. \tag{1}$$

This computation is often performed using Monte Carlo (MC) simulation. An advantage of the procedure is that it grows only linearly with the number of variables involved in the simulation. Therefore, for high dimensional multivariate diffusion problems, it becomes the only feasible approach.¹ One drawback of MC methods is their slower convergence rate, but this shortcoming can be mitigated by choosing an efficient design.² To identify an efficient design one needs to characterize the approximation errors associated with a given Monte Carlo scheme. The purpose of this paper is to study the limit distributions of several Monte Carlo estimators of conditional expectations of diffusions in order to develop efficient Monte Carlo designs and build confidence intervals around the estimated expectations.

When the transition density of the diffusion is unknown, which is often the case in financial models, the MC computation of the conditional expectation involves two types of approximations. The first one is an approximation of the terminal value X_T , of the solution of (1). Several discretization schemes of this SDE can be used for this purpose.³ The most popular, perhaps because of its ease of implementation, is the Euler scheme, an iterative procedure which evaluates the drift and the 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, which is performed by averaging over a finite sample of approximated terminal values $X(T)$. Justification for this averaging is based on the law of large numbers. The combination of these two operations, labeled MCE (*Monte Carlo with Euler discretization*), produces an estimate of $f(t, x)$ which involves two types of errors, (i) a discretization error and (ii) an averaging error. It is the knowledge of these two errors which is important for the design of an efficient MCE scheme. We provide a detailed analysis of these two approximation errors.

Error properties for approximations of solutions of SDEs (the first type of error) have been studied before. For the Euler discretization scheme the asymptotic distribution of the error was found by

¹Partial Differential Equations (PDE) or finite state Markov chains (MKC) are competing procedures which produce numerical approximations of $f(t, x)$. These two methods (PDE and MKC) are deterministic algorithms whose computational cost grows exponentially with the dimension of the problem. In implementations, this cost quickly becomes prohibitive, even for simple problems. The literature on the speed of convergence of finite difference approximations of Bellman equations is sparse. One of the few contributions is a recent paper by Krylov (2000).

²Moreover, the control of errors and the parallelization of the numerical algorithm are easier to perform with a MC technique because innovations in the underlying uncertainty are independent.

³A detailed analysis of discretization schemes available can be found in Kloeden and Platen (1997).

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 (1979)), which reduces the diffusion coefficient of the SDE to unity. This transformation has enjoyed recent popularity in financial econometrics (see, for instance, Ait-Sahalia (2002) and Durham and Gallant (2002)) and has been used for the computation of optimal portfolios by Detemple, Garcia and Rindisbacher (2003). We show that a Doss transformation of the SDE can improve the speed of convergence of the discretization scheme since 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 stands in contrast with the simple Euler scheme applied to the original (non-transformed) SDE which produces an error whose asymptotic law is centered.

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. As a result, the efficient procedure has a second-order bias. In this paper we explicitly characterize the second-order bias as the expected value of a known random variable. Since this random variable can be simulated along with the diffusion, it can be used to design a new approximation that corrects for second-order bias. The bias corrected estimate is shown to be asymptotically equivalent to a Monte Carlo procedure which samples directly from the true distribution of the terminal point of the diffusion.⁴

In applications one often requires approximations of the function $f(X_s, s)$ at future dates $s > t$. For instance, for asset allocation purposes, one needs to evaluate the performance of a portfolio relative to a selected benchmark at a given horizon (market timing experiments). In these instances the unknown function is also evaluated at an unobserved point X_s , which must be estimated along with the function itself. This evaluation can be performed in two ways, using the Doss transformed state variables or the original SDE (1). We derive the asymptotic distributions of the errors involved. Again, we find that the procedure involving a Doss transformation increases the speed of convergence to the true value.

To provide additional perspective we also study the Milshtein second-order scheme (Milshtein (1995)). We show that this scheme does not dominate the Euler scheme with transformation, as far as convergence properties are concerned. This highlights the benefits of the transformation. We also argue that the transformation is computationally less costly. Our results for the weak limit of the solution of SDEs based on the Milshtein scheme and the second-order bias are new and therefore of independent interest.

We illustrate our results with two applications. First, we consider a dynamic portfolio allocation

⁴Our method is simpler than the one in Talay and Tubaro (1991) since 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.

problem. We assume that there is only one state variable, the short rate, and that the market price of risk is constant, for simplicity. To model the evolution of the short rate, we retain the specification introduced in Detemple, Garcia and Rindisbacher (2003) where the short rate is a non-linear diffusion process with mean reversion and a constant elasticity of variance. This specification admits as particular cases the standard CIR and CEV processes. In a dynamic asset allocation problem, the portfolio shares can be expressed as conditional expectations of random variables which depend on the paths of the state variables. We compute the asymptotic laws for the portfolio shares and examine the asymptotic error distribution of future hedging demands when the future value of the state variable is unknown. Second, we examine simulation-based inference for diffusions. In this context we propose an efficient simulation-based estimator of the estimating function and compare it to the simulated maximum likelihood estimators proposed in Pedersen (1995a,b) and Brandt and Santa Clara (2002).

The rest of the paper is organized as follows. In section 2 we study the asymptotic error distribution of approximations of solutions of SDEs and provide numerical illustrations for processes often used in finance. Section 3 describes the asymptotic laws of estimators of conditional expectations, where we distinguish between the case where the conditioning state variable is known and the case where it is unknown. For both cases, we characterize expected approximation errors, second-order discretization biases and bias corrected estimators. New asymptotic convergence results for the Milstein scheme are derived in section 4. Section 5 provides applications of the results to dynamic asset allocation and simulation-based inference. 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. Appendix C reformulates portfolio weights as conditional expectations of solutions of SDEs.

2 Asymptotic laws of estimators of solutions to SDEs

Many continuous-time financial models can be represented by multivariate diffusions with general drift and diffusion functions. The solution of the financial application, whether it is asset pricing, portfolio allocation or risk management, relies on the simulation of a discretized version of the stochastic differential equation. 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 approximations of solutions to SDEs. We first review results on the convergence of the Euler scheme (subsection 2.1). Next we study the Euler scheme applied to a Doss transformation of the state variables (subsection 2.2).⁵ The importance of obtaining an 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 the approximation error in a finite simulation experiment using a

⁵To simplify the presentation we assume homogeneous dynamics of the process to be simulated.

simulated benchmark for the true value X_T , no matter how finely we sample the process.

2.1 Euler approximation without transformation

To set up the stage for convergence results with the transformation and the Milshtein scheme, we recall known results for the Euler scheme without transformation. They also play a key role in finding our explicit expression for the second-order bias. Consider the random variable X_T given by the terminal value of the solution of the SDE, starting at $t = 0$.

$$dX_v = A(X_v)dv + \sum_{j=1}^d B_j(X_v)dW_v^j \quad (2)$$

with $d \times 1$ vectors A and B_j where $A \in \mathcal{C}^1(\mathbb{R}^d)$ and $B_j \in \mathcal{C}^1(\mathbb{R}^d)$ ⁶ and at most of linear growth. Its Euler approximation is

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

where $h = T/N$ and $\Delta W_{nh}^j = W_{(n+1)h}^j - W_{nh}^j$.⁷

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

Theorem 1: *The approximation error $X_T^N - X_T$ converges weakly⁸ at the rate $\frac{1}{\sqrt{N}}$ (i.e. $\sqrt{N}(X_T^N - X_T) \Rightarrow U_T^X$). The asymptotic error is*

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

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

$$\Omega_v = \mathcal{E}^R \left(\int_0^{\cdot} [\partial A](X_s) ds + \sum_{j=1}^d \int_0^{\cdot} \partial B_j(X_s) dW_s^j \right)_v$$

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 $\mathcal{E}^R(\cdot)$ denotes the right stochastic exponential.⁹

⁶The space $\mathcal{C}^1(\mathbb{R}^d)$ denotes the space of once continuously differentiable \mathbb{R}^d valued functions.

⁷To simplify the notation we restrict the error analysis to equidistant discretization schemes.

⁸Let S be a metric space and \mathcal{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 $\int_S f(s) d\mathbf{P}_{X^N}(s) \rightarrow \int_S f(s) d\mathbf{P}_X(s)$ for all continuous and bounded functions f on S .

⁹For a semimartingale M , the right stochastic exponential $Z_v = \mathcal{E}^R(M)_v$ is the unique solution to the $d \times d$ matrix SDE $dZ_v = dM_v Z_v$ with $Z_0 = I_d$.

Theorem 1 says that the error converges in law at the rate $\frac{1}{\sqrt{N}}$ to the random variable U_T^X as the number of discretization points N increases. The asymptotic error U_T^X depends on the coefficients of the SDE and on their derivatives. Surprisingly, it also depends on new Brownian motions ($[Z^{l,j}]_{l,j \in \{k, \dots, d\}}$), 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

$$dX_v = aX_v dv + bX_v dW_v. \quad (4)$$

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

Given the scaling property of the Brownian motions $Z^{l,j}$ we see that the asymptotic distribution, in the univariate case, is always a mixture of normals. For example for a CIR process

$$dX_v = \kappa(\bar{X} - X_v)dv + \sigma\sqrt{X_v}dW_v, \quad (5)$$

we find that: $U_T^X = -\frac{\sigma^2}{2\sqrt{2}}\Omega_T^{-1} \int_0^T \Omega_v^{-1} dZ_v = Z_{\frac{\sigma^4}{8}\Omega_T^2 \int_0^T \Omega_v^{-2} dv}$. However, in this example, we do not know the explicit law of the mixing random variable $\frac{\sigma^4}{8}\Omega_T^2 \int_0^T \Omega_v^{-2} dv$ since $d\Omega_v = (-\kappa dv + (\frac{\sigma}{2}/\sqrt{X_v})dW_v)\Omega_v$ where $\Omega_0 = 1$ and have to rely on numerical procedures. At the end of this section we will provide numerical results for the CIR process as well as for the constant elasticity of variance (CEV) process.

The fact that the asymptotic distribution depends on an independent Brownian motion which 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 the expression for the asymptotic distribution. If we do not know the exact expression for the asymptotic error U_T^X , error analysis using a simulated benchmark $\sqrt{N}(X_T^N - X_T^{N^*})$ with N^* large will always just depend on the original Brownian motions W^j and not on the independent Brownian motions $Z^{l,j}$ which arise in the functional describing the weak limit. Therefore a simulated benchmark cannot be used to assess the true properties of the error distribution. The independent Brownian motions $Z^{l,j}$, characterizing the asymptotic distribution U_T^X , will never emerge with finite N^* .

Theorem 1 illustrates the drawback of Monte Carlo methods, which have a convergence rate of $\frac{1}{\sqrt{N}}$. In the next sub-section we will introduce a useful change of variables which simplifies the volatility coefficient of the underlying SDE, but most importantly which will allow us to construct an approximation of the true value X_T with an improved rate of convergence of $1/N$.

2.2 Euler approximation with Doss transformation

We first introduce the transformation then establish the limit law of the Euler approximation of the terminal value of the transformed SDE. Consider the new state variables $\hat{X}_v = F(X_v)$, where $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the function whose inverse $G : \mathbb{R}^d \rightarrow \mathbb{R}^d$ exists and satisfies

$$\partial_{z_j} G_i(z) = B_{i,j}(G(z)). \quad (6)$$

An application of Ito's lemma shows that the i^{th} component $\hat{X}_{i,v} = F_i(X_v)$ satisfies

$$d\hat{X}_{i,v} = \left(\partial F_i(X_v) A(X_v) + \frac{1}{2} \partial^2 F_i(X_v) \square (B(X_v) B'(X_v)) \right) dv + \left(\sum_{j=1}^d \partial F_i(X_v) B_j(X_v) dW_v^j \right)$$

where $\partial F_i(X_v)$ denotes the $1 \times d$ gradient of F_i , $\partial^2 F_i(X_v)$ is the $d \times d$ Hessian and, for two matrices A, B the operation $A \square B = \sum_{i,j} A_{i,j} B_{i,j}$. Thus, $\partial F_i(X_v) A(X_v)$, $\partial F_i(X_v) B_j(X_v)$ and $\partial^2 F_i(X_v) \square (B(X_v) B'(X_v))$ all have dimension 1×1 . Since $\partial F_i B_j = \mathbf{1}_{\{i=j\}}$ (i.e. $\partial F B$ is the identity matrix) the vector \hat{X}_v evolves according to

$$d\hat{X}_v = \hat{A}(\hat{X}_v) dv + dW_v$$

where $W_v = [W_v^j]_{j=1,\dots,d}$, $\hat{A}(\hat{X}_v) = [\hat{A}_1(\hat{X}_v), \dots, \hat{A}_n(\hat{X}_v)]'$ with $\hat{A}_i(\hat{X}_v)$ given by the drift of $\hat{X}_{i,v}$. For the sake of generality and to cover some specific applications (such as portfolio choice) we consider a slightly more general version of this equation, namely,

$$d\hat{X}_v = \hat{A}(\hat{X}_v) dv + \sum_{j=1}^d \hat{B}_j dW_v^j \quad (7)$$

where $\hat{A}(\hat{X}_v)$ is a $d \times 1$ vector of functions and \hat{B}_j is a $d \times 1$ vector of constants (not necessarily given by the unit vector).

The corresponding Euler approximation for these transformed state variables 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,

Theorem 2: *Suppose that condition (6) holds. The approximation error $\hat{X}_T^N - \hat{X}_T$ converges weakly at the rate $\frac{1}{\sqrt{N}}$ (i.e. $N(\hat{X}_T^N - \hat{X}_T) \Rightarrow U_T^{\hat{X}}$). The asymptotic error is*

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

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

$$\hat{\Omega}_v = \mathcal{E}^R \left(\int_0^{\cdot} \partial \hat{A}(\hat{X}_s) ds \right)_v$$

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

The simplest example is the Ornstein-Uhlenbeck process

$$dX_v = \kappa(\bar{X} - X_v)dv + \sigma dW_v. \quad (9)$$

In this case the asymptotic distribution $U_T^{\hat{X}}$ is given by the sum of two normal random variables

$$U_T^{\hat{X}} = \frac{\kappa}{2} e^{-\kappa T} \int_0^T e^{\kappa v} dX_v + \frac{\kappa \sigma}{\sqrt{12}} e^{-\kappa T} \int_0^T e^{\kappa v} dZ_v = \frac{\kappa^2}{2} (\bar{X} - X_0) e^{-\kappa T} T + W_{\alpha(T)} + Z_{\beta(T)} \quad (10)$$

where $\alpha(T) \equiv \frac{\sigma^2 \kappa (e^{2\kappa T} - 1 + 2\kappa T(1 - \kappa T))}{16e^{2\kappa T}}$ and $\beta(T) \equiv \frac{\kappa \sigma^2}{24} (1 - e^{-2\kappa T})$. If $X_0 \neq \bar{X}$ the asymptotic law is non-centered.

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

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

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

2.3 A numerical example

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

$$dX_v = \kappa(\bar{X} - X_v)dv + \sigma X_v^\gamma dW_v. \quad (11)$$

This specification is often used to model the evolution of the short rate in term structure models, as in Chan et al. (1992). For $\gamma = \frac{1}{2}$ we obtain a CIR process, otherwise it is a standard CEV process.

To characterize the asymptotic error distribution 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 ¹⁰

¹⁰Given two functions f and g , the notation \circ reads $[f \circ g](x) \equiv f(g(x))$.

$$d\hat{X}_v = \left[\left[\frac{\partial A}{B} - \frac{1}{2}\partial B \right] \circ G \right] (\hat{X}_v) + dW_v \quad (12)$$

where $G(x) = (\sigma x)^{-\frac{1}{\gamma}}$. Similarly, expressions for $\partial \hat{A}$ and $\partial^2 \hat{A}$, which appear in $\hat{\Omega}_T$ and $U_T^{\hat{X}}$, are respectively $\partial \hat{A}(x) = \left[\left[\partial A - \frac{A\partial B}{B} - \frac{1}{2}\partial^2 BB \right] \circ G \right] (x)$ and

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

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

For the CEV process we adopt the parameter values in Chan et al. (1992). Parameter values for the CIR process are taken from Broze et al. (1995). Table 1 provides these values.

Figures 1 and 2 present the asymptotic error distribution for CIR and CEV respectively. The simulations are based on $T = 1$ and $N = 365$. The graphs plot the empirical distribution function based on $M = 50000$ replications. We clearly see that the asymptotic distributions using the transformation are non-centered and that the error is considerably smaller with the transformation.

A comparison of figures 3 and 4 for the CIR process and 5 and 6 for the CEV process illustrates the increase in the speed of convergence due to the transformation. In those experiments the benchmark “true” value is computed without the transformation and taking $N = 2^{14}$.¹¹ Approximation errors, relative to this benchmark, are then computed using $N = 2^x$ with $x = 2, \dots, 9$ to measure the convergence. Again the distribution functions are based on $M = 50000$ replications. The results clearly illustrate that both the speed of convergence and the distribution of the asymptotic error are important to assess the properties of the estimation procedure.

The simulation of the discretized version of the process of interest is usually the first step of a Monte Carlo procedure for a financial application such as computing the price of a European option. In the case of a stock option, X_T represents the final value of the stock price given the price today and determines the option payoff under the sampled trajectory. The next step is typically to generate a large number of such discretized trajectories to compute the option price as an average of the payoff obtained for each trajectory. Common wisdom suggests that the precision of this price estimator can be improved by discretizing the process as finely as possible and taking a very large number of independent replications. However, in practice, one faces a limited budget of computation time. Duffie and Glynn (1995) propose a computationally efficient trade-off between reducing the length of the discretization step and 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. As a result, the efficient procedure has a second-order bias. In the next section, we extend their results in several dimensions. First, we explicitly characterize the second-order bias as the expected value of a known random variable. Since this random variable can be simulated along with the diffusion, we design a new

¹¹We take this shortcut to illustrate the relative speed of convergence.

approximation that corrects for second-order bias. We also provide equivalent results for the Doss-transformed process introduced in the previous section. Finally, we provide the error distributions, characterize the second-order bias and derive the bias-corrected conditional estimator when the conditional expectations to be estimated depend on state variables that are not known, a case often found in practical applications.

3 Asymptotic laws of estimators of conditional expectations

We now derive the asymptotic distribution 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 based on a law of large numbers for independent random variates. This error will not disappear even if sampling is from the true distribution. In the following subsections, we characterize both components to obtain an explicit expression for the asymptotic distribution of the error in the approximation of the conditional expectation. We distinguish the case where the conditional expectation depends on a known state variable from the one where it is unknown in order to cover different financial applications.

3.1 Expectations conditional on a known state variable

In this subsection, we wish to calculate $\mathbf{E}[g(X_T)|\mathcal{F}_0] \equiv \mathbf{E}_0[g(X_T)]$ where X solves (2) or (7). To approximate the expectations we rely on a law of large numbers and draw independent replications $X_T^{i,N}$, respectively $\hat{X}_T^{i,N}$, of the terminal points X_T^N , respectively \hat{X}_T^N , of the Euler discretized diffusion with and without Doss transformation. Our next theorem describes the asymptotic laws for these two Euler schemes.

Theorem 3: *Suppose that the assumptions of Theorem 2 hold. Let $g \in \mathcal{C}^1(\mathbb{R}^d)$ such that $g(X_T) \in \mathbb{D}^{1,2}$. For the schemes without and with transformation, we respectively have*

$$\begin{aligned} \sqrt{M} \left(\frac{1}{M} \sum_{i=1}^M g(X_T^{i,N_M}) - \mathbf{E}_0[g(X_T)] \right) &\Rightarrow \epsilon \frac{1}{2} K_T(X_0) + L_T(X_0) \\ \sqrt{M} \left(\frac{1}{M} \sum_{i=1}^M \hat{g}(\hat{X}_T^{i,N_M}) - \mathbf{E}_0[\hat{g}(\hat{X}_T)] \right) &\Rightarrow \epsilon \frac{1}{2} \hat{K}_T(\hat{X}_0) + L_T(\hat{X}_0) \end{aligned}$$

where $\lim_{M \rightarrow \infty} N_M = +\infty$ and $\epsilon = \lim_{M \rightarrow \infty} \frac{\sqrt{M}}{N_M}$, and $L_T(X_0), L_T(\hat{X}_0)$ are the terminal values of centered Gaussian martingales with quadratic variation and conditional variance given by

$$[L, L]_T = \int_0^T \mathbf{E}_0[N_v(N_v)'] dv \equiv \mathbf{var}[g(X_T)|\mathcal{F}_0]$$

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

In these expressions $\mathcal{D}_s X_T$ is the Malliavin derivative of X_T . The deterministic functions K_T and \hat{K}_T are derived in Theorems 4 and 5 of the next subsection.

The random variable $\mathcal{D}_s X_T$ 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 response function which quantifies the sensitivity of the variable X_T to a past uncertainty shock at time s .¹²

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 we could sample directly 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. In our next section we provide intuition for this result. At this stage we remark that the common convergence rate is the rate achieved by the Euler scheme applied to an ODE. Since the rate $\frac{1}{\sqrt{M}}$ is obtained from a central limit theorem, an immediate conclusion is that higher order schemes would fail to improve the convergence speed. They would just reduce the second-order bias.

3.1.1 Expected approximation errors

The second-order discretization biases K_T and \hat{K}_T , which appear in Theorem 3, are obtained by studying the convergence of the expected approximation errors. This section is devoted to this issue.

Euler scheme on the original state variables Our first result describes the convergence of the expected approximation error for the Euler scheme (3), $\mathbf{E}_0[g(X_T^N) - g(X_T)]$, where $g(X_T^N)$ is an approximation of $g(X_T)$ based on the discretization of the process X .

¹²We also assume that $g(X_T)$ is in the domain $\mathbb{D}^{1,2}$ of the Malliavin derivative operator. For a brief introduction to Malliavin calculus see Appendix D in Detemple, Garcia and Rindisbacher (2003). Nualart (1995) provides an exact definition of the domain of the Malliavin derivative operator $\mathbb{D}^{1,2}$.

Define the random variables

$$\begin{aligned}
V_{1,T} &= -\Omega_T \int_0^T \Omega_s^{-1} \left(\partial A(X_s) dX_s + \sum_{j=1}^d [\partial B_j A - \sum_{i=1}^d (\partial B_j)(\partial B_i) B_j](X_s) dW_s^j \right) \\
&\quad + \Omega_T \int_0^T \Omega_s^{-1} \sum_{j=1}^d \left([\partial B_j \partial B_j A](X_s) - \int_0^T \sum_{k,l=1}^d [\partial_k (\partial_l A B_{l,j}) B_{k,j}](X_s) \right) ds \\
&\quad + \Omega_T \int_0^T \Omega_s^{-1} \sum_{i,j=1}^d ([\partial [\partial B_i \partial B_j B_i] B_i - \partial B_i \partial B_i \partial B_j B_i](X_s)) ds \\
V_{2,T} &= - \int_0^T \sum_{i,j=1}^d \nu_{i,j}(X_s, \Omega_s) ds
\end{aligned}$$

where $\partial A, \partial B_j$ are $d \times d$ matrices of Jacobians, Ω is defined in Theorem 1 and

$$\nu_{i,j}(X_t, \Omega_t) dt = d[[(B^i)'(\partial^2 g)(\partial B_i)(\partial B_j) B_i + (\partial g) \Omega_T (\partial B_i) \Omega_T^{-1} (\partial B_i)(\partial B_j) B_i](X), W^i]_t.$$

An explicit expression for $\nu_{i,j}(X, \Omega)$ is given in Appendix B.

With this notation we have

Theorem 4: *Let $g \in \mathcal{C}^3(\mathbb{R}^d)$ be such that*

$$\lim_{r \rightarrow \infty} \limsup_N \mathbf{E}_0 \left[\mathbf{1}_{\{|N(g(X_T^N) - g(X_T))| > r\}} N |g(X_T^N) - g(X_T)| \right] = 0 \quad (13)$$

(**P**-a.s.). *Then we have that*

$$N \mathbf{E}_0 [g(X_T^N) - g(X_T)] \rightarrow \frac{1}{2} K_T(X_0) \equiv \frac{1}{2} \mathbf{E}_0 [\partial g(X_T) V_{1,T} + V_{2,T}] \quad (14)$$

Theorem 4 can be viewed as a full probabilistic counterpart to the results in Talay and Tubaro (1991) and Bally and Talay (1996a,b). These authors show that the expected approximation error of the Euler scheme can be written in terms of the expectations of an unknown function of a random variable, where the function solves a PDE. Our proof uses probabilistic arguments to derive a representation in the form of a conditional expectation of a known random variable whose components depend on the coefficients of the underlying SDE. This characterization is easier to evaluate since it does not require to solve a PDE. As a result, it does not suffer from the curse of dimensionality which affects the numerical solutions of the PDEs characterizing the second-order bias in Talay and Tubaro (1991) and remains applicable for multivariate diffusions. Even in the univariate case, using Monte Carlo simulation in combination with the solution of a PDE is computationally very costly. This may explain why the theoretical results of Talay and Tubaro (1991) and Talay and Bally (1996a,b) have not been used in applications.

Euler scheme on the transformed state variables Let \hat{g} be such that $g(X_T) = [\hat{g} \circ G](\hat{X}_T)$ and define the random variable

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

with $\hat{\Omega}_T = \exp \left(\int_0^v \partial \hat{A}(\hat{X}_s) ds \right)$. For the scheme based on the transformed state variables we obtain,

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

$$\lim_{r \rightarrow \infty} \limsup_N \mathbf{E}_0 \left[\mathbf{1}_{\{|N(\hat{g}(\hat{X}_T^N) - \hat{g}(\hat{X}_T))| > r\}} N |\hat{g}(\hat{X}_T^N) - \hat{g}(\hat{X}_T)| \right] = 0 \quad (15)$$

P-a.s. we have

$$N \mathbf{E}_0 \left[\hat{g}(\hat{X}_T^N) - \hat{g}(\hat{X}_T) \right] \rightarrow \frac{1}{2} \hat{K}_T(\hat{X}_0) \equiv \frac{1}{2} \mathbf{E}_0 \left[\partial \hat{g}(\hat{X}_T) \hat{V}_T \right]. \quad (16)$$

A comparison of (14) with (16) suggests that it will be difficult, in general, to establish the dominance of one method over the other on the basis of the expected approximation error. Indeed, the formulas show that both methods converge at the same speed $1/N$ and, while the second-order biases differ ($K_T(X_0) \neq \hat{K}_T(\hat{X}_0)$), they do not appear to be ordered in any systematic way. To compare the two methods one may want to take the computational cost into account. For instance, in the portfolio application considered later in this paper, it is more costly to calculate K_T than \hat{K}_T .

The difference in the rates of convergence for the limit distribution (Theorems 1 and 2) and the expected approximation error (Theorems 4 and 5) can be explained as follows. The speed of convergence for the limit distribution is determined by the martingale part of the error expansion which 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 approximation 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.¹³

3.1.2 Second-order discretization biases

Going back to Theorem 3 we note that the two procedures have asymptotic second-order discretization biases, respectively given by $\frac{\epsilon}{2} K$ and $\frac{\epsilon}{2} \hat{K}$. Thus, any confidence interval based solely on the Gaussian process L will suffer from a size distortion. In fact, a confidence interval of nominal size α , based on L , will cover the true value $E_0[g(X_T)]$ only with probability $\Phi(\Phi^{-1}(1-\alpha/2)-\delta) - \Phi(\Phi^{-1}(\alpha/2)-\delta)$,

¹³The martingale part of the error converges to the product of a random variable and a stochastic integral with independent integrator.

where Φ is the cumulative Gaussian distribution and $\delta = \epsilon K_T(X_0)/2\sqrt{\mathbf{var}[L_T|\mathcal{F}_0]}$, instead of probability $1 - \alpha$. For the method with transformation the coverage probability of the true value is $\Phi(\Phi^{-1}(1 - \alpha/2) - \hat{\delta}) - \Phi(\Phi^{-1}(\alpha/2) - \hat{\delta})$ with $\hat{\delta} = \epsilon \hat{K}_T(\hat{X}_0)/2\sqrt{\mathbf{var}[L_T|\mathcal{F}_0]}$.¹⁴ The degree of size distortion is then measured by $s(z) \equiv 1 - \alpha - (\Phi(-\Phi^{-1}(\alpha/2) - z) - \Phi(\Phi^{-1}(\alpha/2) - z))$, where $z \in \{\delta, \hat{\delta}\}$. We conclude that these confidence intervals are valid if and only if there is no second-order bias. 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 the formulas (14)-(16) for the second-order biases $K_T(X_0)$ and $\hat{K}_T(\hat{X}_0)$, 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.¹⁵

3.1.3 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 follows. Let

$$g_{cT}^{N,M} = \frac{1}{M} \sum_{i=1}^M \left[g(X_{Nh}^{i,N}) + \frac{1}{2} \partial g(X_{Nh}^{i,N}) C_{1,Nh}^{i,N} + \frac{1}{2} C_{2,Nh}^{i,N} \right] \quad (17)$$

$$\hat{g}_{cT}^{N,M} = \frac{1}{M} \sum_{i=1}^M \left[\hat{g}(\hat{X}_{Nh}^{i,N}) + \frac{1}{2} \partial \hat{g}(\hat{X}_{Nh}^{i,N}) \hat{C}_{Nh}^{i,N} \right] \quad (18)$$

where $C_{1,Nh}^{i,N}$, $C_{2,Nh}^{i,N}$, $\hat{C}_{Nh}^{i,N}$, for $n = 0, \dots, N - 1$, are defined in Appendix B. Our next result shows that $g_{cT}^{N,M}$ and $\hat{g}_{cT}^{N,M}$ are bias-corrected estimators.

¹⁴When $M \rightarrow \infty$ we have

$$\begin{aligned} \mathbf{P}(\mathbf{E}_0[g(X_T)] \in [\frac{1}{\sqrt{M}} \sum_{i=1}^M g(X^{i,N_M}) + \Phi^{-1}(\alpha/2) \frac{\sigma^{M,N_M}}{\sqrt{M}}, \frac{1}{\sqrt{M}} \sum_{i=1}^M g(X^{i,N_M}) - \Phi^{-1}(\alpha/2) \frac{\sigma^{M,N_M}}{\sqrt{M}}]) \\ \downarrow \\ \Phi(-\Phi^{-1}(\alpha/2) - \delta) - \Phi(\Phi^{-1}(\alpha/2) - \delta) \end{aligned}$$

where $(\sigma^{M,N})^2 = \mathbf{VAR}^{M,N}[L_T|\mathcal{F}_0]$ is a convergent estimator of the variance. These confidence intervals have the correct length if and only if the second-order bias vanishes, i.e. $\delta = 0$.

¹⁵Duffie and Glynn (1995) also propose to use the Richardson-Romberg type of estimator $\frac{2}{4M} \sum_{i=1}^{4M} g(X_T^{i,2N}) - \frac{1}{M} \sum_{i=1}^M g(X_T^{i,N})$ to eliminate approximately the second-order bias. To calculate this estimator one must quadruple the number of replications and double the number of discretization points. Since in our method the second-order bias can be simulated along with the state variables, it is computationally less costly than Richardson-Romberg approximation schemes.

Theorem 6: *Suppose that the conditions of Theorem 2 hold. Then, $\sqrt{M}(g_{cT}^{N,M} - \mathbf{E}_0[g(X_T)]) \Rightarrow L_T(X_0)$ and $\sqrt{M}(g_{cT}^{N,M} - \hat{g}_{cT}^{N,M}) \Rightarrow 0$ as $N, M \rightarrow \infty$.*

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

3.2 Expectations conditional on an unknown state

In some applications the conditional expectations to be estimated depend on state variables that are not known. For instance, in market timing experiments, future trajectories of optimal portfolios and hedging terms are simulated. In the area of risk management, measures of riskiness of future imperfectly hedged positions are of interest to a variety of financial participants including speculators, hedgers and issuing firms. Likewise the evolution of the volatility of a stock over time is of interest to equityholders and debtholders of firms. To price American options using Monte Carlo methods, we have to calculate the continuation value backwards. This value is given as the expectation of the remaining pay-offs to maturity conditional on unknown future states.

Formally, all these problems are repeated regression problems. We want to estimate $\mathbf{E}[g(X_T)|X_\tau]$, where X_τ is unknown but X_0 has been observed. Hence, X_τ is replaced by the simulated value $X_\tau^N(X_0)$. The novel difficulty which emerges in these situations is that future estimates are conditional on (future) state variables whose distribution is unknown and requires estimation.

Estimators can be constructed as follows. Suppose that information about the current state is available (X_0 is known) and that we are trying to estimate a conditional expectation at τ , which depends on the unknown value X_τ . An estimate is obtained by first simulating a trajectory up to τ to construct the feasible predictor of the future state $X_\tau^N(X_0)$. The predicted state value then serves as a starting point for M trajectories which are used to estimate the conditional expectation at τ .

Approximation errors for these estimators include, again, two sources of errors, a discretization error and a Monte Carlo error. But now the need to discretize the SDE has two effects. As in the prior case the random variables inside the conditional expectation must be simulated since direct sampling from the true distributions is not feasible. In addition the future state at which the estimation is to be performed can only be approximated.

3.2.1 Error distributions

Consider the estimator

$$g_{\tau,T}^{N,M}(X_\tau^N) = \frac{1}{M} \sum_{i=1}^M g(X_T^{i,N}(X_\tau^{i,N}))$$

where $X_T(X_\tau)$ is the process X started at X_τ with $T > \tau$, which solves

$$X_T(X_\tau) = X_\tau + \int_\tau^T A(X_s(X_\tau))ds + \sum_{j=1}^d \int_\tau^T B_j(X_s(X_\tau))dW_s.$$

Also let $X_\tau^N(X_\tau^N)$ denote the numerical approximation of $X_T(X_\tau)$. We will give the asymptotic distributions of the errors with and without transformation

$$U_{\tau,T}^{g^{N,M}} = \frac{1}{\sqrt{M}} \sum_{i=1}^M \left(g(X_T^{i,N}(X_\tau^N)) - \mathbf{E}_\tau [g(X_T(X_\tau))] \right)$$

$$U_{\tau,T}^{\hat{g}^{N,M}} = \frac{1}{\sqrt{M}} \sum_{i=1}^M \left(g(\hat{X}_T^{i,N}(\hat{X}_\tau^N)) - \mathbf{E}_\tau [g(\hat{X}_T(\hat{X}_\tau))] \right).$$

Theorem 7: *Suppose that the conditions of Theorem 3 hold. Then*

$$\sqrt{M} \left(g_{\tau,T}^{N_M,M}(X_\tau^{N_M}) - \mathbf{E}_\tau [g(X_T)] \right) \Rightarrow L_T(X_\tau) + \varepsilon \mathbf{E}_0 [\partial g(X_T(X_\tau)) \Omega_T(\Omega_\tau)^{-1}] U_\tau^X \quad (19)$$

where N_M and ε are such that $\varepsilon = \lim_{M \rightarrow \infty} \sqrt{M/N_M}$, and the martingale $L_T(X_\tau)$ is defined in Theorem 3 and Ω in Theorem 1. Moreover, $\sqrt{M} (g_{\tau,T}^{N_M,M}(X_\tau^{N_M}) - g_{\tau,T}^{N_M,M}(X_\tau^{N_M})) \Rightarrow 0$ where the bias corrected estimator $g_{\tau,T}^{N_M,M}(X_\tau^{N_M})$ is defined in (17) with $N = N_M$.

As in the case of known state the estimation error has a Gaussian martingale part due to the Monte Carlo approximation error and a bounded variation part associated with the discretization of the state variable process. But since X_τ is unknown the martingale $L_T(X_\tau)$ is not conditionally Gaussian any longer. Theorem 7, then, asserts the convergence to a non-Gaussian random variable.

The corresponding result for the transformation is

Theorem 8: *Suppose that the conditions of Theorem 4 hold. Then*

$$\sqrt{M} \left(\hat{g}_{\tau,T}^{N_M,M}(\hat{X}_\tau^{N_M}) - \mathbf{E}_\tau[\hat{g}(\hat{X}_T)] \right) \Rightarrow \epsilon \frac{1}{2} \hat{K}_T(\hat{X}_\tau) + L_T(X_\tau) + \epsilon \mathbf{E}_0 \left[\partial \hat{g}(\hat{X}_T(\hat{X}_\tau)) \hat{\Omega}_T(\hat{\Omega}_\tau)^{-1} \right] U_\tau^{\hat{X}} \quad (20)$$

where N_M and ϵ are such that $\epsilon = \lim_{M \rightarrow \infty} \sqrt{M}/N_M$, and the martingale $L_T(\hat{X}_\tau)$ is defined in Theorem 4 and $\hat{\Omega}$ in Theorem 2. Moreover,

$$\sqrt{M} \left(\hat{g}_{\tau,T}^{N_M,M}(\hat{X}_\tau^{\hat{N}}) - \mathbf{E}_\tau[\hat{g}(\hat{X}_T)] \right) \Rightarrow L_T(X_\tau) + \epsilon \mathbf{E}_0 \left[\partial \hat{g}(\hat{X}_T(\hat{X}_\tau)) \hat{\Omega}_T(\hat{\Omega}_\tau)^{-1} \right] U_\tau^{\hat{X}},$$

when $\hat{N}, M \rightarrow \infty$, where the bias-corrected estimator $\hat{g}_{\tau,T}^{N_M,M}(\hat{X}_\tau^{\hat{N}})$ is defined in (18) with N_M discretization points on $[\tau, T]$ and \hat{N} discretization points on $[0, t]$.

In contrast to the non-second-order-bias corrected estimator without transformation, the asymptotic law of the non-second-order-bias-corrected estimator with transformation is non-centered, indicating a second-order bias. This fact may decrease the superior performance of estimators with

transformation for fixed M and N . The second-order bias with transformation is a consequence of the approximation of the future value of the conditioning state variable. It follows that this second-order bias does not vanish even when the bias-corrected conditional estimator is used.

Comparing Theorems 7 and 8, one can see that the asymptotic distribution for the bias-corrected estimator differs from the one without bias correction only when the transformation is used. Hence, both discretization errors, i.e. the one from the approximation of the underlying state variable and the one from the approximated random variable in the estimate of the conditional expectation, are of the same order. It also appears that the procedure with Doss transformation has the same convergence speed whether the state is known or unknown.

4 A comparison with Milstein's second-order approximation

While Euler schemes for SDEs are appealing from a computational point of view, they might be judged insufficiently accurate. Indeed, second-order schemes such as Milstein (see Milstein (1995)) or Talay (1984, 1986, 1991) have been proposed to better approximate the solution of an SDE. In this section, we extend our weak convergence analysis to the Milstein second-order approximation.

The Milstein approximation of X_T in (2) is

$$\tilde{X}_T^N = X_0 + \sum_{n=0}^{N-1} \left(A(\tilde{X}_{nh}^N)h + \sum_{j=1}^d B_j(\tilde{X}_{nh}^N)\Delta W_{nh}^j + \sum_{j,l=1}^d [\partial B_l B_j](\tilde{X}_{nh}^N) \int_{nh}^{(n+1)h} \int_{nh}^s dW_v^l dW_s^j \right) \quad (22)$$

where $h = T/N$ 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.¹⁶

The Milstein scheme is often difficult to implement for multivariate diffusions. The increments $\Delta F^{l,j} \equiv \int_{nh}^{(n+1)h} \int_{nh}^v dW_s^l dW_v^j$ in the last term of (22) cannot, in general, be written in a form which is easy to simulate. In fact, for multivariate diffusions without commutative noise ($\partial B_l B_k \neq \partial B_k B_l$ some $k \neq l$), the increment $\Delta F^{l,j}$ can only be simulated using a further discretization of the intervals $[(n+1)h, nh]$ (see Gaines and Lyons (1997)). This leads to 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 Milstein scheme is unclear.¹⁷

¹⁶For details on the stochastic Taylor expansion and how to obtain this result see Kloeden and Platen (1997) or Milstein (1995).

¹⁷The commutative noise condition necessary for this representation is the same condition needed for the Doss transformation. When commutativity fails the transformation cannot be used and the implementation of the Milstein 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.

For commutative noise (i.e. $\partial B_l B_k = \partial B_k B_l$), the last term in (22) simplifies to¹⁸

$$\sum_{j,l=1}^d [\partial B_l B_j](\tilde{X}_{nh}^N) \Delta F^{l,j} = \frac{1}{2} \left(\sum_{j=1}^d [\partial B_j B_j](\tilde{X}_{nh}^N) \left((\Delta W_{nh}^j)^2 - h \right) + \sum_{\substack{j,l=1 \\ j \neq l}}^d [\partial B_l B_j](\tilde{X}_{nh}^N) \Delta W_{nh}^j \Delta W_{nh}^l \right). \quad (23)$$

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 (23).

Our next result describes the asymptotic error of the approximation error associated with (22) and therefore find an explicit expression for the Monte Carlo estimator of conditional expectations based on the Milshtein scheme.

Theorem 9: *The approximation error $\tilde{X}_T^N - X_T$ converges weakly at the rate $\frac{1}{N}$ (i.e. $N(\tilde{X}_T^N - X_T) \Rightarrow \tilde{U}_T^X$). The asymptotic error is*

$$\begin{aligned} \tilde{U}_T^X &= -\frac{1}{2} \Omega_T \int_0^T \Omega_s^{-1} \left(\partial A(X_s) dX_s - \sum_{j=1}^d [(\partial B_j) A B_j](X_s) ds \right) \\ &\quad - \frac{1}{2} \Omega_T \int_0^T \Omega_s^{-1} \left(\sum_{k,l,j=1}^d [((\partial_{k,l} A) B_{k,j} B_{l,j} + \partial_l A (\partial_k B_{l,j}) B_{k,j})](X_s) ds \right) \\ &\quad - \frac{1}{\sqrt{12}} \Omega_T \sum_{j=1}^d \int_0^T \Omega_s^{-1} \left[(\partial A) B_j(X_s) dZ_{1,s}^j - [(\partial B_j) A](X_s) dZ_{2,s}^j \right]. \end{aligned}$$

where for each $k = 1, 2$ the process $[Z_k^j]_{j \in \{1, \dots, d\}}$ is a $d \times 1$ standard Brownian motion independent of W and $Z^{h,j}$. The processes U_T^X and Ω_T are given by the expression in Theorem 1 for the Euler approximation without transformation.

Theorem 9 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 particular case of a constant volatility coefficient the asymptotic error distributions are the same for all schemes (i.e. $\tilde{U}_T^X = U_T^X$).

From Theorem 9 we can easily deduce the expected approximation error

Theorem 10: *Suppose that the assumptions of Theorem 9 hold. Let $g \in \mathcal{C}^1(\mathbb{R}^d)$ such that $g(X_T) \in$*

¹⁸See Milshtein (1995) or Gaines and Lyons (1997).

$\mathbb{D}^{1,2}$. For the Milshtein scheme, we have

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

where $\lim_{M \rightarrow \infty} N_M = +\infty$ and $\epsilon = \lim_{M \rightarrow \infty} \frac{\sqrt{M}}{N_M}$ and $L_T(X_0), L_T(\hat{X}_0)$ are the terminal values of centered Gaussian martingales with quadratic variation and conditional variance defined in Theorem 3. The deterministic function \tilde{K}_T is derived in Theorem 11.

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

In order to describe the expected approximation error we need the random variable

$$\begin{aligned} \tilde{V}_T &= -\frac{1}{2} \Omega_T \int_0^T \Omega_s^{-1} \left(\partial A(X_s) dX_s - \sum_{j=1}^d [(\partial B_j) A B_j](X_s) ds \right) \\ &\quad - \frac{1}{2} \Omega_T \int_0^T \Omega_s^{-1} \left(\sum_{k,l,j=1}^d [((\partial_k A) B_{k,j} B_{l,j} + \partial_l A (\partial_k B_{l,j}) B_{k,j})](X_s) ds \right) \end{aligned}$$

With this definition, we have

Theorem 11: Suppose that the conditions of Theorem 9 hold. For $g \in \mathcal{C}^1(\mathbb{R}^d)$ such that

$$\lim_{r \rightarrow \infty} \limsup_N \mathbf{E}_0 \left[\mathbf{1}_{\{|N(g(\tilde{X}_T^N) - g(X_T))| > r\}} N |g(\tilde{X}_T^N) - g(X_T)| \right] = 0 \quad (24)$$

we have, \mathbf{P} -a.s.,

$$N \mathbf{E}_0 \left[g(\tilde{X}_T^N) - g(X_T) \right] \rightarrow \frac{1}{2} \tilde{K}_T(X_0) \equiv \frac{1}{2} \mathbf{E}_0 \left[\partial g(X_T) \tilde{V}_T \right]. \quad (25)$$

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 before. Bias-corrected estimators of conditional expectations can be constructed by taking

$$\tilde{g}_T^{N,M} = \frac{1}{M} \sum_{i=1}^M \left[g(\tilde{X}_{Nh}^{i,N}) + \frac{1}{2} \partial g(\tilde{X}_{Nh}^{i,N}) \tilde{C}_{Nh}^{i,N} \right] \quad (26)$$

where $\tilde{C}_{1,Nh}^{i,N}$ is a convergent approximation of $-\tilde{V}_T$, defined in Appendix B. Our next result shows that $\tilde{g}_T^{N,M}$ is a bias-corrected estimator of the conditional expectation based on the Milshtein scheme.

Theorem 12: *Suppose that the conditions of Theorem 10 hold. Then, $\sqrt{M}(\tilde{g}_{cT}^{N_M, M} - \mathbf{E}_0[g(X_T)]) \Rightarrow L_T(X_0)$, $\sqrt{M}(g_{cT}^{N_M, M} - \hat{g}_{cT}^{N_M, M}) \Rightarrow 0$ and $\sqrt{M}(g_{cT}^{N_M, M} - \tilde{g}_{cT}^{N_M, M}) \Rightarrow 0$ as $M \rightarrow \infty$.*

It is important to note that 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 correction. For non-commutative noise, our explicit expressions for the second-order bias can be used to develop estimators that are computationally superior to the Milshtein scheme because they do not require further subdivisions of discretization intervals to approximate the increments $\Delta F^{l,j}$.

Our next theorem describes properties of estimates of conditional expectations with unknown initial state. The result shows that the Milshtein scheme has the same qualitative properties as the Euler scheme with Doss transformation.

Theorem 13: *Suppose that the conditions of Theorem 9 and 10 hold. Then*

$$\sqrt{M} \left(\tilde{g}_{\tau, T}^{N_M, M}(\tilde{X}_\tau^{N_M}) - \mathbf{E}_\tau[g(X_T)] \right) \Rightarrow \epsilon \frac{1}{2} \tilde{K}_T(X_\tau) + L_T(X_\tau) + \epsilon \mathbf{E}_0 [\partial g(X_T(X_\tau)) \Omega_T(\Omega_\tau)^{-1}] \tilde{U}_\tau^X \quad (27)$$

as $M \rightarrow \infty$ where $\epsilon = \lim_{M \rightarrow \infty} \sqrt{M}/N_M$, and the martingale $L_T(X_\tau)$ is defined in Theorem 3 and Ω in Theorem 2 for the Euler scheme without transformation. Moreover,

$$\sqrt{M} \left(\tilde{g}_{c\tau, T}^{N_M, M}(\tilde{X}_\tau^N) - \mathbf{E}_\tau[g(X_T)] \right) \Rightarrow L_T(X_\tau) + \epsilon \mathbf{E}_0 [\partial g(X_T(X_\tau)) \Omega_T(\Omega_\tau)^{-1}] \tilde{U}_\tau^X.$$

where the bias-corrected estimator $\tilde{g}_{c\tau, T}^{N, M}(\tilde{X}_\tau^N)$ is defined in (26).

5 Applications

In this section we provide two applications of the theorems developed in the last sections which are particularly relevant for finance. The first application deals with the computation of the optimal portfolio shares in a dynamic asset allocation context. To compute these shares which include hedging terms we follow the Monte Carlo methodology introduced in Detemple, Garcia and Rindisbacher (2003). We are therefore in a framework where computing portfolio shares amounts to computing conditional expectations based on simulating Euler discretized processes and averaging replications of such simulated processes. To illustrate the case where the state is unknown we perform a market timing experiment where, based on today's information, we examine the asymptotic error distribution of the hedging demand at a future date. The second application deals with simulation-based inference. The solution to any financial problem involving diffusions rests on the estimation of the parameters of the drift and diffusion functions. The statistical results developed in this paper are well suited to perform asymptotic error analysis for simulation-based inference methods for diffusions.

5.1 Dynamic asset allocation

We consider a general setting with complete markets.¹⁹ Suppose that d risky stocks and one riskless asset are available. Stock prices, $S_i, i = 1, \dots, d$, satisfy the stochastic differential equations

$$dS_{it} = S_{it}[(\mu_i(t, Y_t) - \delta_i(t, Y_t))dt + \sum_{j=1}^d \sigma_{ij}(t, Y_t)dW_t^j]; \quad i = 1, \dots, d \quad (29)$$

$$dY_t = \mu^Y(t, Y_t)dt + \sum_{j=1}^d \sigma_j^Y(t, Y_t)dW_t^j \quad (30)$$

where W is a d -dimensional vector of Brownian motions capturing the underlying uncertainty and Y is a k -dimensional vector of state variables describing the evolution of the opportunity set. Here μ_i is the gross expected return, δ_i the dividend rate and σ_i the vector of volatility coefficients of security i . Let σ denote the $d \times d$ volatility matrix whose components are $\sigma_{ij}, i, j = 1, \dots, d$. All the functions in (29)-(30) are assumed to satisfy conditions for the existence of a strong solution (S, Y) . The riskless asset pays interest at the rate $r(t, Y_t)$.

We also assume that the market price of risk $\theta(t, Y_t) \equiv \sigma(t, Y_t)(\mu(t, Y_t) - r(t, Y_t)\mathbf{1}_d)$, where $\mathbf{1}_d$ is the d -dimensional unit vector, satisfies the Novikov condition (see Karatzas and Shreve (1987)) so as to ensure the existence of the risk neutral measure. The state price density ξ satisfies the stochastic differential equation $d\xi_t = -\xi_t [r(t, Y_t)dt + \theta(t, Y_t)dW_t]$ subject to the initial condition $\xi_0 = 1$.

In this environment an investor selects a portfolio policy π so as to maximize the expected utility of terminal wealth $E[u(T, X_T)]$ subject to the dynamic budget constraint

$$\begin{cases} dX_t = r(t, Y_t)X_t dt + X_t \pi_t' [(\mu(t, Y_t) - r(t, Y_t)\mathbf{1})dt + \sigma_t dW_t], & X_0 = x \\ X_t \geq 0, \text{ for all } t \in [0, T]. \end{cases} \quad (31)$$

Here X_t represents wealth at date t , x is initial wealth and π_t the vector of proportions invested in the risky assets at t . The non-negativity constraint prevents bankruptcy. We focus on the case of constant relative risk aversion $u(T, X) = (1 - R)^{-1}X^{1-R}$ where $R > 0$ is the relative risk aversion coefficient.

For any function $f(t, Y)$ let $\partial_2 f(t, Y)$ be the gradient of f . With this notation the optimal portfolio policy is given by (see Detemple, Garcia and Rindisbacher (2003)):

$$\pi_t = (\sigma(t, Y_t)')^{-1} \left[\frac{1}{R} \theta(t, Y_t) - \rho \mathbf{E}_t \left[\frac{\xi_{t,T}^\rho}{\mathbf{E}_t [\xi_{t,T}^\rho]} H_{t,T} \right] \right]$$

where $\rho = 1 - 1/R$, $\xi_{t,v}^\rho = \exp(-\rho G_{t,v})$, $G_{t,T} = \int_t^T \lambda(s, Y_s) ds + \int_t^T \theta(s, Y_s) dW_s$ and $H_{t,T} = \int_t^T \partial_2 \lambda(s, Y_s) \mathcal{D}_t Y_s ds + \int_t^T \partial_2 \theta(s, Y_s) \mathcal{D}_t Y_s dW_s$, with $\lambda(t, Y_t) \equiv r(t, Y_t) + \frac{1}{2} \theta(t, Y_t) \theta(t, Y_t)'$.²⁰

¹⁹See Merton (1971), Karatzas, Lehoczky and Shreve (1987), Cox and Huang (1989), Ocone and Karatzas (1991) and Detemple, Garcia and Rindisbacher (2003) for additional details.

²⁰We further suppose that $\mathbf{E}[\xi_T^\rho] < \infty$ and $\xi_T^\rho \in \mathbb{D}^{1,2}$.

The Malliavin derivative $D_t Y_s \equiv (D_{1t} Y_s, \dots, D_{dt} Y_s)$ solves the linear SDE

$$d(\mathcal{D}_{mt} Y_s) = \left[\partial_2 \mu^Y(v, Y_s) ds + \sum_{j=1}^d \partial_2 \sigma_j^Y(s, Y_s) dW_s^j \right] (\mathcal{D}_{mt} Y_s) \quad (32)$$

subject to the boundary condition $\lim_{s \rightarrow t} (\mathcal{D}_{mt} Y_s) = \sigma_m^Y(t, Y_t)$ for $m = 1, \dots, d$.

The optimal portfolio policy π_t has a mean-variance term $(\theta(t, Y_t))$, a hedging term against interest rate fluctuations (the part of $H_{t,v}$ involving $\partial_2 r(t, Y_t)$) and a hedging term against market price of risk fluctuations (the part of $H_{t,v}$ involving $\partial_2 \theta(s, Y_s)$). This formula shows that the hedging terms are conditional expectations of functions of random variables $G_{t,T}, H_{t,T}$ which are themselves path-dependent functions of the state variables Y and their Malliavin derivatives $\mathcal{D}_t Y$. Appendix C shows how to write these conditional expectations as expectations over the terminal values of solutions of SDEs. This will embed the formulas in the general setup of sections 2 and 3.

5.1.1 An illustration of the asymptotic laws for the portfolio shares

This section illustrates the second-order bias, and its effect on the size distortion of asymptotic confidence intervals, for the following one factor model:

$$dY_v = \kappa(\bar{Y} - Y_v) (1 + \phi(\bar{Y} - Y_v)^{2\eta}) dv + \sigma Y_v^\gamma dW_v.$$

This specification was introduced in Detemple, Garcia and Rindisbacher (2003) to model the evolution of the short rate and will hereafter be called a NMRCEV process. It includes a nonlinear term in the drift of the process to capture the faster mean reversion occurring when interest rates depart from their long term value. For $\phi = 0$ and $\gamma = \frac{1}{2}$ we obtain a CIR process.

We assume a constant market price of risk $\theta_t = \bar{\theta} = 0.1$, a stock volatility $\sigma = 0.2$ and a relative risk aversion $R = 4$. We also assume that the short rate follows either a CIR or a NMRCEV process. We then calculate second-order bias-corrected estimators for investment horizons $T = 2, 4, 6, 8, 10$.

Table II shows the relation between the size distortion, which is an increasing function of the parameter $\delta/\epsilon = \frac{1}{2} K_T (\mathbf{var}[g(X_T)|\mathcal{F}_0])^{-1/2}$ (resp. $\hat{\delta}/\epsilon = \frac{1}{2} \widehat{K}_T (\mathbf{var}[g(X_T)|\mathcal{F}_0])^{-1/2}$), and the investment horizon. The values reported are all negative. For the CIR process, the absolute value of the size distortion parameter is larger when the transformation is applied, for all horizons. The opposite is true for the NMRCEV process. The relation across models varies according to the horizon. For short horizons, the parameter is larger for the nonlinear model; the reverse holds for long horizons. For both processes the absolute value increases with the investment horizon, making asymptotic confidence intervals worse for longer horizons.

Figure 7 illustrates the effects of different simulation designs, parametrized by the number of Monte Carlo replications M and the number of discretization points for the Euler scheme N , on size distortion. It shows, for various horizons and both the CIR and NMRCEV models, the size distortion for a 95% confidence interval as a function of $\epsilon = \lim_{M \rightarrow \infty} \sqrt{M}/N_M$. For both CIR and

NMRCEV interest rate processes, asymptotic confidence intervals suffer from an important size distortion when the number of discretization points is small relative to the number of Monte Carlo replications ($\epsilon^M = \sqrt{M}/N_M > 1$). They will, therefore, cover the true value with a probability that is considerably lower than the nominal size of the confidence interval. Figure 7 also shows that for long horizons and small numbers of discretization points, the number of Monte Carlo replications must be severely limited to obtain confidence intervals with coverage probability close to their nominal size. These results emphasize that the construction of valid confidence intervals cannot neglect the second-order bias. For the NMRCEV interest rate process it appears that the transformation reduces the size distortion.

Table III presents results for the second-order bias K_T . The second order bias is small for both processes. Yet, at the same time, the size distortion of asymptotic confidence intervals is large, illustrating that the standardized error distributions of the two estimators differ significantly. It follows that the second-order bias cannot be ignored, when constructing confidence intervals or statistical tests, even if the location of the estimator is not substantially affected by the correction. This follows from the structure of the size distortion parameter $\delta/\epsilon = \frac{1}{2}K_T(\text{var}[g(X_T)|\mathcal{F}_0])^{-1/2}$, which depends on both bias and variance. In our example K_T is small. We then conclude that the size distortion is a consequence of a small variance rather than a large second-order discretization bias. Since the length of a confidence interval measures the efficiency of an estimator this observation has important implications for the use of variance reduction techniques. If variance reduction is applied without taking the second-order bias into account, the value of the size distortion parameter δ can actually increase and this reduces the effective coverage probability of the confidence interval. It follows that an assessment of efficiency based on the length of an asymptotic confidence interval, without taking the second order bias into account, could lead to incorrect conclusions. Our explicit formulas for the second-order bias can be used to eliminate this problem.

Table IV illustrates properties of bias-corrected estimators. First note that the bias correction improves the estimator in the right direction. The benchmark true value is calculated with $N = 2^{12}$ and $M = 200000$ without transformation and without bias correction. Estimators with $N = 2^9$ and $M = 25000$ are then calculated. For the CIR process the bias correction does not play an important role for the precision of the portfolio weight. Bias correction is more important for the NMRCEV process. Since both the second order bias corrected estimator and the estimator without correction converge to the true value at the same speed, their difference must disappear when M and N become sufficiently large. The small difference between the two estimators in Table IV indicates that these asymptotic convergence results apply in our simulation design for (M,N) . The results indicate that second order bias correction is not important if we are only interested in the level of the estimator. However, bias correction cannot be ignored when assessing the relative efficiency of an estimator based on a comparison of confidence intervals. It is important to note that our conclusions are for one set of parameters only. Improvements in the precision of estimators using second-order bias correction may be more important for other parameter values. Our explicit expressions can be used

to verify whether this is indeed the case.

5.1.2 An illustration with a portfolio choice in a market timing experiment

To illustrate the theorems of the previous section we consider an investment horizon of 10 years and examine the asymptotic error distribution of the hedging demand in year 9. Figures 8 and 9 plot the asymptotic error distributions for the CIR process and the NMRCEV process respectively. For both processes we present the asymptotic error distribution using the Euler discretization (no transformation) in the upper panel and two asymptotic error distributions for the process with transformation (with or without bias correction) in the lower panel. Computations are based on $M_1 = 5000$ trajectories to calculate the expected value and $M_2 = 500$ trajectories for the empirical distribution function. The discretization is $N = 2^5$ for the simulation of the state variables and the simulation of the conditional hedging demand.

For the estimators of the hedging demand with transformation, it can be seen that the bias corrected distribution functions are everywhere below the distribution functions of the estimators ignoring the second-order bias. The difference between the two distributions is much more pronounced for the NMRCEV process, due again to the non-linear drift structure. Comparing the estimators with and without transformation we note that the bias corrected error is more concentrated around the median for the estimator with transformation. This is also true for both processes, but again the difference is smaller for the CIR case.

The results obtained for the NMRCEV process show that the use of the transformation with bias correction has considerable benefits. Not only does this scheme increase the speed of convergence but it also produces errors that are more concentrated around the median.

5.2 Simulation-based Inference for diffusions

Standard dynamic models in finance, such as those encountered in asset pricing, involve diffusion processes with coefficients that are not known a priori. Estimation and statistical inference are therefore essential for their implementation. The importance of statistical methods is further enhanced by the fact that parameter uncertainty can have a significant impact on economic results (see the recent paper by Barberis (2000) on optimal portfolio choice).

The absence of explicit formulas for transition densities of general diffusion processes implies that maximum likelihood inference can only proceed with the use of an auxiliary numerical procedure to approximate the transition density.²¹ Some of the numerical approaches that have been proposed are (i) the numerical resolution of the Kolmogorov forward equation (Lo (1988)), (ii) the estima-

²¹Numerical approximations can be avoided if we allow the observation interval to decrease to zero. For this type of fill-in asymptotics, efficient estimators have been obtained by Dacunha-Castelle (1986), Florens-Zmirou (1989, 1993), Genon-Catalot (1990), Genon-Catalot and Jacod (1993) and Bibby and Sørensen (1995). This type of asymptotics seems to be less appropriate for non-experimental setups like those encountered in financial econometrics.

tion of the infinitesimal generator based on a truncated set of eigenfunctions (Kessler and Sørensen (1995), Hansen and Scheinkman (1995)), (iii) the simulated maximum likelihood estimation (Pedersen (1995a,b), Brandt and Santa-Clara (2000), Durham and Gallant (2002)), (iv) orthogonal series expansions (Ait-Sahalia (2002)), and (v) Markov Chain Monte Carlo (MCMC) Bayesian techniques (Eraker (1999), Chib, Elerian and Shephard (2000)). Similarly, the method of moments and other methods using estimating functions require auxiliary numerical techniques, as soon as the moment condition or the estimating function are not available in explicit form. In particular, this is the case if state variables are latent. Simulation-based approaches proposed to handle estimations of this type are (i) the simulated method of moments (Duffie and Singleton (1993)), (ii) indirect inference procedures (Smith (1990), Gouriéroux, Monfort and Renault (1993)) and efficient method of moment procedures (EMM) (Gallant and Tauchen (1996)).²²

Our characterizations of the asymptotic error distributions apply directly to the approximation errors of auxiliary numerical procedures and can therefore be used to find efficient experimental designs for simulation-based inference methods for diffusions. To achieve numerical efficiency for these Monte Carlo procedures, two types of approximation errors must be accounted for. 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 estimating function, by an average over independent replications. Understanding the trade-off between these two errors requires the asymptotic error distribution.

Take indirect inference as an example. This procedure relies on the simulation of diffusion processes at a step which is finer than the observation step to correct the discretization bias. As emphasized by Broze, Scaillet and Zakoian (1999), the use of any fixed time interval to perform the simulation results in an asymptotic bias.²³ Although their theoretical asymptotic results support choosing an arbitrarily small discretization step, their numerical experiments show that there is a trade-off between increasing the length of the simulated data (M times the length T of the observed sample) and the size $h = T/N$ of the discretization step. Durham and Gallant (2002) investigate by experimentation the trade-off between the number of sub-intervals N and the number of Monte Carlo sample paths M in the estimation of diffusion models for various simulation schemes. Our results provide an analytical characterization of these trade-offs for some of their simulation schemes. In particular, we show that the bias is a function of both M and N and our characterization of this bias provides a way to correct the second-order bias of simulation-based estimators.

We briefly sketch a general setup for parameter estimation. Suppose that we observe an equally

²²See Gouriéroux and Montfort (1994) for an overview of simulation-based inference. A particular class of method of moments estimators is based on the empirical characteristic function (see Singleton (1999), Jiang and Knight (2000) and Chacko and Viceira (1999) for the case where the characteristic function is in closed form and Carrasco, Chernov, Gallant and Ghysels (2000) when the characteristic function is obtained by Monte Carlo simulation).

²³Because 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*.

spaced discrete time sample $\{Y_0, \dots, Y_L\}$, where $Y_l = Y_{t_l}$ and $t_{l+1} - t_l = \Delta$, of the diffusion process

$$dY_t = A(Y_t; \theta)dt + \sum_{j=1}^d B_j(Y_t; \theta)dW_t^j \text{ and } Y_{t_0} = y \text{ given,}$$

whose coefficients depend on an unknown parameter $\theta \in \Psi \subset \mathbb{R}^p$. Under this assumption we denote the transition density of the Markov chain $\{Y_0, \dots, Y_L\}$ by $p_\Delta(Y_l, z; \theta_0)\lambda(dz) \equiv \mathbf{P}^{\theta_0}(Y_{t_{l+1}} \in dz | \mathcal{F}_{t_l})$, where we assume that this density is absolutely continuous with respect to the Lebesgue measure λ on \mathbb{R}^d . Also assume that the Markov chain $\{Y_0, \dots, Y_L\}$ is geometrically ergodic.²⁴ This last assumption guarantees the existence of a stationary density, denoted by $\bar{p}(y; \theta_0)$.

We seek to estimate parameters θ using the conditional constraint

$$\int_{\mathbb{R}^d} h_{j,\Delta}(Y_l, z; \theta)p_\Delta(Y_l, x; \theta_0)\lambda(dx) = 0 \text{ for all } j = 1, \dots, p = 0 \Leftrightarrow \theta = \theta_0, . \quad (33)$$

The structure of (33) implies that $h_\Delta(Y_l, z; \theta) \equiv (h_{j,\Delta}(Y_l, z; \theta))_{j=1, \dots, p}$ is a vector of martingale increments. This suggests estimating θ by solutions $\hat{\theta}_\Delta^L$ of martingale estimating equations

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

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

It is well-known²⁵ from Chamberlain (1987,1992) and Newey (1990) that the efficient estimator is obtained for weights given by

$$\begin{aligned} J_\Delta(y; \theta) &= \Psi_\Delta(y; \theta)^{-1}\Gamma_\Delta(y; \theta) \\ \Psi_\Delta(y; \theta) &\equiv \int_{\mathbb{R}^d} h_\Delta(y, z; \theta)(h_\Delta(y, z; \theta))'p_\Delta(y, z)\lambda(dz) \\ \Gamma_\Delta(y; \theta) &\equiv \int_{\mathbb{R}^d} (\partial_\theta h_\Delta(y, z; \theta))'p_\Delta(y, z)\lambda(dz). \end{aligned}$$

The efficient estimator has the following asymptotic distribution

$$\begin{aligned} \sqrt{L}(\hat{\theta}_\Delta^L - \theta_0) &\Rightarrow \Sigma_\Delta(\theta_0)^{-\frac{1}{2}}Z \\ \Sigma_\Delta(\theta_0) &\equiv \left(\int_{\mathbb{R}^d} \Gamma_\Delta(y; \theta_0)'\Psi_\Delta(y; \theta_0)^{-1}\Gamma_\Delta(y; \theta_0)\bar{p}(y; \theta_0)\lambda(dy) \right) \end{aligned}$$

²⁴This 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 $c_L \rightarrow \infty$, as is the case for estimators that are of the locally asymptotically mixed normal (LAMN) or locally asymptotically Brownian functional (LABF) form. See the book of 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 deterministic normalization.

²⁵For a recent summary, see Müller and Wefelmeyer (2002) and references therein. They show that the martingale estimating function framework covers GMM, generalized quasi-likelihood, extended quasi-likelihood estimation, and quasi-likelihood estimation.

where Z is a standard normal random variable.

If the transition density $p_\Delta(y, z; \theta)$ is known, the estimation of the unknown parameters based on (33) is equivalent to the estimation of unknown parameters in a Markovian ergodic time series model.²⁶ It is of particular interest to note that the discrete nature of observations has no impact on the inference procedure.

Unfortunately, the transition density $p_\Delta(y, z; \theta)$ is usually unknown and/or we cannot calculate the optimal weights $J(y; \theta)$ in closed form. The corresponding optimal estimating function cannot 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 $h_{j,\Delta}(Y_l, Y_{l+1}; \theta)$ which identify the parameters through (33) are not known in explicit form and are obtained using Monte-Carlo simulation.²⁷

In a diffusion setting, we can use simulations and numerical solution techniques for SDEs to approximate the optimal weights J and/or the functions h_j defining the constraints which identify the parameters.²⁸ For simulation-based estimators the necessary computations can be performed (i) by discretizing time and using the Euler or the Milstein scheme to approximate the solution of the stochastic differential equation and then (ii) replicating this approximation M times to generate M independent realizations of the terminal point $\{Y_{l+1}^{i,N}(Y_l) : i = 1, \dots, M\}$. 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.

With an unknown transition density function, the estimator $\hat{\theta}^{L,M,N}$ is given in terms of a simulation-based martingale estimating function, i.e. as the unique root of

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

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 different higher-order asymptotic properties than

²⁶The particular choice $h_{j,\Delta}(y, z; \theta) = \partial_{\theta_j} p_\Delta(y, z; \theta)$ leads to a maximum likelihood estimator which attains the Cramér-Rao lower bound $(\int_{\mathbb{R}^d} \Psi_\Delta(y; \theta_0) \bar{p}(y; \theta_0) \lambda(dy))^{-1}$.

²⁷See Gallant and Tauchen (2002) for a survey. For moment-based estimating functions h_j are often of the form $h_{j,\Delta}(Y_l, Y_{l+1}; \theta) = f_j(Y_l, Y_{l+1}) - \int_{\mathbb{R}^d} f_j(Y_l, z) p_\Delta(Y_l, z; \theta) \lambda(dz)$. Hence if the transition density is unknown, in many cases the functions h_j cannot be obtained in closed form either. Similarly, for indirect inference and EMM the functions h_j are given as the distance between some auxiliary simulated model and its sample equivalent.

²⁸Alternative kernel-based estimation methods for conditional expectations converge at a slower rate $(L^{-2/(d+4)})$ which 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 in those cases particularly interesting.

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, we introduce the expression,

$$\kappa_{\Delta}(\theta_0) \equiv \int_{\mathbb{R}^d \times \mathbb{R}^d} K_{\Delta}(y, y, z; \theta_0) p_{\Delta}(y, z; \theta_0) \lambda(dz) \bar{p}(z; \theta_0) \lambda(dy), \quad (34)$$

where $K_{j,\Delta}(\cdot, y, z; \theta_0)$ is given as in Theorem 4, except that we have made explicit that the function being averaged over depends, for simulation-based estimating functions, on the observations Y_l and Y_{l+1} .²⁹

The following result shows that the simulation introduces an additional second-order bias.

Theorem 14: *Assume that the conditions of Theorem 4 for the Euler scheme without transformation are satisfied. Then, for geometrically ergodic diffusions observed over time intervals of equal fixed length Δ , we have*

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

where $\varepsilon_1 = \lim_{L \rightarrow \infty} \frac{\sqrt{L}}{\sqrt{M_L}}$ and $\varepsilon_2 = \lim_{L \rightarrow \infty} \frac{\sqrt{M_L}}{N_L}$. Corresponding estimators for the Euler scheme with Doss transformation (under the assumptions of Theorem 5) and the Milshstein scheme (under the assumptions of Theorem 11), are obtained if $K_{j,\Delta}(\cdot, y, z; \theta_0)$ in (34) 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 11, both adjusted for the dependence of h_j and J on the observations $(Y_l, Y_{l+1}) = (y, z)$.

Theorem 14 shows that the efficient simulated estimator of the estimating function has a second-order bias which depends on the discretization scheme used. The numerical scheme used to solve the SDE becomes irrelevant only if $\varepsilon_1 \varepsilon_2 \kappa_{\Delta}(\theta_0) = 0$. It is therefore necessary that conditions linking the sample size, the number of replications and the number of discretization points (L, M_L, N_L) are such that both ε_1 and ε_2 are finite.

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} which characterize the expected approximation errors of the different approximation methods of the SDE. This follows since 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. As a result, variance reduction without second-order bias correction will increase the size distortion of asymptotic confidence intervals and hypothesis tests.

It is also important to note that the results of Theorem 14 do not directly cover simulated maximum likelihood estimators, obtained from the numerical integration of the Chapman-Kolmogorov equation (see Pedersen (1995a,b) and Brandt and Santa-Clara (2002)). Since Monte Carlo integration of the true transition density kernel is based on a Gaussian kernel approximation with bandwidth

²⁹For MCE estimators these observations do not affect the convergence properties.

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 Milshstein, Schoenmakers and Spokoiny (2001), 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 $M^{-1/2}$. The rate for the score will be $M^{-2/(d+4)}$, and therefore will depend on the dimension of the diffusion. Optimality for these simulated maximum likelihood estimators requires that $\sqrt{L}/M_L^{2/(d+4)} \rightarrow \varepsilon_1$ and $M_L^{2/(d+4)}/N_L \rightarrow \varepsilon_2$. It follows that efficient simulated maximum likelihood estimators based on the numerical integration of the Chapman-Kolmogorov equations are less efficient than maximum likelihood estimators.

Theorem 14 shows that estimators such that $\varepsilon_1\varepsilon_2 = 0$, are inefficient. Their asymptotic variance and therefore the length of confidence intervals are larger than the variance and length of the asymptotic confidence interval of the efficient estimator.³⁰ 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 estimator of Pedersen (1995a,b) and Brandt and Santa Clara (2002) depend in addition on the dimension of the diffusion. These authors assume that $\varepsilon_2 = 0$. This assumption, unfortunately, is not sufficient. To avoid an exploding second-order bias it must be the case that $\sqrt{L}/\sqrt{M_L}$ does not diverge faster than $\sqrt{M_L}/N_L$ vanishes. It follows that the rate of divergence of M_L cannot be chosen independently from 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. The corresponding asymptotic confidence intervals will then cover the true values with probability zero. Hypothesis tests become invalid since their effective size can be considerably smaller than the nominal significance level of the test.

Since Theorems 4, 5 and 11 show how to calculate the second-order bias function for the Euler scheme with and without Doss transformation and for the Milshstein scheme, we are able to develop second-order bias corrected estimators, which are both efficient and have asymptotic confidence intervals that do not suffer from size distortion. It is important to note that to achieve asymptotic efficiency (i.e. estimators with shortest asymptotic confidence intervals) only estimators that correct for second-order bias do not require a refinement of the time discretization in the simulation of SDEs as the sample size increases. In contrast, the computational cost of simulation-based estimators, without second-order bias correction, explodes. Only second-order bias corrected estimators have the same asymptotic distribution as the estimators obtained from the infeasible optimal estimating function.³¹

³⁰This follows since L is too small compared to the efficient estimator.

³¹This does not imply that simulation-based inference methods without second-order bias correction are inferior to

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 analysis provided explicit formulas for error distributions, which can be used to construct asymptotically valid confidence intervals and assess the asymptotic error of the Euler scheme with and without Doss transformation as well as of the Milshtein scheme. Numerical experiments, performed in the context of an application to a dynamic portfolio choice problem, showed that the error distributions differ significantly depending on whether the transformation is used or not. Bias corrections were found to be important in the case of nonlinear state variable processes, such as the NMRCEV process. This result underscores the importance of the error statistics obtained, since processes of relevance in financial economics often have a nonlinear structure. In addition, 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 correction. Our explicit expressions for the second-order bias can be used to derive such estimators.

7 Appendix A: proofs

We start with a series of auxiliary lemmas which are central to prove theorems and corollaries of this paper. They provide the weak limits of all the components arising in the error expansion of estimators without transformation (39) and with Doss transformation (41). They also play a key role in the derivation of our explicit expressions for the second-order bias and the asymptotic distribution of estimators without known initial state. It should be noted that the proofs for the results without transformation are easier since the asymptotic error is exclusively determined by the martingale part of the error. These processes of the martingale part are uniformly tight and therefore do not require Wong-Zakai correction terms to derive the weak limits of stochastic integrals. Since the reader may be unfamiliar with the weak limit theory for stochastic processes and our convergence results are not available in the previous literature, we provide a detailed discussion of the auxiliary convergence results.

The first lemma is important to obtain the asymptotic law of the discretization errors. Let $\eta_t^N = \frac{[Nt]}{N}$ if $Nt \notin \mathbb{N}$ and $\eta_t^N = t - \frac{1}{N}$ otherwise, where $[Nt]$ is the integer part of Nt . The time change η_t^N is such that $\lim_{N \rightarrow \infty} \eta_t^N = t$. We have

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 can only be avoided by increasing the dimensionality of the orthogonal basis as the sample increases. Second-order biases for such estimators are unknown.

Lemma 1: *The following weak convergence results hold*

$$V_t^{1,N} \equiv N \int_0^t (s - \eta_s^N) ds \Rightarrow \frac{1}{2}t \equiv V_t^1 \quad (35)$$

$$V_t^{2,i,N} \equiv N \int_0^t (W_s^i - W^i \circ \eta_s^N) ds \Rightarrow \frac{1}{2}W_t^i + \frac{1}{\sqrt{12}}Z_t^i \equiv V_t^{2,i} \quad (36)$$

$$V_t^{3,i,N} \equiv N \int_0^t (s - \eta_s^N) dW_s^i \Rightarrow \frac{1}{2}W_t^i - \frac{1}{\sqrt{12}}Z_t^i \equiv V_t^{3,i} \quad (37)$$

$$V_t^{4,i,j,N} \equiv \sqrt{N} \int_0^t (W_s^i - W^i \circ \eta_s^N) dW_s^j \Rightarrow \frac{1}{\sqrt{2}}Z_t^{i,j} \equiv V_t^{4,i,j} \quad (38)$$

as $N \rightarrow \infty$, 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 of Lemma 1: Since $\eta_t^N = \frac{[Nt]}{N}$ for $Nt \notin \mathbb{N}$ we have $\int_0^t (s - \eta_s^N) ds = \frac{1}{N^2} \int_0^{Nt} (s - [s]) ds$. We obtain, since $N \int_0^t (s - \eta_s^N) ds = \frac{1}{N} \sum_{k=1}^{[Nt]} \int_{[k-1, k]} (s - [s]) ds + \frac{1}{N} \int_{[Nt]}^{Nt} (s - [s]) ds$

$$N \int_0^t (s - \eta_s^N) ds = \frac{1}{2} \frac{[Nt]}{N} + \frac{1}{2} \frac{1}{N} (Nt - [Nt])^2 \rightarrow \frac{1}{2}t$$

as $N \rightarrow \infty$. The second line above follows since $[s] = k - 1$ for $s \in [k - 1, k]$; the limit in the third line uses the bound $Nt - [Nt] \leq 1$.

Similarly, we can show using the scaling property of Brownian motion that

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

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

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

where Z^i is a standard Brownian motion independent of W^j for all $j \in \{1, \dots, d\}$. This establishes (36) since the continuity of the pathwise integral with respect to the Lebesgue measure gives

$$\mathbf{P} - \lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \int_{[Nt]}^{Nt} (W_s^i - W_{[s]}^i) ds = 0.$$

The same type of argument establishes that

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

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

$$\mathbf{P} - \lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \int_{[Nt]}^{Nt} (s - [s]) dW_s^i = 0.$$

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

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{[Nt]} \int_{[k-1, k[} (s - [s]) dW_s^i \Rightarrow \frac{1}{2} W_t^i - \frac{1}{\sqrt{12}} Z_t^i.$$

This establishes (37).

It remains to show (38). Again by the scaling property of Brownian motion

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

Since the sequence of i.i.d random variables $\int_{[k-1, k[} (W_s^j - W_{[s]}^j) dW_s^i$ has variance of $\frac{1}{2}$ and is independent of W^j , $\int_{[k-1, k[} (W_s^i - W_{[s]}^i) ds$ as well as of $\int_{[k-1, k[} (s - [s]) dW_s^i$ we have again by Donsker's invariance principle that

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

whereas it follows from the continuity of the Itô integral that $\mathbf{P} - \lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \int_{[Nt]}^{Nt} (W_s^j - W_{[s]}^j) dW_s^i = 0$. This completes the proof of (38). ■

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

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

Proof of Lemma 2: From condition A in Duffie and Protter (1992) it follows that $V^{1,N}$ is good if $\sup_N N \int_0^t (s - \eta_s^N) ds < \infty$ 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 $i = 1, \dots, d$ and $\sup_N \mathbf{var}[V_t^{4,i,j,N}] < \infty$ for all $i, j = 1, \dots, d$ and $t \in [0, T]$. But since

$$\mathbf{var}[V_t^{3,i,N}] = N^2 \int_0^t (s - \eta_s^N)^2 ds$$

$$\mathbf{var}[V_t^{4,i,j,N}] = N\mathbf{E} \left[\int_0^t (W_s^i - W_{\eta_s^N}^i)^2 ds \mathbf{1}_{i=j} \right] = N \int_0^t (s - \eta_s^N) ds \mathbf{1}_{i=j},$$

it is sufficient to show that $N^\epsilon \int_0^t (s - \eta_s^N)^\epsilon ds < \infty$ for $\epsilon = 1, 2$. The bound $1 > (t - [t])^\epsilon \geq 0$ implies

$$N^\epsilon \int_0^t (s - \eta_s^N)^\epsilon ds = \frac{1}{N} \int_0^{Nt} (s - [s])^\epsilon ds < \frac{1}{N} \int_0^{Nt} ds = t$$

which shows that $V^{1,N}, V^{3,N}, V^{4,N}$ are good. ■

Our next lemma shows that the total variation of the semimartingale $V^{2,N}$ is $\mathbf{O}_{\mathbf{P}}(\sqrt{N})$.³² As a result $V^{2,N}$ cannot be good.

Lemma 3: *The semimartingale $V^{2,N}$ is not good.*

Proof of Lemma 3: We want to show that $N^{-\frac{1}{2}} N \int_0^t |W_s - W_{\eta_s^N}| ds = \mathbf{O}_{\mathbf{P}}(1)$ and consequently the total variation of $V^{2,N}$ is such that $\int_0^t |dV_s^{2,N}| = \mathbf{O}_{\mathbf{P}}(\sqrt{N})$. Note that

$$\begin{aligned} N^{\frac{1}{2}} \int_0^t |W_s - W_{\eta_s^N}| ds &= \frac{1}{N} \int_0^{Nt} |W_s - W_{[s]}| ds \\ &= \frac{1}{N} \sum_{k=0}^{[Nt]-1} \int_k^{k+1} |W_s - W_k| ds + \frac{1}{N} \int_{[Nt]}^{Nt} |W_s - W_{[Nt]}| ds. \end{aligned}$$

By the continuity of the integral the second term $\frac{1}{N} \int_{[Nt]}^{Nt} |W_s - W_{[Nt]}| ds = \mathbf{O}_{\mathbf{P}}(1)$. As for the first term, note that

$$\frac{1}{N^2} \sum_{k=0}^{[Nt]-1} \mathbf{E} \left[\left(\int_k^{k+1} |W_s - W_k| ds \right)^2 \right] < \frac{1}{N^2} \sum_{k=0}^{[Nt]-1} \int_k^{k+1} (s - k) ds = \frac{1}{N^2} \frac{1}{2} \frac{[Nt]}{N} \rightarrow 0,$$

since $\frac{[Nt]}{N} \rightarrow t$ as $N \rightarrow \infty$. We can then invoke Chebyshev's weak law of large numbers to conclude³³

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

where the last equality follows since $\mathbf{E}[|W_1|] = \sqrt{\frac{2}{\pi}}$ and $\int_0^1 \sqrt{s} ds = \frac{2}{3}$ whereas $\frac{[Nt]}{N} \rightarrow t$ as $N \rightarrow \infty$. This implies $\sup_N \int_0^t |dV_s^{2,N}| = \sup_N N \int_0^t |W_s - W_{\eta_s^N}| ds = \infty$. Theorem 3.2 of Jacod and Protter (1998) then shows that $V_t^{2,N}$ cannot be good. ■

³²We say that X^N is $\mathbf{O}_{\mathbf{P}}(N)$ if $\mathbf{P} - \lim_{N \rightarrow \infty} \frac{1}{N} X^N = K \neq 0$ for some random variable K .

³³Chebyshev's weak law of large numbers states that if $\frac{1}{N} \sum_{i=1}^N Z^i$ is such that $\frac{1}{N^2} \sum_{i=1}^N \mathbf{E}[(Z^i)^2] \rightarrow 0$ then $\mathbf{P} - \lim_{N \rightarrow \infty} (\frac{1}{N} \sum_{i=1}^N (Z^i - \mathbf{E}[Z^i])) = 0$.

Define the errors $U_t^{X^N} = X_t^N - X_t$ and $\bar{U}_t^{X_i^N} = X_{l,t}^N - X_{l,\eta_t^N}^N$. To prove Theorem 1 we use the mean value theorem to write

$$\begin{aligned} A(X_{\eta_t^N}^N) - A(X_t) &= A(X_t^N) - A(X_t) - \left(A(X_t^N) - A(X_{\eta_t^N}^N) \right) \\ &= \partial A \left(X_t + \lambda_{1,l} U_t^{X_i^N} \mathbf{1}_d \right) (X_t^N - X_t) - \partial A \left(X_t^N + \lambda_{3,l} \bar{U}_t^{X_i^N} \mathbf{1}_d \right) (X_t^N - X_{\eta_t^N}^N) \end{aligned}$$

where $\lambda_{\cdot,l} \in]0, 1[$ for all $l = 1, \dots, d$, whereas $\mathbf{1}_d$ is a d -dimensional vector of ones, and a similar expression for $B(X_{\eta_t^N}^N) - B(X_t)$. The expansion of the error $U_t^{X^N}$ of the Euler continuous approximation can then be decomposed as

$$\begin{aligned} U_T^{X^N} &= \int_0^T \sum_{l=1}^d \partial_l A(X_s + \lambda_{1,l} e_l U_s^{X_i^N}) U_s^{X_i^N} ds + \int_0^T \sum_{l=1}^d \sum_{j=1}^d \partial_l B_j(X_s + \lambda_{2,l} e_l U_s^{X_i^N}) U_s^{X_i^N} dW_s^j \\ &\quad - \int_0^T \sum_{l=1}^d \partial_l A(X_s^N + \lambda_{3,l} e_l \bar{U}_s^{X_i^N}) \bar{U}_s^{X_i^N} ds - \int_0^T \sum_{l=1}^d \sum_{j=1}^d \partial_l B_j(X_s^N + \lambda_{4,l} e_l \bar{U}_s^{X_i^N}) \bar{U}_s^{X_i^N} dW_s^j \end{aligned}$$

where $e'_j = [0, \dots, 0, 1, 0, \dots, 0]$ represents the j^{th} coordinate vector. Since $\bar{U}_s^{X_i^N} = A_l(X_{\eta_s^N}^N)(s - \eta_s^N) + \sum_{i=1}^d B_{l,i}(X_s)(W_s^i - W_{\eta_s^N}^i)$ we obtain

$$\begin{aligned} U_T^{X^N} &= \int_0^T \sum_{l=1}^d \partial_l A(X_s + \lambda_{1,l} e_l U_s^{X_i^N}) U_s^{X_i^N} ds + \int_0^T \sum_{l=1}^d \sum_{j=1}^d \partial_l B_j(X_s + \lambda_{2,l} e_l U_s^{X_i^N}) U_s^{X_i^N} dW_s^j \\ &\quad - \frac{1}{N} \int_0^T \sum_{l=1}^d \partial_l A(X_s^N + \lambda_{3,l} e_l \bar{U}_s^{X_i^N}) A_l(X_{\eta_s^N}^N) dV_s^{1,N} \\ &\quad - \frac{1}{N} \int_0^T \sum_{l=1}^d \partial_l A(X_s^N + \lambda_{3,l} e_l \bar{U}_s^{X_i^N}) \sum_{j=1}^d B_{l,j}(X_{\eta_s^N}^N) dV_s^{2,j,N} \\ &\quad - \frac{1}{N} \int_0^T \sum_{l=1}^d \sum_{j=1}^d \partial_l B_j(X_s^N + \lambda_{4,l} e_l \bar{U}_s^{X_i^N}) A_l(X_{\eta_s^N}^N) dV_s^{3,j,N} \\ &\quad - \frac{1}{\sqrt{N}} \int_0^T \sum_{l=1}^d \sum_{j=1}^d \partial_l B_j(X_s^N + \lambda_{4,l} e_l \bar{U}_s^{X_i^N}) \sum_{i=1}^d B_{l,i}(X_{\eta_s^N}^N) dV_s^{4,i,j,N}. \end{aligned} \tag{39}$$

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

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

Proof of Lemma 4: The integration by parts formula combined with the facts that $V^{2,j,N}$ is of bounded total variation and that α^N is good gives

$$\int_0^t \alpha_s^N dV_s^{2,j,N} = \alpha_t^N V_t^{2,j,N} - \int_0^t V_s^{2,j,N} d\alpha_s^N \Rightarrow \alpha_t V_t^{2,j} - \int_0^t V_s^{2,j} d\alpha_s.$$

The result follows upon integrating by parts the limit on the right hand side. ■

The asymptotic error expansion for X^N in (39) involves an integral with respect to the bounded variation process $V^{2,j,N}$, which is not good according to Lemma 3. As a result the limit of this integral does not correspond to the integral of the weak limit of the integrand with respect to the weak limit of the integrator. But, as Lemma 4 shows, if the integrand is good we can obtain the limit by first integrating by parts. To apply this result it remains to be shown that the integrand in the integral with respect to $V^{2,N}$ in (39) is good. For this we need the following lemma.

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

Proof of Lemma 5: By Theorem 3.1 of Jacod and Protter (1991) $\mathbf{P} - \lim_{N \rightarrow \infty} \sup_t \|X_t^N - X_t\| = 0$. It follows that for $\mathcal{F}_{(\cdot)}^{X^N}$ -optional times τ^N and positive constants h_N such that $h_N \rightarrow 0$ as $N \rightarrow \infty$

$$\begin{aligned} \|X_{\tau^N+h_N}^N - X_{\tau^N}^N\| &\leq \|X_{\tau^N+h_N}^N - X_{\tau^N}\| + \|X_{\tau^N}^N - X_{\tau^N}\| \\ &\leq \|X_{\tau^N+h_N}^N - X_{\tau^N+h_N}\| + \|X_{\tau^N+h_N} - X_{\tau^N}\| + \|X_{\tau^N}^N - X_{\tau^N}\|. \end{aligned}$$

Since X is continuous and $\mathbf{P} - \lim_{N \rightarrow \infty} \sup_t \|X_t^N - X_t\| = 0$ we obtain $\mathbf{P} - \lim_{N \rightarrow \infty} \|X_{\tau^N+h_N}^N - X_{\tau^N}^N\| = 0$. This proves optional equicontinuity of X^N .³⁴ By Theorem 14.11 of Kallenberg (1997) this is sufficient for tightness of X^N ,³⁵ which in turn, by Theorem 2.3 in Jacod and Protter (1998), is sufficient for goodness of X^N . ■

The weak limit result (38) in lemma 1 and lemma 2 are sufficient to prove Theorem 1.

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

$$U_T^{\hat{X}^N} = \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s + \lambda_{1,l} e_l U_s^{\hat{X}_s^N}) U_s^{\hat{X}_s^N} ds - \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s^N + \lambda_{3,l} e_l \bar{U}_s^{\hat{X}_{l,t}^N}) \bar{U}_s^{\hat{X}_{l,t}^N} ds$$

³⁴A sequence of stochastic processes X^N is *optional equicontinuous* if for all $\mathcal{F}_{(\cdot)}^{X^N}$ -optional times τ^N and sequences h_N such that $h_N \rightarrow 0$ we have that $\|X_{\tau^N+h_N}^N - X_{\tau^N}^N\| \rightarrow 0$ in probability.

³⁵From Prohorov's theorem we know that tightness is sufficient for relative compactness (a family of probability measures is relatively compact if every sequence of probability measures contains a weakly convergent subsequence), which in turn, together with convergence of finite dimensional distributions, implies weak convergence of probability measures (Billingsley (1968) pp. 35-40). A family of probability measures Π on a metric space is *tight* if for every $\epsilon > 0$ there exist a compact set K such that $\mathbf{P}(K) \geq 1 - \epsilon$ for all $\mathbf{P} \in \Pi$. A sequence of random elements X^N defined on a metric space is tight if and only if the family of induced measures $\mathbf{P}(X^N)^{-1}$ is tight.

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

$$\begin{aligned} U_T^{\hat{X}^N} &= \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s + \lambda_{1,l} e_l U_s^{\hat{X}_l^N}) U_s^{\hat{X}_l^N} ds - \frac{1}{N} \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s^N + \lambda_{3,l} e_l \bar{U}_s^{\hat{X}_l^N}) \hat{A}_l(\hat{X}_{\eta_s^N}^N) dV_s^{1,N} \\ &\quad - \frac{1}{N} \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s^N + \lambda_{3,l} e_l \bar{U}_s^{\hat{X}_l^N}) \sum_{j=1}^d \hat{B}_{l,j} dV_s^{2,j,N}. \end{aligned} \quad (41)$$

As for the error expansion without transformation (39) we have integrals with respect to \hat{X}^N . To find the limit using integration by parts it is necessary that \hat{X}^N be good. Our next lemma verifies this property.

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

Proof of Lemma 6: 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. Taking $\hat{\alpha}_s^{l,j,N} \equiv \partial_l \hat{A}(\hat{X}_s^N + \lambda_{3,l} e_l \bar{U}_s^{\hat{X}_l^N}) \hat{B}_{l,j}$ and applying the same arguments as for $\alpha^{l,j,N}$ in Theorem 1 we find that

$$\int_0^T \hat{\alpha}_s^{l,j,N} dV_s^{2,j,N} \Rightarrow \int_0^T \partial_l \hat{A}(\hat{X}_s) \hat{B}_{l,j} dV_s^{2,j} + \frac{1}{2} \int_0^T \sum_{k=1}^d \partial_{l,k} \hat{A}(\hat{X}_s) \hat{B}_{k,j} \hat{B}_{l,j} ds \quad (42)$$

Then, since $\lim_{N \rightarrow \infty} \eta_s^N = s$ and $\mathbf{P} - \lim_{N \rightarrow \infty} \bar{U}_s^{\hat{X}_l^N} = \mathbf{P} - \lim_{N \rightarrow \infty} U_s^{\hat{X}_l^N} = 0$ and since by Lemma 2 $V^{1,N}$ is good it follows that $NU^{\hat{X}^N} \Rightarrow U^{\hat{X}}$ where

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

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

Proof of Corollary 1: The result follows since $N(G(\hat{X}_T^N) - X_T) = \partial G(\hat{X}_T) \hat{U}_T^{\hat{X}^N} + \mathbf{o}_{\mathbf{P}}(1)$, where $\mathbf{o}_{\mathbf{P}}(1)$ denotes terms that vanish in probability,³⁶ and since $\partial G(z) = B(G(z))$. ■

³⁶ X^N is said to be $\mathbf{o}_{\mathbf{P}}(1)$ if $\mathbf{P} - \lim_{N \rightarrow \infty} X^N = 0$.

We now turn to the proof of the expected approximation error. It follows from the error expansion (41) that the approximation error U^{X^N} satisfies a linear SDE. The solution is

$$\begin{aligned} N(\Omega_T^N)^{-1}U_T^{X^N} &= -\int_0^T (\Omega_s^N)^{-1}dI_{1,s}^N - \int_0^T (\Omega_s^N)^{-1}dI_{2,s}^N - \int_0^T (\Omega_s^N)^{-1}d(I_{3,s}^N - [R^N, I_3^N]_s) \\ &\quad -\sqrt{N}\int_0^T (\Omega_s^N)^{-1}d(I_{4,s}^N - [R^N, I_4^N]_s) \end{aligned} \quad (43)$$

where $\Omega_T^N = \exp(R_T^N)$ with

$$\begin{aligned} R_T^N &= \int_0^T \sum_{l=1}^d \partial_l A(X_s + \lambda_{1,l}e_l U_s^{X_i^N})ds + \int_0^T \sum_{l,j=1}^d \partial_l B_j(X_s + \lambda_{2,l}e_l U_s^{X_i^N})dW_s^j \\ &\quad - \frac{1}{2} \int_0^T \left(\sum_{l,j=1}^d \partial_l B_j(X_s + \lambda_{2,l}e_l U_s^{X_i^N}) \right) \left(\sum_{l,j=1}^d \partial_l B_j(X_s + \lambda_{2,l}e_l U_s^{X_i^N}) \right)' ds \end{aligned} \quad (44)$$

and

$$\begin{aligned} I_{1,T}^N &\equiv \int_0^T \sum_{l=1}^d \partial_l A(X_s^N + \lambda_{3,l}e_l \bar{U}_s^{X_i^N}) A_l(X_{\eta_s^N}^N) dV_s^{1,N} \\ I_{2,T}^N &\equiv \int_0^T \sum_{l=1}^d \partial_l A(X_s^N + \lambda_{3,l}e_l \bar{U}_s^{X_i^N}) \sum_{j=1}^d B_{l,j}(X_{\eta_s^N}^N) dV_s^{2,j,N} \\ I_{3,T}^N &\equiv \int_0^T \sum_{l=1}^d \sum_{j=1}^d \partial_l B_j(X_s^N + \lambda_{4,l}e_l \bar{U}_s^{X_i^N}) A_l(X_{\eta_s^N}^N) dV_s^{3,j,N} \\ I_{4,T}^N &\equiv \int_0^T \sum_{i,j=1}^d \left(\sum_{l=1}^d \partial_l B_j(X_s^N + \lambda_{4,l}e_l \bar{U}_s^{X_i^N}) B_{l,i}(X_{\eta_s^N}^N) \right) dV_s^{4,i,j,N}. \end{aligned}$$

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

Lemma 7: For any \mathcal{F}_T -measurable, $d \times 1$ random vector $H_T^N \Rightarrow H_T$, such that

$$\lim_{r \rightarrow \infty} \limsup_N \mathbf{E}_0 \left[\left| \sqrt{N}(H_T^N)' \int_0^T (\Omega_s^N)^{-1} d(I_{4,s}^N - [R^N, I_4^N]_s) \mathbf{1}_{\{|\sqrt{N}H_T^N \int_0^T (\Omega_s^N)^{-1} d(I_{4,s}^N - [R^N, I_4^N]_s)| > r\}} \right| \right] = 0 \quad (45)$$

(i.e. $(H_T^N)' \int_0^T (\Omega_s^N)^{-1} d(I_{4,s}^N - [R^N, I_4^N]_s)$ is asymptotically uniformly integrable) the cross moment

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

where

$$U_{2,T} = \sum_{i,j=1}^d \left[H_T^i \left(\int_0^T \beta_s^{i,j} dW_s^i + [\beta^{i,j}, W^i]_T \right) + [h^i \beta^{i,j}, W^i]_T \right] \quad (46)$$

and the $d \times 1$ vectors $\beta_t^{i,j}$ and h^l are respectively given by

$$\beta_t^{i,j} = \Omega_t^{-1} [\partial B_i \sum_{l=1}^d \partial_l B_j B_{l,i}] (X_t), \quad \text{for } i, j = 1, \dots, d$$

$$H_T - \mathbf{E}_0 [H_T] = \sum_{l=1}^d \int_0^T h_s^l dW_s^l, \quad \text{for } l = 1, \dots, d.$$

Proof of Lemma 7: To establish this limit recall, first, that X^N is good and that the coefficients are continuous by assumption. It follows that R^N is good as well. Consider then the limit of $\sqrt{N} \mathbf{E}_0 \left[(H_T^N)' \int_0^T (\Omega_s^N)^{-1} d[R^N, I_4^N]_s \right]$. Since

$$\sqrt{N} \int_0^t (\Omega_s^N)^{-1} d[R^N, I_4^N]_s = \sum_{i,j=1}^d \int_0^t \beta_s^{i,j} dV_s^{2,i,N} + \mathbf{o}_{\mathbf{P}}(1)$$

and with the help of Lemma 4 we obtain the weak limit

$$\sqrt{N} \int_0^t (\Omega_s^N)^{-1} d[R^N, I_4^N]_s \Rightarrow \sum_{i,j=1}^d \int_0^t \beta_s^{i,j} dV_s^{2,i} + \sum_{i,j=1}^d [\beta^{i,j}, V^{2,i}]_t.$$

Weak convergence, along with the uniform integrability of $\sqrt{N} (H_T^N)' \int_0^t (\Omega_s^N)^{-1} d[R^N, I_4^N]_s$ in assumption (45), implies convergence in means. Since $\lim_{N \rightarrow \infty} \sqrt{N} \mathbf{E}_0 \left[(H_T^N)' \int_0^t (\Omega_s^N)^{-1} d[R^N, I_4^N]_s \right] = \mathbf{E}_0 \left[H_T' \left(\sum_{i,j=1}^d \int_0^t \beta_s^{i,j} dV_s^{2,i} + \sum_{i,j=1}^d [\beta^{i,j}, V^{2,i}]_T \right) \right] = \frac{1}{2} \mathbf{E}_0 \left[H_T' \left(\sum_{i,j=1}^d \int_0^t \beta_s^{i,j} dW_s^i + \sum_{i,j=1}^d [\beta^{i,j}, W^i]_T \right) \right]$, we obtain that

$$\lim_{N \rightarrow \infty} \sqrt{N} \mathbf{E}_0 \left[(H_T^N)' \int_0^t (\Omega_s^N)^{-1} d[R^N, I_4^N]_s \right] = \frac{1}{2} \sum_{i,j,k=1}^d \mathbf{E}_0 \left[H_T^k \left(\int_0^t \beta_s^{i,j,k} dW_s^i + [\beta^{i,j,k}, W^i]_T \right) \right],$$

where the equality in the second line uses the fact that $\beta^{i,j}$ and H_T are independent from Z^i in $V^{2,i}$ and the third line restates the second line using the fact that we multiply two d -dimensional vectors.

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

$$\begin{aligned} \sum_{l=1}^d \left[\int_0^T h_s^{N,k,l} dW_s^l, \sqrt{N} \int_0^T (\Omega_s^N)^{-1} dI_{4,s}^N \right]_T &= \sum_{i,j=1}^d \int_0^T h_s^{k,i} \beta_s^{i,j,k} dV_s^{2,i,N} + \mathbf{o}_{\mathbf{P}}(1) \\ &\Rightarrow \sum_{i,j=1}^d \int_0^T h_s^{k,i} \beta_s^{i,j,k} dV_s^{2,i} + [h^{k,i} \beta^{i,j,k}, V^{2,i}]_T \quad (47) \end{aligned}$$

where the limit in the third line follows from Lemma 4.

Combining the results above gives with $h_0^N \equiv \sqrt{N} \mathbf{E}_0 \left[(H_T^N)' \int_0^T (\Omega_s^N)^{-1} dI_{4,s}^N \right]$, that

$$\begin{aligned} \lim_{N \rightarrow \infty} h_0^N &= \lim_{N \rightarrow \infty} \sqrt{N} \mathbf{E}_0 \left[\left(\mathbf{E}_0 \left[H_T^k \right] + \sum_{l=1}^d \int_0^T h_s^{k,l} dW_s^l \right)' \left(\int_0^T (\Omega_s^N)^{-1} dI_{4,s}^N \right) \right] \\ &= \lim_{N \rightarrow \infty} \sum_{k=1}^d \mathbf{E}_0 \left[\left[\int_0^T h_s^{k,l} dW_s^l, \int_0^T (\Omega_s^N)^{-1} dI_{4,s}^N \right]_T \right] \\ &= \sum_{i,j,k=1}^d \mathbf{E}_0 \left[\int_0^T h_s^{k,i} \beta_s^{i,j,k} dV_s^{2,i} + [h^{k,i} \beta^{i,j,k}, V^{2,i}]_T \right] = \frac{1}{2} \sum_{i,j,k=1}^d \mathbf{E}_0 \left[[h^{k,i} \beta^{i,j}, W^i]_T \right] \end{aligned}$$

where the second line uses $\mathbf{E}_0 \left[\int_0^T (\Omega_s^N)^{-1} dI_{4,s}^N \right] = 0$, the third line follows from (47) and the last line uses $\mathbf{E}_0 \left[\int_0^T h_s^{k,i} \beta_s^{i,j,k} dV_s^{2,i} \right] = 0$. ■

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

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

By the Lindeberg central limit theorem for i.i.d. random variables we have

$$\frac{1}{\sqrt{M}} \sum_{i=1}^M \left(g(X_T^i) - \mathbf{E}_0 [g(X_T)] \right) \Rightarrow \sqrt{\mathbf{var}[g(X_T)|\mathcal{F}_0]} Z$$

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

If we introduce a sequence of numbers, N_M which depends on M and $\epsilon^M = \frac{\sqrt{M}}{N_M}$ such that $\lim_{M \rightarrow \infty} \epsilon^M = \epsilon < \infty$ we get, by the Kolmogorov's strong law of large numbers,

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

Our previous result on the expected approximation error then gives

$$\frac{1}{\sqrt{M}} \sum_{i=1}^M \left(g(X_T^{i,N_M}) - g(X_T^i) \right) \rightarrow \epsilon K_T(X_0)$$

as $M \rightarrow \infty$. This establishes the result announced ■.

We now proceed with the proof of Theorem 4.

Proof of Theorem 4: By Lemma 7 we know that $\sqrt{N}\mathbf{E}_0 \left[H'_T \int_0^T (\Omega_s^N)^{-1} d(I_{4,s}^N - [R^N, I_4^N]_s) \right] \rightarrow \frac{1}{2}\mathbf{E}_0 [U_{2,T}]$ where $U_{2,T}$ is given in (46), for any \mathcal{F}_T -measurable random vector H_T .

Since all other terms in (43) are of the same order and by assumption we have asymptotic uniform integrability of each terms, we can find the expected approximation error from the weak limit of these terms.

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

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

with $\gamma_{3,s}^{j,N} = \sum_{l=1}^d \partial_l B_j(X_s + \lambda_{4,l} e_l U_s^{X_l^N}) A_l(X_{\eta_s^N}^N)$ we obtain the limit $[R^N, I_3^N]_T \Rightarrow \frac{1}{2} \int_0^T \sum_{j=1}^d [\partial B_j \partial B_j A](X_s) ds$.

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

$$\begin{aligned} I_{1,T} &= \frac{1}{2} \int_0^T \partial A(X_s) A(X_s) ds \\ I_{2,T} &= \int_0^T \sum_{j=1}^d \partial A(X_s) B_j(X_s) \left(\frac{1}{2} dW_s^j + \frac{1}{\sqrt{12}} dZ_s^j \right) + \frac{1}{2} \int_0^T \sum_{k,l,j=1}^d [\partial_k (\partial_l A B_{l,j}) B_{k,j}](X_s) ds \\ I_{3,T} &= \int_0^T \sum_{j=1}^d \partial B_j(X_s) A(X_s) \left(\frac{1}{2} dW_s^j - \frac{1}{\sqrt{12}} dZ_s^j \right). \end{aligned}$$

$I_{2,T}$ is obtained as follows. Lemma 4 with $\alpha_s^N = \sum_{l,j=1}^d \partial_l A(X_s^N + \lambda_{3,l} e_l \bar{U}_s^{X_l^N}) B_{l,j}(X_{\eta_s^N}^N)$ gives

$$\int_0^T \alpha_s^N dV_s^{2,j,N} \Rightarrow \int_0^T \alpha_s dV_s^{2,j} + [\alpha, V^{2,j}]_T.$$

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

$$[\alpha, V^{2,j}] = \frac{1}{2} \int_0^t \sum_{l,j=1}^d [\partial [\partial_l A B_{l,j}]](X_s) d[X, W^j]_s ds = \frac{1}{2} \int_0^t \sum_{l,j,k=1}^d [\partial_k [\partial_l A B_{l,j}] B_{k,j}](X_s) ds,$$

where the second equality follows from $d[X, W^j]_s = B(X_s) ds$. This explains the limit of I_2^N .

Since $g \in \mathcal{C}^3(\mathbb{R}^d)$ and since X^N converges the mean value theorem shows that

$$\begin{aligned} N(g(X_T^N) - g(X_T)) &= \partial g \left(X_T + \text{diag}[\lambda_i] U_T^{X^N} \right) N U_T^{X^N} \\ &= \partial g \left(X_T + \text{diag}[\lambda_i] U_T^{X^N} \right) \Omega_T^N \left(N (\Omega_T^N)^{-1} U_T^{X^N} \right) \end{aligned}$$

for some $\lambda_i \in]0, 1[$, $i = 1, \dots, d$, where $diag[\lambda_i]$ denotes the diagonal matrix with $\lambda_1, \dots, \lambda_d$ on the diagonal and zeros elsewhere, and where $N(\Omega_T^N)^{-1}U_T^{X^N}$ is defined in (43).

Now define $H_T^k = \partial g(X_T)\Omega_{k,T}$ for $k = 1, \dots, d$ where $\Omega_{k,\cdot}$ is the k^{th} column of the matrix process Ω . Since $N(g(X_T^N) - g(X_T))$ is uniformly integrable by assumption (13) we can apply Lemma 7 to conclude that the expected approximation error converges to

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

where $U_{1,T}$ is the limit of the first 3 terms in (43), that is,

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

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

Recall the $d \times 1$ vector process $\beta_t^{i,j} = \Omega_t^{-1}[\partial B_i \sum_{l=1}^d \partial_l B_j B_{l,i}](X_t)$ in the definition of $U_{2,T}$. Ito's lemma gives

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

and $d\Omega_t^{-1} = -\Omega_t^{-1} \sum_j \partial B_j(X_t)dW_t^j + dH_t$, where $d[H, W^i]_t = 0$ for all $i = 1, \dots, d$. Then,

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

Since $\sum_{l=1}^d \partial_l B_j B_{l,i} = \partial B_j B_i$ we can write $\sum_{i,j=1}^d [\beta^{i,j}, W^i]_T = \int_0^T \sum_{i,j=1}^d d[\beta^{i,j}, W^i]_t = \int_0^T \Omega_t^{-1} \eta_t dt$ with $\eta_t = \sum_{i,j=1}^d (\partial_i [\partial B_i \partial B_j B_i] B_i - \partial B_i \partial B_i \partial B_j B_i)$.

Proceeding in a similar manner, we now find an explicit expression for $[(h^i)' \beta^{i,j}, W^i]_T$. By definition h_s^i is the process in the representation of the random variable $H_T \equiv \partial g(X_T)\Omega_T$, i.e. $(h_t^i)' = \mathbf{E}_t [\mathcal{D}_{i,t} (\partial g(X_T)\Omega_T)]$ by the Clark-Ocone formula. Therefore

$$(h_T^i)' = \lim_{t \rightarrow T} \mathcal{D}_{i,t} (\partial g(X_T)\Omega_T) = [(B_i)'(\partial^2 g)\Omega_T + (\partial g)\Omega_T(\partial B_i)](X_T)$$

where $\partial^2 g$ is the Hessian of g . Since $\beta_T^{i,j} = [\Omega_T^{-1}(\partial B_i)(\partial B_j)B_i]$ we have

$$(h_T^i)' \beta_T^{i,j} = [(B_i)'(\partial^2 g)(\partial B_i)(\partial B_j)B_i + (\partial g)\Omega_T(\partial B_i)\Omega_T^{-1}(\partial B_i)(\partial B_j)B_i](X_T) \equiv c_{i,j}(X_T, \Omega_T).$$

The arguments of this matrix function are the $d \times 1$ vector X and the $d \times d$ matrix Ω . Using Ito's lemma gives

$$[c_{i,j}(X, \Omega), W^i]_T = \int_0^T \partial_X c_{i,j}(X_t, \Omega_t) d[X, W^i]_t + \int_0^T \partial_\Omega c_{i,j}(X_t, \Omega_t) d[vec(\Omega), W^i]_t$$

where vec denotes the stack operator (for a matrix A with columns A_1, \dots, A_d we have $vec(A) = ([A'_1, \dots, A'_d])'$),

$$\begin{aligned} \partial_X c_{i,j}(X, \Omega) &= [\partial[(B^i)'(\partial^2 g)(\partial B_i)(\partial B_j)B_i]](X) \\ &+ [(\Omega(\partial B_i)\Omega^{-1}(\partial B_i)(\partial B_j)(B_i))' \partial^2 g + (\partial g)\Omega(\partial B_i)\Omega^{-1} \partial[(\partial B_i)(\partial B_j)B_i]](X) \\ &+ [((\partial g)\Omega \otimes (\Omega^{-1}(\partial B_i)(\partial B_j)B_i)') \partial^2 B_i](X) \end{aligned}$$

$$\partial_\Omega c_{i,j}(X, \Omega) = [(((\partial B_i)\Omega^{-1}(\partial B_i)(\partial B_j)B_i)' \otimes \partial g) - ((\Omega^{-1}(\partial B_i)(\partial B_j)B_i)' \otimes (\partial g)\Omega(\partial B_i)\Omega^{-1})](X)$$

and \otimes denotes the Kronecker product (for two matrices A, B the Kronecker product is $A \otimes B = [a_{i,j}B]_{i,j=1,\dots,d \cdot d}$)³⁷.

We now show that $\int_0^T \partial_\Omega c_{i,j}(X_t, \Omega_t) d[vec(\Omega), W^i]_t = 0$. Since $d\Omega_t = \Omega_t[\partial A(X_t)dt + \sum_{j=1}^d \partial B_j(X_t)dW_t^j]$ we get, using $vec(ABC) = (C' \otimes A)vec(B)$, that

$$dvec(\Omega_t) = \left[((\partial A)' \otimes I_d)(X_t)dt + \sum_{j=1}^d ((\partial B_j)' \otimes I_d)(X_t)dW_t^j \right] vec(\Omega_t)$$

and therefore $d[vec(\Omega), W^i]_t = [((\partial B_i)' \otimes I_d)(X_t)] vec(\Omega_t)dt$. This implies

$$\int_0^T \partial_\Omega c_{i,j}(X_t, \Omega_t) d[vec(\Omega), W^i]_t = \int_0^T [\partial_\Omega c_{i,j}((\partial B_i)' \otimes I_d) vec(\Omega_t)](X_t)dt.$$

But using that $(C' \otimes A)vec(B) = vec(ABC)$ if the product ABC is well defined we get

$$\begin{aligned} \partial_\Omega c_{i,j}((\partial B_i)' \otimes I_d) vec(\Omega) &= (((\partial B_i)(\partial B_i)\Omega^{-1}(\partial B_i)(\partial B_j)B_i)' \otimes \partial g) vec(\Omega) \\ &- (((\partial B_i)\Omega^{-1}(\partial B_i)(\partial B_j)B_i)' \otimes (\partial g)\Omega(\partial B_i)\Omega^{-1}) vec(\Omega) \\ &= (\partial g)\Omega(\partial B_i)(\partial B_i)\Omega^{-1}(\partial B_i)(\partial B_j)B_i \\ &- (\partial g)\Omega(\partial B_i)\Omega^{-1}\Omega(\partial B_i)\Omega^{-1}(\partial B_i)(\partial B_j)B_i = 0. \end{aligned}$$

Therefore, given that $d[X, W^i]_t = B_i(X_t)dt$, we obtain $[c_{i,j}(X, \Omega), W^i]_T = \int_0^T \nu_{i,j}(X_t, \Omega_t)dt$ with $\nu_{i,j}(X, \Omega)$ as defined in the theorem. We conclude $\sum_{i,j=1}^d [(h^i)' \beta^{i,j}, W^i]_T = \int_0^T \sum_{i,j=1}^d \nu_{i,j}(X_t, \Omega_t)dt$.

³⁷For results on matrix differential calculus see Magnus and Neudecker (1988), in particular the identification tables on p. 176 for the first derivative and on p. 190 for the Hessian.

Then, since expectations of stochastic integrals with respect to Z^j vanish (since they are independent from \mathcal{F}_t) we get

$$N\mathbf{E}_0 [g(X_T^N) - g(X_T)] \rightarrow \frac{1}{2}\mathbf{E}_0 [\partial g(X_T)V_{1,T} + V_{2,T}] \quad (48)$$

with $V_{1,T}, V_{2,T}$ as defined in the theorem. This establishes the result announced. ■

We now prove results for the expected approximation error using the Doss transformation.

Proof of Theorem 5: Here all the terms of the error expansion (41) are of order $1/N$ and the limit distribution is non-centered. The result is then obtained, provided the terms in the error expansion are uniformly integrable, by taking the expectation of the solution of the linear SDE for the approximation error

$$NU_T^{\hat{X}_t^N} = -\hat{\Omega}_T^N \left(\int_0^T (\hat{\Omega}_s^N)^{-1} d\hat{I}_{1,s}^N + \int_0^T (\hat{\Omega}_s^N)^{-1} d\hat{I}_{2,s}^N \right) \quad (49)$$

where $\hat{\Omega}_T^N = \exp \left(\int_0^T \partial \hat{A}(\hat{X}_s + \lambda_{1,l} e_l U_s^{\hat{X}_s^N}) ds \right)$ and $\hat{I}_{1,T}^N \equiv \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s^N + \lambda_{3,l} e_l \bar{U}_s^{\hat{X}_s^N}) \hat{A}_l(\hat{X}_{\eta_s^N}^N) dV_s^{1,N}$ as well as $\hat{I}_{2,T}^N \equiv \int_0^T \sum_{l=1}^d \partial_l \hat{A}(\hat{X}_s^N + \lambda_{3,l} e_l \bar{U}_s^{\hat{X}_s^N}) \sum_{j=1}^d \hat{B}_{l,j}(\hat{X}_{\eta_s^N}^N) dV_s^{2,j,N}$. Since $(\hat{\Omega}_t^N, I_{1,t}^N, I_{2,t}^N, \hat{X}_t^N) \Rightarrow (\hat{\Omega}_t, I_{1,t}, I_{2,t}, \hat{X}_t)$ with $\hat{\Omega}_t, \hat{X}_t$ as defined in Theorem 2 and

$$\hat{I}_{1,T} = \frac{1}{2} \int_0^T [(\partial \hat{A}) \hat{A}](\hat{X}_s) ds \quad (50)$$

$$\hat{I}_{2,T} = \int_0^T [\partial A \sum_{j=1}^d \hat{B}_j](\hat{X}_s) \left(\frac{1}{2} dW_s^j + \frac{1}{\sqrt{12}} dZ_s^j \right) + \frac{1}{2} \int_0^T \sum_{j,k,l=1}^d \partial_{l,k} \hat{A}(\hat{X}_s) \hat{B}_{k,j} \hat{B}_{l,j} ds \quad (51)$$

(the expression for $\hat{I}_{2,T}$ follows from (42)), the proof of the theorem follows using the same arguments as in the proof without transformation. ■

Proof of Theorem 6: Consider the Euler approximations of the processes $NC_{j,t}^N$ for $j = 1, 2$ defined in (53)-(54) in Appendix B. The Euler continuous approximation is

$$\begin{aligned} d(NC_{1,s}^N) &= [\partial A(X_{\eta_s^N}^N)h + \sum_{j=1}^d \partial B_j(X_{\eta_s^N}^N) dW_s^j](NC_{1,s}^N) - \left(\sum_{j,k=1}^d [\partial[\partial B_k \partial B_j B_k] B_i](X_{\eta_s^N}^N) \right) ds \\ &+ \left([(\partial A)A + \sum_{j=1}^d \partial B_j \partial A B_j + \sum_{j,k,l=1}^d \partial_k (\partial_l A B_{l,j}) B_{k,j}](X_{\eta_s^N}^N) \right) ds \\ &+ \sum_{j=1}^d [(\partial A)B_j + (\partial B_j)A - \sum_{i=1}^d (\partial B_j)(\partial B_i)B_j](X_{\eta_s^N}^N) dW_s^j \\ d(NC_{2,s}^N) &= \left(\sum_{i,j=1}^d \nu_{i,j}(X_{\eta_s^N}^N, \Omega_{\eta_s^N}^N) \right) ds \end{aligned}$$

with $C_{1,0}^N = C_{2,0}^N = 0$. It follows that $N\mathbf{E}_0 \left[\partial g(X_{Nh}^N) C_{1,Nh}^N + C_{2,Nh}^N \right] \rightarrow -K_T(X_0)$.

Consequently, by the same arguments as in the proof of Theorems 3 and 4 we have, for N_M and $\epsilon^M = \frac{\sqrt{M}}{N_M}$ such that $\lim_{M \rightarrow \infty} \epsilon^M = \epsilon < \infty$, that

$$\frac{1}{2} \frac{1}{\sqrt{M}} \sum_{i=1}^M \left(\partial g(X_{N_M h}^{i,N}) C_{1,N_M h}^{i,N} + C_{2,N_M h}^{i,N} \right) \rightarrow -\frac{1}{2} \epsilon \lim_{N \rightarrow \infty} N\mathbf{E}_0 \left[\partial g(X_T^N) C_{1,T}^N + C_{2,T}^N \right] = -\frac{1}{2} \epsilon K_T(X_0)$$

as $M \rightarrow \infty$ and $\frac{\sqrt{M}}{N_M} \rightarrow \epsilon$. Defining $g_{cT}^{N,M} = \frac{1}{M} \sum_{i=1}^M \left[g(X_{Nh}^{i,N}) + \frac{1}{2} \left(\partial g(X_{Nh}^{i,N}) C_{1,Nh}^{i,N} + C_{2,Nh}^{i,N} \right) \right]$ it then follows that the term $\frac{1}{2} \left(\frac{1}{\sqrt{M}} \sum_{i=1}^M \left(\partial g(X_{N_M h}^{i,N_M}) C_{1,N_M h}^{i,N_M} + C_{2,N_M h}^{i,N_M} \right) \right)$ corrects the asymptotic second-order bias for the estimator without transformation when $M \rightarrow \infty$.

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 Theorem 7: The error can be expanded as follows

$$\begin{aligned} U_{\tau,T}^{g^{N,M}} &= \frac{\sqrt{M}}{N} \left(N \left(\frac{1}{M} \sum_{i=1}^M \left(g(X_T^i(X_\tau^N)) - g(X_T^i(X_\tau)) \right) \right) \right) \\ &\quad + \frac{1}{\sqrt{M}} \sum_{i=1}^M \left(g(X_T^i(X_\tau^N)) - \mathbf{E}_\tau \left[g(X_T(X_\tau^N)) \right] \right) \\ &\quad + \sqrt{M} \mathbf{E}_\tau \left[g(X_T(X_\tau^N)) - g(X_T(X_\tau)) \right]. \end{aligned} \quad (52)$$

Given the proof of Theorem 3 and the fact that $\mathbf{P} - \lim X_\tau^N = X_\tau$ it only remains to analyze the limit of the last term. By the mean value theorem we can write, for some $\lambda_1 \in]0, 1[$,

$$\sqrt{M} \mathbf{E}_\tau \left[U_{\tau,T}^{g(X(X^N))} \right] = \sqrt{M} \sum_{j=1}^d \mathbf{E}_\tau \left[\partial_j g \left(X_T(X_\tau) + \lambda_1 e_j U_{j,\tau,T}^{X(X^N)} \right) U_{j,\tau,T}^{X(X^N)} \right]$$

where $U_{\tau,T}^{g(X(X^N))} = g(X_T(X_\tau^N)) - g(X_T(X_\tau))$ and $U_{j,\tau,T}^{X(X^N)} = X_{j,\tau,T}(X_\tau^N) - X_{j,\tau,T}(X_\tau)$. But again, by the mean value theorem, we have for some $\lambda_2, \lambda_3 \in]0, 1[$ that

$$\begin{aligned} U_{j,\tau,T}^{X(X^N)} &= X_{j,\tau}^N - X_{j,\tau} + \sum_{l=1}^d \int_\tau^T \partial_l A_j \left(X_s(X_\tau) + \lambda_2 e_l U_{l,\tau,T}^{X(X^N)} \right) U_{l,\tau,s}^{X(X^N)} ds \\ &\quad + \sum_{k=1}^d \int_\tau^T \partial_l B_{j,k} \left(X_s(X_\tau) + \lambda_3 e_l U_{l,\tau,T}^{X(X^N)} \right) dW_s^k U_{l,\tau,s}^{X(X^N)} \end{aligned}$$

and, consequently, the limit $\sqrt{M} \mathbf{E}_\tau \left[U_{\tau,T}^{g(X(X^N))} \right] \Rightarrow \frac{\sqrt{M}}{\sqrt{N}} \mathbf{E}_\tau \left[\partial g(X_{\tau,T}) \Omega_{\tau,T} \right] U_\tau^{X^N} + \mathbf{o}_\mathbf{P}(1)$ holds as $N \rightarrow \infty$. Thus, if we choose N_M such that $\frac{M}{N_M} \Rightarrow \epsilon^2 < \infty$ as $M \rightarrow \infty$ we obtain

$$\sqrt{M}\mathbf{E}_\tau \left[U_{\tau,T}^{g(X(X^{N_M}))} \right] \rightarrow \varepsilon \mathbf{E}_\tau [\partial g(X_{\tau,T})\Omega_{\tau,T}] U_\tau^X$$

where $\Omega_{\tau,T}$ and U^X are defined in Theorem 1. Since the first term in (52) is $O(1/\sqrt{N_M})$ when $\frac{M}{N_M}$ is $O(1)$ we conclude that

$$U_{\tau,T}^{g^{N_M,M}} \Rightarrow L_T(X_\tau) + \varepsilon \mathbf{E}_\tau [\partial g(X_{\tau,T})\Omega_{\tau,T}] U_\tau^X.$$

Finally, the equivalence with the error corrected estimator follows since the first term in the error expansion, which converges to the second-order bias, is asymptotically negligible. ■

Proof of Theorem 8: We can expand the error $U_{\tau,T}^{\hat{g}^{N,M}}$ as in the proof of Theorem 7 and obtain for the last term of this expansion $\sqrt{M}\mathbf{E}_0 \left[U_{\tau,T}^{\hat{g}(\hat{X}(\hat{X}^{N_M}))} \right] = \frac{\sqrt{M}}{N} \mathbf{E}_\tau \left[\partial g(\hat{X}_{\tau,T})\hat{\Omega}_{\tau,T} \right] U_\tau^{\hat{X}^N} + \mathbf{o}_P(1)$ as $N \rightarrow \infty$. For N_M such that $\frac{\sqrt{M}}{N_M} \rightarrow \epsilon < \infty$ as $M \rightarrow \infty$, we get

$$\sqrt{M}\mathbf{E}_0 \left[U_{\tau,T}^{\hat{g}(\hat{X}(\hat{X}^{N_M}))} \right] \Rightarrow \epsilon \mathbf{E}_\tau \left[\partial g(\hat{X}_{\tau,T})\hat{\Omega}_{\tau,T} \right] U_\tau^X$$

In this case the first term will not vanish as $\frac{\sqrt{M}}{N_M} \rightarrow \epsilon < \infty$. It follows that

$$U_{\tau,T}^{\hat{g}^{N_M,M}} \Rightarrow \epsilon \hat{K}_T(\hat{X}_\tau) + L_T(\hat{X}_\tau) + \epsilon \mathbf{E}_\tau \left[\partial g(\hat{X}_{\tau,T})\hat{\Omega}_{\tau,T} \right] U_\tau^{\hat{X}}.$$

The equivalent second-order bias corrected estimator is then $\hat{g}_c^{N,M}$ since

$$U_{\tau,T}^{\hat{g}_c^{N_M,M}} \Rightarrow L_T(\hat{X}_\tau) + \epsilon \mathbf{E}_\tau \left[\partial g(\hat{X}_{\tau,T})\hat{\Omega}_{\tau,T} \right] U_\tau^{\hat{X}}.$$

This complete the proof of the theorem. ■

Proof of Theorem 9: Since $N(\tilde{X}_T^N - X_T) = N(\tilde{X}_T^N - X_T) + N(X_T^N - X_T)$ we obtain with

$$\begin{aligned} H_T^N &= \int_0^T \sum_{l=1}^d \partial_l A(X_s^N + \lambda_{3,l} e_l \bar{U}_s^{X_l^N}) A_l(X_{\eta_s^N}^N) dV_s^{1,N} \\ &+ \int_0^T \sum_{l=1}^d \partial_l A(X_s^N + \lambda_{3,l} e_l \bar{U}_s^{X_l^N}) \sum_{j=1}^d B_{l,j}(X_{\eta_s^N}^N) dV_s^{2,j,N} \\ &\int_0^T \sum_{l=1}^d \sum_{j=1}^d \partial_l B_j(X_s^N + \lambda_{4,l} e_l \bar{U}_s^{X_l^N}) A_l(X_{\eta_s^N}^N) dV_s^{3,j,N}, \end{aligned}$$

and $(R_s)' = \int_0^s (\partial A(X_s) ds + \partial B_j(X_s) dW_s^j)$ from the fact that $N(\tilde{X}_T^N - X_T) = N(\tilde{X}_T^N - X_T^N) + N(X_T^N - X_T)$ and the expression for $X_T^N - X_T$ in (39) that

$$\begin{aligned} N(\tilde{X}_T^N - X_T) &= N(\tilde{X}_T^N - X_T^N) - \int_0^T (dR_s)' N(\tilde{X}_s^N - X_s^N) - \sqrt{N} \sum_{i,j=1}^d \int_0^T [\partial B_j B_i](X_s) dV_s^{4,i,j,N} \\ &+ \int_0^T (dR_s)' N(\tilde{X}_s^N - X_s) - H_T^N + \mathbf{o}_P(1) \end{aligned}$$

Since $N(\tilde{X}_T^N - X_T^N) - \int_0^T (dR_s)' N(\tilde{X}_s^N - X_s^N) - \sqrt{N} \sum_{i,j=1}^d \int_0^T [\partial B_j B_i](X_s) dV_s^{4,i,j,N} = -(\tilde{H}_T^N - H_T^N) + o_{\mathbf{P}}(1)$ where \tilde{H}_T^N is obtained by replacing X^N by \tilde{X}^N , and since $\tilde{H}_T^N - H_T^N = o_{\mathbf{P}}(1)$, lemmas 2, 3, and 4 yield

$$\begin{aligned} N(\tilde{X}_T^N - X_T) &\Rightarrow \int_0^T \left[\partial A(X_s) ds + \sum_{j=1}^d \partial B_j(X_s) dW_s^j \right] \tilde{U}_s^X \\ &\quad - \frac{1}{2} \int_0^T \partial A(X_s) dX_s - \frac{1}{2} \int_0^T \sum_{k,l,j=1}^d [((\partial_{k,l} A) B_{k,j} B_{l,j} + \partial_l A (\partial_k B_{l,j}) B_{k,j})](X_s) ds \\ &\quad - \frac{1}{\sqrt{12}} \sum_{j=1}^d \int_0^T [(\partial A) B_j](X_s) dZ_{1,s}^j - [(\partial B_j) A](X_s) dZ_{2,s}^j. \end{aligned}$$

Solving the linear SDE for \tilde{U}_T^X establishes the result announced. ■

Proof of Theorem 10: The proof is the same as for the Euler schemes with and without transformation. ■

Proof of Theorem 11: Note that U_T^X is the product of two independent random variables one of which has null expectation. Since $\mathbf{E} \left[\Omega_T \int_0^T \Omega_s^{-1} [(\partial A) B_j](X_s) dZ_{1,s}^j - [(\partial B_j) A](X_s) dZ_{2,s}^j \right] = 0$, for all $j = 1, \dots, d$ we get the result announced. ■

Proof of Theorem 12: The proof follows the same steps as the proof for the Euler scheme with transformation. In this case the average over independent replications of the random variables $\frac{1}{2} \partial g(\tilde{X}_T^N) \tilde{C}_{nh}^N$ 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 Theorem 13: As for the Euler scheme with transformation we obtain

$$\sqrt{M} \mathbf{E}_0 \left[U_{\tau,T}^{g(X(\tilde{X}^N))} \right] = \frac{\sqrt{M}}{N} \mathbf{E}_\tau [\partial g(X_{\tau,T}) \Omega_{\tau,T}] \tilde{U}_\tau^{\tilde{X}^N} + o_{\mathbf{P}}(1)$$

where $\Omega_{t,T}$ is the same as for the Euler scheme without transformation, as $N \rightarrow \infty$. If $N_M \rightarrow \infty$ and $\frac{\sqrt{M}}{N_M} \rightarrow \epsilon < \infty$ when $M \rightarrow \infty$, then

$$\sqrt{M} \mathbf{E}_0 \left[U_{\tau,T}^{g(X(\tilde{X}^{N_M}))} \right] \Rightarrow \epsilon \mathbf{E}_\tau [\partial g(X_{\tau,T}) \Omega_{\tau,T}] U_\tau^X$$

In this case, however, the first term will not vanish as $\sqrt{M}/N_M \rightarrow \epsilon < \infty$. It follows that

$$U_{\tau,T}^{\tilde{g}^{N_M,M}} \Rightarrow \epsilon \tilde{K}_{\tau,T}(X_\tau) + L_T(X_\tau) + \epsilon \mathbf{E}_\tau [\partial g(X_{\tau,T}) \Omega_{\tau,T}] \tilde{U}_\tau^X.$$

The equivalent second-order bias corrected estimator is then $\tilde{g}_c^{N,M}$ since

$$U_{\tau,T}^{\tilde{g}_c^{N_M,M}} \Rightarrow L_T(X_\tau) + \epsilon \mathbf{E}_\tau [\partial g(X_{\tau,T}) \Omega_{\tau,T}] \tilde{U}_\tau^X.$$

This complete the proof of the theorem. ■

Proof of Theorem 14: The mean value theorem and the fact that $\mathbf{P} - \lim_{L \rightarrow \infty} (\hat{\theta}_\Delta^{L, M_L, N_L} - \hat{\theta}_\Delta^L) = 0$ imply the existence of θ_\star^L such that $\mathbf{P} - \lim_{L \rightarrow \infty} \theta_\star^L = \theta_0$ and

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

where $U^{g^{M_L, N_L}}(Y_l, Y_{l+1}, \hat{\theta}_\Delta^{L, M_L, N_L}, \hat{\theta}_\Delta^L) \equiv \sqrt{M_L} \left(g_\Delta^{M_L, N_L}(Y_l, Y_{l+1}; \hat{\theta}_\Delta^{L, M_L, N_L}) - g_\Delta(Y_l, Y_{l+1}; \hat{\theta}_\Delta^L) \right)$, and $\Gamma_\Delta^{L, M_L, N_L}(Y_l, Y_{l+1}, \theta_\star^L) \equiv \frac{1}{L} \sum_{l=0}^{L-1} \partial_\theta g_\Delta^{M_L, N_L}(Y_l, Y_{l+1}; \theta_\star^L)$. By the law of large numbers for ergodic Markov chains $\mathbf{P} - \lim_{L \rightarrow \infty} \Gamma_\Delta^{L, M_L, N_L}(Y_l, Y_{l+1}, \theta_\star^L) = \Sigma_\Delta(\theta_0)$. Similarly, using Theorems 4, 5, or 11, adjusted for the dependence on the observations Y_l, Y_{l+1} , and the fact that $\mathbf{P} - \lim_{L \rightarrow \infty} (\hat{\theta}_\Delta^{L, M_L, N_L} - \hat{\theta}_\Delta^L) = 0$, we conclude that $\mathbf{P} - \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=0}^{L-1} U^{g^{M_L, N_L}}(Y_l, Y_{l+1}, \hat{\theta}_\Delta^{L, M_L, N_L}, \hat{\theta}_\Delta^L) = \varepsilon_2 \kappa_\Delta(\theta_0)$ if $\lim_{L \rightarrow \infty} \sqrt{M_L}/N_L = \varepsilon_2$. The result announced then follows if $\lim_{L \rightarrow \infty} \sqrt{L}/\sqrt{M_L} = \varepsilon_1 < \infty$. ■

8 Appendix B: correcting for second-order bias

This appendix provides formulas for the terms entering the bias-corrected estimators of the conditional expectations.

Let $\nu_{i,j}(X_t, \Omega_t)$ denote the cross variation

$$\nu_{i,j}(X_t, \Omega_t) dt \equiv d[[(B^i)'(\partial^2 g)(\partial B_i)(\partial B_j)B_i + (\partial g)\Omega_T(\partial B_i)\Omega_T^{-1}(\partial B_i)(\partial B_j)B_i](X), W^i]_t.$$

Simple algebra shows

$$\begin{aligned} \nu_{i,j}(X, \Omega) &= [\partial[(B^i)'(\partial^2 g)(\partial B_i)(\partial B_j)B_i]B_i](X) \\ &\quad + [(B_i)' \partial^2 g \Omega (\partial B_i) \Omega^{-1} (\partial B_i) (\partial B_j) B_i + (\partial g) \Omega (\partial B_i) \Omega^{-1} \partial[(\partial B_i) (\partial B_j) B_i]](X) \\ &\quad + [((\Omega^{-1} (\partial B_i) (\partial B_j) B_i)' \otimes (\partial g) \Omega) \partial^2 B_i B_i](X) \end{aligned}$$

where \otimes is the Kronecker product ($A \otimes B = (a_{i,j} B)_{i,j=1,\dots,d}$).

The terms in the bias-corrected estimators are

$$\begin{aligned}
C_{1,(n+1)h}^{i,N} &= C_{1,nh}^{i,N} + [\partial A(X_{nh}^{i,N})h + \sum_{j=1}^d \partial B_j(X_{nh}^{i,N})\Delta W_{nh}^{j,i}]C_{1,nh}^{i,N} \\
&\quad + \left([(\partial A)A + \sum_{j=1}^d \partial B_j \partial A B_j + \sum_{j,k,l=1}^d \partial_k (\partial_l A B_{l,j}) B_{k,j}](X_{nh}^{i,N}) \right) \frac{h}{N} \\
&\quad - \left(\sum_{j,k=1}^d [\partial [\partial B_k \partial B_j B_k] B_i](X_{nh}^{i,N}) \right) \frac{h}{N} \\
&\quad + \sum_{j=1}^d [(\partial A)B_j + (\partial B_j)A - \sum_{i=1}^d (\partial B_j)(\partial B_i)B_j](X_{nh}^{i,N}) \frac{\Delta W_{nh}^{j,i}}{N} \tag{53}
\end{aligned}$$

$$\begin{aligned}
C_{2,(n+1)h}^{i,N} &= C_{2,nh}^{i,N} + \left(\sum_{i,j=1}^d \nu_{i,j}(X_{nh}^{i,N}, \Omega_{nh}^{i,N}) \right) \frac{h}{N} \\
\Omega_{(n+1)h}^{i,N} - \Omega_{nh}^{i,N} &= \left(\partial A(X_{nh}^{i,N})h + \sum_{j=1}^d \partial B_j(X_{nh}^{i,N})(\Delta W_{nh}^{j,i}) \right) \Omega_{nh}^{i,N} \\
X_{(n+1)h}^{i,N} - X_{nh}^{i,N} &= A(X_{nh}^{i,N})h + \sum_{j=1}^d B_j(X_{nh}^{i,N})(\Delta W_{nh}^{j,i}) \tag{54}
\end{aligned}$$

with $X_0^{i,N} = X_0$, $\Omega_0^{i,N} = I_d$ and $C_{1,0}^{i,N} = C_{2,0}^{i,N} = 0$, for the model without transformation and

$$\begin{aligned}
\hat{C}_{(n+1)h}^{i,N} &= \hat{C}_{nh}^{i,N} + [\partial \hat{A}(\hat{X}_{nh}^{i,N})h] \hat{C}_{nh}^{i,N} + \left([(\partial \hat{A})\hat{A} + \sum_{j,k,l=1}^d \partial_{l,k} \hat{A} \hat{B}_{k,j} \hat{B}_{l,j}](\hat{X}_{nh}^{i,N}) \right) \frac{h}{N} \\
&\quad + \sum_{j=1}^d \partial \hat{A}(\hat{X}_{nh}^{i,N}) \hat{B}_j \frac{\Delta W_{nh}^{j,i}}{N} \tag{55} \\
\hat{X}_{(n+1)h}^{i,N} &= \hat{X}_{nh}^{i,N} + \hat{A}(\hat{X}_{nh}^{i,N})h + \sum_{j=1}^d \hat{B}_j(\Delta W_{nh}^{j,i})
\end{aligned}$$

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

For the Milshtein scheme the process $\tilde{C}_{(n+1)h}$ is given by

$$\begin{aligned}
\tilde{C}_{(n+1)h}^{i,N} &= \tilde{C}_{nh}^{i,N} + \left[\partial A(\tilde{X}_{nh}^{i,N})h + \sum_{j=1}^d \partial B_j(\tilde{X}_{nh}^{i,N})\Delta W_{nh}^j \right] \tilde{C}_{nh}^{i,N} \\
&\quad + \frac{1}{2} \partial A(\tilde{X}_{nh}^{i,N}) \frac{\Delta_h X_{nh}^{i,N}}{N} + \frac{1}{2} \sum_{k,l,j=1}^d [((\partial_{k,l} A) B_{k,j} B_{l,j} + \partial_l A (\partial_k B_{l,j}) B_{k,j})] (\tilde{X}_{nh}^{i,N}) \frac{h}{N} \tag{56}
\end{aligned}$$

$$\tilde{X}_{(n+1)h}^{i,N} = \tilde{X}_{nh}^{i,N} + A(\tilde{X}_{nh}^{i,N})h + \sum_{j=1}^d B_j(\tilde{X}_{nh}^{i,N})(\Delta W_{nh}^{j,i}) + \sum_{l,j=1}^d [(\partial B_l)B_j](\tilde{X}_{nh}^{i,N}) \int_{nh}^{(n+1)h} \int_{nh}^s dW_v^l dW_s^j$$

An alternative is to use the Euler scheme to approximate X in the recursion for \tilde{C} . Since both \tilde{X}^N and X^N converge to X , the resulting approximation is first order asymptotically equivalent.

9 Appendix C: portfolios as expectations of solutions of SDEs

We now show how to rewrite the optimal portfolio components as conditional expectations of functions of the terminal point of a diffusion. This embedding is by expansion of the state space. We describe the method for the cases without and with transformation, then provide an illustrative example.

Case without transformation Define the $(d + d \times k + 1 + d) \times 1$ random vector

$$X'_{t,\cdot} = [X'_{1,t,\cdot}, X'_{2,t,\cdot}, X_{3,t,\cdot}, X'_{4,t,\cdot}] = [Y', \text{vec}((\mathcal{D}_t Y)')', G_{t,\cdot}, H_{t,\cdot}]. \quad (57)$$

Using this enlarged vector of state variables we have

$$\pi_t = (\sigma'_t)^{-1} \left[\frac{1}{R} \theta(t, Y_t) - \rho \frac{\mathbf{E}_t[g_2(X_{t,T})]}{\mathbf{E}_t[g_1(X_{t,T})]} \right] \quad (58)$$

where $g_2(x) = \exp(-\rho x_3)x_4$ and $g_1(x) = \exp(-\rho x_3)$.

A straightforward application of Ito's lemma then shows that $X_{t,\cdot}$ satisfies the SDE $dX_{t,v} = A(t, X_{t,v})dv + \sum_{j=1}^d B_j(t, X_{t,v})dW_v^j$ subject to the initial condition $X'_{t,t} = [Y'_t, \sigma^Y(t, Y_t), 0'_d, 0'_d]$, where the functions $A(t, x)$ and $B_j(t, x)$ are respectively given by,

$$A(t, x) = \begin{bmatrix} \mu^Y(t, x_1) \\ [I_d \otimes \partial_y \mu^Y](t, x_1)x_2 \\ [\lambda](t, x_1) \\ [\partial_2 \lambda](t, x_1)x_2 \end{bmatrix} \quad \text{and} \quad B_j(t, x) = \begin{bmatrix} \sigma_j^Y(t, x_1) \\ [I_d \otimes \partial_y \sigma_j^Y](t, x_1)x_2 \\ \theta_j(t, x_1) \\ [\partial_2 \theta](t, x_1)x_2 \end{bmatrix}$$

where \otimes denotes the Kronecker product.

Case with transformation For the system of transformed state variables we proceed as follows. Fix some constant matrix $\hat{B} \in \mathbb{R}^{2d+d \times k+1} \times \mathbb{R}^d$ and $K \in \mathbb{R}^{d \times k+2d+1} \rightarrow \mathbb{R}^{d \times k+2d+1}$ such that (i)

$$(\partial_{x_j} F_i)'_{i,j=1,\dots,2d+d^2+1} = B(B'B)^{-1}\hat{B}' + (I_{2d+d^2+1} - B(B'B)^{-1}B')K. \quad (59)$$

and (ii) $\text{rank} \left([B(B'B)^{-1}]\hat{B}' + (I_{2d+d^2+1} - B(B'B)^{-1}B')K \right) = d$ for all $x \in \mathbb{R}^{2d+d \times k+1}$ and $t \in [0, T]$. If we find \hat{B} such that the ODE (59) has a solution, we can define the new state variables

$\hat{X}_t = F(t, X_t)$ where X_t is as in the previous section. By (ii) and the implicit function theorem it follows that F has an inverse G .

Ito's formula then gives the dynamics $d\hat{X}_t = \hat{A}(\hat{X}_t)dt + \hat{B}dW_t$ where

$$\hat{A}(t, x) = \left[(H_1 A + \partial_t H_2 + \frac{1}{2} H_3 \square B B') \circ G \right] (t, x)$$

with $H_1(t, x) = [B(B'B)^{-1}\hat{B} + (I_{2d+d^2+1} - B(B'B)^{-1}B')](t, x)$, $H_2(t, x) = \int_{\cdot, x} H^1(t, y) dy$ and $H_3(t, x) = \frac{1}{2} \partial_x \text{vec}(H_1(t, x)') + \frac{1}{2} (\partial_x \text{vec}(H_1(t, x)'))'$ (H_3 is the Hessian of F).³⁸ Given the transformed state variables we obtain the portfolio estimator (58) with $\hat{g}_1(T, \hat{X}_{t,T}), \hat{g}_2(T, \hat{X}_{t,T})$ replacing $g_1(X_{t,T}), g_2(X_{t,T})$ where $\hat{g}_i(t, x) \equiv [g_i \circ G(t, \cdot)](x)$ for $i = 1, 2$.

A one factor example In a one factor model one can find more explicit expressions for the transformation. For $dY_t = \mu^Y(Y_t)dt + \sigma^Y(Y_t)dW_t$ the function $F'(x) = [F_1, F_2, F_3, F_4](x)$ takes the form

$$\begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \end{bmatrix} (x) = \begin{bmatrix} \alpha_{11} \int_{\cdot}^{x_1} \frac{du}{\sigma^Y(u)} + \alpha_{12} \\ \alpha_{21} \left(\frac{x_2}{\sigma^Y(x_1)} \right)^{\alpha_{22}} + \alpha_{23} \\ \alpha_{31} \left(\frac{x_2}{\sigma^Y(x_1)} \right)^{\alpha_{32}} \exp \left(-\alpha_{33} \int_{\cdot}^{x_1} \left[\frac{\theta}{\sigma^Y} \right] (u) du + \alpha_{33} x_3 \right) + \alpha_{34} \\ \alpha_{41} \exp \left(\alpha_{42} \left(\frac{x_2}{\sigma^Y(x_1)} \right)^{\alpha_{43}} + \alpha_{44} \left(x_2 \left[\frac{\theta}{\sigma^Y} \right] (x_1) - x_4 \right) \right) + \alpha_{45} \end{bmatrix}$$

where $\alpha_{ij} \in \mathbb{R}$ are arbitrary constants. It is easily verified that $(\partial F B)' = [\alpha_{11}, 0, 0, 0] = \hat{B}'$. The inverse of the function F is $G'(y) = [G_1, G_2, G_3, G_4](y)$ where

$$\begin{bmatrix} G_1 \\ G_2 \\ G_3 \\ G_4 \end{bmatrix} (y) = \begin{bmatrix} G_1(y) \\ \left(\frac{y_2 - \alpha_{23}}{\alpha_{21}} \right)^{\frac{1}{\alpha_{22}}} [\sigma^Y \circ G_1](y) \\ \log \left\{ \left[\left(\frac{y_3 - \alpha_{34}}{\alpha_{31}} \right) \left(\left[\frac{\sigma^Y \circ G_1}{G_2} \right] (y) \right)^{\alpha_{32}} \exp \left(\alpha_{33} \int_{\cdot}^{G_1(y)} \left[\frac{\theta}{\sigma^Y} \right] (u) du \right) \right]^{\frac{1}{\alpha_{33}}} \right\} \\ \log \left\{ \left[\frac{y_4 - \alpha_{45}}{\alpha_{41}} \right]^{\frac{-1}{\alpha_{44}}} \right\} + \frac{\alpha_{42}}{\alpha_{44}} \left(\left[\frac{G_2}{[\sigma^Y \circ G_1]} \right] (y) \right)^{\alpha_{43}} + \left[\left[\frac{\theta}{\sigma^Y} \right] \circ G_1 \right] (y) G_2(y) \right\} \end{bmatrix}$$

and $G_1(y)$ is the inverse of $F_1(x)$, i.e., G_1 satisfies $\alpha_{11} \int_{\cdot}^{G_1(y)} \frac{du}{\sigma^Y(u)} + \alpha_{12} = y_1$.

For $\alpha_{34} = \alpha_{45} = 0$ and $\alpha_{i,j} = 1$ for all other i, j we obtain the transformed drift function

$$\hat{A} = \begin{bmatrix} \frac{\mu^Y}{\sigma^Y} - \frac{1}{2} \partial \sigma^Y \\ x_2 \left(\frac{\partial \mu^Y}{\sigma^Y} - \frac{\mu^Y \partial \sigma^Y}{(\sigma^Y)^2} - \frac{1}{2} \partial^2 \sigma^Y \right) \\ F_3 \left(- \left(\frac{\partial \sigma^Y}{\sigma^Y} + \frac{\theta}{\sigma^Y} \right) \mu^Y + \partial \mu^Y + \lambda - \frac{1}{2} (\sigma^Y \partial^2 \sigma^Y - \theta \partial \sigma^Y + \sigma^Y \partial \theta) \right) \\ x_2 F_4 \left(\left(\frac{\partial \theta}{\sigma^Y} - \frac{\partial \sigma^Y (1+\theta)}{(\sigma^Y)^2} \right) \mu^Y + \left(\frac{1+\theta}{\sigma^Y} \right) \partial \mu^Y - \partial \lambda + \frac{1}{2} \left(\frac{\partial^2 \theta}{\sigma^Y} - (1+\theta) \partial^2 \sigma^Y \right) \right) \end{bmatrix} \circ G$$

whereas $\hat{B}' = [1 \ 0 \ 0 \ 0]$.

³⁸It is possible to derive more explicit expressions for H_3 .

References

- [1] Ait-Sahalia, Y., “Maximum likelihood estimation of discretely observed diffusions: a closed form approximation approach,” *Econometrica*, 70, 2002: 223-262.
- [2] Bally, V. and D. Talay, “The law of the Euler scheme for stochastic differential equations (I): convergence rate of the distribution function,” *Probability Theory and Related Fields*, 104, 1996a: 43-60.
- [3] Bally, V. and D. Talay, “The law of the Euler scheme for stochastic differential equations (II): convergence rate of the density,” *Monte Carlo Methods and its Applications*, 2, 1996b: 93-128.
- [4] Barberis, N., “Investing for the long run when returns are predictable,” *Journal of Finance*, 55, 2000: 225-264.
- [5] Basawa, I. V. and D. J. Scott, *Asymptotic optimal inference for non-ergodic models*, Lecture Notes in Statistics 17, Springer, New York, 1982.
- [6] Bibby, B. M. and M. Sørensen, “Martingale estimating functions for discretely observed diffusion processes,” *Bernoulli*, 1, 1995: 17-39.
- [7] Billingsley, P., *Convergence of Probability Measures*, J. Wiley & Sons, New York, 1968.
- [8] Brandt, M. and P. Santa-Clara, “Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets,” *Journal of Financial Economics*, 63, 2002: 161-210.
- [9] Broze, L., Scaillet, O. and J-M. Zakoian, “Quasi-indirect inference for diffusion processes,” *Econometric Theory*, 14, 1998: 161-186.
- [10] Carrasco, M., Chernov, M., Gallant, A. R. and E. Ghysels, “Estimating diffusions with a continuum of moment conditions,” Discussion Paper, University of North Carolina.
- [11] Chacko, G. and L. Viceira, “Spectral GMM estimation of continuous-time processes,” Discussion paper, Harvard Business School, 1999.
- [12] Chamberlain, G., “Asymptotic efficiency in estimation with conditional moment restrictions,” *Journal of Econometrics*, 34, 1987: 305-334.
- [13] Chamberlain, G., “Efficiency bounds for semiparametric regression,” *Econometrica*, 60, 1992: 567-596.
- [14] Cox, J. C. and C. Huang, “Optimal consumption and portfolio policies when asset prices follow a diffusion process,” *Journal of Economic Theory*, 49, 1989: 33-83.

- [15] Dacunha-Castelle, D. and D. Florens-Zmirou, "Estimation of the coefficients of a diffusion from discrete observations," *Stochastics* 19, 1986: 263-284.
- [16] Detemple, J. B., Garcia, R. and M. Rindisbacher, "A Monte Carlo method for optimal portfolios," *Journal of Finance*, 58, 2003: 401-446.
- [17] Doss, H., "Liens entre équations différentielles stochastiques et ordinaires," *Annales de l'Institut H. Poincaré*, 13, 1977: 99-125.
- [18] Duffie, D. and P. Glynn, "Efficient Monte Carlo simulation of security prices," *Annals of Applied Probability*, 5, 1995: 897-905.
- [19] Duffie, D. and P. Protter, "From discrete to continuous finance: weak convergence of the financial gain process," *Mathematical Finance*, 2, 1992: 1-15.
- [20] Duffie, D. and K. Singleton, "Simulated moments estimation of Markov models of asset prices," *Econometrica*, 61, 1993: 929-952.
- [21] Durham, G. B. and A. R. Gallant, "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes," *Journal of Business and Economic Statistics*, 20, 2002:297-316.
- [22] Elerian, O., Chib, S. and N. Shepard, "Likelihood inference for discretely observed non-linear diffusions," *Econometrica*, 69, 2001: 959-993.
- [23] Eraker, B., "MCMC analysis of diffusion models with application to finance," *Journal of Business and Economic Statistics*, 19, 2001: 177-191.
- [24] Florens-Zmirou, D., "Approximate discrete-time schemes for statistics of diffusion processes," *Statistics*, 20, 1989: 547-557.
- [25] Florens-Zmirou, D., "On estimating the diffusion coefficient from discrete observations," *Journal of Applied Probability*, 30, 1993:790-804.
- [26] Fournié, E., J-M. Lasry, Lebuchoux, J., Lions, P-L. and N. Touzi, "Applications of Malliavin calculus to Monte Carlo methods in finance," *Finance and Stochastics*, 3, 1999: 391-412.
- [27] Gaines, J. G. and T. J. Lyons, "Variable step size control in the numerical solution of stochastic differential equations," *SIAM Journal of Applied Mathematics*, 57, 1997: 1455-1484.
- [28] Gallant, A. R. and G. Tauchen, "Which moments to match?," *Econometric Theory*, 12, 1996: 657-681.
- [29] Gallant, A. R. and G. Tauchen, "Simulated score methods and indirect inference for continuous-time models," forthcoming in the *Handbook of Financial Econometrics*, 2002.

- [30] Genon-Catalot, V., “Maximum contrast estimation for diffusion processes from discrete observations,” *Statistics*, 21, 1990: 99-116.
- [31] Genon-Catalot, V. and J. Jacod, “On the estimation of the diffusion coefficient for multidimensional diffusion processes,” *Annales de l’I.H.P., Probabilité et Statistiques*, 29, 1993: 119-151.
- [32] Gouriéroux, C., Montfort, A. and E. Renault, “Indirect inference,” *Journal of Applied Econometrics*, 8, 1993: 85-118.
- [33] Gouriéroux, C. and A. Monfort, *Simulation-based econometric methods*, Oxford University Press, Oxford, 1996.
- [34] Heyde, C. C., “On best asymptotic confidence intervals for parameters of stochastic processes,” *Annals of Statistics*, 20, 1992: 603-607.
- [35] Hansen, L. P. and J. A. Scheinkman, “Back to the future: generating moment implications for continuous-time Markov processes,” *Econometrica*, 63, 1995: 767-804.
- [36] Jacod, J. and P. Protter, “Asymptotic error distributions for the Euler method for stochastic differential equations,” *Annals of Probability*, 26, 1998: 267-307.
- [37] Jiang, G. and J. Knight, “Estimation of continuous time processes via the empirical characteristic function,” Discussion paper, York University Canada, 2000.
- [38] Kallenberg, O., *Foundations of modern probability theory*, Springer-Verlag, New York, 1997.
- [39] Karatzas, I., Lehoczky, J. P. and S. E. Shreve, “Optimal portfolio and consumption decisions for a “small investor” on a finite horizon,” *SIAM Journal of Control and Optimization*, 25, 1987: 1557-1586.
- [40] Karatzas, I. and S. E. Shreve, *Brownian motion and stochastic calculus*, Springer-Verlag, New York, 1988.
- [41] Kessler, M. and M. Sørensen, “Estimating equations based on eigenfunctions for a discretely observed diffusion process,” *Bernoulli*, 5(2), 1999: 299-314.
- [42] Kloeden, P. and E. Platen, *Numerical solution of stochastic differential equations*, Springer Verlag, 1997.
- [43] Kurtz, T. G. and P. Protter, “Wong-Zakai corrections, random evolutions and numerical schemes for SDE’s,” in *Stochastic Analysis*, Academic Press, New York, 1991a: 331-346.
- [44] Kurtz, T. G. and P. Protter, “Weak limit theorems of stochastic integrals and stochastic differential equations,” *Annals of Probability*, 19, 1991b: 1035-1070.

- [45] Kurtz, T. G. and P. Protter, “Weak convergence and stochastic integrals and differential equations,” *CIME School in Probability. Lecture Notes in Mathematics*, 1627, Springer Verlag, 1996: 1-41.
- [46] Lo, A., “Maximum likelihood estimation of generalized Ito processes with discretely-sampled data,” *Econometric Theory*, 4, 1988: 231-247.
- [47] Magnus, J. and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics*, John Wiley & Sons Ltd., New York, 1988.
- [48] Merton, R. C., “Optimum consumption and portfolio rules in a continuous time model,” *Journal of Economic Theory*, 3, 1971: 273-413.
- [49] Milshtein, G. N., *Numerical integration of stochastic differential equations*, Kluwer Academic Publishers, New York, 1995.
- [50] Milshtein, G. N., “Weak approximation of solutions of systems of stochastic differential equation,” *Theory of Probability and its Application*, 4 1984: 750-768.
- [51] Milshtein, G. N., Schoenmakers, J. and V. Spokoiny, “Transition density estimation for stochastic differential equations via forward-reverse representations,” Discussion Paper, WIAS, Berlin 2001.
- [52] Müller, U. U. and W. Wefelmeyer, “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, 2002.
- [53] Newey, W. K., “Semiparametric efficiency bounds,” *Journal of Applied Econometrics*, 5, 1990: 90-135.
- [54] Nualart, D., *The Malliavin calculus and related topics*, Springer Verlag, New York, 1995.
- [55] Ocone, D. and I. Karatzas, “A generalized Clark representation formula, with applications to optimal portfolios,” *Stochastics and Stochastic Reports*, 34, 1991: 187-220.
- [56] Pedersen, A. R., “A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations,” *Scandinavian Journal of Statistics*, 22, 1995a: 55-71.
- [57] Pedersen, A. R., “Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes,” *Bernoulli*, 1, 1995b: 257-209.
- [58] Singleton, K., “Estimation of affine pricing models using the empirical characteristic function,” *Journal of Econometrics*, 102, 2001: 111-141.

- [59] Talay, D., “Efficient numerical schemes for the approximation of expectations of functionals of S.D.E.,” in H. Korezioglu, G. Maziotto, and J. Szpirglas, *Filtering and Control of Random Processes*, Lecture Notes in Control and Information Sciences 61, Springer Verlag, New York, 1984: 294-313.
- [60] Talay, D., “Discrétisation d’une équation différentielle stochastique et calcul approché d’espérance de fonctionelles de la solution,” *Mathematical Modelling and Numerical Analysis*, 20, 1986: 141-179.
- [61] Talay, D., “Probabilistic numerical methods for partial differential equations: elements of analysis,” *CIME School in Probability, Lecture Notes in Mathematics*, 1627, Springer Verlag, 1996: 149-196.
- [62] Talay, D. and L. Tubaro, “Expansion of the global error for numerical schemes solving stochastic differential equations,” *Stochastic Analysis and its Application*, 8, 1990: 483-509.
- [63] Van der Vaart, A. W., *Asymptotic Statistics*, Cambridge University Press, Cambridge, 1998.
- [64] Wong, E. and M., Zakai, “On the convergence of ordinary integrals to stochastic integrals,” *Annals of Mathematical Statistics*, 36, 1964: 1560-1564.

Table I: Parameter values for a mean reverting constant elasticity of variance process.

Parameter Values				
$dX_v = \kappa(\bar{X} - X_v)dv + \sigma X_v^\gamma dW_v$				
	κ	\bar{X}	σ	γ
CIR	0.0305	0.0791	-0.0219	0.5000
CEV	0.0171	0.1138	-0.0655	0.9997

Table II: Size distortion parameter $\delta/\epsilon = K_T/\mathbf{var}[g(X_T)|\mathcal{F}_0]$.

Size Distortion Parameter $\frac{\delta}{\epsilon}$				
Horizon	CIR		NMRCEV	
	no transformation	transformation	no transformation	transformation
2	-0.54877	-0.54907	-0.67265	-0.63492
4	-1.44504	-1.46253	-1.63078	-1.57080
6	-2.46890	-2.49207	-2.55984	-2.49120
8	-3.62079	-3.63765	-3.49561	-3.42404
10	-4.74366	-4.77496	-4.36092	-4.29001

Note: The size distortion is increasing in this parameter.

Table III: second-order bias K_T .

Second-order bias K_T				
Horizon	CIR		DGR	
	no transformation	transformation	no transformation	transformation
2	3.6631e-07	3.2710e-05	5.2684e-06	0.00099117
4	7.3048e-07	5.7614e-05	8.4232e-06	0.00125843
6	1.1086e-06	7.6642e-05	1.0451e-05	0.00129968
8	1.5108e-06	9.1076e-05	1.1907e-05	0.00127281
10	1.9314e-06	1.0183e-04	1.3153e-05	0.00122986

Table IV: Bias corrected estimators of hedging demand.

Bias Corrected Estimators of Hedging Demand: CIR					
Horizon	No transformation		With transformation		"True"
	no correction	with correction	no correction	with correction	
2	0.0320	0.0320	0.0320	0.0320	0.0320
4	0.0623	0.0623	0.0623	0.0623	0.0623
6	0.0909	0.0909	0.0909	0.0910	0.0910
8	0.1179	0.1179	0.1179	0.1180	0.1180
10	0.1434	0.1434	0.1433	0.1435	0.1435

Bias Corrected Estimators of Hedging Demand: NMRCEV					
Horizon	No transformation		With transformation		"True"
	no correction	with correction	no correction	with correction	
2	0.0193	0.0193	0.0191	0.0196	0.0193
4	0.0341	0.0341	0.0338	0.0344	0.0341
6	0.0456	0.0456	0.0453	0.0460	0.0457
8	0.0550	0.0550	0.0547	0.0553	0.0551
10	0.0629	0.0629	0.0625	0.0631	0.0630

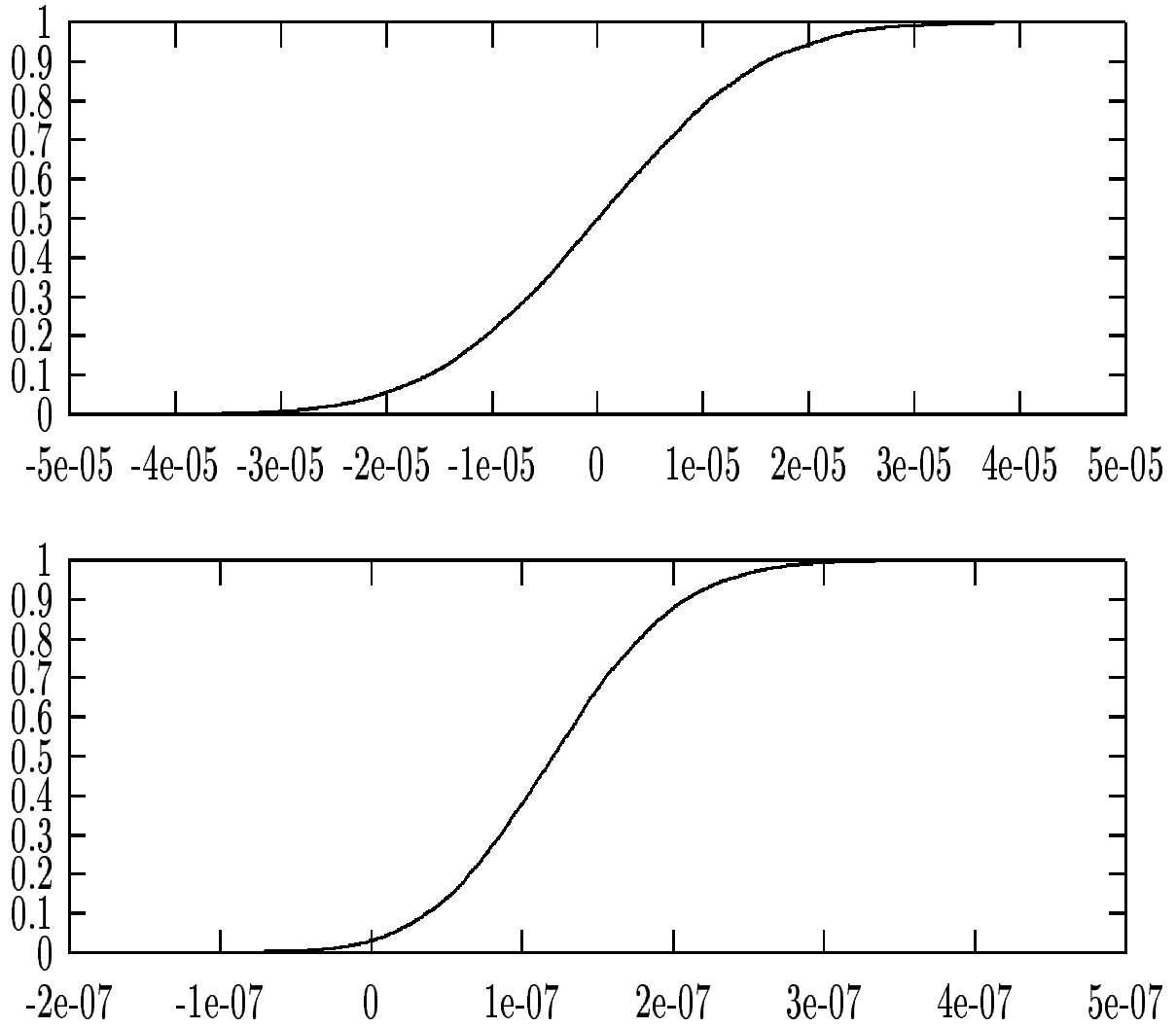


Figure 1: Asymptotic Error distribution function of CIR process, approximated with a Euler scheme without Doss transformation ($\mathbf{P}(U_1^X \leq x)$ upper graph), and with Doss transformation ($\mathbf{P}(U_1^{G(\hat{X})} \leq x)$ lower graph) .

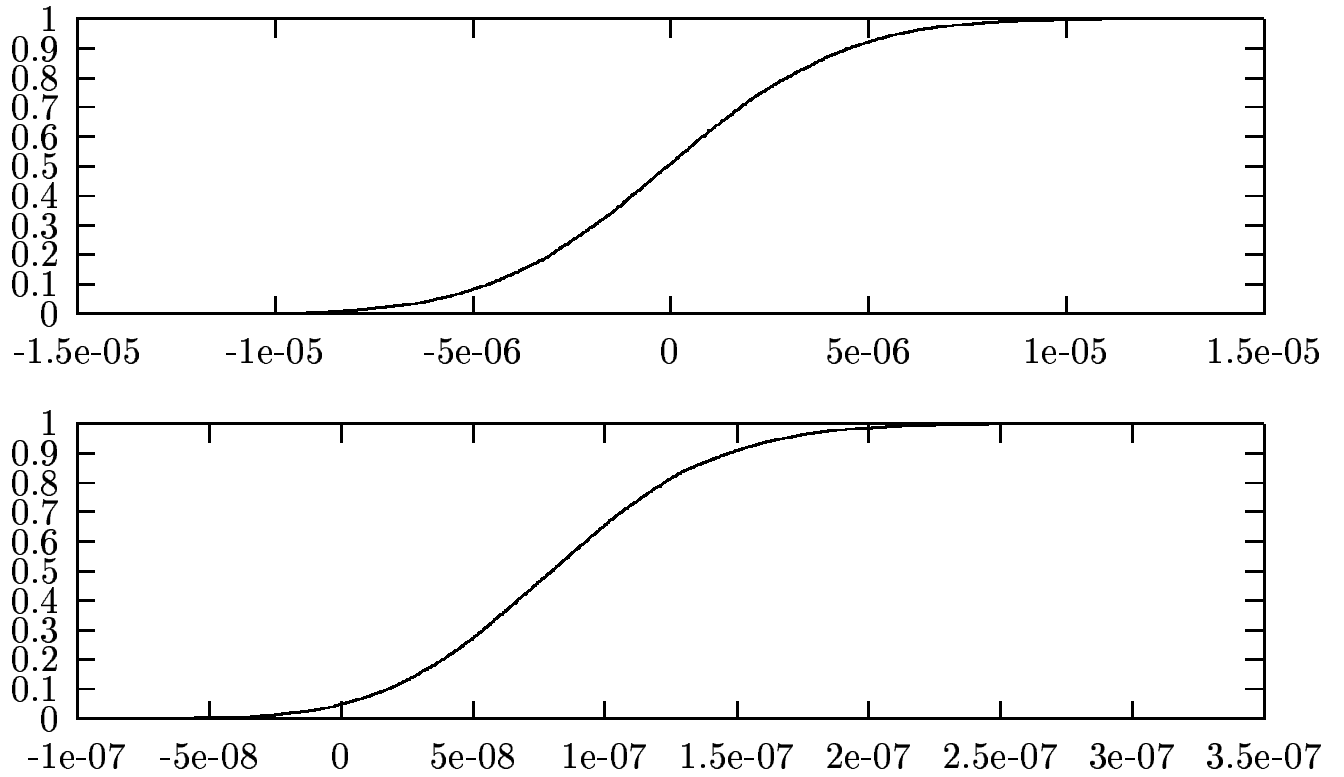


Figure 2: Asymptotic Error distribution function of a CEV process, approximated with a Euler scheme without Doss transformation ($\mathbf{P}(U_1^X \leq x)$ upper graph) , and with Doss transformation ($\mathbf{P}(U_1^{G(\hat{X})} \leq x)$ lower graph)

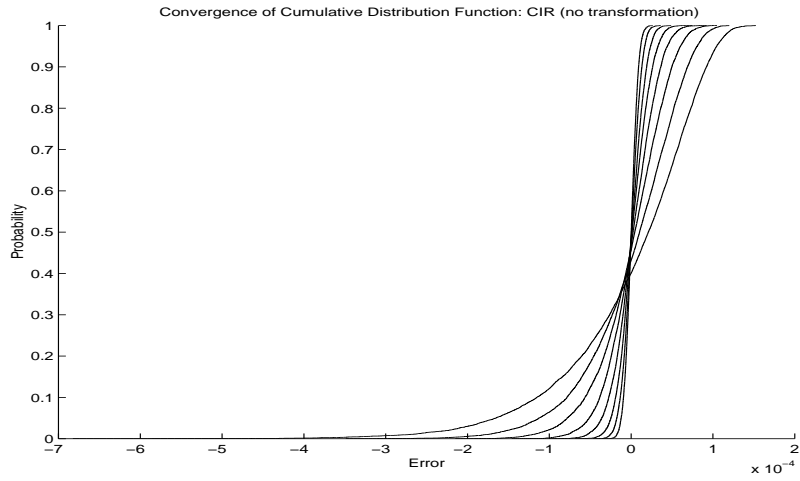


Figure 3: Speed of convergence of distribution function of the approximation error $X_1^N - X_1$ for a CIR process approximated with a Euler scheme for $N = 2^x$ and $x = 2, \dots, 9$.

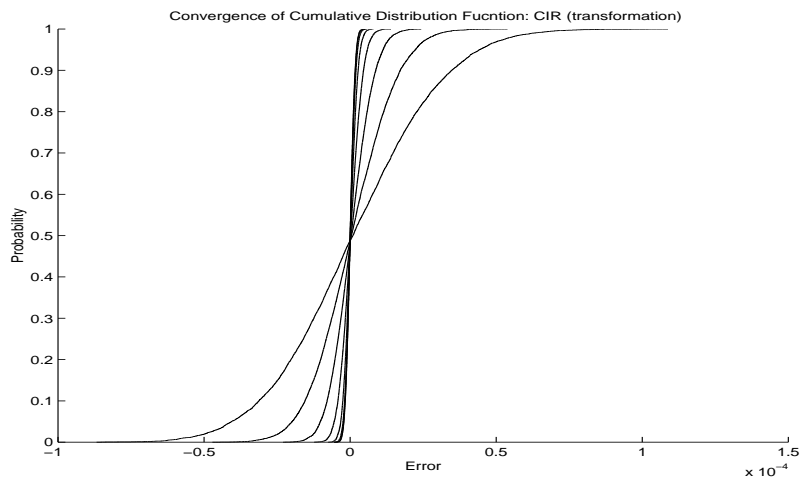


Figure 4: Speed of convergence of the distribution function of the approximation error $G(\hat{X}_1^N) - X_1$ for a Doss-transformed CIR process approximated with a Euler scheme for $N = 2^x$ and $x = 2, \dots, 9$.

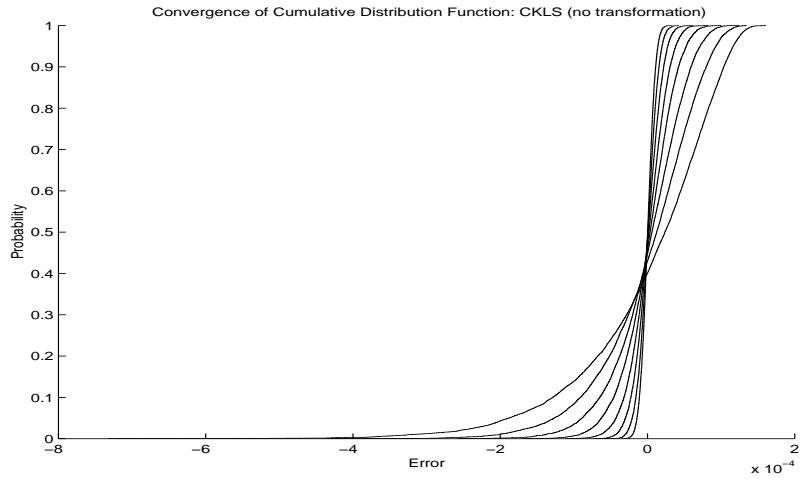


Figure 5: Speed of convergence of the distribution function of the approximation error $X_1^N - X_1$ for a CEV process approximated with a Euler scheme for $N = 2^x$ and $x = 2, \dots, 9$.

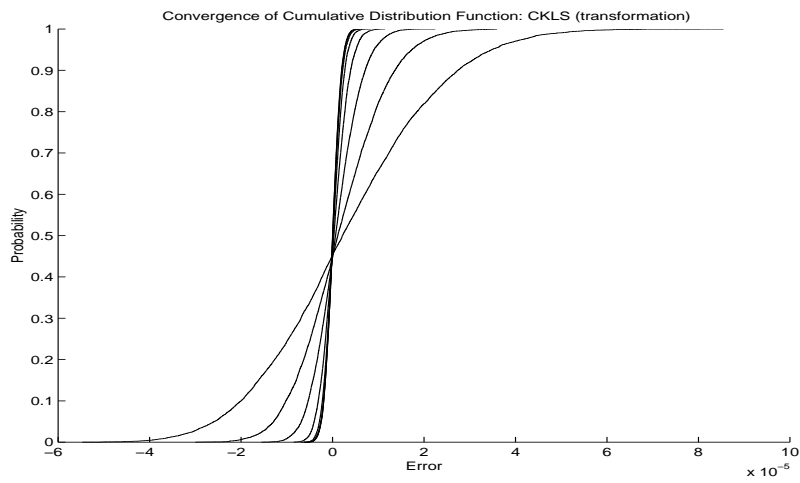


Figure 6: Speed of convergence of the distribution function of the approximation error $G(\hat{X}_T^N) - X_T$ for a Doss-transformed CEV process approximated with a Euler scheme for $N = 2^x$ and $x = 2, \dots, 9$.

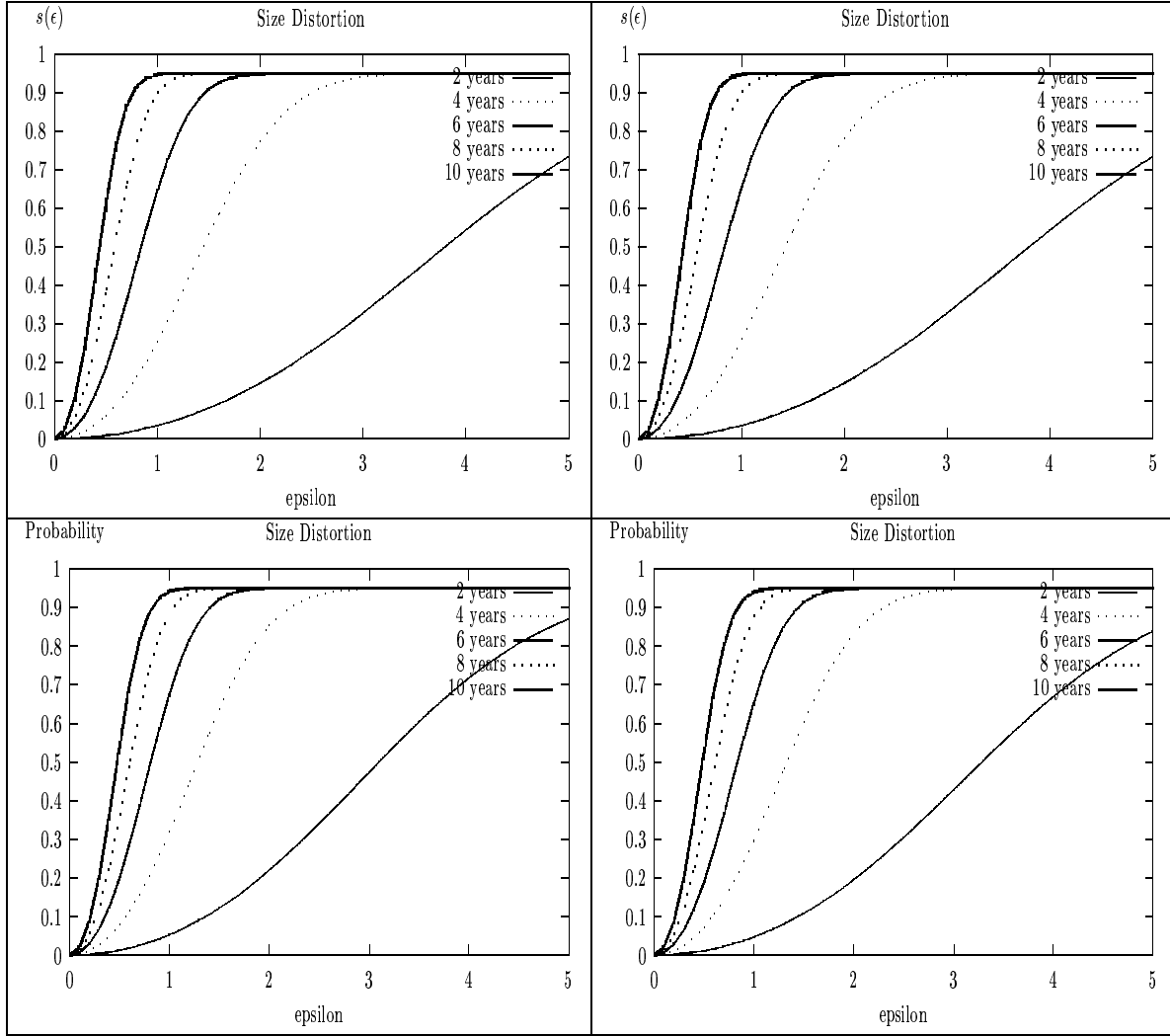


Figure 7: Size Distortion for $(1 - \alpha)\%$ confidence interval: $s(\epsilon) = 1 - \alpha - (\Phi(-\Phi^{-1}(\alpha/2) - \delta(\epsilon)) - \Phi(\Phi^{-1}(\alpha/2) - \delta(\epsilon)))$ as a function of $\epsilon = \lim_{M \rightarrow \infty} \frac{\sqrt{M}}{N_M}$ for $\alpha = 0.05$. The effective coverage probability is not $1 - \alpha$ but $s(\epsilon) - (1 - \alpha)$. CIR, no transformation (upper left); CIR, with transformation (upper right); NMRCEV, no transformation (lower left); NMRCEV, with transformation (lower right)

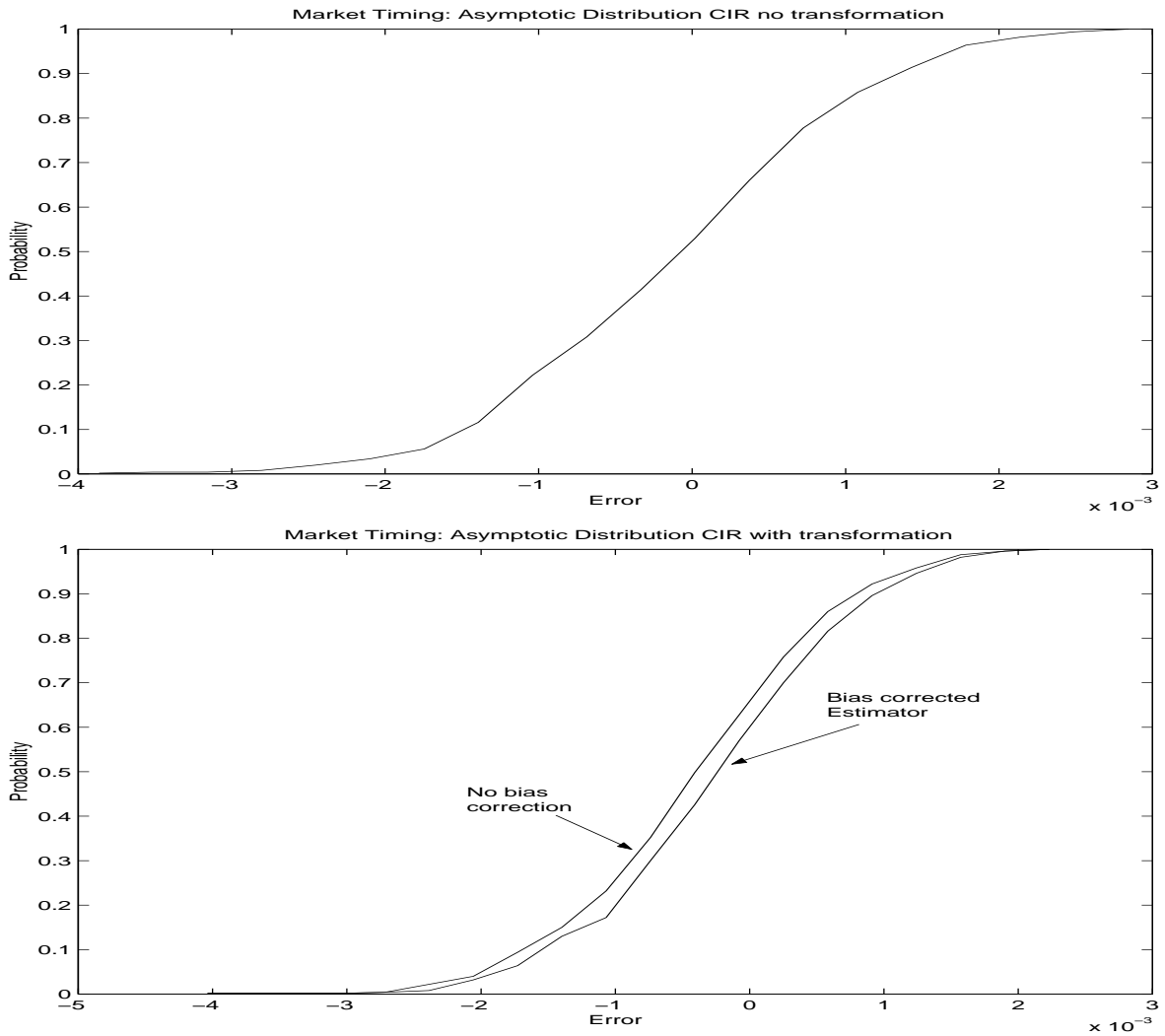


Figure 8: Market Timing CIR: Investment horizon $T = 10$, asymptotic error distribution of hedging demand in year $\tau = 9$ without transformation (top) with transformation (bottom).

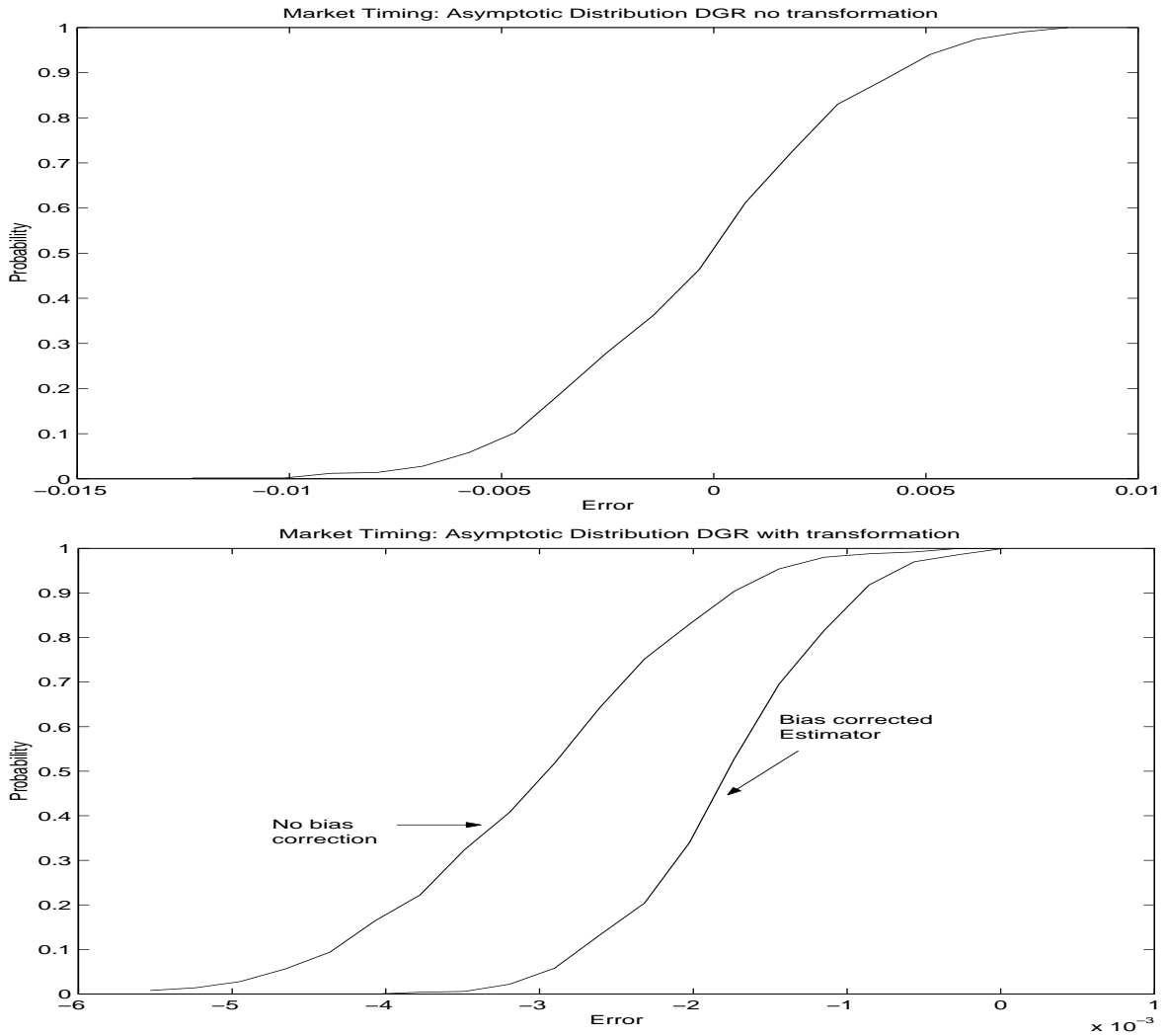


Figure 9: Market Timing DGR: Investment horizon $T = 10$, asymptotic error distribution of hedging demand in year $\tau = 9$ without transformation (top) with transformation (bottom).