Contrast estimation for noisy observations of diffusion processes via closed-form density expansions

Salima El Kolei, Fabien Navarro

Contrast estimation for noisy observations of diffusion processes via closed-form density expansions

Salima El Kolei*, Fabien Navarro†

October 23, 2021

Abstract

When a continuous-time diffusion is observed only at discrete times with measurement noise, in most cases the transition density is not known and the likelihood is in the form of a high-dimensional integral that does not have a closed-form solution and is difficult to compute accurately. Using Hermite expansions and deconvolution strategy, we provide a general explicit sequence of closed-form contrast for noisy and discretely observed diffusion processes. This work allows the estimation of many diffusion processes. We show that the approximation is very accurate and prove that minimizing the sequence results in a consistent and asymptotically normal estimator. Monte Carlo evidence for the Ornstein-Uhlenbeck process reveals that this method works well and outperforms the Euler expansion of the transition density in situations relevant for financial models.

Keywords: M-estimator; Deconvolution; Least Square Method; Parametric approach; Diffusion Processes, Hermite Expansion.

1 Introduction

Statistical inference of one-dimensional diffusion processes with ergodic properties and when the sample path is discretely and observed without measurement errors has been the subject of many papers. The continuous time paradigm has proved to be an immensely useful tool in many applications as for example in financial models where diffusion processes are widely used for instance, to represent the stochastic dynamics of asset returns, exchange rates, interest rates and more generally in economy and biology. Many refinements and extensions are possible, but the basic dynamic model for the variable interest X_t is a stochastic differential equation (SDE)

$$\begin{cases} dX_t = b_\theta(X_t)dt + \sigma_\theta(X_t)dW_t \\ X_0 = \eta, \end{cases} \tag{1}$$

where b and σ are the nonanticipative drift and volatility functions depending on X_t at time t and on an unknown parameter vector θ belonging to Θ, a compact subset of $\mathbb{R}^r$ and η a real random variable independent of the increment of a standard Wiener process dW_t. The

*Salima El Kolei
CREST - ENSAI, France, E-mail: salima.el-kolei@ensai.fr
†Fabien Navarro
SAMM - Université Paris 1 Panthéon-Sorbonne, France, E-mail: fabien.navarro@math.cnrs.fr

process $(X_t)_{t\geq 1}$ is an unobserved real-valued Markov chain belonging to $\mathcal{X} = (l, r)$ where l could possibly be $-\infty$, and r might be $+\infty$. Besides its initial distribution, the chain $(X_t)_{t\geq 1}$ is characterized by its transition, *i.e.*, the distribution of X_{t+1} given X_t and by its stationary density f_θ. The observations consist of the discretized trajectory $(X_{i\Delta})_{1\leq i\leq n}$ observed with measurement errors at dates $\{t_i = i\Delta, i = 0, \ldots, n\}$ where Δ is generally small but fixed as n increases. For instance, the series could be weekly or monthly. Collecting more observations means extending the time period over which data are stored, without shortening the time interval between existing successive observations. Hence, at time $t_i = i\Delta$ we have at disposal noisy data of the sample path, *i.e.*, we observe $Y_{i\Delta} = X_{i\Delta} + \varepsilon_{i\Delta}$ with $(\varepsilon_{i\Delta})_{i\geq 0}$ a sequence of independent and identically distributed (*i.i.d.*) random variables.

Assuming that the model is correctly specified (*i.e.*, θ belongs to the parameter space $\Theta \subset \mathbb{R}^r$, with $r \in \mathbb{N}^*$), we aim to estimate the vector of parameters θ given the measurements $Y_{(1:n)\Delta} = (Y_{1\Delta}, \ldots, Y_{n\Delta})$.

For most estimation strategies relying on discretely sampled data, we need to be able to infer the implications of the infinitesimal time evolution of the process for the finite time intervals at which the process is actually sampled, say weekly or monthly. The transition function plays a key role in that context, but unfortunately, it is, in most cases, unknown. Ignoring the approximation of the transition density leads to inconsistent estimators unless the discretization happens to be an exact one, *i.e.*, the density should be known in closed form. For some of the few exceptions, this transition density is available in closed form, that is if a strong solution of the underlying SDE process is available, meaning that the SDE can be solved analytically in Itô form, one can cite the Black and Scholes model, Ornstein-Uhlenbeck (OU) process and Cox-Ingersoll and Ross (CIR) process (see [52]).

If sampling of the process were continuous and made without measurement noise, the situation would be simpler since the likelihood function for continuous records can be obtained by means of a classical change of measure using Girsanov's Theorem for known volatility function σ. Because diffusions are observed with measurement errors, the log-likelihood function of observations from such a process sampled at finite time intervals do not reduce to the sum of the log-transition function of successive pairs of observations even if the transition density is known, but induces a high-dimensional intractable integral.

This is an important motivation for the study of the theory of estimating functions that provide an alternative method of inference. In fact, inference based on the optimal estimating function can be seen as an approximation of likelihood inference.

Different kinds of estimating functions have been studied for diffusions observed without measurement errors: minimum contrast estimators, simulation based methods, see [20, 41, 32, 49, 2] for large sampling intervals and [53, 21, 31] for small sampling. For noisy observations, in a Bayesian setting, various results are already stated and most of them are based on Monte Carlo (MC) inference (see [8, 47] and [46]). More recently, in [43] the authors propose to use Approximate Bayesian Computation for accelerating inference for nonlinear SDE with measurement errors and large sample size. In this paper, we do not consider the Bayesian approach, the model (1) with measurement noise is known in this case as the so-called convolution diffusion model. If we focus our attention on (semi)-parametric models and low frequency data, few results exist. In [25] the author proposes an estimator based on a contrast method to estimate parameters appearing on the volatility function σ for diffusions processes with additive Gaussian noises. In the same perspective, in [6], the authors propose a Maximum Likelihood Estimator (MLE) for integrated diffusion processes observed with additive noises. They proposed a simulated Expectation-Maximization algorithm to obtain MLE of the parameters in the volatility function. For high frequency data, one can cite the works of [18, 19] where the author proposes a contrast approach to estimate θ inspired by the works of Kessler (see [31]) and Gloter (see [24] and [26]).

In this paper, we propose a new estimation approach which provides a consistent estimator with a parametric rate of convergence for nonlinear diffusions processes as (1) observed discretely with measurement errors. Our approach is to construct appropriate and explicit functions of the observations to replace either the log-likelihood or the score function by a contrast function with the following form: $\sum_{i=2}^{n} g_\theta(Y_{i\Delta}, Y_{(i-1)\Delta})$.

Since the transition density is usually unknown, we construct a closed-form sequence of approximations, hence a sequence of approximations to the true unknown contrast function. These closed-form expressions lead to a computationally efficient alternative to the MLE for diffusions observed with measurement errors. With a given Δ, many methods are available in the literature to compute the transition density. The simplest is based on the Euler approximation of the SDE. The resulting discretization leads to a conditionally Gaussian transition density. A related approach is based on the moment equations for the first two moments (see [48]). Again a conditionally Gaussian scheme is obtained. Alternatively, some of the approaches involve either solving numerically the Kolmogorov partial differential equation known to be satisfied by the transition density (see [40]), or simulating a large number of sample paths along which the process is sampled very finely (see [42], [28], and [10]). Neither of these methods, however, produces a closed-form expression to be minimized over θ, and the calculations for all the pairs (x, y) must be repeated separately every time the value of θ changes. In addition, while these approximations may be useful for small sampling intervals where the transition density deviates only slightly from normality, for larger sampling intervals, corrections are required that take into account the higher order characteristics of the true density. Examples of such approaches include [2, 39, 38] and [12]. In [30] the author compares many different schemes and shows that the Hermite polynomial expansion outperforms these competitors.
In this paper we consider a Hermite expansion with leading Gaussian term but with corrections terms taking into account higher order moments of the true density. This approximation leads to a closed-form expression for the transition density and make possible to minimize the expression in the true contrast with the closed-form approximation. This method has been shown to be very accurate, even when the series are truncated after only a few terms, for a variety of diffusion models, in the univariate case see for example [1, 30, 50, 29] and [3] for multivariate diffusion processes.
Taking these corrections into account is all the more important in the context of data collected with measurement noise. Indeed, we have seen that most approximation schemes lead to a conditionally Gaussian transition density which is not the most favorable situation in any denoising procedure. The approach in practice is facilitated by Gaussian noise since it leads to an analytical expression of estimating functions and their derivatives, but the convergence results are not optimal in this case. In particular in the context of deconvolution, the smoother the function, the slower the convergence. Thus, any approach that allows these approximations schemes to be corrected by taking into account higher order moments of the true transition density will improve the results as we will see in the numerical simulations.

We provide in this paper the expression of our contrast for general diffusion processes, *e.g.*, for SDE with a volatility function $\sigma_\theta(x)$ dependent and independent of the state x. Such a study allows the estimation of many diffusion processes in various fields. We show in this work that in the context of deconvolution, a significant approximation error of the transition density in the time domain is all the more amplified in the frequency domain by using the Fourier transform (FT) and so the numerical simulations show that the Hermite expansion contrast gives better results than the same contrast based on a simple Euler discretization.

The paper is structured as follows. We briefly recall in Section 2 the Hermite expansion of the transition density. Then, in Section 3 we detail the construction of our contrast and explore its asymptotic statistical properties, in particular consistency and central limit theorem (CLT). In Section 4 we shape this contrast for general diffusion processes. A complete analysis of this approach is carried out for the OU process in Section 5. The technical proofs are postponed in Section 6.

2 A sequence of expansion of the transition function

2.1 Notations and assumptions

To define the Hermite polynomials approach, we introduce the following notation. We denote by $\nabla_x g$ the vector of the partial derivatives of g with respect to (w.r.t) x, *i.e.*, $\nabla_x g = (\partial_{x_1} g, \dots, \partial_{x_d} g)'$ where $\partial_{x_i} g$ denotes the partial derivative of g w.r.t. x_i for $i = 1, \dots, d$. The Hessian matrix of g w.r.t x is denoted by $\nabla_x^2 g$. For any matrix $A = A_{i,j}$, the Frobenius norm is defined by $\|A\| = \sqrt{\sum_i \sum_j |A_{i,j}|^2}$.

For ease of exposition, we introduce some technical assumptions which are conventionally proposed in the study of the existence of a solution of the SDE (1) with a transition density that is sufficiently regular to define our approximate contrast estimator shortly thereafter. Weaker conditions on the volatility function close to the boundary of the state space can be considered, *e.g.*, at zero for positive diffusions (see [44] and [2]).

A 1. ***Existence*** & ***Regularity:***

(i) (Smoothness): The functions b and σ in the SDE (1) *are infinitely differentiable in x and three times continuously differentiable in θ.*

(ii) (Non-degeneracy of the diffusion): There exists a constant c such that $\sigma_\theta(x) > c > 0$ for all $x \in \mathcal{X}$ and $\theta \in \Theta$.

Under Assumption **A** 1, the SDE (1) admits a weak solution, unique in probability law, for every distribution of its initial value X_0. Furthermore, X admits a transition density denoted $\Pi_{\theta,X}(\Delta, x_0, x)$ which is continuously differentiable in $\Delta > 0$, infinitely differentiable in x and x_0 and three times continuously differentiable in $\theta \in \Theta$ (see [2]). This latter will serve us for the statistical study of our estimator.

2.2 The Hermite expansion

We follow steps similar to [2]. First, we transform the underlying process X defined in the SDE (1) to another process U, whose transition density becomes closer to the Normal density. Next, we perform another transformation from U to Z, which is sufficiently close to a standard Gaussian random variable. We then find a convergent density expansion for Z with standard normal density serving as a leading term. Further, we obtain the density expansion for X by a change of variable. More specifically, first consider a transformed process also known as the

Lamperti transformed process[1]:

$$\rho_\theta(x) = \int_{x^*}^{x} \frac{1}{\sigma_\theta(s)} ds,$$

where x^* is arbitrary. This transformed process is applied to obtain the diffusion process of $U_t = \rho_\theta(X_t)$. Since $\sigma_\theta > 0$, the transformation ρ_θ is increasing, and by Ito's formula one can easily see that this transformed process satisfies the following unit diffusion:

$$dU_t = \mu_{\theta,U}(U_t)dt + dW_t,$$

where $\mu_{\theta,U}(u) = b_\theta(\rho_\theta^{-1}(u))/\sigma_\theta(\rho_\theta^{-1}(u)) - \partial_x\sigma_\theta(\rho_\theta^{-1}(u))/2$.
Under Assumption **A** 1 and the Jacobi formula, the transition density is given by

$$\Pi_{\theta,X}(\Delta, x_0, x) = \sigma_\theta(x)^{-1}\Pi_{\theta,U}(\Delta, \rho_\theta(x_0), \rho_\theta(x)), \tag{2}$$

and has the same regularity properties as the original SDE (1).

The transform $X \to U$ ensures that the tail of the transition density $\Pi_{\theta,U}(\Delta, u_0, u)$ of U_t will generally vanish exponentially fast so that the Hermite series approximation will converge. However, $\Pi_{\theta,U}(\Delta, u_0, u)$ may get peaked at u_0 when the sample frequency gets smaller. To avoid this, Aït-Sahalia considers a further transformation, that is the pseudo-normalized increment of U and instead of expanding the transition density of U, *i.e.*, the conditional density of U_t given $U_0 = u_0$, we expand the conditional density of the normalized increment

$$Z_t = \Delta^{-1/2}(U_t - u_0),$$

given $U_0 = u_0$. Thus, we have

$$\Pi_{\theta,U}(\Delta, u_0, u) = \Delta^{-1/2}\Pi_{\theta,Z}(\Delta, \Delta^{-1/2}(u - u_0)|u_0), \tag{3}$$

where $\Pi_{\theta,Z}(\Delta, z|u_0)$ is the conditional density of Z given that $U_0 = u_0$.

We can now obtain an approximation to the transition density of X, by expanding the conditional density $\Pi_{\theta,Z}$ of Z given $U_0 = u_0$ in terms of Hermite polynomials up to order K:

$$\Pi^K_{\theta,Z}(\Delta, z|u_0) = \varphi(z)\sum_{k=0}^{K} \eta_{\theta,k}(\Delta, u_0)H_k(z), \tag{4}$$

where $\varphi(z)$ is the standard and centered normal density, and $H_k(z)$ is the kth Hermite polynomial defined as

$$H_k(z) = (-1)^k e^{z^2}\frac{\partial^k}{\partial z^k}e^{-z^2}, \quad k = 0, 1, \dots, K,$$

and computed sequentially from the recursive relation $H_{k+1}(z) = zH_k(z) - kH_{k-1}(z), k > 0$. These polynomials are orthogonal on the real line with respect to the weight function e^{-z^2}:

$$\int_{\mathbb{R}} H_l(z)H_m(z)e^{-z^2}dz = 2^l l!\sqrt{\pi}\delta_{l-m},$$

where δ_{l-m} corresponds to the Kronecker function. Using this orthogonality property, the coefficients $\eta_{\theta,k}$'s in (4) are obtained from the following conditional expectation

$$\eta_{\theta,k}(\Delta, u_0) = \frac{1}{k!}\mathbb{E}\left[H_k\left(\Delta^{-1/2}(U_\Delta - u_0)\right)\middle|U_0 = u_0\right].$$

[1] In this paper we consider univariate diffusion processes so reducible processes since it can be transformed into a unit diffusion by Lamperti transformation. Hence, the Hermite method in [2] can be applied to get the equivalent approximate transition density. However for multivariate processes one can generalize our approach, using the approximate transition expansions obtained by the irreducible method established in [3] and [16].

By inserting the expansion (4) in (3) and (2), we obtain the following approximations of the transition densities $\Pi_{\theta,U}$ and $\Pi_{\theta,X}$

$$\Pi^K_{\theta,U}(\Delta,u_0,u)=\Delta^{-1/2}\varphi(\Delta^{-1/2}(u-u_0))\sum_{k=0}^{K}\eta_{\theta,k}(\Delta,u_0)H_k(\Delta^{-1/2}(u-u_0)),$$

and

$$\Pi^K_{\theta,X}(\Delta,x_0,x)=\frac{\varphi\left(\frac{\rho_\theta(x)-\rho_\theta(x_0)}{\sqrt{\Delta}}\right)}{\sqrt{\Delta}\sigma_\theta(x)}\sum_{k=0}^{K}\eta_{\theta,k}(\Delta,\rho_\theta(x_0))H_k\left(\frac{\rho_\theta(x)-\rho_\theta(x_0)}{\sqrt{\Delta}}\right). \tag{5}$$

Under Assumption **A** 1, the convergence as K tends to infinity of the approximation $\Pi^K_{\theta,X}$ to the exact transition density $\Pi_{\theta,X}$ is proved in [2].

3 A sequence of approximations to the minimum contrast estimator

3.1 Notations

In what follows, we denote by θ_0 the true value of the parameter, *i.e.*, the value of the parameter which rules the observation. The FT of an integrable function u is denoted by $u^*(t)=\int e^{-itx}u(x)dx$. We set $\langle u,v\rangle_f=\int u(x)\overline{v}(x)f_{\theta_0}(x)dx$ with $v\overline{v}=|v|^2$. The norm of the operator T is defined by $\|T\|_f=\left(\iint|T(x,y)|^2f_{\theta_0}(x)dxdy\right)^{1/2}$. Let us recall that, by the properties of the FT, we have $(u^*)^*(x)=2\pi u(-x)$. We set $\mathbf{Y}_{i\Delta}=(Y_{i\Delta},Y_{(i+1)\Delta})$ and $\mathbf{y}_{i\Delta}=(y_{i\Delta},y_{(i+1)\Delta})$ is a given realization of $\mathbf{Y}_{i\Delta}$ and we denote $G_{\mathbf{Y}}$ the joint distribution of $\mathbf{Y}_{i\Delta}$.

In the following, for the sake of conciseness, $\mathbb{P},\mathbb{E},\mathbb{V}ar$ and $\mathbb{C}ov$ denote respectively the probability $\mathbb{P}_{\theta_0}$, the expected value $\mathbb{E}_{\theta_0}$, the variance $\mathbb{V}ar_{\theta_0}$ and the covariance $\mathbb{C}ov_{\theta_0}$ when the true parameter is θ_0.

For the purposes of this study, we work with $\Pi(\Delta,x,y)$ the transition density of $X_{(i+1)\Delta}=y$ given $X_{i\Delta}=x$ on a compact subset $A=A_1\times A_2$. For more clarity, we omit the subscript X in $\Pi_{\theta,X}$ and write Π_θ instead of $\Pi_\theta\mathbf{1}_A$. This restriction of the transition function on the compact subset is purely theoretical since in practice the computation of all integrals and FTs are truncated and thus carried out on compact of $\mathbb{R}$ of large size.

3.2 Minimum contrast estimation

Hereafter, we propose explicit estimators of θ, based on the minimization of suitable functions of the observations, called contrasts. We refer to [13, Chapter 3] for a general account of this notion. For the purpose of this study, we consider the contrast function initially introduced by [34] in a nonparametric setting, inspired by regression-type contrasts and later used in various works (see, *e.g.*, [11, 35, 36, 37]), that is

$$\gamma_{\theta,n}(\Delta)=\frac{1}{n}\sum_{i=1}^{n}g_\theta(\mathbf{Y}_{i\Delta}),$$

where $g_\theta(\mathbf{y})$ is a real function defined on $\Theta\times\mathbb{R}^2$ as follows

$$g_\theta(\mathbf{y}_{i\Delta})=\mathcal{K}^1_{\Pi^2_{\theta,X}}(y_{i\Delta})-2\mathcal{K}^2_{\Pi_{\theta,X}}(\mathbf{y}_{i\Delta}), \tag{6}$$

where the functions $\mathcal{K}^1$ and $\mathcal{K}^2$ are two operators defined for any function $h \in L_1(\mathbb{R})$ as

$$\mathcal{K}_h^1(x) = \frac{1}{2\pi}\int e^{ixu}\frac{h^*(u,0)}{f_\varepsilon^*(-u)}du \quad \text{and} \quad \mathcal{K}_h^2(x,y) = \frac{1}{4\pi^2}\iint e^{i(xu+yv)}\frac{h^*(u,v)}{f_\varepsilon^*(-u)f_\varepsilon^*(-v)}dudv, \tag{7}$$

and they are chosen to verify the following Lemma (see [35, 6.1. Proof of Lemma 2] for the proof):

Lemma 3.1. *For all $i \in \{1,\dots,n+1\}$ and any function $h \in L_1(\mathbb{R})$, we have*

1. $\mathbb{E}[\mathcal{K}_h^2(\mathbf{Y}_{i\Delta})|X_{1\Delta},\dots,X_{(n+1)\Delta}] = h(X_{i\Delta},X_{(i+1)\Delta})$.
2. $\mathbb{E}[\mathcal{K}_h^1(Y_{i\Delta})|X_{1\Delta},\dots,X_{(n+1)\Delta}] = \int h(X_{i\Delta},y)dy$.
3. $\mathbb{E}[\mathcal{K}_h^2(\mathbf{Y}_{i\Delta})] = \iint h(x,y)\Pi_{\theta,X}(\Delta,x,y)f_\theta(x)dxdy$.
4. $\mathbb{E}[\mathcal{K}_h^1(Y_{i\Delta})] = \iint h(x,y)f_\theta(x)dxdy$.

We study the associated minimum contrast estimators $\hat{\theta}_n$ as any solution of

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta}\gamma_{\theta,n}(\Delta). \tag{8}$$

Before explaining such a choice of contrast, we define the assumptions necessary for its existence and its asymptotic study.

3.3 Assumptions

Let us consider the one-dimensional process which is defined by the SDE in (1). For the statistical study, the key assumption is that the diffusion (1) has to be a strictly stationary ergodic process on an interval (l,r) with stationary distribution f_θ. Assumption **A** 2 gives conditions on functions b_θ and σ_θ, ensuring it.

A 2. ***Stationary & ergodicity:***

(i) *For all θ, $\exists K_\theta, \forall x \in (l,r), |b_\theta(x)| \le K_\theta(1+|x|)$ and $|\sigma_\theta^2(x)| \le K_\theta(1+|x^2|)$. For $x_0 \in (l,r)$ and for all $\theta \in \Theta$, define the scale and the speed densities of diffusion (X_t),*

$$\varsigma_\theta(x) = \exp\left(-2\int_0^x \frac{b_\theta(z)}{\sigma_\theta^2(z)}dz\right), \quad m_\theta(x) = \left(\varsigma_\theta(x)\sigma_\theta^2(x)\right)^{-1}.$$

(ii) *For all $\theta \in \Theta$, $\int_l^\infty \varsigma_\theta(x)dx = \int_{-\infty}^r \varsigma_\theta(x)dx = +\infty$ and $\int_l^r m_\theta(x)dx = C_\theta < +\infty$. **A** 2(i) leads to ensure the uniqueness-in-law of the solution of (1) and we denote this law by $\mathbb{P}_\theta$ (see [45]). Furthermore, under **A** 2(ii), the process X is ergodic on $\mathbb{R}$ and, with respect to the Lebesgue measure, its invariant measure has density denoted by f_θ and given by*

$$f_\theta : x \in (l,r) \mapsto \frac{1}{C_\theta}m_\theta(x).$$

(iii) *The initial random variable η has distribution $f_\theta(dx) = f_\theta(x)dx$.*

*Assumption **A** 2(iii) provides the strict stationarity.*

(iv) $\lim \sigma_\theta(x) m_\theta(x) = 0$ *as* $x \downarrow l$ *and* $x \uparrow r$ *and* $1/\beta_\theta(x)$ *has a finite limit as* $x \downarrow l$ *and* $x \uparrow r$ *where* $\beta_\theta(x) = \partial_x \sigma_\theta(x) - 2b_\theta(x)/\sigma_\theta(x)$.

Assumption ***A*** *2(iv) is needed to study the mixing property. Note that, in view of* ***A*** *2(ii),* ***A*** *2(iv) is not a strong assumption (see [23]). This assumption is needed for the further central limit theorem of the paper. In particular, under* ***A*** *2(i)-****A*** *2(iv), there exists a positive constant* c_2 *such that for all* $t > 0$, $\alpha_X(t) \leq e^{-c_2 t}/4$, *meaning that the process is geometrically* α*-mixing (see, [22] and [23]).*

Using the Lebesgue dominated convergence and [23, Theorem 2.4] we also have that the discretely sampled process $(X_{i\Delta}, i \geq 0)$ inherits the same properties.

Remark 3.1. *A lot of diffusion processes satisfy this mixing assumption. There are simple conditions for geometric* α*-mixing for one dimensional diffusions. These conditions rely on the eigenvalues of the generator of the diffusion process (see [27]). In particular, ergodic diffusions with a linear drift* $-\zeta(x-\eta), \zeta > 0$ *as for example OU, CIR processes and Pearson's diffusions are geometrically* α*-mixing with* $c_2 = \zeta$.

According to the classical deconvolution scheme, the following assumptions are needed.

A 3. ***Regularity Assumptions:***

(i) $(\varepsilon_{i\Delta})_{i\geq 0}$ *a sequence of i.i.d. random variables satisfying* $\mathbb{E}[\varepsilon_{i\Delta}] = 0$ *and* $\mathbb{E}[\varepsilon_{i\Delta}^2] = \sigma_\varepsilon^2$ *and independent of* $(X_{i\Delta})_{i\geq 1}$. *For identifiability issues, we assume that* $\varepsilon_{1\Delta}$ *admits a known density with respect to the Lebesgue measure denoted by* f_ε[2].

(ii) the density of $\varepsilon_{i\Delta}$, f_ε *belongs to* $\mathbb{L}_2(\mathbb{R})$, *and for all* $x \in \mathbb{R}$, $f_\varepsilon^*(x) \neq 0$;

(iii) the function Π_θ *belongs to* $\mathbb{L}_1(A) \cap \mathbb{L}_2(A)$ *and is twice continuously differentiable w.r.t* $\theta \in \Theta$ *for any* x *and measurable w.r.t* x *for all* θ *in* Θ. *Additionally, each coordinate of* $\nabla_\theta \Pi_\theta$ *and each coordinate of* $\nabla_\theta^2 \Pi_\theta$ *need to belong to* $\mathbb{L}_1(A) \cap \mathbb{L}_2(A)$.

(iv) The functions $\mathcal{K}^1_{\Pi_\theta^2}$ *and* $\mathcal{K}^2_{\Pi_\theta}$ *defined in* (7) *are integrable.*

The assumption **A** 3(*ii*) on f_ε is quite common when considering deconvolution estimation and is essential for identifiability of the model. On the other hand, as we show in Section 5, the variance of measurement noises σ_ε^2 can be estimated from the observations $(Y_{i\Delta})_{1\leq i\leq n}$. Furthermore, **A** 3(*iii*) allows us to define the FT and to exchange the integral and the expectation and **A** 3(*iv*) can be understood as "Π_θ^* (resp. $(\Pi_\theta^2)^*$) has to be smooth enough compared to f_ε^*".

Let us explain the choice of such contrast: under assumptions **A** 2(*ii*) and **A** 3(*i*)–(*iii*) we have that our empirical contrast $\gamma_{\theta,n}(\Delta)$ converges in probability as n tends to infinity to the true unknown contrast $\gamma_\theta(\Delta)$. And, by using Lemma 3.1 we have

$$\begin{aligned}
\gamma_\theta(\Delta) &= \mathbb{E}[g_\theta(\mathbf{Y}_{1\Delta})] = \mathbb{E}[\Pi_\theta^2(\Delta, X_{1\Delta})] - 2\mathbb{E}[\Pi_\theta(\Delta, \mathbf{X}_{1\Delta})] \\
&= \iint \Pi_\theta^2(\Delta, x, y) f_{\theta_0}(x) dx dy - 2 \iint \Pi_\theta(\Delta, x, y)\Pi_{\theta_0}(\Delta, x, y) f_{\theta_0}(x) dx dy \\
&= \|\Pi_\theta\|_f^2 - 2\langle \Pi_\theta, \Pi_{\theta_0}\rangle_f = \|\Pi_\theta - \Pi_{\theta_0}\|_f^2 - \|\Pi_{\theta_0}\|_f^2 .
\end{aligned}$$

Then, this contrast is an empirical counterpart of the distance $\|\Pi_\theta - \Pi_{\theta_0}\|_f$. Since $\Pi_\theta(\Delta, ., .)$ is known in rare instances, we construct a closed-form sequence $\gamma_{\theta,n}^{(K)}(\Delta)$ of approximations to

[2] Note that the variance of the measurement error σ_ε^2 is an extra parameter that can be consistently estimated by using $(1/2n)\sum_{i=1}^{n-1}(Y_{(i+1)\Delta} - Y_{i\Delta})^2$ which is half of the quadratic variation of the observations.

the contrast function $\gamma_{\theta,n}(\Delta)$ where we use the sequence $\Pi_\theta^{(K)}(\Delta,.,.)$ of approximations to the transition density defined by (5) in Section 2. We define the Hermite approximate contrast estimator as

$$\hat{\theta}_n^{(K)} = \arg\min_{\theta\in\Theta} \gamma_{\theta,n}^{(K)}(\Delta). \tag{9}$$

We will see in Section 4 that this contrast is greatly simplified for constant volatility function where in this case the contrast reduces only to the computation of the operator $\mathcal{K}^2$.

3.4 Asymptotic properties of the minimum contrast estimator

We study the properties of the sequence of the minimum contrast estimators $\hat{\theta}_n^{(K)}$ derived from minimizing over θ in Θ the approximate contrast function. We will then show that $\hat{\theta}_n^{(K)}$ converges as K tends to infinity to the true but unknown contrast estimator $\hat{\theta}_n$. The strategy we employ to study the asymptotic properties of $\hat{\theta}_n^{(K)}$ is to first determine those of $\hat{\theta}_n$ (see Theorem 3.1) and then show that $\hat{\theta}_n^{(K)}$ share the same asymptotic properties as $\hat{\theta}_n$ provided that K tends to infinity with n (see Theorem 3.2).

To establish the asymptotic study of our estimator, we define the following estimating equation:

$$S_{\theta,n}(\Delta) = \frac{1}{n}\sum_{i=1}^{n} s_\theta(\mathbf{Y}_{i\Delta}),$$

where s_θ is a r-dimensional function given by,

$$\mathbf{y}_{i\Delta} \mapsto s_\theta(\mathbf{y}_{i\Delta}) = \mathcal{K}^1_{\nabla_\theta \Pi_\theta^2}(y_{i\Delta}) - 2\mathcal{K}^2_{\nabla_\theta \Pi_\theta}(\mathbf{y}_{i\Delta}).$$

To make an analogy with the likelihood estimator, one can see the quantity $S_{\theta,n}(\Delta)$ as the score vector. Under assumptions **A** 2(*ii*) and **A** 3(*i*)–(*iii*) we have that $S_{\theta,n}(\Delta)$ converges in probability as n tends to infinity to $S_\theta(\Delta)$. And, by using Lemma 3.1 and the same strategy of deconvolution we obtain

$$S_\theta(\Delta) = \mathbb{E}[s_\theta(\mathbf{Y}_1)] = 2\big\langle \nabla_\theta \Pi_\theta, \Pi_\theta - \Pi_{\theta_0}\big\rangle_f.$$

Let us now introduce the matrix $\Sigma_\theta(\Delta)$ given by

$$\Sigma_\theta(\Delta) = \mathcal{V}_\theta^{-1}(\Delta)\Omega_\theta(\Delta)\mathcal{V}_\theta^{-1'}(\Delta) \quad \text{with} \quad \Omega_\theta(\Delta) = \Omega_{\theta,0}(\Delta) + 2\sum_{j=2}^{+\infty}\Omega_{\theta,j-1}(\Delta),$$

where $\mathcal{V}_\theta(\Delta)$ denotes the Hessian matrix of $\gamma_\theta(\Delta)$, *i.e.*, $\mathcal{V}_\theta(\Delta) = \mathbb{E}[\nabla_\theta s_\theta(\mathbf{Y}_{1\Delta})]$ and $\Omega_{\theta,0}(\Delta) = \mathbb{V}ar\,(s_\theta(\mathbf{Y_{1\Delta}}))$ and $\Omega_{\theta,j-1}(\Delta) = \mathbb{C}ov\,(s_\theta(\mathbf{Y_{1\Delta}}), s_\theta(\mathbf{Y_{j\Delta}}))$.

To obtain our asymptotic results, that is, the consistency and the asymptotic distribution of our estimator, we further assume that the following assumptions hold true.

A 4. ***Asymptotic statistical assumptions:*** *The parameter value* $\theta_0 \in \mathrm{int}(\Theta)$ *and a neighborhood* $\mathcal{U}$ *of* $\theta_0 \in \Theta$ *exists such that*

(*i*) *The mapping* $\theta \mapsto \gamma_\theta(\Delta)$ *admits a unique minimum at* $\theta = \theta_0$ *and its Hessian matrix* $\mathcal{V}_\theta(\Delta)$ *computed at* θ_0 *is invertible.*

(*ii*) *The function* s_θ *has to be continuously differentiable on* $\mathcal{U}$ *for all* $\mathbf{y}_{1\Delta}$ *and integrable with respect to* $G_{\mathbf{Y}}$ *for all* $\theta \in \mathcal{U}$.

(*iii*) *There exists an integrable function h such that $\|s_\theta(\mathbf{y}_{1\Delta})\| \leq h(\mathbf{y}_{1\Delta})$ for all $\theta \in \mathcal{J}$ a compact subset of Θ.*

(*iv*) *The function $\mathbf{y}_{1\Delta} \mapsto \|\nabla_\theta s_\theta(\mathbf{y}_{1\Delta})\|$ is dominated for all $\theta \in \mathcal{U}$ by a function which is integrable with respect to $G_{\mathbf{Y}}$.*

(*v*) *(CLT assumption): $\sqrt{n}S_{\theta,n}(\Delta)$ converges in law to $\mathcal{N}(0, \Omega_\theta(\Delta))$ under $\mathbb{P}_{\theta_0}$ for any $\theta \in \Theta$ for which $S_\theta(\Delta) = 0$.*

Let us explain what Assumptions **A** 4(*ii*) imply for the integrability of the two operators $\mathcal{K}^1$ and $\mathcal{K}^2$. Each coordinate of $\mathcal{K}^1_{\nabla_\theta \Pi^2_\theta}$ and $\mathcal{K}^1_{\nabla^2_\theta \Pi^2_\theta}$ have to be integrable. In the same way, each coordinate of $\mathcal{K}^2_{\nabla_\theta \Pi_\theta}$ and $\mathcal{K}^2_{\nabla^2_\theta \Pi_\theta}$ have to be integrable as well. This assumption is the analogous of Assumption **A** 3(*iii*). Furthermore, Assumptions **A** 4(*iii*) and (*iv*) imply local dominance conditions of the first two derivatives of $\mathcal{K}^1$ and $\mathcal{K}^2$.

A sufficient condition ensuring the CLT assumption **A** 4(*v*) with $\Omega_\theta(\Delta)$, given in Corollary 3.1, is that the diffusion process (1) is stationary ergodic and geometrically α-mixing, which is given by assumptions **A** 2(*i*)-**A** 2(*iv*), that $\Omega_\theta(\Delta)$ converges and is strictly positive definite and for some $\epsilon > 0$, $\mathbb{E}[(\nabla_\theta s_\theta(\mathbf{Y}_{1\Delta}))^{2+\epsilon}] < \infty$ (see [15] for more details).

For models where integrability assumptions **A** 3(*iii*) and **A** 4(*ii*) are not satisfied, we propose to insert a weight function φ or a truncation Kernel as in [14, p. 285] to circumvent the issue of integrability. More precisely, we define the operators as follows

$$\mathcal{K}^1_{h\star N_{B_n}}(x) = \frac{1}{2\pi}\int e^{ixu}\frac{(h\star N_{B_n})^*(u,0)}{f^*_\varepsilon(-u)}du, \qquad (\mathcal{K}^2)^*_{h\star N_{B_n}} = \frac{(h\star N_{B_n})^*}{f^*_\varepsilon \otimes f^*_\varepsilon},$$

where $N^*_{B_n}$ is the FT of a density deconvolution kernel with compact support and satisfies $|1 - N^*_{B_n}(t)| \leq \mathbf{1}_{|t|>1}$ and B_n is a sequence which tends to infinity with n. The contrast is then defined as

$$\gamma_{\theta,n}(\Delta) = \frac{1}{n}\sum_{i=1}^{n-1} \mathcal{K}^1_{\Pi^2_\theta \star N}(Y_{i\Delta}) - 2\mathcal{K}^2_{\Pi_\theta \star N}(\mathbf{Y}_{i\Delta}).$$

This contrast still works under Assumptions **A** 2–4 by taking $N_{B_n}(t)^* = \mathbf{1}_{|t|\leq B_n}$ with $B_n \to +\infty$.

The identifiability assumption **A** 4(*i*) is based on the strictly convex character of the contrast function $\gamma_\theta(\Delta)$ and on the following condition on the drift and volatility functions:

$$b_\theta(x) = b_{\theta_0}(x) \text{ and } \sigma_\theta(x) = \sigma_{\theta_0}(x) \text{ for } f_\theta \text{ almost all } x \text{ imply } \theta = \theta_0.$$

The invertibility of $\mathcal{V}_\theta(\Delta)$ requires to study the invertibility of the matrix $(\langle \partial_{\theta_k}\Pi_\theta, \partial_{\theta_j}\Pi_\theta\rangle)_{j,k}$ for $j, k \in \{1, \ldots, r\}$. This suggests that the transition function Π_θ must not be uniformly flat in the direction of any one of the parameters θ_k. Otherwise, $\partial\Pi_\theta/\partial\theta_k \equiv 0$ for all (x, y) and the vector of parameters cannot be identified.

Theorem 3.1. *Under Assumptions **A** 2–**A** 4, and for $\Delta \in (0, \overline{\Delta})$ with $\overline{\Delta} > 0$, let $\hat{\theta}_n$ be the contrast estimator defined in (8). Then we have*

$$\hat{\theta}_n \to \theta_0 \text{ in probability as } n \to \infty.$$

Moreover,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to \mathcal{N}(0, \Sigma_{\theta_0}(\Delta)) \text{ in law as } n \to \infty.$$

Furthermore, if the function s_θ is locally dominated integrable[3] w.r.t. $G_{\mathbf{Y}}$ and $S_\theta(\Delta) \neq 0$ for all $\theta \neq \theta_0$, then the estimator $\hat{\theta}_n$ is the unique estimator on any bounded subset of Θ containing θ_0 with probability approaching one as $n \to \infty$.

[3] Locally dominated integrability w.r.t. $G_{\mathbf{Y}}$ means that, for all $\theta_1 \in \Theta$ there exists a neighborhood $\mathcal{U}_{\theta_1}$ of θ_1 and a non-negative $G_{\mathbf{Y}}$-integrable function h_{θ_1} such that $|s_\theta(\mathbf{y}_{1\Delta})| \leq h_{\theta_1}(\mathbf{y}_{1\Delta})$ for all $(\mathbf{y}_{1\Delta}, \theta_1)$.

The proof of Theorem 3.1 is provided in Subsection 6.1.

The following corollary gives an expression of the matrices $\Omega_{\theta_0}(\Delta)$ and $\mathcal{V}_{\theta_0}(\Delta)$ defined in $\Sigma_{\theta_0}(\Delta)$ of Theorem 3.1.

Corollary 3.1. *Under assumptions* ***A*** *2–****A*** *4, the matrix $\Omega_{\theta_0}(\Delta)$ is given by*

$$\Omega_{\theta_0}(\Delta) = \Omega_{\theta_0,0}(\Delta) + 2\sum_{j=2}^{+\infty}\Omega_{\theta_0,j-1}(\Delta),$$

where

$$\begin{aligned}\Omega_{\theta_0,0}(\Delta) =&\mathbb{E}[(\mathcal{K}^1_{\nabla_\theta\Pi^2_\theta}(Y_{1\Delta}))^2] + 4\mathbb{E}[(\mathcal{K}^2_{\nabla_\theta\Pi_\theta}(\mathbf{Y}_{1\Delta}))^2] - \left(\mathbb{E}\left[\int \nabla_\theta\Pi^2_\theta(\Delta, X_{1\Delta}, y)dy\right]^2\right.\\ &\left.+4\mathbb{E}\left[\nabla_\theta\Pi_\theta(\Delta,\mathbf{X}_{1\Delta})\right]^2 - 4\mathbb{E}\left[\int\nabla_\theta\Pi^2_\theta(\Delta,X_{1\Delta},y)dy\right]\mathbb{E}[\nabla_\theta\Pi_\theta(\Delta,\mathbf{X}_{1\Delta})]\right),\end{aligned}$$

and, the covariance terms are given by

$$\begin{aligned}\Omega_{\theta_0,j-1}(\Delta) =&\mathbb{C}ov\left(\int\nabla_\theta\Pi^2_\theta(\Delta,X_{1\Delta},y)dy, \int\nabla_\theta\Pi^2_\theta(\Delta,X_{j\Delta},y)dy\right)\\ &+4\mathbb{C}ov\left(\nabla_\theta\Pi_\theta(\Delta,\mathbf{X}_{1\Delta}),\nabla_\theta\Pi_\theta(\Delta,\mathbf{X}_{j\Delta})\right)\\ &-4\mathbb{C}ov\left(\int\nabla_\theta\Pi^2_\theta(\Delta,X_{1\Delta},y)dy,\nabla_\theta\Pi_\theta(\Delta,\mathbf{X}_{j\Delta})\right),\end{aligned}$$

where the differential $\nabla_\theta\Pi_\theta$ is taken at point $\theta=\theta_0$.
Furthermore, the Hessian matrix $\mathcal{V}_{\theta_0}(\Delta)$ is given by

$$\left([\mathcal{V}_{\theta_0}]_{j,k}(\Delta)\right)_{1\le j,k\le r} = 2\left(\left\langle\partial_{\theta_k}\Pi_\theta,\partial_{\theta_j}\Pi_\theta\right\rangle\right)_{j,k} \text{ at point } \theta=\theta_0.$$

The proof of Corollary 3.1 stems mainly from Lemma 3.1 and assumptions **A** 2–**A** 4. For a complete detail see [17].

If one wants to use Corollary 3.1 to build confidence sets, one needs to construct consistent estimator of the corresponding matrix $\Sigma_{\theta_0}(\Delta)$. Since, by assumption $\mathcal{V}_\theta(\Delta)$ is a continuous function of θ, $\mathcal{V}_{\hat\theta_n}(\Delta)$ is a consistent estimator of $\mathcal{V}_{\theta_0}(\Delta)$ under our assumptions. This result is given in Corollary 3.2. Another possible way to estimate $\Sigma_\theta(\Delta)$ is to use a bootstrap method, following, for instance [7] when the hidden variables form a Markov chain.

Corollary 3.2. *Under Assumptions* ***A*** *4(i)–(v), we have*

$$\mathcal{V}_{\theta,n}(\Delta) = \frac{1}{n}\sum_{i=1}^n \partial_\theta s_\theta(\mathbf{Y}_{i\Delta}) \to \mathcal{V}_{\theta_0}(\Delta) \text{ in probability as } n\to+\infty, \tag{10}$$

where $\mathcal{V}_{\theta,n}(\Delta)$ is computed at $\hat\theta_n$ a consistent estimator of θ_0. The probability that $\mathcal{V}_{\theta,n}(\Delta)$ is invertible approaches one as $n\to\infty$. If, moreover, the function $\mathbf{y}_{1\Delta}\mapsto\|s_\theta(\mathbf{y}_{1\Delta})\|$ is dominated for all $\theta\in\mathcal{U}$ where $\mathcal{U}$ is a neighborhood of $\theta_0\in\Theta$ by a function which is square integrable with respect to $G_{\mathbf{Y}}$, then

$$\Omega_{\theta,0,n}(\Delta) = \frac{1}{n}\sum_{i=1}^n s_\theta(\mathbf{Y}_{i\Delta})s_\theta(\mathbf{Y}_{i\Delta})' \to \Omega_{\theta_0,0}(\Delta) \text{ in probability as } n\to+\infty,$$

where $\Omega_{\theta,0,n}(\Delta)$ is computed at $\hat\theta_n$ a consistent estimator of θ_0.

The proof of Corollary 3.2 is given in Subsection 6.2 and results from the Uniform Law of Large Numbers (ULLN).

The following theorem shows that $\hat{\theta}_n^{(K)}$ inherits the asymptotic properties of the uncomputable $\hat{\theta}_n$.

Theorem 3.2. *Under Assumptions **A** 2–**A** 4, and for $\Delta \in (0, \overline{\Delta})$ with $\overline{\Delta} > 0$,*

- *Fix the sample size n. Then as $K \to \infty$, $\quad \hat{\theta}_n^{(K)} \to \hat{\theta}_n$ in probability.*
- *As $n \to \infty$, there exists K_n such that $\hat{\theta}_n^{(K_n)} - \hat{\theta}_n = o_p\left(\Sigma_{\theta_0}^{-1/2}(\Delta)\right)$ under $\mathbb{P}_{\theta_0}$ which then makes $\hat{\theta}_n^{(K_n)}$ and $\hat{\theta}_n$ share the same asymptotic distribution.*

The proof is given in Subsection 6.3.

4 Expression of the Hermite contrast for general diffusions

Let us consider the SDE given in (1). We give a general expression of the FT of the Hermite approximation density, which allows us to calculate our contrast approach. We consider the two following cases:

1. constant volatility function, meaning that the function σ_θ is independent of the state x;
2. state dependent volatility function.

Theorem 4.1. *The FT of the Hermite transition density approximation in (5) are given by:*

$$((\Pi_\theta^K)^2(\Delta, x, 0))^* = \frac{1}{\sqrt{\pi\Delta}} \sum_{k=0}^{K} 2^{k-1} k! (\eta_{\theta,k}^2(\Delta, \rho_\theta(x)))^* \quad \textit{for case 1 and 2,}$$

and,

$$(\Pi_\theta^K(\Delta, x, y))^* = \begin{cases} \varphi(y, 0, 1) \sum_{k=0}^{K} (-i)^k H_k(\sqrt{\Delta\sigma_\theta^2} y) [\eta_{\theta,k}(\Delta, \rho_\theta(x+y))]^* & \textit{for case 1} \\ \sum_{k=0}^{K} \mathcal{L}_{\theta,k}(\Delta, x, y) & \textit{for case 2} \end{cases}$$

with

$$\mathcal{L}_{\theta,k}(\Delta, x, y) = \frac{1}{\sqrt{2\pi\Delta}} \iint e^{-iux} \eta_{\theta,k}(\Delta, \rho_\theta(u)) e^{-iy\rho_\theta^{-1}(\sqrt{\Delta}t + \rho_\theta(u))} \varphi(t, 0, 1) H_k(t) dt du,$$

where $\varphi(t, 0, 1)$ corresponds to the standard zero mean Gaussian density and ρ_θ the Lamperti function defined in Subsection 2.2.

The proof is given in 6.4.

For SDE with a constant volatility function (case 1), the Lamperti function ρ_θ is linear w.r.t. x which involves that the Hermite contrast can be easily obtained and quickly computed.

5 Numerical simulation

In this section, we study empirical behavior of our strategy of estimation for parametric models on the hidden diffusion X_t commonly used in several applications. We consider the OU model where the transition is known in closed form in order to compare the performance of the Hermite contrast estimator with the true contrast and with the Euler contrast which consists in a simple discretization of the continuous-time SDE. This process is widely used in many applications in finance or biology and satisfies the following SDE

$$dX_t = -\phi_0 X_t dt + \sigma_{0,X} dW_t, \tag{11}$$

where $\theta_0 = (\phi_0, \sigma_{0,X})$ and $\mathcal{X} = (-\infty, +\infty)$. The diffusion term is independent of the state, so we are in the first case of Theorem 4.1. The transition density $\Pi_{\theta_0}(\Delta, x, y)$ is Gaussian with mean $m_{\theta_0,X,\Delta}(x) = xe^{-\phi_0\Delta}$ and variance $\gamma^2_{\theta_0,X,\Delta}(x) = (1-e^{-2\phi_0\Delta})\sigma^2_{0,X}/(2\phi_0)$ (independent of x) and is continuously infinitely differentiable w.r.t. x and θ. Furthermore, for $\phi_0 > 0$ the process is ergodic and the speed of convergence to equilibrium is exponential, meaning that the process is geometrically ergodic. Indeed, let us denote by $\|.\|_{\mathrm{TV}}$ the total variation norm and by $D(g||h)$ the Kullback-Leibler divergence, we have

$$2\|\Pi_{\theta_0} - f_{\theta_0}\|^2_{\mathrm{TV}} \leq D(\Pi_{\theta_0}||f_{\theta_0}),$$

where f_{θ_0} corresponds to the invariant density of the OU process (11), a Gaussian density with zero mean and variance $\sigma^2_{0,X}/2\phi_0$. Furthermore, for two Gaussian densities g, h with mean and variance respectively m_1 (resp. m_2) and v_1 (resp. v_2) we have

$$D(g||h) = \frac{(m_1 - m_2)^2}{2v_2} + \frac{1}{2}\left(\frac{v_1}{v_2} - 1 - \ln\frac{v_1}{v_2}\right).$$

Applying this to the OU process, we obtain that

$$\begin{aligned} 2\|\Pi_{\theta_0} - f_{\theta_0}\|^2_{TV} &\leq \frac{\phi_0}{\sigma^2_{0,X}} x^2 e^{-\phi_0\Delta} - \frac{1}{2}\left(e^{-2\phi_0\Delta} + \ln(1 - e^{-\phi_0\Delta})\right) \\ &\leq C_{\theta_0} x e^{-\phi_0\Delta} + O(e^{-\phi_0\Delta}). \end{aligned}$$

All moment conditions are satisfied for the Gaussian density of the OU process but integrability assumptions **A** 3*(iii)* and **A** 4*(ii)* are not satisfied. For the simulations, we have opted for the introduction of a weight function φ as detailed in Section 6.7.

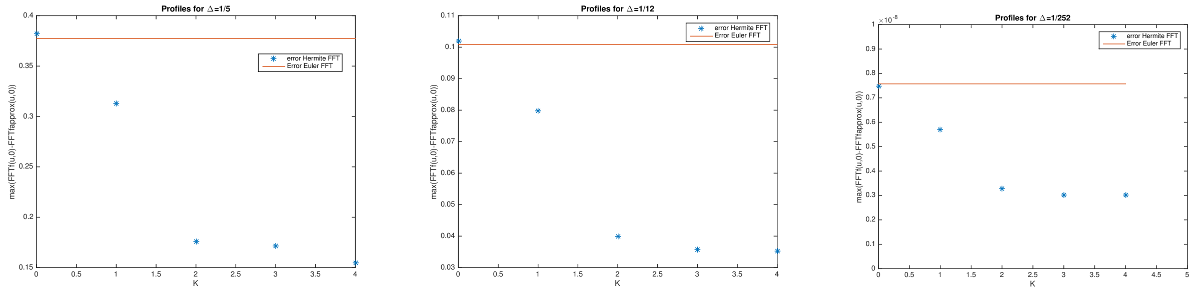
5.1 Comparison of the FTs of the Hermite and Euler approximations of the closed-form density

Before comparing the different contrasts obtained by different approximations of the transition density, it is interesting to first compare the FTs of the approximate transition densities since our contrast is essentially based on this deconvolution strategy.

The Euler approximation corresponds to a simple discretization of the continuous-time SDE where the differential equation (11) is replaced by

$$X_{t+\Delta} - X_t = (1 - \phi\Delta)X_t + \sigma_X\sqrt{\Delta}\eta_{t+\Delta},$$

with $\eta_{t+\Delta} \sim \mathcal{N}(0,1)$. Hence the Euler transition density is also Gaussian with mean $(1-\phi\Delta)x$ and variance $\Delta\sigma^2_X$. The Hermite transition density is given in the following Lemma.



(a) $\Delta = 1/5$ (b) $\Delta = 1/12$ (c) $\Delta = 1/252$

Figure 1: Comparison of the uniform error of the FT of the approximations densities with the FT of the closed form density for different sampling interval $\Delta \in \{1/5, 1/12, 1/252\}$ and different order $K \in \{0, \dots, 4\}$.

Lemma 5.1. *For the OU process defined in (11), the Hermite approximation of the transition density has the following form*

$$\Pi_\theta^K(\Delta, x, y) = \frac{1}{\sqrt{2\pi\Delta}\sigma_X} \exp\left(-\frac{(y-x)^2}{2\Delta\sigma_X^2}\right)\left(\sum_{k=0}^{K} \eta_{\theta,k}(\Delta, x) H_k\left(\frac{y-x}{\sigma_X\sqrt{\Delta}}\right)\right).$$

The proof and expressions of terms $\eta_{\theta,k}, k = 1$ up to 3 are postponed in Subsection 6.5, their expressions for general diffusions are given in [1].

To examine the accuracy of the expansion of the FT of Π_θ^K with the FT of the true transition probability density of the underlying process we use as in [1] and [39] the uniform error between the K-th order expansion and the true density as a measure of approximation error defined as $\max_y |e_\theta^K(\Delta, x, y)|$ over the range $\mathcal{D} = [-5, 5]$ where $e_\theta^K(\Delta, x, y) = (\Pi_\theta^K(\Delta, x, y))^* - (\Pi_\theta(\Delta, x, y))^*$.

Figure 1 displays the approximation errors of the expansion for the OU process. Two general patterns arise in the experiment results. First, for a fixed order K, the error of the FT of the density approximation decreases as the observational time interval Δ shrinks. When the observation frequency changes from monthly to daily (Δ from $1/12$ to $1/252$), the uniform error of the density approximation reduced very significantly. Second, when we fix the observation frequency Δ, the expansion with a larger K will lead to a smaller approximation error. These patterns for the FT of the density approximation corroborate the theoretical statements in [2, Theorem 1 p.232] and [3]. Moreover, whatever the frequency of observation Δ, the Hermite approximation gives better results than the Euler approximation from $K \geq 1$.

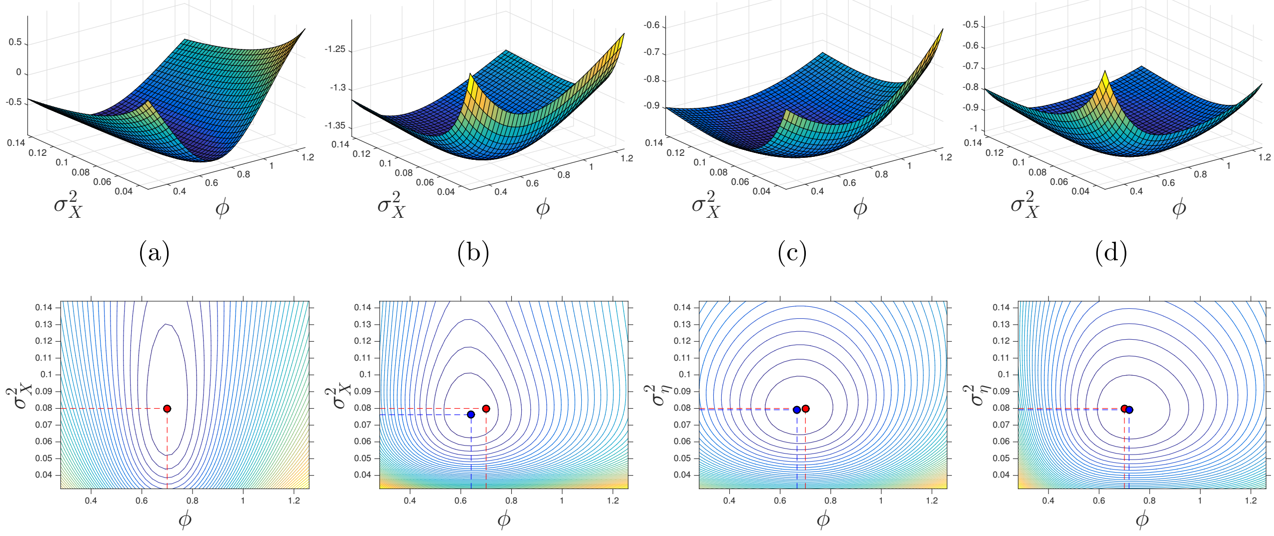
5.2 Comparison of the Hermite and Euler approximation contrast estimators with the true contrast

The following proposition gives the theoretical unknown contrast $\gamma_\theta(\Delta)$ defined in (6) for the OU process (11).

Proposition 5.1. *For the OU process defined in (11), the theoretical contrast function is*

$$\gamma_\theta(\Delta) = -1 + \frac{\gamma_{\theta,X,\Delta} + \gamma_{\theta_0,X,\Delta}}{2\sqrt{\pi\Delta\gamma_{\theta,X,\Delta}^2\gamma_{\theta_0,X,\Delta}^2}} - \sqrt{\frac{2}{\pi\Delta(\gamma_{\theta_0,X,\Delta}^2 + \gamma_{\theta,X,\Delta}^2)}} \exp\left(-\frac{\Delta}{2}\frac{(\phi_0 - \phi)^2}{(\gamma_{\theta_0,X,\Delta}^2 + \gamma_{\theta,X,\Delta}^2)}\right),$$

with $\gamma_{\theta,X,\Delta}^2$ the variance of the OU process transition density.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 2: Contrast functions with Gaussian noises f_ε. (a): γ as a function of the parameters ϕ and σ_X^2 for one realization of (12), with $n = 2000, \Delta = 1$. (b): γ_n. (c): γ_n^E. (d): γ_n^K with $K = 2$. (e)-(h): Corresponding contour lines. The red circle represents the global minimizer θ_0 of $\gamma(\theta)$ and the blue circle, the one of γ_n, γ_n^E and γ_n^K respectively.

Proof is provided in Section 6.6.

Proposition 5.2. *From Theorem 4.1, it is straightforward to see that the empirical Hermite contrast can be obtained from the following quantities, whatever the FT of the noises density* f_ε^*.

$$((\Pi_\theta^K)^2(\Delta, x, 0))^* = \frac{1}{\sqrt{\pi\Delta\sigma_X^2}} \sum_{k=0}^{K} 2^k k! \left(\eta_k^2\left(\frac{x}{\sigma_X}\right)\right)^*. \tag{12}$$

And,

$$(\Pi_\theta^K(\Delta, x, y))^* = \varphi(y, 0, 1) \sum_{k=0}^{K} (-i)^k H_k(\sqrt{\Delta\sigma_X^2}y)\left(\eta_k\left(\frac{x+y}{\sigma_X}\right)\right)^*.$$

Proof. The proof is essentially a consequence of Theorem 4.1 and the fact that the Lamperti transform function ρ for the OU process is $\rho_\theta(x) = x/\sigma_X$, which is linear in x (case 1). □

5.3 Numerical results

Study of the contrasts: We study the behavior of the true contrast estimator of θ_0 (which is computable in this example since the transition function is known in closed form) and compare it with the Euler and Hermite contrast estimators. r, we do not report the results here.

Figure 2 (resp. Figure 3) plots the true theoretical contrast γ defined in (12), the empirical true contrast γ_n, the Euler contrast γ_n^E and γ_n^K computed with the Hermite transition density given in Proposition 5.2, when we take a Gaussian density noise f_ε (resp. Laplace density). We illustrate the corresponding estimators $\hat\theta_n, \hat\theta_n^E$ and $\hat\theta_n^K$ with the true value θ_0. For the simulation we take $n = 2000$, $\theta_0 = (0.7, 0.08)$ and $\Delta = 1$. For measurement errors, we consider a signal-to-noise ratio denoted SNR and defined as $\sigma_X^2/\sigma_\varepsilon^2$ equal to 10, so a high noise level and we make only one MC replication for these illustrations. For integrability condition, all these

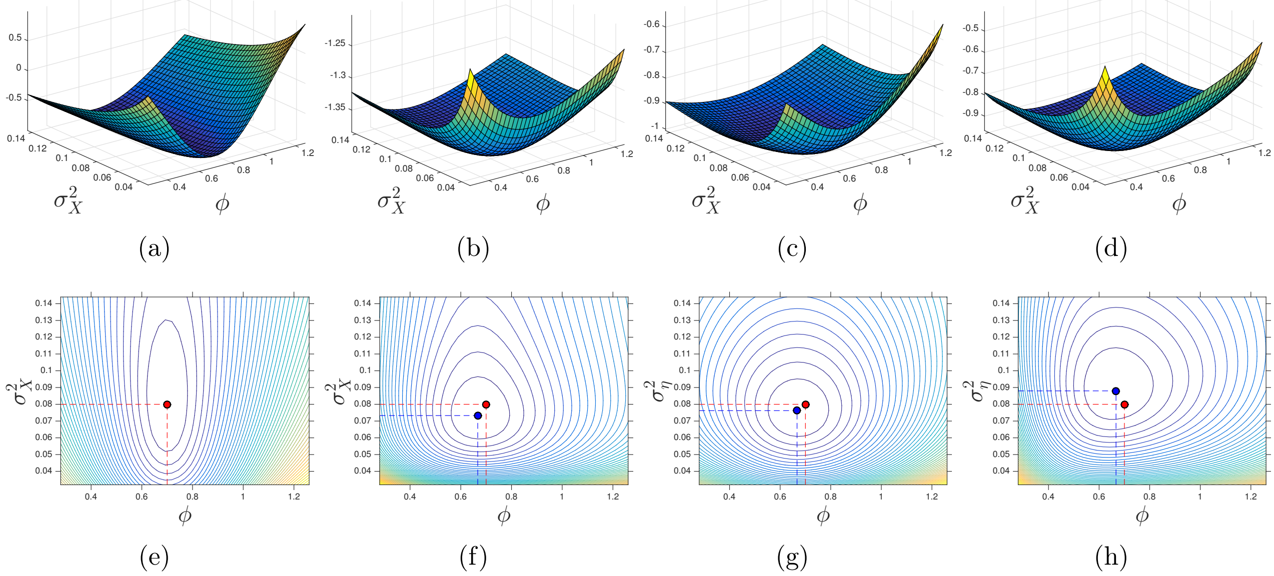


Figure 3: Contrast functions with Laplace noises f_ε. (a): γ as a function of the parameters ϕ and σ_X^2 for one realization of (12), with $n = 2000, \Delta = 1$. (b): γ_n. (c): γ_n^E. (d): γ_n^K with $K = 2$. (e)-(h): Corresponding contour lines. The red circle represents the global minimizer θ_0 of $\gamma(\theta)$ and the blue circle, the one of γ_n, γ_n^E and γ_n^K respectively.

contrasts are computed on a large truncated intervals. For the Gaussian case, $[-10, 10]$ yields a calculation that is both accurate and fast. From these simulations, we note a much larger bias can already be noted with the Euler scheme. This bias is not surprising since, on the one hand, we have seen that the FT approximation was less accurate than that of the Hermite scheme. Unlike Euler's scheme, Hermite's approximation takes into account higher order moments than the first two moments. On the other hand, it can be noted that for Laplace measurement noise, the results are still as good with the Hermite estimator and remain unchanged, contrary to the Euler estimator which presents better results than those when the measurement errors are Gaussian (see Figure 3).

Monte-Carlo experiments: Considering the following decomposition

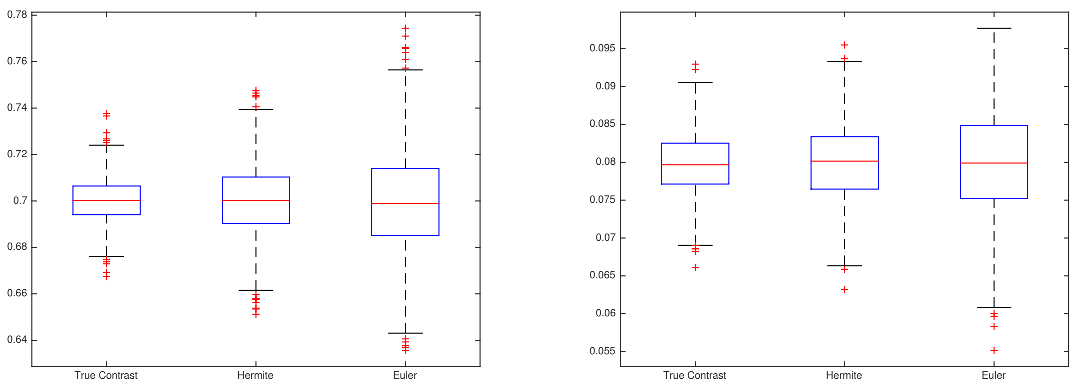
$$(\theta_n^{(K)} - \theta_0) = (\theta_n^{(K)} - \hat{\theta}_n) + (\hat{\theta}_n - \theta_0).$$

We identify two sources of errors contributing to the estimation error of the Hermite expansion. The first one is $(\hat{\theta}_n - \theta_0)$, the discrepancy between the true contrast estimator $\hat{\theta}_n$ and the parameter θ_0. This quantity measures the error caused intrinsically by the contrast approach, which is independent of the proposed Hermite approximation. The other one is $(\hat{\theta}_n^{(K)} - \hat{\theta}_n)$ which is affected by the accuracy of the density approximation. Since the true density for the OU process is explicitly known, we can compute the true contrast $\hat{\theta}_n$ and thus distinguish the impact of these two error sources. We tabulate the mean and the standard deviations of these errors in Table 1. This Table shows that when the time step Δ gets smaller, or as the order of approximation K increases, the approximate contrast $\hat{\theta}_n^{(K)}$ defined in (9) gets closer to the true uncomputable contrast, and thus to the true parameter. While keeping the time step Δ fixed, the approximation error $(\theta_n^{(K)} - \hat{\theta}_n)$ decreases and is dominated by the intrinsic estimation error $(\hat{\theta}_n - \theta_0)$ as K increases.

Sensibility of the results w.r.t. σ_ε^2: In Figure 4 we report the comparison of the various approximations when one considers the half of the quadratic variation of the observations as

Table 1: Monte-Carlo Evidence for the OU process with different order of expansion. The number of simulation trials is $MC = 100$ and the number of observations is $n = 1000$.

θ_0	$\hat{\theta}_n - \theta_0$	$\hat{\theta}_n^{(1)} - \theta_0$	$\hat{\theta}_n^{(2)} - \theta_0$	$\hat{\theta}_n^{(3)} - \theta_0$
$\mathbf{\Delta = 1/252}$				
$\phi_0 = 0.7$	0.219	0.032	0.022	0.0134
	(0.314)	(0.045)	(0.031)	(0.0141)
$\sigma_{0,X}^2 = 0.08$	0.00041	0.000032	0.000021	0.000015
	(0.015)	(0.00038)	(0.000014)	(0.000008)
$\mathbf{\Delta = 1/12}$				
$\phi_0 = 0.7$	0.191	0.028	0.013	0.009
	(0.124)	(0.049)	(0.021)	(0.009)
$\sigma_{0,X}^2 = 0.08$	0.0008	0.00007	0.000011	0.000006
	(0.015)	(0.00024)	(0.000054)	(0.000051)
$\mathbf{\Delta = 1/5}$				
$\phi_0 = 0.7$	0.374	0.056	0.042	0.031
	(0.222)	(0.06)	(0.038)	(0.025)
$\sigma_{0,X}^2 = 0.08$	0.0045	0.00032	0.00028	0.00015
	(0.047)	(0.0054)	(0.0037)	(0.0014)



(a) (b)

Figure 4: (a) Boxplot of $\hat{\phi}_n$, $\hat{\phi}_n^K$ with $K = 2$ and $\hat{\phi}_n^E$. (b) Boxplot of $\hat{\sigma}_{X,n}$, $\hat{\sigma}_{X,n}^K$ with $K = 2$ and $\hat{\sigma}_{X,n}^E$

an estimator of σ_ε^2 (see [51])

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{2n} \sum_{i=2}^{n} (Y_{i\Delta} - Y_{(i-1)\Delta})^2.$$

We report the Boxplot of each estimator of ϕ and σ_X^2 for a number of observations $n = 2000$, $\Delta = 1$ and $N = 100$ replications. For the Hermite contrast we take an order $K = 2$. We note that our approach still works when the variance of measurement noise is estimated. The results are better for the Hermite contrast where the boxplot size is smaller than the Euler contrast. Moreover, we note for the Euler approximation that the drift seems to be significantly more difficult to estimate. This may in part be explained by the large sampling interval and would support the known results highlighting the significant bias in drift estimation when the sampling interval does not tend towards zero for this approach (see [20]).

Conclusion

A new approach based on Hermite expansions is presented in this paper and makes an alternative practical option to the intractable MLE for the estimation of parameters in discretely sampled diffusion models observed with measurement errors. This approach is useful in many applications where the underlying process is unobservable or partially observed, for example application of this approach to derivative pricing under stochastic volatility models is interesting since the estimation of the parameters of the dynamics of the unobservable volatility is done without using a proxy variable of the latter as in [4] and more precisely without the need to have option prices upstream of the estimation of the parameters to calculate the proxy. The consistency and CLT is proved and a thorough comparison of this approach is made with the same construction of the contrast when it is computed from a Euler-Maruyama time discretization of the continuous time process. We show the superiority of the Hermite's expansion whatever the observation frequency and for a reasonable order of approximation K. Several extensions of this work are possible. On the one hand, the investigation of a rule of thumb allowing to choose the parameter K based on a penalized criterion. On the other hand, an extension to the multivariate framework by considering the approach proposed in [5] which allows to avoid the computation of the Lamperti transformation, which is not always possible in a multidimensional setting.

6 Proofs

6.1 Proof of Theorem 3.1

For the reader convenience we split the proof of Theorem 3.1 into two parts: in Subsection 6.1.1, we give the proof of the existence of our contrast estimator defined in (6). In Subsection 6.1.2, we prove the consistency and the asymptotic normality of our estimator.

6.1.1 Proof of the existence and measurability of the M-estimator

By assumption, the function $g_\theta(\mathbf{y}_{i\Delta}) = \mathcal{K}^1_{\Pi^2_\theta}(y_{i\Delta}) - 2\mathcal{K}^2_{\Pi_\theta}(\mathbf{y}_{i\Delta})$ is continuous w.r.t θ. Hence, the function $\gamma_{\theta,n}(\Delta) = \frac{1}{n}\sum_{i=1}^n g_\theta(\mathbf{Y}_{i\Delta})$ is continuous w.r.t θ belonging to the compact subset Θ. So, there exists $\tilde{\theta}$ belongs to Θ such that $\inf_{\theta\in\Theta}\gamma_{\theta,n}(\Delta) = \gamma_{\tilde{\theta},n}(\Delta)$. □

6.1.2 Proof of the consistency and the asymptotic normality

The proof of the consistency relies on [33, Theorem 1.58 p 87]. It remains to verify that under our assumptions, the function $\theta \mapsto \mathcal{V}_\theta$ defined in Assumption **A** 4(*i*) and explicitly given in Corollary 3.1 is continuous and verifies the following ULLN

$$\sup_{\theta\in\mathcal{J}} \|\frac{1}{n}\sum_{i=1}^{n-1} \nabla_\theta s_\theta(\mathbf{Y}_{i\Delta}) - \mathcal{V}_\theta)\| \to_{\mathbb{P}_{\theta_0}} 0,$$

where $\mathcal{J}$ is a compact subset of Θ.

The continuity of $\mathcal{V}_\theta$ is given by the dominated convergence theorem and in particular $\mathcal{V}_\theta$ is uniformly continuous on the compact set $\mathcal{J}$. To prove the ULLN, define for $\xi > 0$:

$$\kappa(\xi;\mathbf{y}_{1\Delta}) = \sup_{\theta_1,\theta_2\in\mathcal{J}:\|\theta_2-\theta_1\|\le\xi} \|\nabla_\theta s_{\theta_2}(\mathbf{y}_{1\Delta}) - \nabla_\theta s_{\theta_1}(\mathbf{y}_{1\Delta})\|.$$

For convenience, we will denote $\kappa(\xi)$ the function $(\mathbf{y}_{1\Delta}) \mapsto \kappa(\xi;\mathbf{y}_{1\Delta})$.
By Assumption **A** 4(*iv*), $\kappa(\xi) \le 2h$ and thanks to the dominated convergence theorem $\mathbb{E}[\kappa(\xi)] \to 0$ as $\xi \to \infty$. Since $\mathcal{V}_\theta(\Delta)$ is uniformly continuous w.r.t. θ, for any given $\epsilon > 0$, we can choose $\xi > 0$ such that $\mathbb{E}[\kappa(\xi)] \le \epsilon$ and such that $\|\theta_2 - \theta_1\| \le \xi$ implies that $\|\mathcal{V}_{\theta_1}(\Delta) - \mathcal{V}_{\theta_2}(\Delta)\| \le \epsilon$ for $\theta_1, \theta_2 \in \mathcal{J}$. Define the Balls $B_{\theta_1,\xi} = \{\theta_2 \text{ such that } \|\theta_2 - \theta_1\| \le \xi\}$. Since $\mathcal{J}$ is compact, there exists a finite covering

$$\mathcal{J} \subseteq \bigcup_{j=1}^m B_{\theta_j,\xi},$$

where $\theta_1,\dots,\theta_m \in \mathcal{J}$, hence we can find $\theta_r, r \in \{1,\dots,m\}$ such that $\theta \in B_{\theta_r,\xi}$. Hence, let

$$H_n(\theta,\Delta) = \frac{1}{n}\sum_{i=1}^{n-1} \nabla_\theta s_\theta(\mathbf{Y}_{i\Delta}).$$

We have

$$\begin{aligned}\|H_{\theta,n}(\Delta)-\mathcal{V}_\theta(\Delta)\| &\le \|H_{\theta,n}(\Delta)-H_{\theta_r,n}(\Delta)\|+\|H_{\theta_r,n}(\Delta)-\mathcal{V}_{\theta_r}(\Delta)\|+\|\mathcal{V}_{\theta_r}(\Delta)-\mathcal{V}_\theta(\Delta)\|\\
&\le \frac{1}{n}\sum_{i=1}^{n-1}\kappa(\xi,\mathbf{Y}_{i\Delta},)+\|H_{\theta_r,n}(\Delta)-\mathcal{V}_{\theta_r}(\Delta)\|+\epsilon\\
&\le \left|\frac{1}{n}\sum_{i=1}^{n-1}\kappa(\xi,\mathbf{Y}_{i\Delta})-\mathbb{E}[\kappa(\xi)]\right|+|\mathbb{E}[\kappa(\xi)]|+\|H_{\theta_r,n}(\Delta)-\mathcal{V}_{\theta_r}(\Delta)\|+\epsilon\\
&\le T_n+2\epsilon,\end{aligned}$$

where

$$T_n=\left|\frac{1}{n}\sum_{i=1}^{n-1}\kappa(\xi,\mathbf{Y}_{i\Delta})-\mathbb{E}[\kappa(\xi)]\right|+\max_{1\le r\le m}\|H_{\theta_r,n}(\Delta)-\mathcal{V}_{\theta_r}(\Delta)\|.$$

By the Ergodic Theorem, we have $\mathbb{P}_{\theta_0}(T_n>\epsilon)\to 0$ as $n\to\infty$. Hence,

$$\mathbb{P}_{\theta_0}\big(\sup_{\theta\in\mathcal{J}}\|H_{\theta,n}(\Delta)-\mathcal{V}_\theta(\Delta)\|>3\epsilon\big)\to 0,$$

for all $\epsilon>0$.

Furthermore, by the local dominated integrability of s_θ we have that s_θ satisfied the ULLN:

$$\sup_{\theta\in\mathcal{J}}|S_{\theta,n}(\Delta)-S_\theta(\Delta)|\to 0\quad\text{under }\mathbb{P}_{\theta_0}.$$

Hence, by Assumption **A** 4 (the local dominated integrability, the dominated convergence and identifiability assumption) we obtain

$$\inf_{\mathcal{J}\backslash\overline{B}_{\theta_0,\epsilon}}|S_\theta(\Delta)|>0,$$

for all $\epsilon>0$, where $\overline{B}_{\theta,\epsilon}$ is the closed ball with a radius ϵ centered at θ. Hence, for any sequence $(\hat{\theta}_n)$ of estimators

$$\mathbb{P}_{\theta_0}\big(\hat{\theta}_n\in\mathcal{J}\backslash\overline{B}_{\theta_0,\epsilon}\big)\to 0\quad\text{as }n\to\infty,\quad\text{for every }\epsilon>0. \tag{13}$$

Let $\hat{\theta}'_n$ be an estimator, and define another estimator $\hat{\theta}''_n$ as $\hat{\theta}''_n=\hat{\theta}'_n\mathbf{1}_{\hat{\theta}'_n\in\mathcal{J}}+\hat{\theta}_n\mathbf{1}_{\hat{\theta}'_n\notin\mathcal{J}}$ where $\hat{\theta}_n$ is the consistent estimator. Hence, by (13) $\hat{\theta}''_n$ is consistent and thanks to [33, Theorem 1.58], $\mathbb{P}_{\theta_0}(\hat{\theta}_n\neq\hat{\theta}''_n)\to 0$ as $n\to\infty$. So, $\hat{\theta}_n$ is eventually the unique estimator on $\mathcal{J}$.

The asymptotic normality follows from [33, Theorem 1.60 p.88] and Assumption **A** 4(v). □

6.2 Proof of Corollary 3.2

Let D be a compact subset of $\mathcal{U}$ such that $\theta_0\in int(D)$. By the ULLN,

$$\frac{1}{n}\sum_{i=1}^{n}\nabla_\theta s_\theta(\mathbf{Y}_{i\Delta})\to\mathcal{V}_{\theta_0}(\Delta)\text{ in probability as }n\ \to+\infty,$$

uniformly for $\theta\in D$. This implies (10) because $\hat{\theta}_n$ converges in probability to θ_0. The invertibility of $\mathcal{V}_{\theta,n}(\Delta)$ follows from the invertibility assumption **A** 4(i) of $\mathcal{V}_\theta(\Delta)$. Also, the uniform convergence in probability for $\theta\in D$ of

$$\frac{1}{n}\sum_{i=1}^{n}s_\theta(\mathbf{Y}_{i\Delta})s_\theta(\mathbf{Y}_{i\Delta})'\to_{\mathbb{P}_{\theta_0}}\Omega_{\theta_0,0}(\Delta),$$

follows from the ULLN. □

6.3 Proof of Theorem 3.2

The proof is based on the following Lemma.

Lemma 6.1. *Under Assumption* ***A*** *1 there exists* $\overline{\Delta} > 0$ *such that for every* $\Delta \in (0, \overline{\Delta}), \theta \in \Theta$ *and* (x, x_0) *in* $\mathcal{X}^2$*:*

$$\Pi^K_{\theta,X}(\Delta, x_0, x) \to \Pi_{\theta,X}(\Delta, x_0, x) \qquad \text{as } K \to +\infty.$$

Moreover, the convergence is uniform in $\theta \in \Theta$*, in* x_0 *over compact subsets of* $\mathcal{X}$ *and uniform in* x *over the entire domain* $\mathcal{X}$*.*

For the proof of Lemma 6.1 see [2, Theorem 1 p.232]. □

Assumptions **A** 1–**A** 3 and Lemma 6.1 lead to

$$\gamma^{(K)}_{\theta,n}(\Delta) \to \gamma_{\theta,n}(\Delta) \text{ in probability as } K \to +\infty \text{ uniformly in } \theta \in \Theta.$$

So, the convergence in probability of the respective arg min $\hat{\theta}^K_n$ towards $\hat{\theta}_n$ as $K \to +\infty$ is then an application of standard methods since the respective contrasts $\gamma^{(K)}_{\theta,n}(\Delta), \gamma_{\theta,n}(\Delta)$ and their derivatives are both continuous in θ for all n and K. This proves the first part of Theorem 3.2. For the second part, that is, for the convergence as $n \to +\infty$, one can see that from the first part, one can find a value K_n for each n to render $|\hat{\theta}^{K_n}_n - \hat{\theta}_n|$ arbitrarily small in probability. Hence, one can select $K_n \to +\infty$ such that $\hat{\theta}^{K_n}_n - \hat{\theta}_n = o_p(\Sigma^{-1/2}_{\theta_0})$ as $n \to +\infty$.

6.4 Proof of Theorem 4.1

For the proof we will use the following Lemma about some orthogonality properties of Hermite polynomials w.r.t. the weight factor e^{-s^2} on the interval $\mathbb{R}$ (see [9]).

Lemma 6.2.

$$\int_{\mathbb{R}} H_l(s)H_k(s)e^{-s^2}ds = \begin{cases} 0 & \textit{if } l \neq k \\ 2^k k! \int_{\mathbb{R}} e^{-s^2}ds = 2^k k! \sqrt{\pi} & \textit{for } l = k \end{cases}$$

Proof. Let $l < k$, using the definition of $H_k(s)$ given in Section 2.2 and integration by part we have

$$\begin{aligned} (-1)^k \int_{\mathbb{R}} H_l(s)H_k(s)e^{-s^2}ds &= \int_{\mathbb{R}} H_l(s)\frac{\partial^k}{\partial s^k}e^{-s^2}ds \\ &= \left[H_l(s)\frac{\partial^{k-1}}{\partial s^{k-1}}e^{-s^2}\right]_{\mathbb{R}} - \int \partial_s H_l(s)\frac{\partial^{k-1}}{\partial s^{k-1}}e^{-s^2}ds \\ &= -\int \partial_s H_l(s)\frac{\partial^{k-1}}{\partial s^{k-1}}e^{-s^2}ds. \end{aligned}$$

Now, using the fact that $\partial_s H_l(s) = 2lH_{l-1}(s)$ and repeating the integration by part l times we obtain

$$\begin{aligned} \int_{\mathbb{R}} H_l(s)H_k(s)e^{-s^2}ds &= (-1)^{1+k}2l\int_{\mathbb{R}} H_{l-1}(s)\frac{\partial^{k-1}}{\partial s^{k-1}}e^{-s^2}ds \\ &= (-1)^{l+k}2^l l! \int_{\mathbb{R}} H_0(s)\frac{\partial^{k-l}}{\partial s^{k-l}}e^{-s^2}ds \\ &= (-1)^{l+k}2^l l! \left[\frac{\partial^{k-l-1}}{\partial s^{k-l-1}}e^{-s^2}\right]_{\mathbb{R}} = 0, \end{aligned}$$

whereas in the case $l = k$ the result is given by

$$\int_{\mathbb{R}} H_k^2(s)e^{-s^2}ds = 2^k k! \int_{\mathbb{R}} e^{-s^2}ds = 2^k k! \sqrt{\pi}.$$

□

The proof of Theorem 4.1 is given in turn for case 1 and case 2.
Case 1 (independent state diffusion) We have that $\rho_\theta(x) = x/\sigma_\theta$ for all x and so

$$(\Pi_\theta^K(\Delta, x, y))^2 = \frac{1}{2\sqrt{\pi\Delta\sigma_\theta^2}}\varphi(\rho_\theta(y), \rho_\theta(x), \frac{\Delta}{2})\left(\sum_{k=0}^K \eta_{\theta,k}(\Delta, \rho_\theta(x))H_k\left(\frac{\rho_\theta(y)-\rho_\theta(x)}{\sqrt{\Delta}}\right)\right)^2,$$

where $\varphi(z, y, a)$ denotes the Gaussian density with mean y, variance a and is evaluated at z. Hence,

$$((\Pi_\theta^K(\Delta, x, 0))^2)^* = \int e^{-ixu}\tilde{\Pi}_\theta^K(\Delta, u)du,$$

with

$$\begin{aligned}
\tilde{\Pi}_\theta^K(\Delta, u) = &\int (\Pi_\theta^K)^2(\Delta, u, v)dv \\
=&\frac{1}{2\sqrt{\pi\Delta}}\int \varphi(\rho_\theta(v), \rho_\theta(u), \frac{\Delta}{2})\frac{1}{\sigma_\theta}\left(\sum_{k=0}^K \eta_{\theta,k}(\Delta, \rho_\theta(u))H_k\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)\right)^2 dv \\
=&\frac{1}{2\pi\Delta}\left\{\int \frac{1}{\sigma_\theta}e^{-\frac{(\rho_\theta(v)-\rho_\theta(u))^2}{\Delta}}\left(\sum_{k=0}^K \eta_{\theta,k}(\Delta, \rho_\theta(u))H_k\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)\right)^2 dv\right\}. \\
=&\frac{1}{2\pi\Delta\sigma_\theta}\left\{\sum_{k=0}^K \eta_{\theta,k}^2(\Delta, \rho_\theta(u))\int e^{-\frac{(\rho_\theta(v)-\rho_\theta(u))^2}{\Delta}}H_k^2\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)dv\right. \\
&+2\sum_{1\le k<l<K}\eta_{\theta,k}(\Delta, \rho_\theta(u))\eta_{\theta,l}(\Delta, \rho_\theta(u))\times \\
&\left.\int e^{-\frac{(\rho_\theta(v)-\rho_\theta(u))^2}{\Delta}}H_k\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)H_l\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)dv\right\}.
\end{aligned}$$

Let the following change of variable $s = \frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}$ and $ds = \frac{\partial_v\rho_\theta(v)}{\sqrt{\Delta}}dv = \frac{dv}{\sigma_\theta\sqrt{\Delta}}$ (for the case 1 ρ_θ is linear in v), we obtain

$$\begin{aligned}
\tilde{\Pi}_\theta^K(\Delta, u) = &\frac{1}{2\pi\sqrt{\Delta}}\left\{\sum_{k=0}^K \eta_{\theta,k}^2(\Delta, \rho_\theta(u))\int e^{-s^2}H_k^2(s)ds\right. \\
&\left.+2\sum_{1\le k<l<K}\eta_{\theta,k}(\Delta, \rho_\theta(u))\eta_{\theta,l}(\Delta, \rho_\theta(u))\int e^{-s^2}H_k(s)H_l(s)ds\right\}.
\end{aligned}$$

Using Lemma 6.2, the last term vanishes and we have

$$\begin{aligned}
\tilde{\Pi}_\theta^K(\Delta, u) &= \frac{1}{\sqrt{\pi\Delta}}\sum_{k=0}^K 2^{k-1}k!\eta_{\theta,k}^2(\Delta, \rho_\theta(u)) \\
(\tilde{\Pi}_\theta^K(\Delta, u))^* &= \frac{1}{\sqrt{\pi\Delta}}\sum_{k=0}^K 2^{k-1}k!(\eta_{\theta,k}^2(\Delta, \rho_\theta(u)))^*.
\end{aligned}$$

Let's now consider the terms involved in the operator $\mathcal{K}^2$. We have

$$(\Pi_\theta^K(\Delta, x, y))^* = \iint e^{-i(xu+yv)} \frac{\varphi(\rho_\theta(v), \rho_\theta(u), \Delta)}{\sigma_\theta} \sum_{k=0}^{K} \eta_{\theta,k}(\Delta, \rho_\theta(u)) H_k\left(\frac{\rho_\theta(u) - \rho_\theta(u)}{\sqrt{\Delta}}\right) dudv.$$

with σ_θ independent of v and ρ_θ linear in v. By Fubini and change of variable $z = v - u$, this integral becomes

$$\begin{aligned}
&\int e^{-ixu} \sum_{k=0}^{K} \eta_{\theta,k}(\Delta, \rho_\theta(u)) \Big(\int e^{-iyv} \frac{\varphi(\rho_\theta(v), \rho_\theta(u), \Delta)}{\sigma_\theta} H_k\left(\frac{\rho_\theta(v) - \rho_\theta(u)}{\sqrt{\Delta}}\right) dv \Big) du \\
&= \frac{1}{\sqrt{2\pi\Delta}\sigma_\theta} \int e^{-iu(x+y)} \sum_{k=0}^{K} \eta_{\theta,k}(\Delta, \rho_\theta(u)) \int e^{-iy(v-u)} e^{-\frac{(\rho_\theta(v)-\rho_\theta(u))^2}{2\Delta}} H_k\left(\frac{\rho_\theta(v) - \rho_\theta(u)}{\sqrt{\Delta}}\right) dvdu \\
&= \frac{1}{\sqrt{2\pi\Delta}\sigma_\theta} \int e^{-iu(x+y)} \sum_{k=0}^{K} \eta_{\theta,k}(\Delta, \rho_\theta(u)) \int e^{-iyz} e^{-\frac{(\rho_\theta(z+u)-\rho_\theta(u))^2}{2\Delta}} H_k\left(\frac{\rho_\theta(z+u) - \rho_\theta(u)}{\sqrt{\Delta}}\right) dzdu.
\end{aligned}$$

By the change of variable $t = \frac{z}{\sqrt{\Delta}\sigma_\theta}$ and since ρ_θ is linear in v we obtain

$$\begin{aligned}
(\Pi_\theta^K(\Delta, x, y))^* &= \frac{1}{\sqrt{2\pi\Delta\sigma_\theta^2}} \int e^{-iu(x+y)} \sum_{k=0}^{K} \eta_{\theta,k}(\Delta, \rho_\theta(u)) \Big(\int e^{-iyz} e^{-\frac{z^2}{2\Delta\sigma_\theta^2}} H_k\big(\frac{z}{\sqrt{\Delta\sigma_\theta^2}}\big) dz \Big) du \\
&= \frac{1}{\sqrt{2\pi}} \int e^{-iu(x+y)} \sum_{k=0}^{K} \eta_{\theta,k}(\Delta, \rho_\theta(u)) \Big(\int e^{-iy\sqrt{\Delta\sigma_\theta^2}t} e^{-\frac{t^2}{2}} H_k(t) dt \Big) du
\end{aligned}$$

By property of the FT of the Hermite functions $h_k(t) = (2^k k! \sqrt{\pi})^{-1/2} H_k(t) e^{-\frac{t^2}{2}}$ and in particular $h_k^*(t) = \sqrt{2\pi}(-i)^k h_k(t)$, we obtain

$$\int e^{-iy\sqrt{\Delta\sigma_\theta^2}t} e^{-\frac{t^2}{2}} H_k(t) dt = \sqrt{2\pi}(-i)^k H_k\Big(y\sqrt{\Delta\sigma_\theta^2}\Big) e^{-\frac{y^2}{2}},$$

and so

$$\begin{aligned}
(\Pi_\theta^K(\Delta, x, y))^* &= \sum_{k=0}^{K} (-i)^k H_k\Big(y\sqrt{\Delta\sigma_\theta^2}\Big) e^{-\frac{y^2}{2}} \int e^{-iu(x+y)} \eta_k(\Delta, \rho_\theta(u)) du \\
&= \varphi(y, 0, 1) \sum_{k=0}^{K} (-i)^k H_k\Big(y\sqrt{\Delta\sigma_\theta^2}\Big) (\eta_k(\Delta, \rho_\theta(x+y)))^*,
\end{aligned}$$

with $\varphi(y, 0, 1)$ the standard Gaussian density.

Case 2 (dependent state diffusion) We have

$$
\begin{aligned}
\tilde{\Pi}_\theta^K(\Delta,u) &= \int (\Pi_\theta^K)^2(\Delta,u,v)dv \\
&= \frac{1}{2\sqrt{\pi\Delta}} \int \varphi(\rho_\theta(v),\rho_\theta(u),\Delta/2)\frac{1}{\sigma_\theta(v)}\left(\sum_{k=0}^K \eta_{\theta,k}(\Delta,\rho_\theta(u))H_k\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)\right)^2 dv \\
&= \frac{1}{2\pi\Delta}\left\{\int \frac{1}{\sigma_\theta(v)}e^{-\frac{(\rho_\theta(v)-\rho_\theta(u))^2}{\Delta}}\left(\sum_{k=0}^K \eta_{\theta,k}(\Delta,\rho_\theta(u))H_k\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)\right)^2 dv\right\}. \\
&= \frac{1}{2\pi\Delta}\left\{\sum_{k=0}^K \eta_{\theta,k}^2(\Delta,\rho_\theta(u))\int \frac{1}{\sigma_\theta(v)}e^{-\frac{(\rho_\theta(v)-\rho_\theta(u))^2}{\Delta}} H_k^2\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)dv\right. \\
&\quad + 2\sum_{1\le k<l<K} \eta_{\theta,k}(\Delta,\rho_\theta(u))\eta_{\theta,l}(\Delta,\rho_\theta(u)) \\
&\qquad \left.\times \int \frac{1}{\sigma_\theta(v)}e^{-\frac{(\rho_\theta(v)-\rho_\theta(u))^2}{\Delta}} H_k\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)H_l\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)dv\right\}.
\end{aligned}
$$

Let the following change of variable $s = \frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}$ and $ds = \frac{\partial_v\rho_\theta(v)}{\sqrt{\Delta}}dv = \frac{dv}{\sigma_\theta(v)\sqrt{\Delta}}$ for the case 2 and using Lemma 6.2 we obtain as in the case 1

$$
\begin{aligned}
\tilde{\Pi}_\theta^K(\Delta,u) =& \frac{1}{2\pi\sqrt{\Delta}}\left\{\sum_{k=0}^K \eta_{\theta,k}^2(\Delta,\rho_\theta(u))\int e^{-s^2}H_k^2(s)ds\right. \\
&\left. + 2\sum_{1\le k<l<K} \eta_{\theta,k}(\rho_\theta(u))\eta_l(\Delta,\rho_\theta(u))\int e^{-s^2}H_k(s)H_l(s)ds\right\}. \\
=& \frac{1}{\sqrt{\pi\Delta}}\sum_{k=0}^K 2^{k-1}k!\eta_{\theta,k}^2(\Delta,\rho_\theta(u)),
\end{aligned}
$$

and the FT $(\tilde{\Pi}_\theta^K(\Delta,u))^*$ is the same as the one obtained in case 1. For the computation of $(\Pi_\theta^K(\Delta,x,y))^*$ we have

$$
(\Pi_\theta^K(\Delta,x,y))^* = \iint e^{-i(xu+yv)}\frac{\varphi(\rho_\theta(v),\rho_\theta(u),\Delta)}{\sigma_\theta(v)}\sum_{k=0}^K \eta_{\theta,k}(\Delta,\rho_\theta(u))H_k\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)dudv,
$$

with σ_θ dependent of v. By Fubini and change of variable $t = \frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}$ such that $dt = \frac{\partial_v\rho_\theta(v)}{\sqrt{\Delta}}dv = \frac{1}{\sqrt{\Delta}\sigma_\theta(v)}dv$, this integral becomes

$$
\begin{aligned}
(\Pi_\theta^K(x,y))^* &= \int e^{-ixu}\sum_{k=0}^K \eta_{\theta,k}(\Delta,\rho_\theta(u))\left(\int e^{-iyv}\frac{\varphi(\rho_\theta(v),\rho_\theta(u),\Delta)}{\sigma_\theta(v)}H_k\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)dv\right)du \\
&= \frac{1}{\sqrt{2\pi\Delta}}\int e^{-iux}\sum_{k=0}^K \eta_{\theta,k}(\Delta,\rho_\theta(u))\int \frac{1}{\sigma_\theta(v)}e^{-iyv}e^{-\frac{(\rho_\theta(v)-\rho_\theta(u))^2}{2\Delta}}H_k\left(\frac{\rho_\theta(v)-\rho_\theta(u)}{\sqrt{\Delta}}\right)dvdu \\
&= \frac{1}{\sqrt{2\pi\Delta}}\int e^{-iux}\sum_{k=0}^K \eta_{\theta,k}(\Delta,\rho_\theta(u))\int e^{-iy\rho_\theta^{-1}(\sqrt{\Delta}t+\rho_\theta(u))}e^{-\frac{t^2}{2}}H_k(t)dtdu \\
&= \sum_{k=0}^K \mathcal{L}_{\theta,k}(\Delta,x,y).
\end{aligned}
$$

6.5 Proof of Lemma 5.1

Proof. To compute the Hermite polynomial approximation of the transition density we need to compute the coefficient $\eta_k, k = 1, \ldots, K$ defined in Subsection 2.2 from the transformed process $\mu_{\theta,U}(u)$ and the Hermite polynomial functions H_k. For the OU process, the transformed process $\mu_{\theta,U}$ is equal to $-\phi u$.

From this transformed process we can easily compute the coefficients $\eta_k, k = 0, \ldots, K$ which are obtained by a Taylor expansion at the point $\rho_\theta(x)$ defined in Section (2). For the OU process, this function is equal to x/σ_X so we have all ingredients to compute the Hermite density approximation. We give below the coefficients η_k for $K = 1$ up to $K = 3$ and the associated Hermite polynomials.

$$
\begin{aligned}
&\eta_{\theta,0}(\Delta, x_0) = 1; \quad H_0 = 1,\\
&\eta_{\theta,1}(\Delta, x_0) = \Delta^{1/2}\phi\frac{x_0}{\sigma_X}\left((1-\phi\frac{\Delta}{2}+\phi^2\frac{\Delta^2}{6})\right) - \Delta^{5/2}\frac{x_0^2}{6\sigma_X^2}\phi^3; \quad H_1 = \frac{(x-x_0)}{\sigma_X\sqrt{\Delta}},\\
&\eta_{\theta,2}(\Delta, x_0) = \frac{\Delta}{2}\phi^2\frac{x_0^2}{\sigma_X^2}\left(1+\Delta+\frac{7}{12}\phi^4\Delta^2\right) + \Delta\phi^2\left(\frac{1}{2}-\Delta\frac{\phi}{3}-\phi\frac{\Delta^2}{6}\right); \quad H_2 = H_1^2 - 1,\\
&\eta_{\theta,3}(\Delta, x_0) = \phi^2\frac{\Delta^{3/2}}{6}\frac{x_0^3}{\sigma_X^3}\left(1-3\phi^2\frac{\Delta}{2}\right) + \phi^2\frac{\Delta^{3/2}}{6}\left(\frac{7}{2}\phi\Delta - 3\right); \quad H_3 = H_1^3 - 3H_1.
\end{aligned}
$$

Furthermore, the quantity $1/(\sqrt{\Delta}\sigma_\theta(x))\varphi((\rho_\theta(x)-\rho_\theta(x_0))/\sqrt{\Delta})$ for the OU process corresponds to a Gaussian density with mean x_0 and variance $\Delta\sigma_X^2$. Hence, the Hermite approximation of the transition density has the following form

$$
\Pi_\theta^{(K)}(\Delta, x_0, x) = \frac{1}{\sqrt{2\pi\Delta}\sigma_X}\exp\left(-\frac{(x-x_0)^2}{2\Delta\sigma_X^2}\right)\left(\sum_{k=0}^{K} D_{\theta,k}(\Delta, x_0)(x-x_0)^k\right),
$$

with $D_{\theta,k}(\Delta, x_0)$ parameters depending on Δ, θ and x_0 and corresponding to the coefficients terms $\eta_{\theta,k}, k = 1, \ldots, K$ up to some constants. □

6.6 Proof of Proposition 5.1

The calculation of the theoretical contrast defined as

$$
\gamma_\theta(\Delta) = \|\Pi_\theta\|_f^2 - 2\langle \Pi_\theta, \Pi_{\theta_0}\rangle_f,
$$

was obtained with the help of a symbolic calculation language (Mathematica). The Mathematica version has been simplified and reduced to the same denominator. In this section we give the computation of $\|\Pi_\theta\|_f^2$ and only the main lines of the calculation of the second term $\langle \Pi_\theta, \Pi_{\theta_0}\rangle_f$. In practice, this contrast is not known since it depends on θ_0, so only empirical and approximate contrasts calculations are necessary. They are detailed in the following.

For the OU process, let $m_{\theta,X,\Delta}$ and $\gamma_{\theta,X,\Delta}^2$ the mean and the variance of the true transition density $\Pi_\theta(\Delta, x, y)$ respectively given by

$$
m_{\theta,X,\Delta}(x) = xe^{-\phi\Delta} := a_{\theta,\Delta}x, \qquad \gamma_{\theta,X,\Delta}^2 = \frac{(1-e^{-2\phi\Delta})\sigma_{X,\Delta}^2}{2\phi}.
$$

The square of the transition density is also Gaussian up to the parameter $1/(2\sqrt{\pi}\gamma_{\theta,X,\Delta})$ with mean $m_{X,\Delta}(x)$ and variance $\gamma_{\theta,X,\Delta}^2/2$. The stationary density f_{θ_0} is also Gaussian with zero mean and variance $\mathcal{B}^2_{\theta_0} = \sigma^2_{X,0}/2\phi_0$. Hence,

$$\begin{aligned}\|\Pi_\theta\|^2_f &= \iint |\Pi_\theta(\Delta, x, y)|^2 f_{\theta_0}(x) dx dy = \int f_{\theta_0}(x) \Big(\int |\Pi_\theta(\Delta, x, y)|^2 dy \Big) dx \\ &= \int f_{\theta_0}(x) \tilde{\Pi}_\theta(\Delta, x) dx,\end{aligned}$$

with

$$\tilde{\Pi}_\theta(\Delta, x) = \int |\Pi_\theta(\Delta, x, y)|^2 dy = \frac{1}{2\sqrt{\pi\gamma^2_{\theta,X,\Delta}}} \frac{1}{\sqrt{\pi\gamma^2_{\theta,X,\Delta}}} \int \exp^{-\frac{(y-m_{\theta,X,\Delta}(x))^2}{\gamma^2_{\theta,X,\Delta}}} dy = \frac{1}{2\sqrt{\pi\gamma^2_{\theta,X,\Delta}}}.$$

Thus,

$$\|\Pi_\theta\|^2_f = \frac{1}{2\sqrt{\pi\gamma^2_{\theta,X,\Delta}}} \int f_{\theta_0}(x) dx = \frac{1}{2\sqrt{\pi\gamma^2_{\theta,X,\Delta}}}.$$

For the second part of the theoretical contrast we have

$$\begin{aligned}\langle \Pi_\theta, \Pi_{\theta_0} \rangle_f &= \iint \Pi_\theta(\Delta, x, y) \Pi_{\theta_0}(\Delta, x, y) f_{\theta_0}(x) dx dy = \int f_{\theta_0}(x) \int \Pi_\theta(\Delta, x, y) \Pi_{\theta_0}(\Delta, x, y) dy dx \\ &= \frac{1}{\sqrt{2\pi\gamma^2_{\theta,X,\Delta}}} \frac{1}{\sqrt{2\pi\gamma^2_{\theta_0,X,\Delta}}} \int f_{\theta_0}(x) \int e^{-\left(\frac{(y-m_{\theta,X,\Delta}(x))^2}{2\gamma^2_{X,\Delta}} + \frac{(y-m_{\theta_0,X,\Delta}(x))^2}{2\gamma^2_{\theta_0,X,\Delta}}\right)} dy dx \\ &= \frac{1}{\sqrt{2\pi\gamma^2_{\theta,X,\Delta}}} \frac{1}{\sqrt{2\pi\gamma^2_{\theta_0,X,\Delta}}} \int e^{(\frac{\gamma^2_{\theta_0,X,\Delta} m_{X,\Delta}(x) + \gamma^2_{\theta,X,\Delta} m_{0,X,\Delta}(x)}{(\gamma^2_{\theta,X,\Delta} + \gamma^2_{\theta_0,X,\Delta})})^2 - (\frac{\gamma^2_{\theta_0,X,\Delta} m^2_{\theta,X,\Delta}(x) + \gamma^2_{\theta,X,\Delta} m^2_{\theta_0,X,\Delta}(x)}{2\gamma^2_{\theta,X,\Delta}\gamma^2_{\theta_0,X,\Delta}})^2} \\ &\quad \times f_{\theta_0}(x) \Bigg(\int e^{-\left(\sqrt{\gamma^2_{\theta,X,\Delta} + \gamma^2_{\theta_0,X,\Delta}}\, y - \frac{\gamma^2_{\theta_0,X,\Delta} m_{\theta,X,\Delta}(x) + \gamma^2_{\theta,X,\Delta} m_{\theta_0,X,\Delta}(x)}{\sqrt{(\gamma^2_{\theta,X,\Delta} + \gamma^2_{\theta_0,X,\Delta})}}\right)^2} dy \Bigg) dx.\end{aligned}$$

Now, let $u = \sqrt{\gamma^2_{\theta,X,\Delta} + \gamma^2_{\theta_0,X,\Delta}}\, y$ we note that

$$\begin{aligned}&\int e^{-\left(\sqrt{\gamma^2_{\theta,X,\Delta} + \gamma^2_{\theta_0,X,\Delta}}\, y - \frac{\gamma^2_{\theta_0,X,\Delta} m_{\theta,X,\Delta}(x) + \gamma^2_{\theta,X,\Delta} m_{\theta_0,X,\Delta}(x)}{\sqrt{(\gamma^2_{\theta,X,\Delta} + \gamma^2_{\theta_0,X,\Delta})}}\right)^2} dy \\ &= \frac{1}{\sqrt{\gamma^2_{\theta,X,\Delta} + \gamma^2_{\theta_0,X,\Delta}}} \int e^{-\left(u - \frac{\gamma^2_{\theta_0,X,\Delta} m_{\theta,X,\Delta}(x) + \gamma^2_{\theta,X,\Delta} m_{\theta_0,X,\Delta}(x)}{\sqrt{(\gamma^2_{\theta,X,\Delta} + \gamma^2_{\theta_0,X,\Delta})}}\right)^2} du \\ &= \frac{1}{\sqrt{\pi(\gamma^2_{\theta,X,\Delta} + \gamma^2_{\theta_0,X,\Delta})}}.\end{aligned}$$

Hence,

$$\begin{aligned}\langle \Pi_\theta, \Pi_{\theta_0} \rangle_f &= \frac{1}{\sqrt{\pi(\gamma^2_{\theta,X,\Delta} + \gamma^2_{\theta_0,X,\Delta})}} \frac{1}{\sqrt{2\pi\gamma^2_{\theta,X,\Delta}}} \frac{1}{\sqrt{2\pi\gamma^2_{\theta_0,X,\Delta}}} \\ &\quad \times \int e^{(\frac{\gamma^2_{\theta_0,X,\Delta} m_{X,\Delta}(x) + \gamma^2_{\theta,X,\Delta} m_{\theta_0,X,\Delta}(x)}{(\gamma^2_{\theta,X,\Delta} + \gamma^2_{\theta_0,X,\Delta})})^2 - (\frac{\gamma^2_{\theta_0,X,\Delta} m^2_{\theta,X,\Delta}(x) + \gamma^2_{\theta,X,\Delta} m^2_{\theta_0,X,\Delta}(x)}{2\gamma^2_{\theta,X,\Delta}\gamma^2_{\theta_0,X,\Delta}})^2} f_{\theta_0}(x) dx.\end{aligned}$$

6.7 Details on the construction of the contrasts

Contrasts for the simulations: To compute the different contrasts defined in Section 5 and taking into account the non-integrability assumption, the following quantities are essentially required: $(\Pi_\theta^2(\Delta,x,0))^*, \Pi_\theta^*(\Delta,x,y)$ and $f_\varepsilon^*(x)$ and the introduction of a weight function is necessary.

We detail the calculation for the true transition density of the OU process but the reasoning remains valid for the Euler and for the Hermite approximation transition density one can use Theorem 4.1 with the Hermite coefficient given in Lemma 5.1.

For the OU process, we note that the square of the transition density is also Gaussian up to the parameter $1/(2\sqrt{\pi\gamma_{\theta,X,\Delta}^2})$ with mean $m_{\theta,X,\Delta}(x) := a_{\theta,\Delta}x$ and variance $\gamma_{\theta,X,\Delta}^2/2$. Hence, we are interested in computing the following FT:

$$(\Pi_\theta^2(\Delta,x,0))^* = \int e^{-ixu}\Big(\int \Pi_\theta^2(\Delta,u,v)dv\Big)du = \int e^{-ixu}\tilde{\Pi}_\theta(\Delta,u)du = (\tilde{\Pi}_\theta(\Delta,x))^*.$$

By integration of the Gaussian density, we have that $\tilde{\Pi}_\theta(\Delta,x) = 1/(2\sqrt{\pi}\gamma_{\theta,X,\Delta})$ $\forall x$, which is integrable on $\mathbb{L}_1(A)$. Nevertheless, for Gaussian noises (super-smooth noises), Assumptions **A** 3(*iii*) and **A** 4(*ii*) are not satisfied since $x \mapsto (\tilde{\Pi}_\theta(\Delta,x))^*/f_\varepsilon^*(x)$ is not integrable despite the fact that the numerator and denominator taken separately can be integrated. In this case, we introduce a weight function φ belongs to $\mathcal{S}(\mathbb{R})$, where $\mathcal{S}(\mathbb{R})$ is the Schwartz space of functions defined by $\mathcal{S}(\mathbb{R}) = \{f \in \mathcal{C}^\infty(\mathbb{R}),\quad \forall\alpha, N$ there exists $C_{N,\alpha}$ such that $|\nabla_x^\alpha f(x)| \leq C_{N,\alpha}(1+|x|)^{-N}\}$.

Hence, $\forall\varphi \in \mathcal{S}(\mathbb{R})$, we have

$$\begin{aligned}\Big\langle \varphi, \tilde{\Pi}_\theta^*\Big\rangle &= \int \varphi(x)dx \int \tilde{\Pi}_\theta(\Delta,u)e^{-ixu}du = \frac{1}{2\sqrt{\pi\gamma_{\theta,X,\Delta}^2}}\int \varphi(x)dx\int e^{-ixu}du \\ &= \frac{1}{2\sqrt{\pi\gamma_{\theta,X,\Delta}^2}}\Big\langle \varphi, \delta_0\Big\rangle,\end{aligned}$$

where δ_x is the Dirac distribution at point x. Hence, by taking $\varphi : u \mapsto \tilde{\varphi}(u)e^{ixu}/f_\varepsilon^*(-u) \in \mathcal{S}(\mathbb{R})$ with $\tilde{\varphi} : u \mapsto 2\pi e^{-\sigma_\varepsilon^2 u^2}$, we obtain the operator Q as follows

$$Q_{\Pi_\theta^2}(x) = \frac{1}{2\pi}\int e^{ixu}\frac{\tilde{\varphi}(u)\tilde{\Pi}_\theta^*(u)}{f_\varepsilon^*(-u)}du = \Big\langle \varphi, \tilde{\Pi}_\theta^*\Big\rangle = \frac{1}{2\sqrt{\pi\gamma_{\theta,X,\Delta}^2}}\Big\langle \varphi, \delta_0\Big\rangle = \frac{1}{2\sqrt{\pi\gamma_{\theta,X,\Delta}^2}}\varphi(0),$$

where $\varphi(0) = 1$ for all cases (Gaussian and Laplace noises) in Section 5. Here, we take $\tilde{\varphi}$ dependent of σ_ε^2 since we assume that this variance is known but one can take any function $\tilde{\varphi}$ such that $\tilde{\varphi}/f_\varepsilon^*$ is in $\mathcal{S}$.

For $\Pi_\theta^*(\Delta,x,y)$ we make the same analogy, that is let $\Pi_{u,\theta}(\Delta,v)$ the function $v \mapsto \Pi_\theta(\Delta,u,v)$ $\forall u$. For the Gaussian transition density Π_θ we have $\forall u$,

$$\begin{aligned}(\Pi_{u,\theta}(\Delta,y))^* &= \int e^{-iyv}\Pi_\theta(\Delta,u,v)dv = \int e^{-iyv}\frac{1}{\sqrt{2\pi\gamma_{\theta,X,\Delta}^2}}e^{-\frac{(v-a_{\theta,\Delta}u)^2}{2\gamma_{\theta,X,\Delta}^2}}dv \\ &= e^{-ia_{\theta,\Delta}uy-\frac{\gamma_{\theta,X,\Delta}^2}{2}y^2}.\end{aligned}$$

Let $\Pi_{y,\theta}(\Delta,u)$ be the function $u\mapsto(\Pi_{u,\theta}(\Delta,y))^*$ $\forall y$. Then, we have $\forall\varphi\in\mathcal{S}$ and $\forall y$

$$\begin{aligned}\Big\langle\varphi,\Pi_{y,\theta}\Big\rangle &= \int\varphi(x)dx\int e^{-ixu}(\Pi_{u,\theta}(\Delta,y))^*du\\ &= \int\varphi(x)dx\int e^{-ixu}e^{-ia_{\theta,\Delta}uy-\frac{\gamma^2_{\theta,X,\Delta}}{2}y^2}du\\ &= e^{-\frac{\gamma^2_{\theta,X,\Delta}}{2}y^2}\int\varphi(z-a_{\theta,\Delta}y)dz\int e^{-iuz}du\\ &= e^{-\frac{\gamma^2_{\theta,X,\Delta}}{2}y^2}\Big\langle\varphi(.a_{\theta,\Delta}y),\mathbf{1}^*\Big\rangle\\ &= e^{-\frac{\gamma^2_{\theta,X,\Delta}}{2}y^2}\Big\langle\varphi(.a_{\theta,\Delta}y),\delta_0\Big\rangle = e^{-\frac{\gamma^2_{\theta,X,\Delta}}{2}y^2}\varphi(-a_{\theta,\Delta}y).\end{aligned}$$

Hence, the operator V_{Π_θ} is obtained as follows for Gaussian noises, *i.e.*,

$$\begin{aligned}V_{\Pi_\theta}(x,y) &= \frac{1}{4\pi^2}\int\int\tilde{\varphi}_1(u)\tilde{\varphi}_2(v)e^{i(xu+yv)}\frac{\Pi^*_\theta(\Delta,u,v)}{f^*_\varepsilon(-u)f^*_\varepsilon(-v)}dudv\\ &= \frac{1}{4\pi^2}\int\int\frac{\tilde{\varphi}_1(u)\tilde{\varphi}_2(v)}{f^*_\varepsilon(-u)f^*_\varepsilon(-v)}e^{i(xu+yv)}\left(\int e^{-iwu}(\Pi_{w,\theta}(\Delta,v))^*dw\right)dudv\\ &= \frac{1}{4\pi^2}\int\frac{\tilde{\varphi}_2(v)}{f^*_\varepsilon(-v)}e^{iyv}\left(\int\frac{\tilde{\varphi}_1(u)}{f^*_\varepsilon(-u)}e^{ixu}\left(\int e^{-iwu-iva_{\theta,\Delta}w-\frac{\gamma^2_{\theta,X,\Delta}}{2}v^2}dw\right)du\right)dv\\ &= \frac{1}{4\pi^2}\int\frac{\tilde{\varphi}_2(v)}{f^*_\varepsilon(-v)}e^{iyv-\frac{\gamma^2_{\theta,X,\Delta}}{2}v^2}\left(\int\frac{\tilde{\varphi}_1(u)}{f^*_\varepsilon(-u)}e^{ixu}\left(\int e^{-iw(u+a_{\theta,\Delta}v)}dw\right)du\right)dv\\ &= \frac{1}{4\pi^2}\int\frac{\tilde{\varphi}_2(v)}{f^*_\varepsilon(-v)}e^{iyv-\frac{\gamma^2_{\theta,X,\Delta}}{2}v^2}\left(\int\frac{\tilde{\varphi}_1(z-a_{\theta,\Delta}v)}{f^*_\varepsilon(\phi v-z)}e^{ix(z-a_{\theta,\Delta}v)}\left(\int e^{-iwz}dw\right)dz\right)dv\\ &= \frac{1}{2\pi}\int\frac{\tilde{\varphi}_2(v)}{f^*_\varepsilon(-v)}e^{iyv-\frac{\gamma^2_{\theta,X,\Delta}}{2}v^2}\left(\Big\langle\varphi(.-a_{\theta,\Delta}v),\mathbf{1}^*\Big\rangle\right)dv\\ &= \frac{1}{2\pi}\int\frac{\tilde{\varphi}_2(v)}{f^*_\varepsilon(-v)}e^{iyv-\frac{\gamma^2_{\theta,X,\Delta}}{2}v^2}\varphi(-a_{\theta,\Delta}v)dv\\ &= \frac{1}{2\pi}\int e^{-ivy}\left(e^{iv(y-a_{\theta,\Delta}x)-\frac{v^2}{2}(\gamma^2_{\theta,X,\Delta}-\sigma^2_\varepsilon(1+a^2_{\theta,\Delta}))}\right)dv\\ &= \frac{1}{2\pi}\int e^{ivy}\left(e^{-iv(y-a_{\theta,\Delta}x)-\frac{v^2}{2}(\gamma^2_{\theta,X,\Delta}-\sigma^2_\varepsilon(1+a^2_{\theta,\Delta}))}\right)dv\\ &= \frac{1}{\sqrt{2\pi(\gamma^2_{\theta,X,\Delta}-\sigma^2_\varepsilon(1+a^2_{\theta,\Delta}))}}\exp\left(-\frac{(y-a_{\theta,\Delta}x)^2}{2(\gamma^2_{\theta,X,\Delta}-\sigma^2_\varepsilon(1+a^2_{\theta,\Delta}))}\right),\end{aligned}$$

where $\varphi:u\mapsto e^{ixu}\tilde{\varphi}_1(u)/f^*_\varepsilon(-u)$ with $\tilde{\varphi}_1:u\mapsto 2\pi e^{-\sigma^2_\varepsilon u^2}$ and $\tilde{\varphi}_2:v\mapsto e^{-ivy-\sigma^2_\varepsilon v^2}$ and such that φ,φ_1 and $\varphi_2\in\mathcal{S}$. For Laplace noises, one can make the same computations by replacing f^*_ε by its expression $\frac{1}{1+\sigma^2_\varepsilon x^2/2}$. □

Acknowledgements

The authors are very grateful to the Editor, the Associate Editor and the two anonymous referees for their careful reading and valuable comments that have led to an improvement of the paper.

References

[1] Y. Aït-Sahalia. Transition densities for interest rate and other nonlinear diffusions. *The journal of finance*, 54(4):1361–1395, 1999.

[2] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70(1):223–262, 2002.

[3] Y. Aït-Sahalia. Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics*, 36(2):906–937, 2008.

[4] Y. Aït-Sahalia and R. Kimmel. Maximum likelihood estimation of stochastic volatility models. *Journal of financial economics*, 83(2):413–452, 2007.

[5] G. Bakshi and N. Ju. A refinement to Aït-Sahalia's (2002) "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach". *The Journal of Business*, 78(5):2037–2052, 2005.

[6] F. Baltazar-Larios and M. Sørensen. Maximum likelihood estimation for integrated diffusion processes. In *Contemporary quantitative finance*, pages 407–423. Springer, 2010.

[7] P. Bertail and S. Clémençon. Approximate regenerative-block bootstrap for markov chains. *Computational Statistics & Data Analysis*, 52(5):2739–2756, 2008.

[8] A. Beskos, O. Papaspiliopoulos, G. O. Roberts, and P. Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382, 2006.

[9] J. P. Boyd. *Dynamics of the equatorial ocean.* Springer, 2018.

[10] M. W. Brandt and P. Santa-Clara. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of financial economics*, 63(2):161–210, 2002.

[11] E. Brunel, F. Comte, and C. Lacour. Adaptive estimation of the conditional density in the presence of censoring. *Sankhyā: The Indian Journal of Statistics*, pages 734–763, 2007.

[12] S. Choi. Explicit form of approximate transition probability density functions of diffusion processes. *Journal of econometrics*, 187(1):57–73, 2015.

[13] D. Dacunha-Castelle and M. Duflo. *Probabilités et statistiques: problèmes à temps fixe*, volume 1. Masson, 1982.

[14] J. Dedecker, A. Samson, and M.-L. Taupin. Estimation in autoregressive model with measurement error. *ESAIM Prob. Stat.*, 18:277–307, 2014.

[15] P. Doukhan. *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994. Properties and examples.

[16] A. V. Egorov, H. Li, and Y. Xu. Maximum likelihood estimation of time-inhomogeneous diffusions. *Journal of Econometrics*, 114(1):107–139, 2003.

[17] S. El Kolei and F. Pelgrin. Parametric inference of autoregressive heteroscedastic models with errors in variables. *Statist. Probab. Lett.*, 130:63–70, 2017.

[18] B. Favetto. Parameter estimation by contrast minimization for noisy observations of a diffusion process. *Statistics*, 48(6):1344–1370, 2014.

[19] B. Favetto. Estimating functions for noisy observations of ergodic diffusions. *Statistical Inference for Stochastic Processes*, 19(1):1–28, 2016.

[20] D. Florens-Zmirou. Approximate discrete-time schemes for statistics of diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 20(4):547–557, 1989.

[21] V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. In *Annales de l'IHP Probabilités et statistiques*, volume 29, pages 119–151, 1993.

[22] V. Genon-Catalot, T. Jeantheau, and C. Laredo. Parameter estimation for discretely observed stochastic volatility models. *Bernoulli*, 5(5):855–872, 1999.

[23] V. Genon-Catalot, T. Jeantheau, and C. Larédo. Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*, 6(6):1051–1079, 2000.

[24] A. Gloter. Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM: Probability and Statistics*, 4:205–227, 2000.

[25] A. Gloter. Estimation du coefficient de diffusion de la volatilité d'un modèle à volatilité stochastique. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, 330(3):243–248, 2000.

[26] A. Gloter. Parameter estimation for a discretely observed integrated diffusion process. *Scandinavian Journal of Statistics*, 33(1):83–104, 2006.

[27] L. P. Hansen, J. A. Scheinkman, and N. Touzi. Spectral methods for identifying scalar diffusions. *Journal of Econometrics*, 86(1):1–32, 1998.

[28] P. Honoré. Maximum likelihood estimation of non-linear continuous-time term-structure models. *Available at SSRN 7669*, 1997.

[29] A. Hurn, J. Jeisman, and K. Lindsay. Ml estimation of the parameters of sdeâĂŹs by numerical solution of the fokker–planck equation. In *MODSIM 2005: international congress on modelling and simulation: advances and applications for management and decision making*, pages 849–855. Citeseer, 2005.

[30] B. Jensen and R. Poulsen. Transition densities of diffusion processes: numerical comparison of approximation techniques. *The Journal of Derivatives*, 9(4):18–32, 2002.

[31] M. Kessler. Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24(2):211–229, 1997.

[32] M. Kessler. Simple and explicit estimating functions for a discretely observed diffusion process. *Scandinavian Journal of Statistics*, 27(1):65–82, 2000.

[33] M. Kessler, A. Lindner, and M. Sorensen. *Statistical methods for stochastic differential equations*. Chapman and Hall/CRC, 2012.

[34] C. Lacour. Adaptive estimation of the transition density of a markov chain. In *Annales de l'IHP Probabilités et statistiques*, volume 43, pages 571–597, 2007.

[35] C. Lacour. Adaptive estimation of the transition density of a particular hidden markov chain. *Journal of Multivariate Analysis*, 99(5):787–814, 2008.

[36] C. Lacour. Least squares type estimation of the transition density of a particular hidden Markov chain. *Electron. J. Stat.*, 2:1–39, 2008.

[37] C. Lacour. Nonparametric estimation of the stationary density and the transition density of a markov chain. *Stochastic Processes and their Applications*, 118(2):232–260, 2008.

[38] Y. D. Lee, S. Song, and E.-K. Lee. The delta expansion for the transition density of diffusion models. *Journal of econometrics*, 178:694–705, 2014.

[39] C. Li. Maximum-likelihood estimation for diffusion processes via closed-form density expansions. *The Annals of Statistics*, 41(3):1350–1380, 2013.

[40] A. W. Lo. Maximum likelihood estimation of generalized itô processes with discretely sampled data. *Econometric Theory*, 4(2):231–247, 1988.

[41] A. R. Pedersen. Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, pages 257–279, 1995.

[42] A. R. Pedersen. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian journal of statistics*, pages 55–71, 1995.

[43] U. Picchini and A. Samson. Coupling stochastic em and approximate bayesian computation for parameter inference in state-space models. *Computational Statistics*, 33(1):179–212, 2018.

[44] L. Rogers. Stochastic differential equations and diffusion processes: Nobuyuki ikeda and shinzo watanabe north-holland, amsterdam, 1981, xiv+ 464 pages, dfl. 175.00, 1982.

[45] L. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales*, volume 2 of *Cambridge Mathematical Library*. Cambridge University Press, 2 edition, 2000.

[46] S. Särkkä, J. Hartikainen, I. S. Mbalawata, and H. Haario. Posterior inference on parameters of stochastic differential equations via non-linear gaussian filtering and adaptive mcmc. *Statistics and Computing*, 25(2):427–437, 2015.

[47] G. Sermaidis, O. Papaspiliopoulos, G. O. Roberts, A. Beskos, and P. Fearnhead. Markov chain monte carlo for exact inference for diffusions. *Scandinavian Journal of Statistics*, 40(2):294–321, 2013.

[48] H. Singer. Moment equations and hermite expansion for nonlinear stochastic differential equations with application to stock price models. *Computational Statistics*, 21(3-4):385–397, 2006.

[49] M. Sørensen. *Parametric Inference for Discretely Sampled Stochastic Differential Equations*, pages 531–553. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[50] O. Stramer and J. Yan. On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation. *Journal of Computational and Graphical Statistics*, 16(3):672–691, 2007.

[51] J. Von Neumann. Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics*, 12(4):367–395, 1941.

[52] E. Wong. The construction of a class of stationary markoff processes. *Stochastic processes in mathematical physics and engineering*, (17):264–276, 1964.

[53] N. Yoshida. Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, 41(2):220–242, 1992.