PHYSICAL REVIEW E **80**, 031103 (2009)

Arbitrary-order corrections for finite-time drift and diffusion coefficients

C. Anteneodo* and R. Riera†

Department of Physics, PUC-Rio and National Institute of Science and Technology for Complex Systems, CP 38071, 22452-970 Rio de Janeiro, RJ, Brazil

(Received 15 June 2009; published 3 September 2009)

We address a standard class of diffusion processes with linear drift and quadratic diffusion coefficients. These contributions to dynamic equations can be directly drawn from data time series. However, real data are constrained to finite sampling rates and therefore it is crucial to establish a suitable mathematical description of the required finite-time corrections. Based on Itô-Taylor expansions, we present the exact corrections to the finite-time drift and diffusion coefficients. These results allow to reconstruct the real hidden coefficients from the empirical estimates. We also derive higher-order finite-time expressions for the third and fourth conditional moments that furnish extra theoretical checks for this class of diffusion models. The analytical predictions are compared with the numerical outcomes of representative artificial time series.

DOI: 10.1103/PhysRevE.80.031103 PACS number(s): 02.50.Ey, 05.10.Gg, 05.40.−a

I. INTRODUCTION

Many fluctuating random phenomena can be modeled by a univariate Itô-stochastic differential equation (SDE) of the form below, characterizing a diffusion model:

$$dX_t = D_1(X_t)dt + \sqrt{2D_2(X_t)}dW_t, \qquad (1)$$

where W_t is a Wiener process, $D_1(X_t)$ is the coefficient of the slowly varying component (called drift coefficient), and $D_2(X_t)$ is the coefficient of the rapid one (called diffusion coefficient).

For sufficiently smooth and bounded drift and diffusion coefficients, the associated probability density function (PDF) $P(x,t) \equiv P(X_t = x, t)$ is governed by the corresponding Fokker-Planck equation [1]

$$\partial_t P(x,t) = -\partial_x[D_1(x)P(x,t)] + \partial_{xx}[D_2(x)P(x,t)]. \qquad (2)$$

Here, we are concerned with the empirical access to unknown drift and diffusion coefficients of stochastic processes. For an ideal time series X_t generated by Eq. (1) and sampled with a sufficiently high resolution on a long time period, the original coefficients can be perfectly reconstructed. For stationary processes, the coefficients $D_k(x)$ with $k=1,2$ can be directly estimated from the conditional moments [1] as

$$D_k(x) = \lim_{\tau\to 0} \widetilde{D}_k(x,\tau), \qquad (3)$$

where

$$\widetilde{D}_k(x,\tau) = \frac{1}{\tau k!}\langle [X_{t+\tau} - X_t]^k\rangle|_{X_t=x}, \qquad (4)$$

with $\langle \cdots \rangle$ denoting statistical average and $|_{X_t=x}$ meaning that at time t the stochastic variable assumes the value x.

Conversely, for general Markovian stochastic processes, the time evolution of PDFs is governed by a generalization of Eq. (2), namely,

$$\partial_t P(x,t) = \sum_{k\geq 0} (-\partial_x)^k [D_k(x)P(x,t)], \qquad (5)$$

with coefficients $D_k(x)$ given by Eqs. (3) and (4) for any integer $k \geq 1$. For diffusion processes, Eq. (5) reduces to Eq. (2). Therefore, processes governed by the Itô-Langevin Eq. (1) must furnish null coefficients $\widetilde{D}_k$ for $k \geq 3$. Pawula theorem [1] simplifies this task by stating that if D_4 is null, all other coefficients with $k \geq 3$ are null as well. The coefficient D_4 is then a key coefficient to be investigated in order to establish the validity of the modeling of data series by Eqs. (1) and (2).

However, due to the finite sampling rate of real data, numerical estimations of $\widetilde{D}_k$ cannot always be straightforwardly extrapolated to the limit $\tau \to 0$ in Eq. (3). In such cases, one accesses only the finite-τ estimation of the coefficients given by Eq. (4), which may significantly differ from the true coefficients D_k. This is especially relevant when τ is large compared to the characteristic time scales of the process.

Some authors [2] have introduced finite sampling rate corrections to the coefficients $D_1(x,\tau)$ and $D_2(x,\tau)$ by deriving expansions for the conditional moments up to some specified low order of τ directly from the Fokker-Planck equations. Applications of this approach have already been implemented for those coefficients up to second order [3]. The error in the finite-τ estimated coefficients $\widetilde{D}_k$ can also be derived from the stochastic Itô-Taylor expansion [4] of the integrated form of Eq. (1). Within this line, the first-order expansion of drift and diffusion coefficients was recently presented in Ref. [5]. However, low-order corrections may be inappropriate when the convergence of the limit in Eq. (3) is slow [2,6]. Moreover, there is no *a priori* knowledge of whether the sampling rate is fine enough to justify the use of the lowest order approximation.

The most common diffusion models of the theoretical literature defined by Eq. (1) have linear drift coefficient, namely, $D_1(x) = -a_1 x$, representing a harmonic restoring

*celia@fis.puc-rio.br
†rrif@fis.puc-rio.br

mechanism and quadratic state-dependent diffusion coefficient, namely, $D_2(x)=b_0+b_2x^2$. In fact, this equation is frequently found in a diversity of processes from turbulence to finance [7,8]. This ubiquity is mainly due to the fact that those functional forms represent first-order Taylor expansions of more general drift and diffusion coefficients.

In the present work, we investigate finite-time corrections for this important class of diffusion models. For discretely sampled data at intervals τ, we will derive, from the stochastic Itô-Taylor expansion, finite-τ expressions for the parameters $\{a_1,b_0,b_2\}$ up to infinite order. These exact expressions will allow us to reconstruct the true drift and diffusion coefficients from their empirical finite-time estimates. As a corollary, one can determine up to which value of τ a given order of truncation is reliable (within a fixed tolerance), or reciprocally, which is the suitable order for a given τ.

Furthermore, as empirical estimates suffer from finite-τ effects, one always gets non-null D_4. Therefore, the evaluation of the corrections for this coefficient is crucial for a suitable probe of the modeling. In this work, we also derive finite-τ expressions for coefficients D_3 and D_4, which furnish extra theoretical tests of consistency for the diffusion models considered. Our theoretical findings are corroborated by the outcomes of exemplary artificial time series generated by Eq. (1).

II. EXACT CORRECTIONS FOR DRIFT AND DIFFUSION COEFFICIENTS

Let us consider the Itô formula [4] for a given function F of the stochastic variable X_t,

$$\begin{aligned} dF &= (\partial_t F + D_1\partial_X F + D_2\partial_{XX}F)dt + \sqrt{2D_2}\partial_X F dW \\ &\equiv L^0 F dt + L^1 F dW, \end{aligned} \tag{6}$$

and its integrated form

$$F(X_{t+\tau}) = F(X_t) + \int_t^{t+\tau} L^0 F(X_s)ds + \int_t^{t+\tau} L^1 F(X_s)dW_s. \tag{7}$$

Let τ be the sampling interval of state space observations. By applying Itô formula (7) to the functions $D_1(X_s)$ and $\sqrt{2D_2(X_s)}$ in the integral form of Eq. (1),

$$X_{t+\tau} = X_t + \int_t^{t+\tau} D_1(X_s)ds + \int_t^{t+\tau}\sqrt{2D_2(X_s)}dW_s, \tag{8}$$

one finds

$$\begin{aligned} X_{t+\tau} = X_t &+ \int_t^{t+\tau}\Big[D_1(X_t) + \int_t^s L^0 D_1(X'_s)ds' \\ &+ \int_t^s L^1 D_1(X_{s'})dW_{s'}\Big]ds + \int_t^{t+\tau}\Big[\sqrt{2D_2(X_t)} \\ &+ \int_t^s L^0\sqrt{2D_2(X'_s)}ds' + \int_t^s L^1\sqrt{2D_2(X_{s'})}dW_{s'}\Big]dW_s. \end{aligned} \tag{9}$$

After iterated applications of Itô formula, one gets an expression for the increment of the stochastic variable in terms of multiple stochastic integrals [4]:

$$X_{t+\tau} - X_t = \sum_{\alpha_k} c_{\alpha_k}(D_1,D_2)I_{\alpha_k}, \tag{10}$$

where $\alpha_k=(j_1,j_2,\ldots,j_k)$, with $j_i=0,1$ for all i, $c_{\alpha_k}(D_1,D_2)=L^{j_1}L^{j_2}\ldots L^{j_{k-1}}L^{j_k}$ and I_{α_k} are multiple stochastic integrals of the form $I_{\alpha_k} = \int_t^{t+\tau}\int_t^{t+t_k}\int_t^{t+t_{k-1}}\ldots\int_t^{t+t_2}dt_1^{j_1}\ldots dt_{k-1}^{j_{k-1}}dt_k^{j_k}$, with $dt_i^0\equiv dt_i$ and $dt_i^1\equiv dW_i$.

By inserting Eq. (10) into Eq. (4) and performing the averaging, for $k=1,2$, we achieve analytical expressions for the finite-τ drift and diffusion coefficients up to arbitrary order in powers of τ. The resulting expressions preserve the linear and quadratic x dependence, respectively, and can be written as

$$\tilde{D}_1(x,\tau) = -\tilde{a}_1(\tau)x,$$

$$\tilde{D}_2(x,\tau) = \tilde{b}_0(\tau) + \tilde{b}_2(\tau)x^2. \tag{11}$$

Hence, we are led to the theoretical relation between the finite-τ coefficients $\{\tilde{a}_1,\tilde{b}_0,\tilde{b}_2\}$ and the true ones $\{a_1,b_0,b_2\}$, namely,

$$\tilde{a}_1(\tau) = a_1\sum_{j\geq 0}\frac{[-a_1]^j}{(j+1)!}\tau^j, \tag{12}$$

$$\tilde{b}_0(\tau) = b_0\sum_{j\geq 0}\frac{[-2(a_1-b_2)]^j}{(j+1)!}\tau^j, \tag{13}$$

$$\tilde{b}_2(\tau) = \sum_{j\geq 0}\frac{\frac{1}{2}[-2(a_1-b_2)]^{j+1} - [-a_1]^{j+1}}{(j+1)!}\tau^j. \tag{14}$$

Details of the derivation of Eqs. (12)–(14) can be found in the Appendix.

By restricting expansions (12)–(14) to some common finite power n, one gets the respective nth-order approximation. This result extends previous findings of first- [5] and second- [3] order terms.

Notice that Eq. (12) is uncoupled meaning that the estimated harmonic stiffness $\tilde{a}_1$ is not affected by the exact noise components. Moreover, from Eq. (14), the estimated multiplicative noise parameter $\tilde{b}_2$ does not depend on the exact additive noise component.

Summing the series in Eqs. (12)–(14) up to infinite order, and defining $Z\equiv\exp(-a_1\tau)$ and $W\equiv\exp(2b_2\tau)$, we find the exact finite-τ expressions:

$$\tilde{a}_1 = \frac{1 - Z}{\tau},$$

$$\tilde{b}_0 = \frac{b_0}{a_1 - b_2} \frac{1 - Z^2 W}{2\tau},$$

$$\tilde{b}_2 = \frac{1 - Z}{\tau} - \frac{1 - Z^2 W}{2\tau}. \qquad (15)$$

Notice that $\lim_{\tau\to 0}\{\tilde{a}_1, \tilde{b}_0, \tilde{b}_2\} = \{a_1, b_0, b_2\}$ holds.

From Eq. (15), we obtain an invariant relation among the estimated and exact parameters, namely,

$$\frac{\tilde{a}_1 - \tilde{b}_2}{\tilde{b}_0} = \frac{a_1 - b_2}{b_0}. \qquad (16)$$

The meaning of this invariance can be drawn, for instance, from the stationary PDF $P^*(x)$ associated to the corresponding Fokker-Planck equation given by Eq. (2). With the present choice of drift and diffusion coefficients, for a_1, $b_0 > 0$, $b_2 \geq 0$, one has

$$P^*(x) = P_o / \left[1 + \frac{b_2}{b_0} x^2\right]^{(a_1/2b_2)+1}, \qquad (17)$$

with P_o as a normalization constant. This solution is of the q-Gaussian form [9], for which, if $a_1 - b_2 > 0$, the variance is finite with value $\sigma^2 = b_0/(a_1 - b_2)$. Hence, Eq. (16) represents the uphold of the data variance under changes in sampling intervals. For $b_2 = 0$, one recovers the Gaussian stationary solution and its variance relation. Notice also that, from Eqs. (12)–(14), Eq. (16) still holds if one considers partial corrections of the parameters up to any common order n of truncation of the sums.

Let us remark that the results presented in Eqs. (11)–(15) are valid even when the variance is infinite. However, we will deal only with finite variance cases and consider normalized data (with unitary variance), which only implies a rescaling of $b_0 \to b_0/\sigma^2$. Then,

$$a_1 = b_0 + b_2. \qquad (18)$$

Taking into account constraint (18), from Eq. (15), the exact finite-τ expressions are

$$\tilde{a}_1 = \frac{1 - \exp(-a_1 \tau)}{\tau}, \qquad (19)$$

$$\tilde{b}_0 = \frac{1 - \exp(-2 b_0 \tau)}{2\tau}. \qquad (20)$$

Equations (19) and (20) can be readily inverted to extract the true parameters from their finite-τ estimates:

$$a_1 = \frac{\ln(1 - \tilde{a}_1 \tau)}{-\tau}, \qquad (21)$$

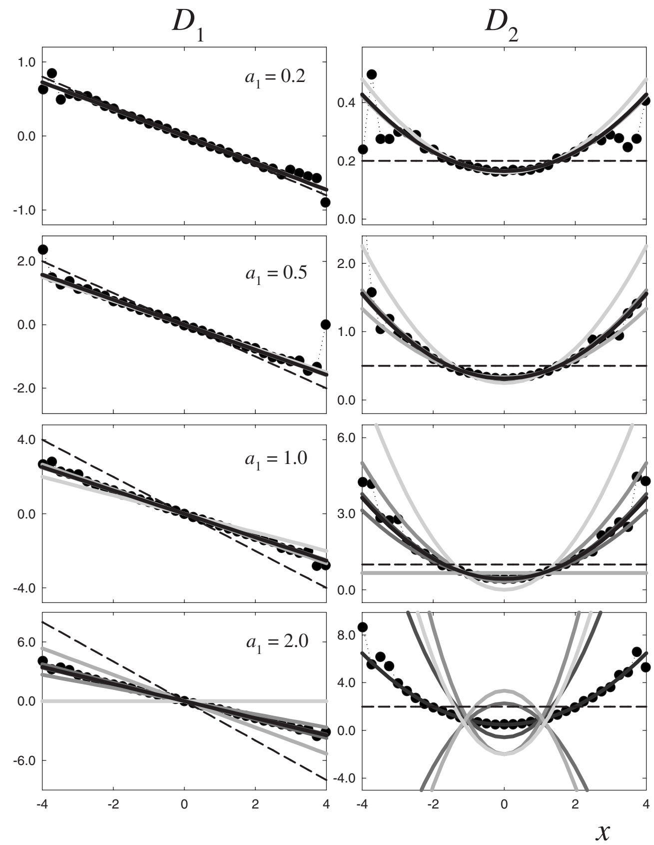

FIG. 1. Drift and diffusion coefficients for the O-U process. Symbols correspond to the numerical computation for artificial series (10^5 data), synthetized with the values of $a_1 = b_0$ [$b_2 = 0$, in accord with constraint (18)] indicated on each panel. Lines represent the coefficients given by Eq. (11) using the theoretical τ expansions (12)–(14) at different orders of truncation. The darker the color, the higher the order, from first up to fifth order. The infinite order (exact expression) is represented in thick black lines. The zeroth order, corresponding to the true values, is plotted in dashed lines.

$$b_0 = \frac{\ln(1 - 2\tilde{b}_0 \tau)}{-2\tau}. \qquad (22)$$

Notice that $\tilde{a}_1 \tau$ (and also $2\tilde{b}_0 \tau$) cannot be greater than unit.

In what follows, we fix the time scale $\tau = 1$. A different choice would simply lead to a rescaling of the parameters $(a_1, b_0, b_2) \to (\tau a_1, \tau b_0, \tau b_2)$.

Now we investigate the importance of finite-τ effects for discretely sampled realizations of representative known diffusion processes. To this end, we generated artificial time series through numerical integration of Eq. (1) by means of an Euler algorithm with time step $dt = 10^{-3}$, recording the data at each $1/dt$ time steps, in accord with our choice $\tau = 1$. Our theoretical results for D_1 and D_2 will be compared to the ones numerically computed from the time series through Eq. (4).

The particular case $b_2 = 0$ corresponding to the Orstein-Uhlenbeck (O-U) process and the general case with multiplicative component $b_2 > 0$ will be investigated separately. Fig-

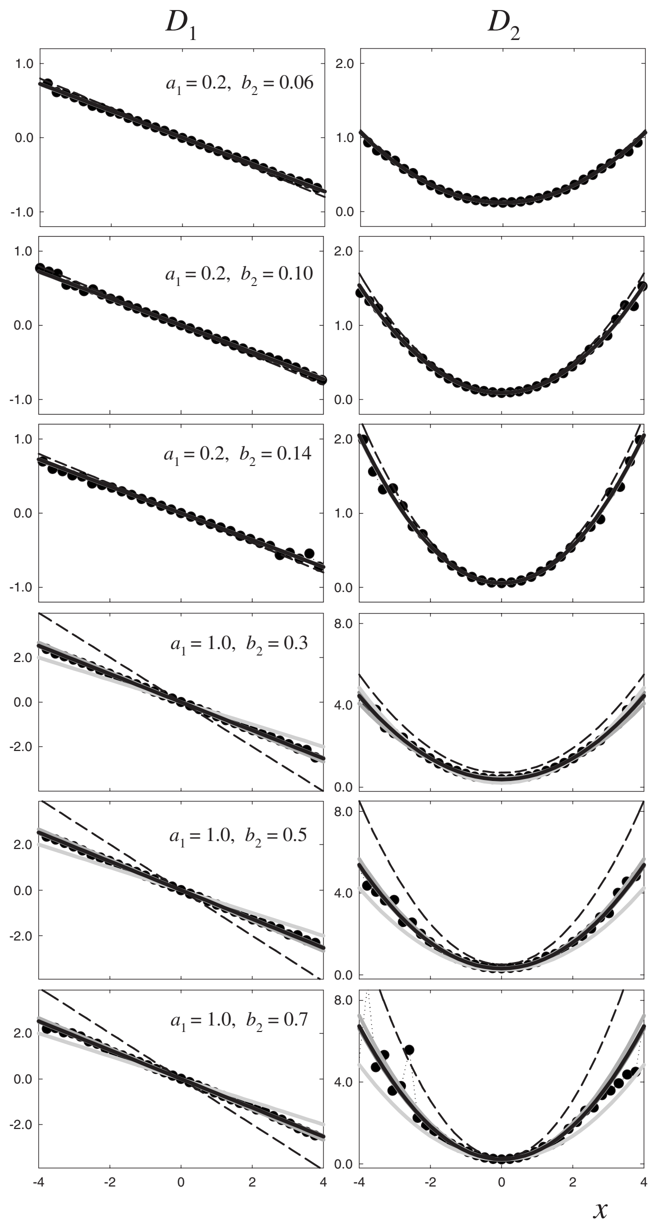

FIG. 2. Drift and diffusion coefficients for the general process with multiplicative noise. Symbols correspond to the numerical computation for artificial series (10^6 data), synthetized with the values of a_1 [$=b_0+b_2$, in accord with constraint (18)] and b_2 indicated on each panel. Lines are as in Fig. 1.

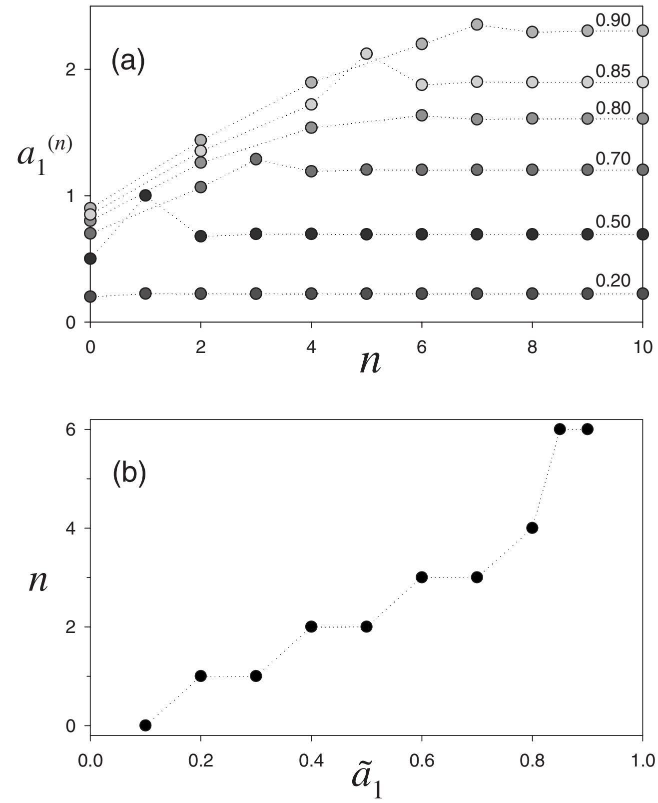

FIG. 3. Dependence of $a_1^{(n)}$ on the order n of the approximation given by Eq. (12) for different values of $\tilde{a}_1$ [panel (a)]. Dotted lines correspond to the respective true values (a_1) and missing points denote the absence of real solutions. Panel (b) exhibits the order at which the limiting value is attained (within 5%) as a function of $\tilde{a}_1$.

ure 1 shows the results for the artificial series with known values of the parameters $a_1=b_0$ ($b_2=0$) together with our theoretical predictions. The exact theoretical expressions reproduce the numerical (finite-time) outcomes. Comparing the panels in Fig. 1, it is clear that the larger a_1, the slower the convergence to the observed coefficients. The results for $a_1>1$ also illustrate the entanglement one may find in large-τ measurements, specifically, an oscillatory convergence of $\tilde{a}_1$ and an alternating signal of $\tilde{b}_2$.

In Fig. 2, we plot the numerical computations for artificial time series together with analytical predictions for $b_2>0$. Again, the theoretical approximations present slower convergence as a_1 increases while the exact theoretical expressions agree with finite-time estimates directly obtained from the time series. Moreover, the actual value of b_2 sets the convergence rate of $\tilde{b}_2$.

All these results raise a question about the domain of validity of lower order approximations presented before in the literature. Let us investigate this issue quantitatively. Given $\tilde{a}_1$, obtained from numerical (finite-time) evaluation, the exact value of a_1 can be recovered from Eq. (21). Approximate values $a_1^{(n)}$ can be found by inversion of Eq. (12) truncated at order n. Figure 3 illustrates $a_1^{(n)}$ as a function of n for different values of $\tilde{a}_1$. Clearly, convergence to the true value a_1 is attained (within a given tolerance) at different orders that depend on the value of $\tilde{a}_1$. For instance, for $\tilde{a}_1>0.5$, an order larger than 2 is required. Convergence is faster for smaller $\tilde{a}_1$, that is, as soon as $1/\tilde{a}_1$ becomes large compared to the time scale $\tau=1$. For $\tilde{b}_0$ we obtained a very similar convergence scheme (not shown). In Refs. [3,5], the fitness of low-order expressions for O-U processes results from the particular employment of $\tilde{a}_1\tau<0.5$. However, this may not be the case when dealing with generic empirical data.

III. HIGHER-ORDER COEFFICIENTS

Inserting the Itô-Taylor expansion Eq. (10) into Eq. (4) and performing the average for $k=3,4$, we also computed the finite-τ expansion for $\widetilde{D}_3(x,\tau)$ and $\widetilde{D}_4(x,\tau)$. The resulting expressions are invariant functions of x, namely,

$$\widetilde{D}_3(x,\tau) = -\widetilde{c}_1(\tau) + \widetilde{c}_3(\tau)x^3, \tag{23}$$

$$\widetilde{D}_4(x,\tau) = \widetilde{d}_0(\tau) + \widetilde{d}_2(\tau)x^2 + \widetilde{d}_4(\tau)x^4. \tag{24}$$

For the particular case $b_2=0$, we were able to derive the infinite order expansion for the τ parameters:

$$\widetilde{c}_1(\tau) = b_0 \sum_{j\geq 0} \frac{3^{j+1}-2^{j+1}-1}{2} \frac{[-a_1]^j}{(j+1)!}\tau^j,$$

$$\widetilde{c}_3(\tau) = \sum_{j\geq 0} \frac{3^{j}-2^{j+1}+1}{2} \frac{[-a_1]^{j+1}}{(j+1)!}\tau^j, \tag{25}$$

$$\widetilde{d}_0(\tau) = b_0^2 \sum_{j\geq 0} \frac{4^{j}-2^{j}}{2} \frac{[-a_1]^{j-1}}{(j+1)!}\tau^j,$$

$$\widetilde{d}_2(\tau) = b_0 \sum_{j\geq 0} \frac{2\times 4^{j}-3^{j+1}+1}{2} \frac{[-a_1]^{j}}{(j+1)!}\tau^j,$$

$$\widetilde{d}_4(\tau) = \sum_{j\geq 0} \frac{4^{j}-3^{j+1}-1}{6} \frac{[-a_1]^{j+1}}{(j+1)!}\tau^j. \tag{26}$$

Notice that $\lim_{\tau\to 0}\{\widetilde{c}_1,\widetilde{c}_3,\widetilde{d}_0,\widetilde{d}_2,\widetilde{d}_4\}=0$ as expected and that the relevant parameter for the rate of series convergence is a_1. Summing series (25) and (26) up to infinite order, and recalling that $Z\equiv\exp(-a_1\tau)$, one gets

$$\widetilde{c}_1(\tau) = -\frac{b_0}{a_1}\frac{(1-Z)^2(1+Z)}{2\tau},$$

$$\widetilde{c}_3(\tau) = -\frac{(1-Z)^3}{6\tau}, \tag{27}$$

$$\widetilde{d}_0(\tau) = \frac{b_0^2}{a_1^2}\frac{(1-Z)^2(1+Z)^2}{8\tau},$$

$$\widetilde{d}_2(\tau) = \frac{b_0}{a_1}\frac{(1-Z)^3(1+Z)}{4\tau},$$

$$\widetilde{d}_4(\tau) = \frac{(1-Z)^4}{24\tau}. \tag{28}$$

Figure 4 shows the numerical computation of $\widetilde{D}_3(x,\tau)$ and $\widetilde{D}_4(x,\tau)$ for the same artificial series as in Fig. 1. For comparison, the theoretical estimates at different orders of truncation of the series in Eqs. (25) and (26) are shown. Notice that, although D_3, $D_4=0$ for the diffusion processes considered here; their finite-time counterparts have cubic and quadratic forms. Indeed, the exact theoretical expressions given by Eqs. (27) and (28) reproduce the numerical outcomes validating our approach as furnishing meaningful tests for O-U models. However, for $a_1\gtrsim 1$, rich pictures for the low-order approximations of $\widetilde{D}_3$ and $\widetilde{D}_4$ arise, which hinder the asymptotic estimation.

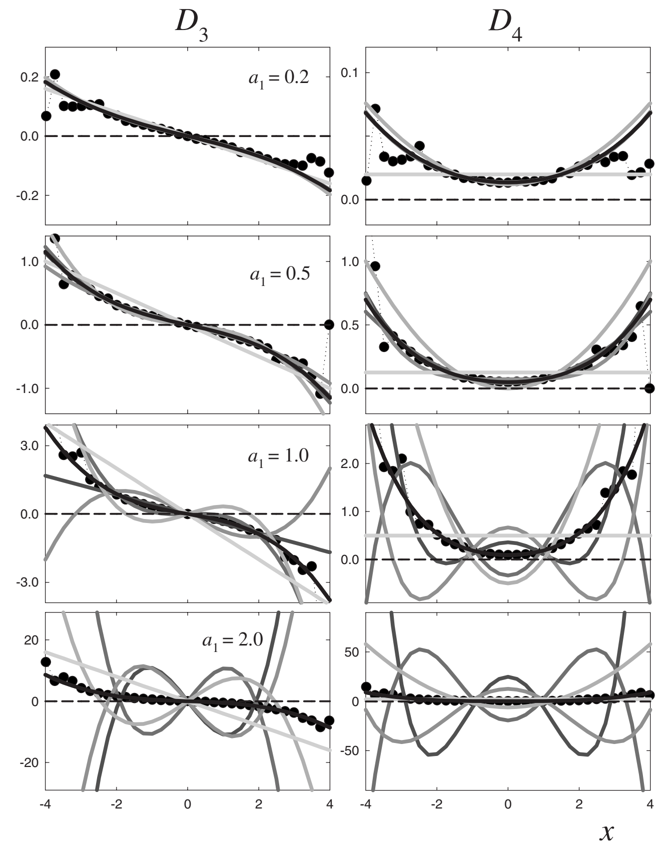

FIG. 4. Third- and fourth-order coefficients for the O-U process. Symbols correspond to the numerical computation for the same artificial series of Fig. 1. Lines represent the coefficients given by Eqs. (23) and (24), using the theoretical τ expansions (25) and (26), at different orders of truncation. Colors as in Fig. 1.

For the general case with $b_2\geq 0$, we computed the third-order τ expansions for $\widetilde{D}_3(x,\tau)$ and $\widetilde{D}_4(x,\tau)$. Each power of τ of order $j\leq 3$ has a prefactor denoted by $\widetilde{D}_3^{(j)}$ and $\widetilde{D}_4^{(j)}$, respectively. We find for $\widetilde{D}_3$:

$$\widetilde{D}_3^{(0)} = 0,$$

$$\widetilde{D}_3^{(1)} = -b_0\alpha x - b_2\alpha x^3,$$

$$\widetilde{D}_3^{(2)} = \frac{1}{6}b_0\alpha(9a_1-16b_2)x - \frac{1}{6}\alpha(a_1^2-13a_1b_2+16b_2^2)x^3,$$

$$\widetilde{D}_3^{(3)} = -\frac{1}{6}b_0\alpha^2(8a_1-13b_2)x + \frac{1}{12}\alpha(3a_1^3-32a_1^2b_2+74a_1b_2^2-52b_2^3)x^3, \tag{29}$$

where $\alpha=a_1-2b_2$. At all orders, $\widetilde{D}_3$ vanishes if $a_1=2b_2$.

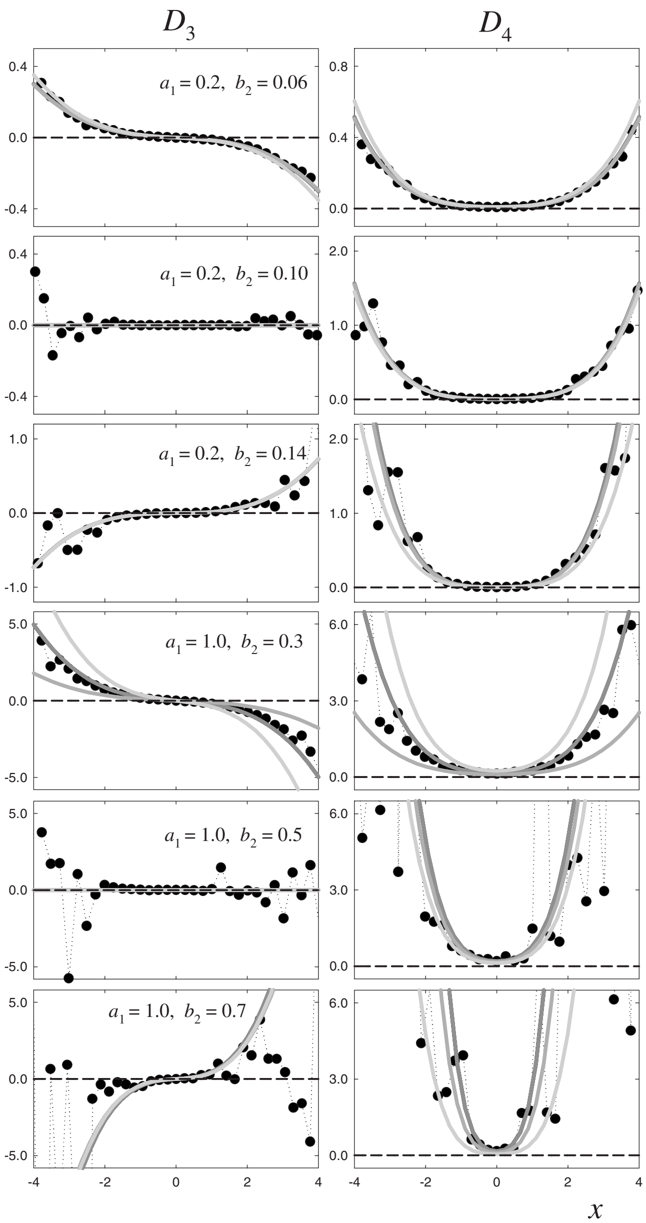

FIG. 5. Third- and fourth-order coefficients for the general process with multiplicative noise. Symbols correspond to the numerical computation for the same artificial series of Fig. 2. Lines represent the coefficients given by Eqs. (23) and (24), using the theoretical τ expansions (29) and (30), at different orders of truncation up to third order. Colors as in previous figures.

For $\widetilde{D}_4$, we have

$$\widetilde{D}_4^{(0)}=0,$$

$$\widetilde{D}_4^{(1)}=\frac{1}{2}b_0^2+b_0b_2x^2+\frac{1}{2}b_2^2x^4,$$

$$\widetilde{D}_4^{(2)}=-\frac{1}{3}b_0^2(3a_1-7b_2)+\frac{1}{6}b_0(3a_1^2-30a_1b_2+52b_2^2)x^2+\frac{1}{6}b_2(3a_1^2-24a_1b_2+38b_2^2)x^4,$$

$$\widetilde{D}_4^{(3)}=\frac{1}{6}b_0^2(7a_1^2-34a_1b_2+43b_2^2)-\frac{1}{6}b_0(6a_1^3-65a_1^2b_2+206a_1b_2^2-206b_2^3)x^2+\frac{1}{24}(a_1^4-36a_1^3b_2+276a_1^2b_2^2-736a_1b_2^3+652b_2^4)x^4. \quad (30)$$

In Fig. 5, we show the numerical computation of $\widetilde{D}_3$ and $\widetilde{D}_4$ for the same artificial series as in Fig. 2. According to our theoretical results, $a_1=2b_2$ is a threshold between positive and negative slopes of $\widetilde{D}_3$ as illustrated in Fig. 5. For comparison, we also show up to the third order theoretical estimates $\Sigma_{j\geq 1}\widetilde{D}_3^{(j)}$ and $\Sigma_{j\geq 1}\widetilde{D}_4^{(j)}$ according to Eqs. (29) and (30).

Coefficients $\widetilde{D}_3$ and $\widetilde{D}_4$ provide further tests of validity of the present model. Third-order estimates furnish suitable forecast of the numerical finite-τ measurements for small enough a_1 as illustrated in Fig. 2. In such cases, once obtained a_1, b_0 from Eqs. (21) and (22), these values can be used in theoretical equations for $\widetilde{D}_3$ and $\widetilde{D}_4$ to check if the corresponding non-null coefficients can be attributed to finite-τ effects.

IV. SUMMARY AND FINAL COMMENTS

Numerical computation of drift and diffusion coefficients from a real time series frequently yields linear and quadratic forms, respectively. In fact, these functional dependencies correspond to the harmonic approximation to a general potential and to a state-dependent diffusion coefficient. For this important class of diffusion models, we have derived exact formulas that connect the empirical discrete-time estimates with the actual values of the parameters of drift and diffusion coefficients. Additionally, we also provided theoretical expressions for higher-order coefficients which serve as a further probe for the validity of this class of diffusion models.

Our results allow us to access the generating stochastic process. A possible procedure to identify it and its parameters can be summarized as follows:

(i) Linear and quadratic curve fitting for D_1 and D_2 furnish the values of $\{\widetilde{a}_1,\widetilde{b}_0,\widetilde{b}_2\}$. For consistency of the linear-quadratic model, it must hold $\widetilde{a}_1\tau<1$ and $\widetilde{a}_1=\widetilde{b}_0+\widetilde{b}_2$ for normalized data (with unitary variance). Let us recall that non-normalized data will simply require rescaling b_0 by the variance.

(ii) Once $\widetilde{a}_1$ and $\widetilde{b}_0$ are known, Eqs. (21) and (22) allow to obtain, exactly, the original parameters a_1 and b_0 (hence also $b_2=a_1-b_0$).

(iii) A further check of the modeling consists in the analysis of higher-order coefficients, e.g., to see whether a non-null D_4 can be attributed to its finite-time corrections.

Let us remark that the detection of a quadratic $\widetilde{D}_2$ does not imply the existence of multiplicative components in the actual process. In fact, by analyzing the O-U process, we found that a low sampling rate would significantly affect the diffusion coefficient estimate by adding an extra quadratic term. Moreover, in this case, estimations of D_3 and D_4 from low-order approximations would lead to results inconsistent with the empirical outcomes.

Notice also that our results are not only valid for SDE (1) with linear-quadratic coefficients but also for the statistically equivalent class of SDEs with additive-multiplicative noises [9,10] given by

$$dX_t = -a_1 X_t dt + \sqrt{2b_0} dW_t + \sqrt{2b_2} X_t dW'_t, \tag{31}$$

where W_t and W'_t are uncorrelated Wiener processes.

The theoretical expressions also allow to quantify the errors induced by a finite sampling rate τ in the numerically estimated coefficients. The analytical results indicate that, in order to grasp the true values of the parameters from the knowledge of the observed ones, the required correction depends strongly on the (hidden) inverse time a_1. Our work shows that one should be careful when applying low-order finite-τ corrections for diffusion models. Furthermore, as shown in Fig. 3, our results provide a criterion, from the knowledge of $\tilde{a}_1$, to determine the required order n, or equivalently, up to which value of τ the respective approximation is reliable.

ACKNOWLEDGMENTS

We acknowledge Brazilian agencies Faperj and CNPq for partial financial support.

APPENDIX

Rewriting Eq. (4) according to the notation introduced in Eq. (10), one has

$$\tau \tilde{D}_1 = \langle \Delta X \rangle = \sum_{\alpha_k} c_{\alpha_k}(D_1, D_2) \langle I_{\alpha_k} \rangle.$$

Only multiple stochastic integrals I_{α_k} such that $\alpha_k = (0,\ldots,0)_k$ have non-null average, being $\langle I_{(0,\ldots,0)_k} \rangle = \tau^k / k!$. From the iterated application of Itô formula to Eq. (8), $c_{(0,\ldots,0)_k}(D_1,D_2) = (L^0)^{k-1} D_1$. These are general results independent of the particular form of D_1 and D_2. Noticing that, from Eq. (6), $L^0 = \partial_t + D_1 \partial_x + D_2 \partial_{xx}$, and that D_1 is time independent and linear in x, then, $c_{\alpha_k} = D_1 (D'_1)^{k-1} = (-a_1)^k x$. Finally,

$$\tilde{D}_1 = \langle \Delta X \rangle / \tau = \sum_{k \ge 1} \frac{1}{k!} (-a_1)^k \tau^{k-1} x,$$

which is of the same functional form of the true D_1 and can be identified with $-\tilde{a}_1 x$, so that

$$\tilde{a}_1 = -\sum_{k \ge 1} \frac{1}{k!} (-a_1)^k \tau^{k-1},$$

which gives Eq. (12).

For the second conditional moment, one has

$$2\tau \tilde{D}_2 = \langle (\Delta X)^2 \rangle = \sum_{\alpha_n, \beta_m} c_{\alpha_n} c_{\beta_m} \langle I_{\alpha_n} I_{\beta_m} \rangle. \tag{A1}$$

From the definition of c_{α_k} in Eq. (10), if $b_2 = 0$ (then D_2 is constant), only two classes of terms in Eq. (A1) are non-null; those with (i) $\alpha_n = (0,\ldots,0)_n$ and $\beta_m = (0,\ldots,0)_m$, and (ii) $\alpha_n = (1,0,\ldots,0)_n$ and $\beta_m = (1,0,\ldots,0)_m$.

In those cases, the products $c_{\alpha_n} c_{\beta_m}$ take the following values: (i) $(D_1)^2 (D'_1)^{k-1} = (-a_1)^{k+1} x^2$ (with $k = m+n-1$); (ii) $2D_2 (D'_1)^k = 2b_0 (-a_1)^k$ (with $k = m+n-2$).

In order to evaluate the averages of products of multiple stochastic integrals, it is useful to recall that $\langle I_{(0,\ldots,0)_n} I_{(0,\ldots,0)_m} \rangle = \tau^{n+m} / (n! m!)$ and that $\langle I_{(1,\ldots,0)_n} I_{(1,\ldots,0)_m} \rangle = \tau^{n+m-1} / [(n+m-1)(n-1)!(m-1)!]$ [4]. Then, summing over all the pairs (n,m) contributing to the order τ^{k+1}, one gets (i) $\Sigma \langle I_{\alpha_n} I_{\beta_m} \rangle / \tau^{k+1} = 1/(k+1)! \Sigma_{n=1}^{k} \binom{k+1}{n} = 2[(2^k - 1)/(k+1)!]$ and (ii) $\Sigma \langle I_{\alpha_n} I_{\beta_m} \rangle / \tau^{k+1} = 1/(k+1)! \Sigma_{j=0}^{k} \binom{k}{j} = 2^k/(k+1)!$.

Finally, from Eq. (A1), we arrive at

$$\tilde{D}_2 = \langle (\Delta X)^2 \rangle / (2\tau) = \frac{1}{2} \sum_{k \ge 0} \left[2 \frac{2^k - 1}{(k+1)!} (-a_1)^{k+1} x^2 + \frac{2^k}{(k+1)!} 2b_0 (-a_1)^k \right] \tau^k, \tag{A2}$$

which can be cast in the form $\tilde{b}_2 x^2 + \tilde{b}_0$, allowing to identify $\tilde{b}_2$ and $\tilde{b}_0$ with functions of the true parameters, as

$$\tilde{b}_0 = \frac{1}{2} \sum_{k \ge 0} \frac{2^k}{(k+1)!} 2b_0 (-a_1)^k \tau^k,$$

$$\tilde{b}_2 = \sum_{k \ge 0} \frac{2^k - 1}{(k+1)!} (-a_1)^{k+1} \tau^k. \tag{A3}$$

For the general case $b_2 \ge 0$, a similar but tricky derivation leads to Eqs. (13) and (14) that generalize expressions (A3).

Proceeding with the third order, products of three multiple integrals appear. For $b_2 = 0$, there are two types of products $I_{\alpha_n} I_{\beta_m} I_{\gamma_l}$ contributing to $\langle (\Delta X)^3 \rangle$:

(i) $\alpha_n = (0,\ldots,0)_n$, $\beta_m = (0,\ldots,0)_m$, and $\gamma_l = (0,\ldots,0)_l$;

(ii) $\alpha_n = (0,\ldots,0)_n$, $\beta_m = (1,0,\ldots,0)_m$, and $\gamma_l = (1,0,\ldots,0)_l$, with prefactors proportional to $(D_1)^3 (D'_1)^{k-2} = (-a_1)^{k+1} x^3$ and to $2D_2 D_1 (D'_1)^{k-1} = 2b_0 (-a_1)^k x$, respectively. Evaluating the averages of products of three multiple stochastic integrals and summing over all the triplets (n,m,l) contributing to the same order in τ, as done for the second-order coefficient, one gets $\tilde{D}_3 = -\tilde{c}_1(\tau) + \tilde{c}_3(\tau) x^3$ as in Eq. (23).

Analogously, at fourth order, considering the relevant products of four multiple integrals, for $b_2 = 0$, three types of contributions appear, yielding $\tilde{D}_4(x,\tau) = \tilde{d}_0(\tau) + \tilde{d}_2(\tau) x^2 + \tilde{d}_4(\tau) x^4$, as in Eq. (24).

Let us recall that the averages of products of n multiple stochastic integrals appearing in the nth-order term of the coefficients can be expressed, in general, as multinomial terms, whose summation over all the products has the form $\mu_1 1^k + \mu_2 2^k + \mu_3 3^k + \cdots \mu_n n^k$ (with rational μ_i) for the order k in τ. Moreover, products of multiple stochastic integrals can be readily simplified by means of useful relations between multiple Itô integrals [4]. Then, although at the cost of increasing the number of indices, the number of factors can be reduced.

[1] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer-Verlag, Berlin, 1984).

[2] R. Friedrich, Ch. Renner, M. Siefert, and J. Peinke, Phys. Rev. Lett. **89**, 149401 (2002).

[3] J. Gottschall and J. Peinke, New J. Phys. **10**, 083034 (2008) and references therein.

[4] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, Berlin, 1992).

[5] P. Sura and J. Barsugli, Phys. Lett. A **305**, 304 (2002).

[6] M. Ragwitz and H. Kantz, Phys. Rev. Lett. **87**, 254501 (2001); **89**, 149402 (2002).

[7] A. Naert, R. Friedrich, and J. Peinke, Phys. Rev. E **56**, 6719 (1997); St. Lück, J. Peinke, and R. Friedrich, Phys. Rev. Lett. **83**, 5495 (1999).

[8] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge, Nature (London) **381**, 767 (1996).

[9] C. Anteneodo and C. Tsallis, J. Math. Phys. **44**, 5194 (2003).

[10] C. Anteneodo and R. Riera, Phys. Rev. E **72**, 026106 (2005).