Noname manuscript No.
(will be inserted by the editor)

Approximate discrete-time schemes for the estimation of diffusion processes from complete observations

J.C. Jimenez

Received: date / Accepted: date

Abstract In this paper, a modification of the conventional approximations to the quasi-maximum likelihood method is introduced for the parameter estimation of diffusion processes from discrete observations. This is based on a convergent approximation to the first two conditional moments of the diffusion process through discrete-time schemes. It is shown that, for finite samples, the resulting approximate estimators converge to the quasi-maximum likelihood one when the error between the discrete-time approximation and the diffusion process decreases. For an increasing number of observations, the approximate estimators are asymptotically normal distributed and their bias decreases when the mentioned error does it. A simulation study is provided to illustrate the performance of the new estimators. The results show that, with respect to the conventional approximate estimators, the new ones significantly enhance the parameter estimation of the test equations. The proposed estimators are intended for the recurrent practical situation where a nonlinear stochastic system should be identified from a reduced number of complete observations distant in time.

Keywords system identification · quasi-maximum likelihood estimator · diffusion process · stochastic differential equation · numerical integrator · local linear approximation

1 Introduction

The statistical inference for diffusion processes described by Stochastic Differential Equations (SDEs) is currently a subject of intensive researches. A basic difficulty of this statistical problem is that, except for a few simple examples, the joint distribution of the discrete-time observations of the process has unknown closed-form. To overcome this, a number of estimators based on analytical and simulated approximations have been developed during more than three decades. Such methods are the focus of a growing literature. See, for instance, the review papers by Bibby and Sorensen (1996), Prakasa-Rao (1999), Nielsen et al. (2000) and Jimenez et al. (2006).

In particular, the present paper deals with the class of quasi-maximum likelihood (QML) estimators for the parameter estimation of SDEs given a time series of complete observations. These are the estimators obtained by maximizing a normal log-likelihood function when the assumption of normality is not satisfied and all the components of the diffusion process are discretely observed. The simplest approximations to this class of estimators are derived, for SDEs with additive noise, from the likelihood of the discrete-time process defined by a numerical integrator. Typically, they are derived from the Euler-Maruyama scheme (Prakasa Rao 1983, Yoshida 1992, Florens-Zmirou 1989) or from the Local Linearization (LL) one (Ozaki 1985, 1992; Shoji & Ozaki 1997, 1998). It have been demonstrated that, when the distance between two consecutive observations remains fixed, these approximate QML estimators are asymptotically biased

Instituto de Cibernética, Matemática y Física
Departamento de Matemática Interdisciplinaria
Calle 15, no. 551, Vedado, La Habana, Cuba
e-mail: jcarlos@icimaf.cu

when the number of observations increases (Florens-Zmirou 1989). However, a number of comparative studies among different estimation methods have shown that, in practical situations in which the distance between observations is small enough, the approximate QML estimators based on LL integrators display the better performance due to their simplicity, computational efficiency and negligible bias (see, e.g., Shoji & Ozaki 1997, Durham & Gallant 2002, Singer 2002, Hurn et al. 2007). Therefore, any modification to these approximate QML methods that yields a bias reduction will be useful. In this direction, two methods have early been proposed. For the estimators based on the Euler-Maruyama integrator, Clement (1995) introduced a correction to the bias by means of simulations. Whereas, on the basis of Taylor expansions of the first two conditional moments of the discrete-time process, Kessler (1997) archives similar results. Depending of the specific SDE to be estimated, the first method could be computationally time demanding, while the second one could be affected by numerical instabilities resulting from a high order Taylor expansion. More recently, Huang (2011) proposed new estimators based on high-order numerical integrators, which improve the accuracy of the approximation for the first two conditional moments and can be straightforward applied to SDEs with multiplicative noise.

A common feature of the approximate QML methods mentioned above is that, once the observations are given, the error between the approximate and the exact moments of the diffusion is fixed and completely determined by the distance between observations. Clearly, this fixes the bias of the QML estimation for finite samples and obstructs its asymptotic correction when the number of observations increases.

In this paper, an alternative modification of the conventional approximations to the QML estimator for diffusion processes is introduced, which is oriented to reduce and control the estimation bias. This is based on a recursive computation of the first two conditional moments of discrete-time approximations converging to the diffusion process between two consecutive observations. It is shown that, for finite samples, the resulting approximate estimators converge to the exact QML estimator when the error between the discrete-time approximation and the diffusion process decreases. For an increasing number of observations, the approximate estimators are asymptotically normal distributed and their bias decreases when the above mentioned error does it. As a particular instance, the approximate QML estimators designed with the well-known Local Linear approximations for SDEs are presented. Their convergence, practical algorithms and performance in simulations are also considered in detail. The simulations show that, with respect to the conventional QML estimators, the new approximate estimators significantly enhance the parameter estimation of the test equations given a reduced number of discrete observations distant in time, which is a typical situation in many practical inference problems.

The paper is organized as follows. In section 2, basic notations and definitions are presented. In section 3, the new approximate estimators are defined and some of their properties are studied. As example, the order-β QML estimator based on the Local Linearization schemes is presented in Section 4, as well as algorithms for its practical implementation. In the last section, the performance of the new estimators is illustrated with various examples.

2 Notation and preliminary

Let $(\Omega, \mathcal{F}, P)$ be the underlying complete probability space and $\{\mathcal{F}_t,\ t \geq t_0\}$ be an increasing right continuous family of complete sub σ-algebras of $\mathcal{F}$. Consider a d-dimensional diffusion process $\mathbf{x}$ defined by the following stochastic differential equation

$$d\mathbf{x}(t) = \mathbf{f}(t, \mathbf{x}(t); \boldsymbol{\theta})dt + \sum_{i=1}^{m} \mathbf{g}_i(t, \mathbf{x}(t); \boldsymbol{\theta})d\mathbf{w}^i(t) \tag{1}$$

for $t \geq t_0 \in \mathbb{R}$, where $\mathbf{f}$ and $\mathbf{g}_i$ are differentiable functions, $\mathbf{w} = (\mathbf{w}^1, .., \mathbf{w}^m)$ is an m-dimensional $\mathcal{F}_t$-adapted standard Wiener process, $\boldsymbol{\theta} \in \mathcal{D}_\theta$ is a vector of parameters, and $\mathcal{D}_\theta \subset \mathbb{R}^p$ is a compact set. Linear growth, uniform Lipschitz and smoothness conditions on the functions $\mathbf{f}$ and $\mathbf{g}_i$ that ensure the existence and uniqueness of a strong solution of (1) with bounded moments are assumed for all $\boldsymbol{\theta} \in \mathcal{D}_\theta$.

Denote by $\mathbf{z}$ the diffusion process defined by (1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0 \in \mathcal{D}_\theta$, and suppose that M observations of the process $\mathbf{z}$ on an increasing sequence of time instants $\{t\}_M = \{t_k : t_k < t_{k+1},\ k = 0, 1, .., M-1\}$

are given. More precisely, denote by $\mathbf{z}_k$ the observation of the process $\mathbf{z}$ at t_k for all $t_k \in \{t\}_M$ and by $Z = \{\mathbf{z}_0, .., \mathbf{z}_{M-1}\}$ the sequence of these observations.

The inference problem to be consider here is the estimation of the parameter $\boldsymbol{\theta}_0$ of the SDE (1) given the time series Z. In particular, let us consider the quasi-maximum likelihood estimator defined by

$$\widehat{\boldsymbol{\theta}}_M = \arg\{\min_{\boldsymbol{\theta}} U_M(\boldsymbol{\theta}, Z)\} \tag{2}$$

where

$$U_M(\boldsymbol{\theta}, Z) = (M-1)\ln(2\pi) + \sum_{k=1}^{M-1} \ln(\det(\boldsymbol{\Sigma}_k)) + (\mathbf{z}_k - \boldsymbol{\mu}_k)^\intercal (\boldsymbol{\Sigma}_k)^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_k),$$

and $\boldsymbol{\mu}_k = E(\mathbf{x}(t_k)|\mathbf{z}_{k-1})$ and $\boldsymbol{\Sigma}_k = E(\mathbf{x}(t_k)\mathbf{x}^\intercal(t_k)|\mathbf{z}_{k-1}) - \boldsymbol{\mu}_k\boldsymbol{\mu}_k^\intercal$ denote the conditional mean and variance of the diffusion process $\mathbf{x}$ at t_k given $\mathbf{z}_{k-1}$, for all $t_{k-1}, t_k \in \{t\}_M$ and $\boldsymbol{\theta} \in \mathcal{D}_\theta$. Because the first two conditional moments of $\mathbf{x}$ are correctly specified, the score of the normal log-likelihood satisfies the martingale difference property, and so the QML estimator (2) is consistent and has an asymptotically normal distribution. See Bollerslev & Wooldridge (1992) and Wooldridge (1994) for ergodic and no ergodic processes, respectively.

In general, since the conditional mean and variance of equation (1) have not explicit formulas, approximations to them are needed. If $\widetilde{\boldsymbol{\mu}}_k$ and $\widetilde{\boldsymbol{\Sigma}}_k$ are approximations to $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$, then the estimator

$$\widehat{\boldsymbol{\vartheta}}_M = \arg\{\min_{\boldsymbol{\theta}} \widetilde{U}_M(\boldsymbol{\theta}, Z)\},$$

with

$$\widetilde{U}_M(\boldsymbol{\theta}, Z) = (M-1)\ln(2\pi) + \sum_{k=1}^{M-1} \ln(\det(\widetilde{\boldsymbol{\Sigma}}_k)) + (\mathbf{z}_k - \widetilde{\boldsymbol{\mu}}_k)^\intercal (\widetilde{\boldsymbol{\Sigma}}_k)^{-1} (\mathbf{z}_k - \widetilde{\boldsymbol{\mu}}_k)$$

provides an approximation to the quasi-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_M$.

Approximate estimators of this type have early been considered in a number of papers (Prakasa Rao 1983; Florens-Zmirou 1989; Yoshida 1992; Ozaki 1985, 1992; Shoji & Ozaki 1997, 1998) and recently in Huang (2011). In all of them, the approximate mean $\widetilde{\boldsymbol{\mu}}_k = E(\mathbf{y}_k|\mathbf{z}_{k-1})$ and variance $\widetilde{\boldsymbol{\Sigma}}_k = E((\mathbf{y}_k - \widetilde{\boldsymbol{\mu}}_k)(\mathbf{y}_k - \widetilde{\boldsymbol{\mu}}_k)^\intercal|\mathbf{z}_{k-1})$ are derived from a discrete-time scheme $\mathbf{y}_k = \mathbf{z}_{k-1} + \phi(t_{k-1}, \mathbf{z}_{k-1}, t_k - t_{k-1})$ that approximate to $\mathbf{x}(t_k)$ in just one step of size $t_k - t_{k-1}$ from the observation $\mathbf{z}_{k-1}$. Indistinctly, these estimators are called pseudo-likelihood estimators, or minimum contrast estimators or prediction error estimators depending of the inferential considerations that want to be emphasized. It has been proved (Florens-Zmirou, 1989) that, for the time partition $\{t\}_M = \{t_k = k\boldsymbol{\delta} : k = 0, 1, \cdots, M-1\}$ with $\boldsymbol{\delta} > 0$ fixed, these estimators are biased as $M\boldsymbol{\delta} \to \infty$. Contrary, they are asymptotically unbiased on the time partition $\{t\}_M = \{t_k = k\boldsymbol{\delta}_M : k = 0, 1, \cdots, M-1\}$ in the case that $M\boldsymbol{\delta}_M \to \infty$, but with $M\boldsymbol{\delta}_M^3 \to 0$ (or more accurately with $M\boldsymbol{\delta}_M^2 \to 0$ as in Yoshida, 1992). However, last restriction on M and $\boldsymbol{\delta}_M$ imposes too strong relation among the number of observations and the time distance between them, which is very inconvenient from a practical viewpoint. Further note that, once the data Z are given (and so the time partition $\{t\}_M$ is specified), the error between $\mathbf{y}_k$ and $\mathbf{x}(t_k)$ is completely settled by $t_k - t_{k-1}$ and can not be reduced. In this way, the difference between the approximate quasi-maximum likelihood estimator $\widehat{\boldsymbol{\vartheta}}_M$ and the exact one $\widehat{\boldsymbol{\theta}}_M$ can not be reduced neither.

Denote by $\mathcal{C}_P^l(\mathbb{R}^d, \mathbb{R})$ the space of l time continuously differentiable functions $g : \mathbb{R}^d \to \mathbb{R}$ for which g and all its partial derivatives up to order l have polynomial growth.

3 Order-β quasi-maximum likelihood estimator

Let $(\tau)_{h>0} = \{\tau_n : \tau_{n+1} - \tau_n \leq h, \ n = 0, 1, \ldots, N\}$ be a time discretization of $[t_0, t_{M-1}]$ such that $(\tau)_h \supset \{t\}_M$, and $\mathbf{y}_n$ be the approximate value of $\mathbf{x}(\tau_n)$ obtained from a discretization of the equation (1) for all $\tau_n \in (\tau)_h$. Let us consider the continuous time approximation $\mathbf{y} = \{\mathbf{y}(t), t \in [t_0, t_{M-1}] : \mathbf{y}(\tau_n) = \mathbf{y}_n$ for all $\tau_n \in (\tau)_h\}$ of $\mathbf{x}$ with initial conditions

$$E\left(\mathbf{y}(t_0)\middle|\mathcal{F}_{t_0}\right) = E\left(\mathbf{x}(t_0)\middle|\mathcal{F}_{t_0}\right) \quad \text{and} \quad E\left(\mathbf{y}(t_0)\mathbf{y}^\intercal(t_0)\middle|\mathcal{F}_{t_0}\right) = E\left(\mathbf{x}(t_0)\mathbf{x}^\intercal(t_0)\middle|\mathcal{F}_{t_0}\right);$$

satisfying the bound condition

$$E\left(\left|\mathbf{y}(t)\right|^{2q}\left|\mathcal{F}_{t_k}\right.\right) \leq L \tag{3}$$

for all $t \in [t_k, t_{k+1}]$; and the weak convergence criteria

$$\sup_{t_k \leq t \leq t_{k+1}} \left|E\left(g(\mathbf{x}(t))\left|\mathcal{F}_{t_k}\right.\right) - E\left(g(\mathbf{y}(t))\left|\mathcal{F}_{t_k}\right.\right)\right| \leq L_k h^{\beta} \tag{4}$$

for all $t_k, t_{k+1} \in \{t\}_M$ and $\boldsymbol{\theta} \in \mathcal{D}_\theta$, where $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$, L and L_k are positive constants, $\beta \in \mathbb{N}_+$, and $q = 1, 2....$ The process $\mathbf{y}$ defined in this way is typically called order-β approximation to $\mathbf{x}$ in weak sense (Kloeden & Platen, 1999). In addition, the second conditional moment of $\mathbf{y}$ is assumed to be positive definite and continuous for all $\boldsymbol{\theta} \in \mathcal{D}_\theta$.

When an order-β approximation to the solution of equation (1) is chosen, the following approximate quasi-maximum likelihood estimator can be naturally defined.

Definition 1 Given a time series Z of M observations of the SDE (1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ on $\{t\}_M$, the order-β quasi-maximum likelihood estimator for the parameters of (1) is defined by

$$\widehat{\boldsymbol{\theta}}_M(h) = \arg\{\min_{\boldsymbol{\theta}} U_{M,h}(\boldsymbol{\theta}, Z)\}, \tag{5}$$

where

$$U_{M,h}(\boldsymbol{\theta}, Z) = (M-1)\ln(2\pi) + \sum_{k=1}^{M-1} \ln(\det(\boldsymbol{\Sigma}_{h,k})) + (\mathbf{z}_k - \boldsymbol{\mu}_{h,k})^{\intercal}(\boldsymbol{\Sigma}_{h,k})^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_{h,k}),$$

$\mu_{h,k} = E(\mathbf{y}(t_k)|\mathbf{z}_{k-1})$, $\boldsymbol{\Sigma}_{h,k} = E(\mathbf{y}(t_k)\mathbf{y}^{\intercal}(t_k)|\mathbf{z}_{k-1}) - \boldsymbol{\mu}_{h,k}\boldsymbol{\mu}_{h,k}^{\intercal}$, $\mathbf{y}$ is an order-β approximation to the solution of (1) in weak sense such that $E(\mathbf{y}(t_k)|\mathbf{z}_k) = \mathbf{z}_k$ and $E(\mathbf{y}(t_k)\mathbf{y}^{\intercal}(t_k)|\mathbf{z}_k) = \mathbf{z}_k\mathbf{z}_k^{\intercal}$ for all $t_k \in \{t\}_M$, and h is the maximum stepsize of the time discretization $(\tau)_h \supset \{t\}_M$ associated to $\mathbf{y}$.

In principle, according to the above definition, any kind of approximation $\mathbf{y}$ converging to $\mathbf{x}$ in a weak sense can be used to construct an approximate order-β quasi-maximum likelihood estimator, e.g., those considered in Kloeden & Platen (1999). In this way, the Euler-Maruyama, the Local Linearization and any high order numerical scheme for SDEs might be used as well, but the approximations $\mu_{h,k}$ and $\boldsymbol{\Sigma}_{h,k}$ will be now derived from the conditional moments of the numerical scheme after various iterations with stepsizes lower than $t_k - t_{k-1}$. Note that, when $(\tau)_h \equiv \{t\}_M$, the so defined order-β quasi-maximum likelihood estimator reduces to the corresponding approximate quasi-maximum likelihood estimator mentioned in Section 2. That is, to one of those considered in Prakasa Rao (1983), Yoshida (1992), Florens-Zmirou (1989), Ozaki (1985,1992), Shoji & Ozaki (1997,1998), or Huang (2011).

Note that the goodness of the approximation $\mathbf{y}$ to $\mathbf{x}$ is measured (in weak sense) by the left hand side of (4). Thus, the inequality (4) gives a bound for the errors of the approximation $\mathbf{y}$ to $\mathbf{x}$, for all $t \in [t_k, t_{k+1}]$ and all pair of consecutive observations $t_k, t_{k+1} \in \{t\}_M$. Moreover, this inequality states the convergence (in weak sense and with rate β) of the approximation $\mathbf{y}$ to $\mathbf{x}$ as the maximum stepsize h of the time discretization $(\tau)_h \supset \{t\}_M$ goes to zero. Clearly this includes, as particular case, the convergence of the first two conditional moments of $\mathbf{y}$ to those of $\mathbf{x}$. Since the approximate estimator in Definition 1 is designed in terms of the first two conditional moments of the approximation $\mathbf{y}$, the weak convergence of $\mathbf{y}$ to $\mathbf{x}$ should imply the convergence of the approximate QML estimator to the exact one and the similarity of their asymptotic properties, as h goes to zero. Next results deal with these matters.

3.1 Convergence

For a finite sample Z of M observation of (1), the following convergence results are useful.

Theorem 1 *Let Z be a time series of M observations of the SDE (1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ on the time partition $\{t\}_M$. Let $\widehat{\boldsymbol{\theta}}_M$ and $\widehat{\boldsymbol{\theta}}_M(h)$ be, respectively, the quasi-maximum likelihood and an order-β quasi-maximum likelihood estimator for the parameters of (1) given Z. Then*

$$\left|\widehat{\boldsymbol{\theta}}_M(h) - \widehat{\boldsymbol{\theta}}_M\right| \to 0$$

as $h \to 0$. Moreover,

$$E(\left|\widehat{\boldsymbol{\theta}}_M(h) - \widehat{\boldsymbol{\theta}}_M\right|) \to 0$$

as $h \to 0$, where the expectation is with respect to the measure on the underlying probability space generating the realizations of the SDE (1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proof Defining $\Delta\boldsymbol{\Sigma}_{h,k} = \boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}_{h,k}$, it follows that

$$\begin{aligned}\det(\boldsymbol{\Sigma}_{h,k}) &= \det(\boldsymbol{\Sigma}_k - \Delta\boldsymbol{\Sigma}_{h,k}) \\ &= \det(\boldsymbol{\Sigma}_k)\det(\mathbf{I} - \boldsymbol{\Sigma}_k^{-1}\Delta\boldsymbol{\Sigma}_{h,k})\end{aligned} \tag{6}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{h,k}^{-1} &= (\boldsymbol{\Sigma}_k - \Delta\boldsymbol{\Sigma}_{h,k})^{-1} \\ &= \boldsymbol{\Sigma}_k^{-1} + \boldsymbol{\Sigma}_k^{-1}\Delta\boldsymbol{\Sigma}_{h,k}(\mathbf{I} - \boldsymbol{\Sigma}_k^{-1}\Delta\boldsymbol{\Sigma}_{h,k})^{-1}\boldsymbol{\Sigma}_k^{-1}.\end{aligned} \tag{7}$$

By using these two identities and the identity

$$\begin{aligned}(\mathbf{z}_k - \boldsymbol{\mu}_{h,k})^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_{h,k}) &= (\mathbf{z}_k - \boldsymbol{\mu}_k)^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_k) \\ &+(\mathbf{z}_k - \boldsymbol{\mu}_k)^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k}) \\ &+(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k})^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_k) \\ &+(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k})^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k})\end{aligned} \tag{8}$$

it is obtained that

$$U_{M,h}(\boldsymbol{\theta}, Z) = U_M(\boldsymbol{\theta}, Z) + R_{M,h}(\boldsymbol{\theta}), \tag{9}$$

where U_M and $U_{M,h}$ are defined in (2) and (5), respectively, and

$$\begin{aligned}R_{M,h}(\boldsymbol{\theta}) = \sum_{k=1}^{M-1} &\ln(\det(\mathbf{I} - \boldsymbol{\Sigma}_k^{-1}\Delta\boldsymbol{\Sigma}_{h,k})) + (\mathbf{z}_k - \boldsymbol{\mu}_k)^\intercal\mathbf{M}_{h,k}(\mathbf{z}_k - \boldsymbol{\mu}_k) \\ &+(\mathbf{z}_k - \boldsymbol{\mu}_k)^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k}) + (\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k})^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_k) \\ &+(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k})^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k})\end{aligned}$$

with $\mathbf{M}_{h,k} = \boldsymbol{\Sigma}_k^{-1}\Delta\boldsymbol{\Sigma}_{h,k}(\mathbf{I} - \boldsymbol{\Sigma}_k^{-1}\Delta\boldsymbol{\Sigma}_{h,k})^{-1}\boldsymbol{\Sigma}_k^{-1}$.

For the functions $g(\mathbf{x}(t)) = \mathbf{x}^i(t)$ and $g(\mathbf{x}(t)) = \mathbf{x}^i(t)\mathbf{x}^j(t)$ belonging to the function space $\mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$, for all $i, j = 1..d$, condition (4) directly implies that

$$|E(\mathbf{x}(t_k)|\mathbf{z}_{k-1}) - E(\mathbf{y}(t_k)|\mathbf{z}_{k-1})| \le \sqrt{d}L_{k-1}h^\beta \tag{10}$$

and

$$|E(\mathbf{x}(t_k)\mathbf{x}^\intercal(t_k)|\mathbf{z}_{k-1}) - E(\mathbf{y}(t_k)\mathbf{y}^\intercal(t_k)|\mathbf{z}_{k-1})| \le dL_{k-1}h^\beta. \tag{11}$$

From this and the finite bound for the conditional mean of $\mathbf{x}$ and $\mathbf{y}$, it is obtained that

$$\left|\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k}\right| \to \mathbf{0} \qquad \text{and} \qquad |\boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}_{h,k}| \to \mathbf{0}$$

as $h \to 0$ for all $\boldsymbol{\theta} \in \mathcal{D}_\theta$ and $k = 1, .., M-1$. This and the finite bound for the first two conditional moments of $\mathbf{x}$ and $\mathbf{y}$ imply that $R_{M,h}(\boldsymbol{\theta}) \to \mathbf{0}$ as well with h. From this and (9),

$$\left|\widehat{\boldsymbol{\theta}}_M(h) - \widehat{\boldsymbol{\theta}}_M\right| = \left|\arg\{\min_{\boldsymbol{\theta}}\{U_M(\boldsymbol{\theta}, Z) + R_{M,h}(\boldsymbol{\theta})\}\} - \arg\{\min_{\boldsymbol{\theta}} U_M(\boldsymbol{\theta}, Z)\}\right| \to 0 \tag{12}$$

as $h \to 0$, which implies the first assertion of the theorem.

On the other hand, since the value of the constant L_{k-1} in (10) and (11) does not depend of a specific realization of the SDE (1), from these inequalities follows that

$$E(|E(\mathbf{x}(t_k)|\mathbf{z}_{k-1}) - E(\mathbf{y}(t_k)|\mathbf{z}_{k-1})|) \leq \sqrt{d} L_{k-1} h^{\beta}$$

and

$$E(|E(\mathbf{x}(t_k)\mathbf{x}^{\intercal}(t_k)|\mathbf{z}_{k-1}) - E(\mathbf{y}(t_k)\mathbf{y}^{\intercal}(t_k)|\mathbf{z}_{k-1})|) \leq d L_{k-1} h^{\beta},$$

where the new expectation here is with respect to the measure on the underlying probability space generating the realizations of the SDE (1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. From this and (12) follows that $E(\left|\widehat{\boldsymbol{\theta}}_M(h) - \widehat{\boldsymbol{\theta}}_M\right|) \to 0$ as $h \to 0$, which concludes the proof.

The the first assertion of this theorem states that, for each given data Z, the order-β QML estimator $\widehat{\boldsymbol{\theta}}_M(h)$ converges to the exact one $\widehat{\boldsymbol{\theta}}_M$ as h goes to zero. Because h controls the weak convergence criteria (4) is then clear that the order-β QML estimator (5) converges to the exact one (2) when the error (in weak sense) of the order-β approximation $\mathbf{y}$ to $\mathbf{x}$ decreases. On the other hand, the second assertion implies that the average of the errors $\left|\widehat{\boldsymbol{\theta}}_M(h) - \widehat{\boldsymbol{\theta}}_M\right|$ corresponding to different realizations of (1) decreases when h does.

Next theorem deals with error between the averages of the estimators $\widehat{\boldsymbol{\theta}}_M(h)$ and $\widehat{\boldsymbol{\theta}}_M$ computed for different realizations of the SDE.

Theorem 2 *Let Z be a time series of M observations of the SDE (1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ on the time partition $\{t\}_M$. Let $\widehat{\boldsymbol{\theta}}_M$ and $\widehat{\boldsymbol{\theta}}_M(h)$ be, respectively, the quasi-maximum likelihood and an order-β quasi-maximum likelihood estimator for the parameters of (1) given Z. Then,*

$$\left|E(\widehat{\boldsymbol{\theta}}_M(h)) - E(\widehat{\boldsymbol{\theta}}_M)\right| \to 0$$

as $h \to 0$, where the expectation is with respect to the measure on the underlying probability space generating the realizations of the SDE (1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proof Trivially,

$$\begin{aligned} \left|E(\widehat{\boldsymbol{\theta}}_M(h)) - E(\widehat{\boldsymbol{\theta}}_M)\right| &= \left|E(\widehat{\boldsymbol{\theta}}_M(h) - \widehat{\boldsymbol{\theta}}_M)\right| \\ &\leq E(\left|\widehat{\boldsymbol{\theta}}_M(h) - \widehat{\boldsymbol{\theta}}_M\right|), \end{aligned}$$

where the expectation here is taken with respect to the measure on the underlying probability space generating the realizations of the SDE (1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. From this and the second assertion of Theorem 1, the proof is completed.

Here, it is worth to remak that the conventional approximate QML estimators mentioned in Section 2 do not have the desired convergence properties stated in the theorems above for the order-β QML estimator. Further note that, either in Definition 1 nor in Theorems 1 and 2 some restriction on the time partition $\{t\}_M$ for the data has been assumed. Thus, there are not specific constraints about the time distance between two consecutive observations, which allows the application of the order-β QML estimator in a variety of practical problems with a reduced number of not close observations in time, with sequential random measurements, or with multiple missing data. Neither there are restrictions on the time discretization $(\tau)_h \supset \{t\}_M$ on which the order-β QML estimator is defined. Thus, $(\tau)_h$ can be set by the user by taking into account some specifications or previous knowledge on the inference problem under consideration, or automatically designed by an adaptive strategy as it will be shown in the section concerning the numerical simulations.

3.2 Asymptotic properties

In this section, asymptotic properties of the approximate quasi-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_M(h)$ will be studied by using a general result obtained in Ljung and Caines (1979) for prediction error estimators. According to that, the relation between the estimator $\widehat{\boldsymbol{\theta}}_M(h)$ and the global minimum $\boldsymbol{\theta}_M^*$ of the function

$$W_M(\boldsymbol{\theta}) = E(U_M(\boldsymbol{\theta}, Z)) \text{ with } \boldsymbol{\theta} \in \mathcal{D}_\theta \tag{13}$$

should be considered, where U_M is defined in (2) and the expectation is taken with respect to the measure on the underlying probability space generating the realizations of the SDE (1). Here, it is worth to remark that $\boldsymbol{\theta}_M^*$ is not an estimator of $\boldsymbol{\theta}$ since the function W_M does not depend of a given data Z. In fact, $\boldsymbol{\theta}_M^*$ indexes the best predictor, in the sense that the average prediction error loss function W_M is minimized at this parameter (Ljung & Caines, 1979).

In what follows, regularity conditions for the unique identifiability of the SDE (1) are assumed, which are typically satisfied by stationary and ergodic diffusion processes (see, e.g., Bollerslev & Wooldridge (1992) or Ljung & Caines (1979)).

Lemma 1 *If $\boldsymbol{\Sigma}_k$ is positive definite for all $k = 1, .., M-1$, then the function $W_M(\boldsymbol{\theta})$ defined in (13) has an unique minimum and*

$$\arg\{\min_{\boldsymbol{\theta}\in\mathcal{D}_\theta} W_M(\boldsymbol{\theta})\} = \boldsymbol{\theta}_0. \tag{14}$$

Proof Since $\boldsymbol{\Sigma}_k$ is positive definite for all $k = 1, .., M-1$, Lemma A.2 in Bollerslev & Wooldridge (1992) ensures that $\boldsymbol{\theta}_0$ is the unique minimum of the function

$$l_k(\boldsymbol{\theta}) = E(\ln(\det(\boldsymbol{\Sigma}_k)) + (\mathbf{z}_k - \boldsymbol{\mu}_k)^\intercal(\boldsymbol{\Sigma}_k)^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_k)|\mathbf{z}_{k-1})$$

on $\mathcal{D}_\theta$ for all k. Consequently and under the assumed unique identifiability of the SDE (1), $\boldsymbol{\theta}_0$ is then the unique minimum of

$$W_M(\boldsymbol{\theta}) = (M-1)\ln(2\pi) + \sum_{k=1}^{M-1} E(l_k(\boldsymbol{\theta}))$$

on $\mathcal{D}_\theta$.

Here, it is worth to remark that the result of this Lemma is restricted to the QML estimator (2) for SDEs. However, for other types of stochastic processes, a similar result can be found in the proof of Theorem 2.1 of Bollerslev & Wooldridge (1992) concerning the asymptotic properties of the QML estimator under more general framework.

Denote by $U'_{M,h}$ the derivative of $U_{M,h}$ with respect to $\boldsymbol{\theta}$, and by W''_M the second derivative of W_M with respect to $\boldsymbol{\theta}$.

Theorem 3 *Let Z be a time series of M observations of the SDE (1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ on the time partition $\{t\}_M$. Let $\widehat{\boldsymbol{\theta}}_M(h)$ be an order-β quasi-maximum likelihood estimator for the parameters of (1) given Z. Then*

$$\widehat{\boldsymbol{\theta}}_M(h) - \boldsymbol{\theta}_0 \to \Delta\boldsymbol{\theta}_M(h) \tag{15}$$

w.p.1 as $M \to \infty$, where $\Delta\boldsymbol{\theta}_M(h) \to 0$ as $h \to 0$. Moreover, if for some $M_0 \in \mathbb{N}$ there exists $\epsilon > 0$ such that

$$W''_M(\boldsymbol{\theta}) > \epsilon\mathbf{I} \quad \textit{and} \quad \mathbf{H}_{M,h}(\boldsymbol{\theta}) = ME(U'_{M,h}(\boldsymbol{\theta}, Z)(U'_{M,h}(\boldsymbol{\theta}, Z))^\intercal) > \epsilon\mathbf{I} \tag{16}$$

for all $M > M_0$ and $\boldsymbol{\theta} \in \mathcal{D}_\theta$, then

$$\sqrt{M}\mathbf{P}_{M,h}^{-1/2}(\widehat{\boldsymbol{\theta}}_M(h) - \boldsymbol{\theta}_0) \sim \mathcal{N}(\Delta\boldsymbol{\theta}_M(h), \mathbf{I}) \tag{17}$$

as $M \to \infty$, where $\mathbf{P}_{M,h} = (W''_M(\boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}_M(h)))^{-1}\mathbf{H}_{M,h}(\boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}_M(h))(W''_M(\boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}_M(h)))^{-1} + \Delta\mathbf{P}_{M,h}$ with $\Delta\mathbf{P}_{M,h} \to \mathbf{0}$ as $h \to 0$.

Proof Let $W_{M,h}(\boldsymbol{\theta}) = E(U_{M,h}(\boldsymbol{\theta}, Z))$ and $\boldsymbol{\alpha}_M(h) = \arg\{\min_{\boldsymbol{\theta}\in\mathcal{D}_\theta} W_{M,h}(\boldsymbol{\theta})\}$, where $U_{M,h}$ is defined in (5).

For a h fixed, Theorem 1 in Ljung & Caines (1979) implies that

$$\widehat{\boldsymbol{\theta}}_M(h) - \boldsymbol{\alpha}_M(h) \to 0 \tag{18}$$

w.p.1 as $M \to \infty$; and

$$\sqrt{M}\mathbf{P}_{M,h}^{-1/2}(\boldsymbol{\alpha}_M(h))(\widehat{\boldsymbol{\theta}}_M(h) - \boldsymbol{\alpha}_M(h)) \sim \mathcal{N}(0, \mathbf{I}) \tag{19}$$

as $M \to \infty$, where

$$\mathbf{P}_{M,h}(\boldsymbol{\theta}) = (W''_{M,h}(\boldsymbol{\theta}))^{-1}\ \mathbf{H}_{M,h}(\boldsymbol{\theta})\ (W''_{M,h}(\boldsymbol{\theta}))^{-1}$$

with $\mathbf{H}_{M,h}(\boldsymbol{\theta}) = ME(U'_{M,h}(\boldsymbol{\theta}, Z)(U'_{M,h}(\boldsymbol{\theta}, Z))^\intercal)$.

By using the identities (6)-(8), the function

$$W_{M,h}(\boldsymbol{\theta}) = (M-1)\ln(2\pi) + \sum_{k=1}^{M-1} E(\ln(\det(\boldsymbol{\Sigma}_{h,k})) + (\mathbf{z}_k - \boldsymbol{\mu}_{h,k})^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_{h,k}))$$

can be written as

$$W_{M,h}(\boldsymbol{\theta}) = W_M(\boldsymbol{\theta}) + E(R_{M,h}(\boldsymbol{\theta})), \tag{20}$$

where W_M is defined in (13) and

$$\begin{aligned} R_{M,h}(\boldsymbol{\theta}) = \sum_{k=1}^{M-1} & E(\ln(\det(\mathbf{I} - \boldsymbol{\Sigma}_k^{-1}\Delta\boldsymbol{\Sigma}_{h,k}))|\mathcal{F}_{t_{k-1}}) + E((\mathbf{z}_k - \boldsymbol{\mu}_k)^\intercal\mathbf{M}_{h,k}(\mathbf{z}_k - \boldsymbol{\mu}_k)|\mathcal{F}_{t_{k-1}}) \\ & + E((\mathbf{z}_k - \boldsymbol{\mu}_k)^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k})|\mathcal{F}_{t_{k-1}}) + E((\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k})^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_k)|\mathcal{F}_{t_{k-1}}) \\ & + E((\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k})^\intercal(\boldsymbol{\Sigma}_{h,k})^{-1}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k})|\mathcal{F}_{t_{k-1}}) \end{aligned}$$

with $\mathbf{M}_{h,k} = \boldsymbol{\Sigma}_k^{-1}\Delta\boldsymbol{\Sigma}_{h,k}(\mathbf{I} - \boldsymbol{\Sigma}_k^{-1}\Delta\boldsymbol{\Sigma}_{h,k})^{-1}\boldsymbol{\Sigma}_k^{-1}$ and $\Delta\Sigma_{h,k} = \boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}_{h,k}$.

Denote by $W''_{M,h}$ and $R''_{M,h}$ the second derivative of $W_{M,h}$ and $R_{M,h}$ with respect to $\boldsymbol{\theta}$.

Taking into account that

$$\begin{aligned} (W''_{M,h}(\boldsymbol{\theta}))^{-1} &= (W''_M(\boldsymbol{\theta}) + E(R''_{M,h}(\boldsymbol{\theta})))^{-1} \\ &= (W''_M(\boldsymbol{\theta}))^{-1} + \mathbf{K}_{M,h}(\boldsymbol{\theta}) \end{aligned}$$

with

$$\mathbf{K}_{M,h}(\boldsymbol{\theta}) = -(W''_M(\boldsymbol{\theta}))^{-1}E(R''_{M,h}(\boldsymbol{\theta}))(\mathbf{I} + (W''_M(\boldsymbol{\theta}))^{-1}E(R''_{M,h}(\boldsymbol{\theta})))^{-1}(W''_M(\boldsymbol{\theta}))^{-1},$$

it is obtained that

$$\mathbf{P}_{M,h}(\boldsymbol{\theta}) = (W''_M(\boldsymbol{\theta}))^{-1}\mathbf{H}_{M,h}(\boldsymbol{\theta})(W''_M(\boldsymbol{\theta}))^{-1} + \Delta\mathbf{P}_{M,h}(\boldsymbol{\theta}), \tag{21}$$

where

$$\Delta\mathbf{P}_{M,h}(\boldsymbol{\theta}) = \mathbf{K}_{M,h}(\boldsymbol{\theta})\mathbf{H}_{M,h}(\boldsymbol{\theta})(W''_M(\boldsymbol{\theta}))^{-1} + (W''_M(\boldsymbol{\theta}))^{-1}\mathbf{H}_{M,h}(\boldsymbol{\theta})\mathbf{K}_{M,h}(\boldsymbol{\theta}) + \mathbf{K}_{M,h}(\boldsymbol{\theta})\mathbf{H}_{M,h}(\boldsymbol{\theta})\mathbf{K}_{M,h}(\boldsymbol{\theta}).$$

For the functions $g(\mathbf{x}(t)) = \mathbf{x}^i(t)$ and $g(\mathbf{x}(t)) = \mathbf{x}^i(t)\mathbf{x}^j(t)$ belonging to the function space $\mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$, for all $i, j = 1..d$, condition (4) directly implies that

$$|E(\mathbf{x}(t_k)|\mathbf{z}_{k-1}) - E(\mathbf{y}(t_k)|\mathbf{z}_{k-1})| \le \sqrt{d}L_{k-1}h^\beta$$

and

$$|E(\mathbf{x}(t_k)\mathbf{x}^\intercal(t_k)|\mathbf{z}_{k-1}) - E(\mathbf{y}(t_k)\mathbf{y}^\intercal(t_k)|\mathbf{z}_{k-1})| \le dL_{k-1}h^\beta.$$

From this and the finite bound for the conditional mean of $\mathbf{x}$ and $\mathbf{y}$, it is obtained that

$$\left|\boldsymbol{\mu}_k - \boldsymbol{\mu}_{h,k}\right| \to \mathbf{0} \qquad \text{and} \qquad |\ \boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}_{h,k}| \to \mathbf{0}$$

as $h \to 0$ for all $\boldsymbol{\theta} \in \mathcal{D}_\theta$ and $k = 1, .., M-1$. This and the finite bound for the first two conditional moments of $\mathbf{x}$ and $\mathbf{y}$ imply that $|R_{M,h}(\boldsymbol{\theta}, Z)| \to 0$ and $\left|R''_{M,h}(\boldsymbol{\theta}, Z)\right| \to 0$ as well with h. From this and (20), it is obtained that

$$W_{M,h}(\boldsymbol{\theta}) \to W_M(\boldsymbol{\theta}) \qquad \text{and} \qquad W''_{M,h}(\boldsymbol{\theta}) \to W''_M(\boldsymbol{\theta}) \quad \text{as} \qquad h \to 0. \tag{22}$$

In addition, left (22) and Lemma 1 imply that

$$\Delta\boldsymbol{\theta}_M(h) = \boldsymbol{\alpha}_M(h) - \boldsymbol{\theta}_0 = \arg\{\min_{\boldsymbol{\theta}\in\mathcal{D}_\theta} W_{M,h}(\boldsymbol{\theta})\} - \arg\{\min_{\boldsymbol{\theta}\in\mathcal{D}_\theta} W_M(\boldsymbol{\theta})\} \to 0 \quad \text{as} \qquad h \to 0, \tag{23}$$

whereas from right (22) follows that

$$\Delta\mathbf{P}_{M,h}(\boldsymbol{\theta}) \to 0 \quad \text{as} \qquad h \to 0. \tag{24}$$

Finally, (23)-(24) together (18), (19) and (21) imply that (15) and (17) hold, which completes the proof.

Theorem 3 states that, for an increasing number of observations, the order-β QML estimator $\widehat{\boldsymbol{\theta}}_M(h)$ is asymptotically normal distributed and its bias decreases when h goes to zeros. This is a predictable result due to the known asymptotic properties of the exact QML estimator $\widehat{\boldsymbol{\theta}}_M$ stated in Bollerslev and Wooldridge (1992) and the convergence of the approximate estimator $\widehat{\boldsymbol{\theta}}_M(h)$ to $\widehat{\boldsymbol{\theta}}_M$ given by Theorem 1 when $h \to 0$. Further note that, when $h = 0$, the Theorem 3 reduces to Theorem 1 in Ljung & Caines (1979) for the exact QML estimator $\widehat{\boldsymbol{\theta}}_M$. This is other expected result since the order-β QML estimator $\widehat{\boldsymbol{\theta}}_M(h)$ reduces to the exact one $\widehat{\boldsymbol{\theta}}_M$ when $h = 0$. Further note that, neither in Theorem 3 there are restrictions on the time partition $\{t\}_M$ for the data or on the time discretization $(\tau)_h \supset \{t\}_M$ on which the approximate estimator is defined. Therefore, the comments about them at the end of the previous subsection are valid here as well.

4 Order-β QML estimator based on Local Linear approximations

Since, in principle, any type of approximation converging to the solution of (1) in a weak sense can be used to construct an order-β QML estimator, some additional criterions could be considered for the selection of one of them. For instance, high order of convergence, efficient algorithm for the computation of the moments, and so on. In this paper, we elected the order-β Local Linear approximations (see, e.g., Jimenez & Biscay, 2002, and Jimenez & Ozaki, 2003) for the following reasons: 1) their first two conditional moments have simple explicit formulas that can be computed by means of efficient algorithm (including high dimensional equations) as in Jimenez & Ozaki (2002,2003) and Jimenez (2012a); 2) their first two conditional moments are exact for linear equations in all the possible variants (with additive and/or multiplicative noise, autonomous or not), see Jimenez & Ozaki (2002); 3) they have an adequate order $\beta = 1, 2$ of weak convergence (Carbonell et al., 2006 and Jimenez, 2012b); and 4) the better performance of the conventional QML estimators based on Local Linearization schemes due to their simplicity, computational efficiency and negligible bias (see, e.g., Shoji & Ozaki 1997, Durham & Gallant 2002, Singer 2002, Hurn et al. 2007).

It is known that the first two conditional moments of the Local Linear approximations satisfy a set ordinary differential equations. Explicit formulas for the solution of these equations can be found in various papers as it was mentioned before. In what follows, the simplified expressions derived in Jimenez (2012a) are presented.

Denote by $\mathbf{y}_{\tau_n/t_k} = E(\mathbf{y}(\tau_n)|\mathbf{z}_k)$ and $\mathbf{P}_{\tau_n/t_k} = E(\mathbf{y}(\tau_n)\mathbf{y}^\intercal(\tau_n)|\mathbf{z}_k)$ the first two conditional moment of the order-β Local Linear approximation $\mathbf{y}$ at τ_n given the observation $\mathbf{z}_k$, for all $\tau_n \in \{(\tau)_h \cap [t_k, t_{k+1}]\}$ and $k = 0, .., M-2$. Clearly, $\mathbf{y}_{t_{k+1}/t_k}$ and $\mathbf{V}_{t_{k+1}/t_k} = \mathbf{P}_{t_{k+1}/t_k} - \mathbf{y}_{t_{k+1}/t_k}\mathbf{y}^\intercal_{t_{k+1}/t_k}$ provide approximations to the exact conditional mean $\boldsymbol{\mu}_{k+1}$ and variance $\boldsymbol{\Sigma}_{k+1}$, respectively, for all $t_k, t_{k+1} \in \{t\}_M$. Moreover, $\mathbf{y}_{t_k/t_k} = \mathbf{z}_k$ and $\mathbf{P}_{t_k/t_k} = \mathbf{z}_k\mathbf{z}_k^\intercal$ for all $t_k \in \{t\}_M$. Let $n_t = \max\{n = 0, 1, \ldots : \tau_n \leq t \text{ and } \tau_n \in (\tau)_h\}$ for all $t \in [t_0, t_{M-1}]$.

According to Jimenez (2012b), the approximate moments $\mathbf{y}_{t_{k+1}/t_k}$ and $\mathbf{P}_{t_{k+1}/t_k}$ are obtained by evaluating the recursive formulas

$$\mathbf{y}_{t/t_k} = \mathbf{y}_{\tau_{n_t}/t_k} + \mathbf{L}_2 e^{\mathbf{M}(\tau_{n_t})(t-\tau_{n_t})}\mathbf{u}_{\tau_{n_t},t_k} \tag{25}$$

and

$$vec(\mathbf{P}_{t/t_k}) = \mathbf{L}_1 e^{\mathbf{M}(\tau_{n_t})(t-\tau_{n_t})}\mathbf{u}_{\tau_{n_t},t_k} \tag{26}$$

at $t = t_{k+1}$, where the vector $\mathbf{u}_{\tau,t_k}$ and the matrices $\mathbf{M}(\tau)$, $\mathbf{L}_1$, $\mathbf{L}_2$ are defined as

$$\mathbf{M}(\tau) = \begin{bmatrix} \mathcal{A}(\tau) & \mathcal{B}_5(\tau) & \mathcal{B}_4(\tau) & \mathcal{B}_3(\tau) & \mathcal{B}_2(\tau) & \mathcal{B}_1(\tau) \\ \mathbf{0} & \mathbf{C}(\tau) & \mathbf{I}_{d+2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}(\tau) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 2 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 0 & 1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{u}_{\tau,t_k} = \begin{bmatrix} vec(\mathbf{P}_{\tau/t_k}) \\ \mathbf{0} \\ \mathbf{r} \\ 0 \\ 0 \\ 1 \end{bmatrix} \in \mathbb{R}^{(d^2+2d+7)}$$

and

$$\mathbf{L}_1 = \begin{bmatrix} \mathbf{I}_{d^2} & \mathbf{0}_{d^2\times(2d+7)} \end{bmatrix}, \qquad \mathbf{L}_2 = \begin{bmatrix} \mathbf{0}_{d\times(d^2+d+2)} & \mathbf{I}_d & \mathbf{0}_{d\times 5} \end{bmatrix}$$

in terms of the matrices and vectors

$$\mathcal{A}(\tau) = \mathbf{A}(\tau)\oplus\mathbf{A}(\tau) + \sum_{i=1}^{m} \mathbf{B}_i(\tau)\otimes\mathbf{B}_i^{\intercal}(\tau),$$

$$\mathbf{C}(\tau) = \begin{bmatrix} \mathbf{A}(\tau) & \mathbf{a}_1(\tau) & \mathbf{A}(\tau)\mathbf{y}_{\tau/t_k} + \mathbf{a}_0(\tau) \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{(d+2)\times(d+2)},$$

$$\mathbf{r}^{\intercal} = \begin{bmatrix} \mathbf{0}_{1\times(d+1)} & 1 \end{bmatrix}$$

$\mathcal{B}_1(\tau) = vec(\boldsymbol{\beta}_1(\tau)) + \boldsymbol{\beta}_4(\tau)\mathbf{y}_{\tau/t_k}$, $\mathcal{B}_2(\tau) = vec(\boldsymbol{\beta}_2(\tau)) + \boldsymbol{\beta}_5(\tau)\mathbf{y}_{\tau/t_k}$, $\mathcal{B}_3(\tau) = vec(\boldsymbol{\beta}_3(\tau))$, $\mathcal{B}_4(\tau) = \boldsymbol{\beta}_4(\tau)\mathbf{L}$ and $\mathcal{B}_5(\tau) = \boldsymbol{\beta}_5(\tau)\mathbf{L}$ with

$$\begin{aligned}
\boldsymbol{\beta}_1(\tau) &= \sum_{i=1}^{m} \mathbf{b}_{i,0}(\tau)\mathbf{b}_{i,0}^{\intercal}(\tau) \\
\boldsymbol{\beta}_2(\tau) &= \sum_{i=1}^{m} \mathbf{b}_{i,0}(\tau)\mathbf{b}_{i,1}^{\intercal}(\tau) + \mathbf{b}_{i,1}(\tau)\mathbf{b}_{i,0}^{\intercal}(\tau) \\
\boldsymbol{\beta}_3(\tau) &= \sum_{i=1}^{m} \mathbf{b}_{i,1}(\tau)\mathbf{b}_{i,1}^{\intercal}(\tau) \\
\boldsymbol{\beta}_4(\tau) &= \mathbf{a}_0(\tau) \oplus \mathbf{a}_0(\tau) + \sum_{i=1}^{m} \mathbf{b}_{i,0}(\tau) \otimes \mathbf{B}_i(\tau) + \mathbf{B}_i(\tau) \otimes \mathbf{b}_{i,0}(\tau) \\
\boldsymbol{\beta}_5(\tau) &= \mathbf{a}_1(\tau) \oplus \mathbf{a}_1(\tau) + \sum_{i=1}^{m} \mathbf{b}_{i,1}(\tau) \otimes \mathbf{B}_i(\tau) + \mathbf{B}_i(\tau) \otimes \mathbf{b}_{i,1}(\tau),
\end{aligned}$$

$\mathbf{L} = \begin{bmatrix} \mathbf{I}_d & \mathbf{0}_{d\times 2} \end{bmatrix}$, and the d-dimensional identity matrix $\mathbf{I}_d$. Here,

$$\mathbf{A}(\tau) = \frac{\partial \mathbf{f}(\tau, \mathbf{y}_{\tau/t_k})}{\partial \mathbf{y}} \qquad \text{and} \qquad \mathbf{B}_i(\tau) = \frac{\partial \mathbf{g}_i(\tau, \mathbf{y}_{\tau/t_k})}{\partial \mathbf{y}}$$

are matrices, and the vectors $\mathbf{a}_0(\tau_{n_t})$, $\mathbf{a}_1(\tau_{n_t})$, $\mathbf{b}_{i,0}(\tau_{n_t})$ and $\mathbf{b}_{i,1}(\tau_{n_t})$ satisfy the expressions

$$\mathbf{a}^{\beta}(t;\tau_{n_t}) = \mathbf{a}_0(\tau_{n_t}) + \mathbf{a}_1(\tau_{n_t})(t-\tau_{n_t}) \qquad \text{and} \qquad \mathbf{b}_i^{\beta}(t;\tau_{n_t}) = \mathbf{b}_{i,0}(\tau_{n_t}) + \mathbf{b}_{i,1}(\tau_{n_t})(t-\tau_{n_t})$$

for all $t \in [t_k, t_{k+1}]$ and $\tau_{n_t} \in (\tau)_h$, where

$$\mathbf{a}^{\beta}(t;\tau) = \begin{cases} \mathbf{f}(\tau, \mathbf{y}_{\tau/t_k}) - \frac{\partial \mathbf{f}(\tau,\mathbf{y}_{\tau/t_k})}{\partial \mathbf{y}} \mathbf{y}_{\tau/t_k} + \frac{\partial \mathbf{f}(\tau,\mathbf{y}_{\tau/t_k})}{\partial \tau}(t-\tau) & \text{for } \beta = 1 \\ \mathbf{a}^1(t;\tau) + \frac{1}{2}\sum\limits_{j,l=1}^{d} [\mathbf{G}(\tau, \mathbf{y}_{\tau/t_k})\mathbf{G}^{\intercal}(\tau, \mathbf{y}_{\tau/t_k})]^{j,l} \ \frac{\partial^2 \mathbf{f}(\tau,\mathbf{y}_{\tau/t_k})}{\partial \mathbf{y}^j \partial \mathbf{y}^l}(t-\tau) & \text{for } \beta = 2 \end{cases}$$

and

$$\mathbf{b}_i^{\beta}(t;\tau) = \begin{cases} \mathbf{g}_i(\tau, \mathbf{y}(\tau)) - \frac{\partial \mathbf{g}_i(\tau,\mathbf{y}_{\tau/t_k})}{\partial \mathbf{y}} \mathbf{y}_{\tau/t_k} + \frac{\partial \mathbf{g}_i(\tau,\mathbf{y}_{\tau/t_k})}{\partial \tau}(t-\tau) & \text{for } \beta = 1 \\ \mathbf{b}_i^1(t;\tau) + \frac{1}{2}\sum\limits_{j,l=1}^{d} [\mathbf{G}(\tau, \mathbf{y}_{\tau/t_k})\mathbf{G}^{\intercal}(\tau, \mathbf{y}_{\tau/t_k})]^{j,l} \ \frac{\partial^2 \mathbf{g}_i(\tau,\mathbf{y}(\tau))}{\partial \mathbf{y}^j \partial \mathbf{y}^l}(t-\tau) & \text{for } \beta = 2 \end{cases}$$

are functions associated to the order-β Ito-Taylor expansions for the drift and diffusion coefficients of (1) in the neighborhood of $(\tau, \mathbf{y}_{\tau/t_k})$, respectively, and $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_m]$ is an $d \times m$ matrix function. The symbols vec, $\oplus$ and $\otimes$ denote the vectorization operator, the Kronecker sum and product, respectively.

Under general conditions, Lemma 7 and Theorem 9 in Jimenez (2012b) state that the order-β Local Linear approximation $\mathbf{y}$ satisfies the bound condition (3) and the weak convergence criteria (4). Hence, Theorem 1 implies that the order-β QML estimator

$$\widehat{\boldsymbol{\theta}}_M(h) = \arg\{\min_{\boldsymbol{\theta}} U_{M,h}(\boldsymbol{\theta}, Z)\}, \tag{27}$$

with

$$U_{M,h}(\boldsymbol{\theta}, Z) = (M-1)\ln(2\pi) + \sum_{k=0}^{M-2} \ln(\det(\mathbf{V}_{t_{k+1}/t_k})) + (\mathbf{z}_{k+1} - \mathbf{y}_{t_{k+1}/t_k})^{\intercal}(\mathbf{V}_{t_{k+1}/t_k})^{-1}(\mathbf{z}_{k+1} - \mathbf{y}_{t_{k+1}/t_k}),$$

converges to the exact one (2) as h goes to zero for all given Z. For the same reason, the order-β QML estimator (27) has the asymptotic properties stated in Theorem 3, and the average of their values for different realizations of the SDE satisfies the convergence property of Theorem 2.

For one-dimensional SDEs with additive noise, the order-β QML estimator (27) reduces to the conventional estimators of Ozaki (1985, 1992) and Shoji & Ozaki (1997, 1998) when $(\tau)_h \equiv \{t\}_M$. It is worth to emphasize that, for each data $\mathbf{z}_k$, the formulas (25)-(26) are recursively evaluated at all the time instants $\tau_n \in \{(\tau)_h \cap (t_k, t_{k+1}]\}$ for the first estimator, whereas they are evaluated only at $t_{k+1} = \{\{t\}_M \cap (t_k, t_{k+1}]\}$ for the conventional ones.

From computational viewpoint, each evaluation of the formulas (25)-(26) at τ_n requires the computation of just one exponential matrix whose matrix depends of the drift and diffusion coefficients of (1) at $(\tau_{n-1}, \mathbf{y}_{\tau_{n-1}/t_k})$. This exponential matrix can the efficiently computed through the well known Padé method (Moler & Van Loan, 2003) or, alternatively, by means of the Krylov subspace method (Moler & Van Loan, 2003) in the case of high dimensional SDEs. Even more, low order Padé and Krylov methods as suggested in Jimenez & de la Cruz (2012) can be used as well for reducing the computation cost, but preserving the order-β of the approximate moments. Alternatively, simplified formulas for the moments can be used when the equation to be estimate is autonomous or has additive noise (see Jimenez, 2012a). All this makes simple and efficient the evaluation of the approximate predictors $\mathbf{y}_{t_{k+1}/t_k}$ and $\mathbf{V}_{t_{k+1}/t_k}$.

In practical situations, it is convenient to write a code that automatically determines the time discretization $(\tau)_h$ for achieving a prescribed absolute $(atol_{\mathbf{y}}, atol_{\mathbf{P}})$ and relative $(rtol_{\mathbf{y}}, rtol_{\mathbf{P}})$ error tolerance in the computation of $\mathbf{y}_{t_{k+1}/t_k}$ and $\mathbf{P}_{t_{k+1}/t_k}$. With this purpose the adaptive strategy proposed in Jimenez (2012b) is useful.

5 Simulation study

In this section, the performance of the new approximate estimators is illustrated, by means of simulations, with four test SDEs. To do so, four types of QML estimators are computed and compared: 1) the exact one (2), when it is possible; 2) the conventional one based on the LL scheme. That is, the estimator defined by (27) with $(\tau)_h \equiv \{t\}_M$ and $\beta = 1$; 3) the order-1 QML estimator (27) with various uniform time discretizations $(\tau)_{h,T}^{u}$; and 4) the adaptive order-1 QML estimator (27) with the adaptive selection of time discretizations $(\tau)_{\cdot,T}$ proposed in Jimenez (2012b). For each example, histograms and confidence limits for the estimators are computed from various sets of discrete observations taken with different time distances (sampling periods) on time intervals with distinct lengths.

5.1 Test equations

Example 1. Equation with multiplicative noise

$$dx = \alpha t x dt + \sigma\sqrt{t} x dw_1 \tag{28}$$

where $\alpha = -0.1$ and $\sigma = 0.1$ are parameters to be estimated, and $x(t_0) = 1$ is the initial value of x at $t_0 = 0.5$. For this equation, the conditional mean and variance of x at t_{k+1} given the observation z_k of x at t_k are

$$\boldsymbol{\mu}_{k+1} = z_k e^{\alpha(t_{k+1}^2 - t_k^2)/2} \quad \text{and} \quad \boldsymbol{\Sigma}_{k+1} = z_k^2 e^{(\alpha+\sigma^2/2)(t_{k+1}^2 - t_k^2)} - \boldsymbol{\mu}_{k+1}^2,$$

respectively, for all $t_{k+1} > t_k \geq t_0$.

Example 2. Equation with two additive noise

$$dx = \alpha t x dt + \sigma t^2 e^{\alpha t^2/2} dw_1 + \rho\sqrt{t} dw_2 \tag{29}$$

where $\alpha = -0.25$, $\sigma = 5$, and $\rho = 0.1$ are parameters to be estimated, and $x(t_0) = 10$ is the initial value of x at $t_0 = 0.01$. For this equation, the conditional mean and variance of x at t_{k+1} given the observation z_k of x at t_k are

$$\boldsymbol{\mu}_{k+1} = z_k e^{\alpha(t_{k+1}^2 - t_k^2)/2}$$

and

$$\boldsymbol{\Sigma}_{k+1} = \frac{\rho^2}{2\alpha} e^{\alpha(t_{k+1}^2 - t_k^2)} + \frac{\sigma^2}{5}(t_{k+1}^5 - t_k^5) e^{\alpha t_{k+1}^2} - \frac{\rho^2}{2\alpha},$$

respectively, for all $t_{k+1} > t_k \geq t_0$.

Example 3. Van der Pool oscillator with random input (Gitterman, 2005)

$$dx_1 = x_2 dt \tag{30}$$

$$dx_2 = (-(x_1^2 - 1)x_2 - x_1 + \alpha)dt + \sigma dw \tag{31}$$

where $\alpha = 0.5$ and $\sigma^2 = (0.75)^2$ are, respectively, the intensity and the variance of the random input that should be estimated. In addition, $t_0 = 0$, and $\mathbf{x}^\intercal(t_0) = [1\ 1]$.

Example 4. Van der Pool oscillator with random frequency (Gitterman, 2005)

$$dx_1 = x_2 dt \tag{32}$$

$$dx_2 = (-(x_1^2 - 1)x_2 - \alpha x_1)dt + \sigma x_1 dw \tag{33}$$

where $\alpha = 1$ and $\sigma^2 = 1$ are, respectively, the frequency mean value and variance that should be estimated. In addition, $t_0 = 0$, and $\mathbf{x}^\intercal(t_0) = [1\ 1]$.

In these examples, autonomous or non autonomous, linear or nonlinear, one or two dimensional equations with additive or multiplicative noise are considered for the estimation of two or three parameters. Note that, since the first two conditional moments of the Examples 1 and 2 have explicit expressions, the exact QML estimator (2) can be computed.

These four equations have previously been used in Jimenez (2012b) to illustrate the convergence of the approximate moments (25)-(26) by means of simulations. Tables with the errors between the approximate moments and the exact ones as a function of h were given for the Examples 1 and 2. Tables with the estimated rate of convergence were provided for the fours examples.

5.2 Simulations with one-dimensional equations

For the first two examples, 100 realizations of the solution were computed by means of the Euler (Kloeden & Platen, 1999) or the Local Linearization scheme (Jimenez et al., 1999) for the equation with multiplicative or additive noise, respectively. For each example, the realizations where computed over the thin time partition $\{t_0 + 10^{-4}n : n = 0, .., 30 \times 10^4\}$ to guarantee a precise simulation of the stochastic solutions on the time interval $[t_0, t_0 + 30]$. Twelve subsamples of each realization at the time instants $\{t\}_{M,T} = \{t_k = t_0 + kT/M : k = 0, .., M-1\}$ were taken as observation Z of $\mathbf{x}$ for making inference with various values of M and T. In particular, the values $T = 10, 20, 30$ and $M = T/\delta$ with $\delta = 1, 0.1, 0.01, 0.001$ were used. In this way, twelve sets of 100 time series $Z^i_{\delta,T} = \{z^i_k : k = 0, .., M-1, M = T/\delta\}$, with $i = 1, .., 100$, of M observations z^i_k each one were finally available for each example with the twelve values of (δ, T) mentioned above. This will allow us to explore and compare the performance of each estimator from observations taken with different sampling periods δ on time intervals with distinct lengths T.

Figure 1 shows the histograms and the confidence limits for both, the exact $(\widehat{\alpha}^E_{\delta,T})$ and the conventional $(\widehat{\alpha}_{\delta,T})$ QML estimators of α computed from the twelve sets of 100 time series $Z^i_{\delta,T}$ available for the example 1. Figure 2 shows the same but, for the exact $(\widehat{\sigma}^E_{\delta,T})$ and the conventional $(\widehat{\sigma}_{\delta,T})$ QML estimators of σ. As it was expected, for the samples $Z^i_{\delta,T}$ with largest sampling periods, the parameter estimation is distorted by the well-known lowpass filter effect of signals sampling (see, e.g., Oppenheim & Schafer, 2010). This is the reason of the under estimation of the variance $\widehat{\sigma}^E_{\delta,T}$ from the samples $Z^i_{\delta,T}$, with $\delta = 1$ and $T = 10, 20, 30$, when the parameter α in the drift coefficient of (28) is better estimated by $\widehat{\alpha}^E_{\delta,T}$. Contrarily, from these samples, the conventional QML estimators $\widehat{\alpha}_{\delta,T}$ can not provided a good approximation to α, and so the whole unexplained component of the drift term of (28) included in the samples is interpreted as noise by the conventional QML estimators. For this reason, $\widehat{\sigma}_{\delta,T}$ over estimates the value of the parameter σ. Further, note that when the sampling period δ decreases and the length T of the observation time increases, the difference between the exact $(\widehat{\alpha}^E_{\delta,T}, \widehat{\sigma}^E_{\delta,T})$ and the conventional $(\widehat{\alpha}_{\delta,T}, \widehat{\sigma}_{\delta,T})$ QML estimators decreases, as well as the bias of both estimators. This is also other expected result. Here, the bias is estimated by the difference between the parameter value and the estimator average, whereas the difference between estimators refers to the histogram shape and confidence limits.

$\delta = 1$		$h = \delta$	$h = \delta/2$	$h = \delta/8$	$h = \delta/32$
α	$T = 10$	$5.7 \pm 3.7 \times 10^{-3}$	$1.3 \pm 0.9 \times 10^{-3}$	$2.2 \pm 1.6 \times 10^{-4}$	$5.4 \pm 4.0 \times 10^{-5}$
	$T = 20$	$3.7 \pm 1.6 \times 10^{-3}$	$6.6 \pm 3.8 \times 10^{-4}$	$8.2 \pm 7.6 \times 10^{-5}$	$2.3 \pm 2.1 \times 10^{-5}$
	$T = 30$	$5.8 \pm 1.6 \times 10^{-3}$	$8.1 \pm 3.4 \times 10^{-4}$	$5.4 \pm 4.6 \times 10^{-5}$	$1.7 \pm 1.4 \times 10^{-5}$
σ	$T = 10$	$2.8 \pm 2.0 \times 10^{-2}$	$8.6 \pm 5.2 \times 10^{-3}$	$1.7 \pm 0.9 \times 10^{-3}$	$4.0 \pm 2.1 \times 10^{-4}$
	$T = 20$	$1.4 \pm 1.3 \times 10^{-2}$	$4.5 \pm 2.9 \times 10^{-3}$	$9.4 \pm 5.1 \times 10^{-4}$	$2.2 \pm 1.1 \times 10^{-4}$
	$T = 30$	$8.4 \pm 8.7 \times 10^{-3}$	$2.6 \pm 2.1 \times 10^{-3}$	$6.2 \pm 3.6 \times 10^{-4}$	$1.5 \pm 8.4 \times 10^{-5}$

Table I. Confidence limits for the error between the exact and the approximate QML estimators of the equation (28). $h = \delta$, for the conventional; and $h = \delta/2, \delta/8, \delta/32$, for the order-1 on $(\tau)^u_{h,T}$.

$\delta = 1$	α			σ		
h	$T = 10$	$T = 20$	$T = 30$	$T = 10$	$T = 20$	$T = 30$
δ	-0.0030	-0.0035	-0.0058	-0.0286	-0.0150	-0.0071
$\delta/2$	-0.0006	-0.0005	-0.0008	-0.0086	-0.0045	-0.0027
$\delta/8$	0	0.0001	0	-0.0017	-0.0009	-0.0006
$\delta/32$	0	0	0	-0.0004	-0.0002	-0.0002
$\cdot$	0	0	-0.0003	-0.0002	-0.0002	-0.0013

Table II: Difference between the averages of the exact and the approximate QML estimators for the equation (28). $h = \delta$, for the conventional; $h = \delta/2, \delta/8, \delta/32$, for the order-1 on $(\tau)^u_{h,T}$; and $h = \cdot$, for the adaptive order-1 on $(\tau)_{\cdot,T}$.

For the data of (28) with largest sampling period $\delta = 1$, the order-1 QML estimators $(\widehat{\alpha}^u_{h,\delta,T}, \widehat{\sigma}^u_{h,\delta,T})$ and $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T})$ on uniform $(\tau)^u_{h,T} = \{\tau_n = t_0 + nh : n = 0, .., T/h\} \supset \{t\}_{T/\delta,T}$ and adaptive $(\tau)_{\cdot,T} \supset$

$\{t\}_{T/\delta,T}$ time discretizations, respectively, were computed with $h = \delta/2, \delta/8, \delta/32$ and tolerances $rtol_{\mathbf{y}} = rtol_{\mathbf{P}} = 5 \times 10^{-6}$ and $atol_{\mathbf{y}} = 5 \times 10^{-9}$, $atol_{\mathbf{P}} = 5 \times 10^{-12}$. For each data $Z^i_{\delta,T}$, with $i = 1, .., 100$, the errors

$$\varepsilon_i(\alpha, h, \delta, T) = \left|\widehat{\alpha}^E_{\delta,T} - \widehat{\alpha}^u_{h,\delta,T}\right| \text{ and } \varepsilon_i(\sigma, h, \delta, T) = \left|\widehat{\sigma}^E_{\delta,T} - \widehat{\sigma}^u_{h,\delta,T}\right|$$

between the exact $(\widehat{\alpha}^E_{\delta,T}, \widehat{\sigma}^E_{\delta,T})$ and the approximate $(\widehat{\alpha}^u_{h,\delta,T}, \widehat{\sigma}^u_{h,\delta,T})$ QML estimators were computed. Average and standard deviation of these 100 errors were calculated for each set of values h, δ, T specified above, which are summarized in Table I. Note as, for fixed T, the average of the errors decreases as h does it. This clearly illustrates the convergence of the order-1 QML estimators to the exact one stated in Theorem 1 when h goes to zero. In addition, Figure 3 shows the histograms and the confidence limits for the order-1 QML estimators $(\widehat{\alpha}^u_{h,\delta,T}, \widehat{\sigma}^u_{h,\delta,T})$ and $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T})$ for each set of values h, δ, T. By comparing the results of this figure with the corresponding in the previous ones, the decreasing difference between the order-1 QML estimators $(\widehat{\alpha}^u_{h,\delta,T}, \widehat{\sigma}^u_{h,\delta,T})$ and the exact one $(\widehat{\alpha}^E_{\delta,T}, \widehat{\sigma}^E_{\delta,T})$ is observed as h decreases, which is consistent with the convergence results of Table I. Similarly, for $T = 10, 20$, the difference between the order-1 QML estimators $(\widehat{\alpha}^u_{h,\delta,T}, \widehat{\sigma}^u_{h,\delta,T})$ and the adaptive QML estimators $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T})$ decreases when h does it, due to the negligible difference between the adaptive QML estimators $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T})$ with the exact ones $(\widehat{\alpha}^E_{\delta,T}, \widehat{\sigma}^E_{\delta,T})$. This illustrates the usefulness of the adaptive strategy for improving the QML parameter estimation for finite samples with large sampling periods. These findings are more precisely summarized in Table II, which shows the difference between the averages of the exact and the approximate QML estimators. Observe the lightly higher difference between the averages of the exact and the adaptive estimators for both parameters when $T = 30$. The reason is that, for $t_k > 21$, the variance $\mathbf{\Sigma}_k$ of the diffusion (28) becomes almost indistinguishable of zero. This is so small that the roundoff errors becomes significant in such a way that the adaptive strategy under estimates the integration errors and so over estimates the length of the step sizes. This can be seem in Figure 4, which shows the average of accepted and fail steps of the adaptive QML estimators at each $t_k \in \{t\}_{T/\delta,T}$. Note that, for $t_k > 21$, the number of accepted steps is lower that 8. Therefore, the adaptive estimator can not be so good as those on uniform time discretization with $h = 1/8, 1/32$ in Table II. Contrary, for $t_k < 21$, the number of accepted steps is larger that 32, and the adaptive estimator performs better than all the others. Further, note that the results of Table II illustrate the convergence findings of Theorem 2.

$\delta = 0.1$	α			σ			ρ		
h	$T = 10$	$T = 20$	$T = 30$	$T = 10$	$T = 20$	$T = 30$	$T = 10$	$T = 20$	$T = 30$
δ	0.00048	0.00045	0.00043	-0.0405	-0.0397	-0.0396	-0.00015	4.4×10^{-5}	3.9×10^{-5}
$\delta/2$	0.00014	0.00013	0.00011	-0.0110	-0.0100	-0.0100	-0.00009	0.9×10^{-5}	1.4×10^{-5}
$\delta/4$	0.00003	0.00004	0.00003	-0.0025	-0.0023	-0.0022	-0.00002	0.8×10^{-5}	1.0×10^{-5}
$\delta/8$	0	0.00001	0	-0.0004	-0.0004	-0.0004	0	0.5×10^{-5}	0.6×10^{-5}
$\cdot$	-0.00008	0.00004	-0.00006	0.0049	0.0010	0.0092	0.00011	0	0.4×10^{-5}

Table III: Difference between the averages of the exact and the approximate QML estimators for the equation (29). $h = \delta$, for the conventional; $h = \delta/2, \delta/4, \delta/8$, for the order-1 on $(\tau)^u_{h,T}$; and $h = \cdot$, for the adaptive order-1 on $(\tau)_{\cdot,T}$.

$\delta = 0.1$		$h = \delta$	$h = \delta/2$	$h = \delta/4$	$h = \delta/8$
α	$T = 10$	$6.6 \pm 4.9 \times 10^{-4}$	$1.6 \pm 1.2 \times 10^{-4}$	$3.9 \pm 2.9 \times 10^{-5}$	$9.9 \pm 8.4 \times 10^{-6}$
	$T = 20$	$6.8 \pm 4.7 \times 10^{-4}$	$1.4 \pm 0.9 \times 10^{-4}$	$3.4 \pm 2.1 \times 10^{-5}$	$9.3 \pm 5.8 \times 10^{-6}$
	$T = 30$	$6.7 \pm 4.6 \times 10^{-4}$	$1.4 \pm 0.9 \times 10^{-4}$	$3.3 \pm 2.1 \times 10^{-5}$	$8.2 \pm 5.9 \times 10^{-6}$
σ	$T = 10$	$5.9 \pm 4.4 \times 10^{-2}$	$1.3 \pm 0.8 \times 10^{-2}$	$2.6 \pm 1.8 \times 10^{-3}$	$5.1 \pm 5.1 \times 10^{-4}$
	$T = 20$	$6.0 \pm 4.2 \times 10^{-2}$	$1.2 \pm 0.8 \times 10^{-2}$	$2.4 \pm 1.6 \times 10^{-3}$	$4.9 \pm 4.0 \times 10^{-4}$
	$T = 30$	$5.9 \pm 4.1 \times 10^{-2}$	$1.1 \pm 0.7 \times 10^{-2}$	$2.4 \pm 1.6 \times 10^{-3}$	$4.4 \pm 3.8 \times 10^{-4}$
ρ	$T = 10$	$1.1 \pm 1.3 \times 10^{-3}$	$2.9 \pm 4.0 \times 10^{-4}$	$6.2 \pm 8.3 \times 10^{-5}$	$1.5 \pm 1.9 \times 10^{-5}$
	$T = 20$	$1.6 \pm 1.5 \times 10^{-4}$	$5.1 \pm 4.6 \times 10^{-5}$	$1.5 \pm 0.8 \times 10^{-5}$	$6.2 \pm 2.4 \times 10^{-6}$
	$T = 30$	$9.0 \pm 7.4 \times 10^{-5}$	$3.1 \pm 2.2 \times 10^{-5}$	$1.1 \pm 0.4 \times 10^{-5}$	$5.8 \pm 1.9 \times 10^{-6}$

Table IV: Confidence limits for the error between the exact and the approximate QML estimators of the equation (29). $h = \delta$, for the conventional; and $h = \delta/2, \delta/4, \delta/8$, for the order-1 on $(\tau)^u_{h,T}$.

Figure 5 shows the histograms and the confidence limits for both, the exact $(\widehat{\alpha}^E_{\delta,T})$ and the conventional $(\widehat{\alpha}_{\delta,T})$ QML estimators of α computed from the twelve sets of 100 time series $Z^i_{\delta,T}$ available for the example

2. Figure 6 shows the same but, for the exact ($\widehat{\sigma}^{E}_{\delta,T}$) and the conventional ($\widehat{\sigma}_{\delta,T}$) QML estimators of σ, whereas Figure 7 does it for the estimators $\widehat{\rho}^{E}_{\delta,T}$ and $\widehat{\rho}_{\delta,T}$ of ρ. Note that, for this example, the diffusion parameters σ and ρ can not be estimated from the samples $Z^{i}_{\delta,T}$ with the largest sampling period $\delta = 1$. From the other data with sampling period $\delta < 1$, the tree parameters can be estimated and, the bias of the exact and the conventional QML estimators is not so large as in the previous example. Nevertheless, in this extreme situation of low information in the data, the order-1 QML estimators is able to improve the accuracy of the parameter estimation when h decreases. This is shown in Figure 8 for the samples $Z^{i}_{\delta,T}$ with $\delta = 0.1$ and $T = 10, 20, 30$, and summarized in Table III. The order-1 QML estimators $(\widehat{\alpha}^{u}_{h,\delta,T}, \widehat{\sigma}^{u}_{h,\delta,T}, \widehat{\rho}^{u}_{h,\delta,T})$ and $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T}, \widehat{\rho}_{\cdot,\delta,T})$ are again computed on uniform $(\tau)^{u}_{h,T} \supset \{t\}_{T/\delta,T}$ and adaptive $(\tau)_{\cdot,T} \supset \{t\}_{T/\delta,T}$ time discretizations, respectively, with $T = 10, 20, 30$, $h = \delta/2, \delta/4, \delta/8$ and tolerances $rtol_{\mathbf{y}} = rtol_{\mathbf{P}} = 10^{-6}$ and $atol_{\mathbf{y}} = 10^{-9}$, $atol_{\mathbf{P}} = 10^{-12}$. The average of accepted and fail steps of the adaptive QML estimators at each $t_k \in \{t\}_{T/\delta,T}$ are shown in Figure 4. Note that, the average of the accepted steps of the adaptive algorithm is not bigger than 8 for each $t_k \in \{t\}_{T/\delta,T}$. Therefore, as it is shown in Table III, the average of the adaptive estimator is not so good than that of some other estimators on uniform time discretizations $(\tau)^{u}_{h,T}$ with $h \geq \delta/8$. For this example, Table IV gives the confidence limits for the error between the exact and the order-1 QML estimators for different values of h. Note that, Table III and IV illustrate the convergence results of Theorems 2 and 1, respectively.

5.3 Simulations with two-dimensional equations

For the examples 3 and 4, 100 realizations of the equation were similarly computed by means of the Local Linearization and the Euler scheme, respectively. For each example, the realizations where computed over the thin time partition $\{t_0 + 10^{-4}n : n = 0, .., 30 \times 10^4\}$ for guarantee a precise simulation of the stochastic solutions on the time interval $[t_0, t_0 + 30]$. Two subsamples of each realization at the time instants $\{t\}_{M,T} = \{t_k = t_0 + kT/M : k = 0, .., M-1\}$ were taken as observation Z of $\mathbf{x}$ for making inference with $T = 30$ and two values of M. In particular, $M = 30, 300$ were used, which correspond to the sampling periods $\delta = 1, 0.1$. In this way, two sets of 100 time series $Z^{i}_{\delta,T} = \{\mathbf{z}^{i}_{k} : k = 0, .., M-1, M = T/\delta\}$, with $i = 1, .., 100$, of M observations $\mathbf{z}^{i}_{k}$ each one were available for each example with the two values of (δ, T) mentioned above.

$T = 30$	α		σ	
h	$\delta = 1$	$\delta = 0.1$	$\delta = 1$	$\delta = 0.1$
δ	-0.2500	-0.1428	-0.7328	-0.0052
$\delta/16$	-0.0965	-0.0044	-0.0893	0.0012
$\delta/64$	-0.0333	0.0029	-0.0757	0.0013
$\cdot$	-0.0096	0.0068	-0.0739	0.0013

Table V: Bias of the approximate QML estimators for the equation (30)-(31). $h = \delta$, for the conventional; $h = \delta/16, \delta/64$, for the order-1 on $(\tau)^{u}_{h,T}$; and $h = \cdot$, for the adaptive order-1 on $(\tau)_{\cdot,T}$.

$T = 30$	α		σ	
h	$\delta = 1$	$\delta = 0.1$	$\delta = 1$	$\delta = 0.1$
δ	-0.8000	-0.2507	-0.8000	-0.3748
$\delta/8$	-0.4234	-0.0481	-0.2451	0.0005
$\delta/32$	-0.2210	-0.0219	-0.1910	0.0015
$\cdot$	-0.1611	-0.0046	-0.1898	0.0001

Table VI: Bias of the approximate QML estimators for the equation (32)-(33). $h = \delta$, for the conventional; $h = \delta/8, \delta/32$, for the order-1 on $(\tau)^{u}_{h,T}$; and $h = \cdot$, for the adaptive order-1 on $(\tau)_{\cdot,T}$.

For both examples, the order-1 QML estimators $(\widehat{\alpha}^{u}_{h,\delta,T}, \widehat{\sigma}^{u}_{h,\delta,T})$ and $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T})$ on uniform $(\tau)^{u}_{h,T} \supset \{t\}_{T/\delta,T}$ and adaptive $(\tau)_{\cdot,T} \supset \{t\}_{T/\delta,T}$ time discretizations, respectively, were computed from the two sets of 100 data $Z^{i}_{\delta,T}$ with $T = 30$ and $\delta = 1, 0.1$. The values of h were set as $h = \delta, \delta/16, \delta/64$ for the example 3, and as $h = \delta, \delta/8, \delta/32$ for the example 4. The tolerances for the adaptive estimators were set as in the first example. Figures 9 and 11 show the histograms and the confidence limits for the estimators $(\widehat{\alpha}^{u}_{h,\delta,T}, \widehat{\sigma}^{u}_{h,\delta,T})$ and $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T})$ corresponding to each example. For the two examples, the difference between the order-1 QML estimator $(\widehat{\alpha}^{u}_{h,\delta,T}, \widehat{\sigma}^{u}_{h,\delta,T})$ and the adaptive one $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T})$ decreases when h does it. This is, according Theorem 1, an expected result by assuming that the difference between the adaptive and the exact QLM estimators is negligible for $(\tau)_{\cdot,T}$ thin enough. In addition, Table V and VI show the bias of the approximate QML estimators for these examples. Observe as the adaptive $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T})$ and the order-1 QML estimator $(\widehat{\alpha}^{u}_{h,\delta,T}, \widehat{\sigma}^{u}_{h,\delta,T})$ with $h < \delta$ provide much less biased estimation of the parameters (α, σ) than the conventional QML estimator $(\widehat{\alpha}^{u}_{\delta,\delta,T}, \widehat{\sigma}^{u}_{\delta,\delta,T})$, which is in fact unable to identify the parameters of the examples. Clearly, this illustrates the usefulness of the order-1 QML estimator and its adaptive implementation. However, as it is shown in Table V for $\delta = 0.1$, no always the adaptive estimator $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T})$ is less unbiased than the order-1 QML estimator $(\widehat{\alpha}^{u}_{h,\delta,T}, \widehat{\sigma}^{u}_{h,\delta,T})$ for some $h < \delta$. This can happen for one of following reasons: 1) the bias of the exact QML estimator when the adaptive estimator is close enough to it, or 2) an insufficient number of accepted steps of the adaptive estimator for a given tolerance. In our case, since $(\widehat{\alpha}^{u}_{h,\delta,T}, \widehat{\sigma}^{u}_{h,\delta,T})$ converges to $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T})$ as h decreases (Figure 9 with $\delta = 0.1$) and the average of accepted steps of the adaptive estimators is acceptable (Figure 10 with $\delta = 0.1$), the first explanation is more suitable. Figures 10 and 12 show the average of accepted and fail steps of the adaptive QML estimators at each $t_k \in \{t\}_{T/\delta,T}$ for each example. Note how the average of accepted steps corresponding to the estimators from samples with $\delta = 0.1$ is ten time lower than that of the estimators from samples with $\delta = 1$, which is an expected result as well.

6 Conclusions

A modification of the conventional approximations to the quasi-maximum likelihood (QML) method was introduced for the parameter estimation of diffusion processes given a time series of complete observations. This is based on a recursive approximation to the first two conditional moments of the diffusion process through discrete-time schemes. For finite samples, the convergence of the modified QML estimators to the exact one was proved when the error between the discrete-time approximation and the diffusion process decreases. It was also demonstrated that, for an increasing number of observations, they are asymptotically normal distributed and their bias decreases when the above mentioned error does it. As particular instance, the order-β QML estimators based on Local Linearization schemes were proposed. For them, practical algorithms were also provided and their performance in simulation illustrated with various examples. Simulations shown that: 1) with thin time discretizations between observations, the order-1 QML estimator provides satisfactory approximations to the exact QML estimator; 2) the convergence of the order-1 QML estimator to the exact one when the maximum stepsize of the time discretization between observations decreases; 3) with respect to the conventional QML estimator, the order-1 QML estimator gives much better approximation to the exact QML estimator, and has less bias and higher efficiency; 4) with an adequate tolerance, the adaptive order-1 QML estimator gives an automatic, suitable and computational efficient approximation to the exact QML estimator; and 5) the effectiveness of the order-1 QML estimator for the identification of SDEs from a reduced number of complete observations distant in time. Further note that new estimators can also be easily applied to a variety of practical problems with sequential random measurements or with multiple missing data.

Acknowledgements The numerical simulations of this paper were concluded on July 2012 within the framework of the Associateship Scheme of the Abdus Salam International Centre for Theoretical Physics (ICTP), Trieste, Italy. The author thanks to the ICTP for the partial support to this work.

References

Bibby B.M. and Sorensen M. (1996) Estimation for discrete observed diffusion processes: A review, Theory Stoch. Processes, 2, 49-56.

Bollerslev T. and Wooldridge J.M. (1992) Quasi-maximun likelihood estimation and inference in dynamic models with time-varying covariances. Econom. Rev., 11, 143-172.

Carbonell F., Jimenez J.C. and Biscay R.J. (2006) Weak local linear discretizations for stochastic differential equations: convergence and numerical schemes, J. Comput. Appl. Math., 197, 578-596.

Clement E. (1995) Bias correction for the estimation of discretized diffusion processes from an approximate likelihood function, Theory Prob. Appl., 42, 283-288.

Durham G.B. and Gallant A.R. (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes, J. Business Econom. Statist., 20, 297-316.

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

Gitterman M., The noisy oscillator, World Scientific, 2005.

Huang X. (2011) Quasi-maximun likelihood estimation of discretely observed diffusions. Econom. J., 14, 241-256.

Hurn A.S., Jeisman I.J. and Lindsay, K.A. (2007) Seeing the wood for the trees: a critical evaluation of methods to estimate the parameters of stochastic differential equations. J. Financial Econometrics, 5, 390-455.

Jimenez J.C. (2012a) Simplified formulas for the mean and variance of linear stochastic differential equations. To appear in J. Comput. Appl. Math. http://arxiv.org/pdf/1207.5067.pdf.

Jimenez J.C. (2012b) Approximate linear minimum variance filters for continuous-discrete state space models: convergence and practical algorithms. To appear in Int. J. Control. http://arxiv.org/pdf/1207.6023v2.pdf.

Jimenez J.C. and Biscay R. (2002) Approximation of continuous time stochastic processes by the Local Linearization method revisited. Stochast. Anal. & Appl., 20, 105-121.

Jimenez J.C., Biscay R. and Ozaki T. (2006) Inference methods for discretely observed continuous-time stochastic volatility models: A commented overview, Asia-Pacific Financial Markets, 12, 109-141.

Jimenez J.C. and de la Cruz H. (2012) Convergence rate of strong Local Linearization schemes for stochastic differential equations with additive noise, BIT, 52, 357-382.

Jimenez J.C. and Ozaki T. (2002) Linear estimation of continuous-discrete linear state space models with multiplicative noise, Systems & Control Letters, 47, 91-101.

Jimenez J.C. and Ozaki T. (2003) Local Linearization filters for nonlinear continuous-discrete state space models with multiplicative noise. Int. J. Control, 76, 1159-1170.

Jimenez J.C., Shoji I. and Ozaki T. (1999) Simulation of stochastic differential equations through the Local Linearization method. A comparative study, J. Statist. Physics, 94, 587-602.

Kessler M. (1997). Estimation of an ergodic diffusion from discrete observations, Scand. J. Statist., 24, 211-229.

Kloeden P.E. and Platen E. (1999) Numerical Solution of Stochastic Differential Equations, Springer-Verlag, Berlin, Third Edition.

Ljung L. and Caines P.E. (1979) Asymptotic normality of prediction error estimators for approximate system models, Stochastics, 3, 29-46.

Moler C. and Van Loan C. (2003) Nineteen dubious ways to compute the exponential of a matrix, SIAM Review, 45, 3-49.

Nielsen J.N., Madsen H. and Young, P. C. (2000) Parameter estimation in stochastic differential equations: an overview. Annual Review of Control, 24, 83-94.

Oppenheim A.V. and Schafer R.W. (2010) Discrete-Time Signal Processing, Prentice Hall, Third Edition.

Ozaki T. (1985) Statistical identification of storage model with application to stochastic hydrology, Water Resourse Bulletin, 21, 663-675.

Ozaki T. (1992) A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach, Statistica Sinica, 2, 113-135.

Prakasa-Rao B.L.S. (1983) Asymptotic theory for nonlinear least squares estimator for diffusion processes, Mathematische Operationsforschang Statistik Serie Statistik, 14, 195-209.

Prakasa-Rao B.L.S. (1999) Statistical inference for diffusion type processes. Oxford University Press.

Shoji I. and Ozaki T. (1997) Comparative study of estimation methods for continuous time stochastic processes, J. Time Ser. Anal., 18, 485–506.

Shoji I. and Ozaki T. (1998) A statistical method of estimation and simulation for systems of stochastic differential equations, Biometrika, 85, 240–243.

Singer H. (2002) Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood versus extended Kalman filter and Ito-Taylor expansion, J. Comput. Graphical Statist., 11, 972-995.

Wooldridge J.M. (1994) Estimation and inference for dependent processes. In Engle R.F. and McFadden D. (Eds.) Handbook of Econometrics, 4, 2639-2738.

Yoshida N. (1992) Estimation for diffusion processes from discrete observation. J. Multivariate Anal., 41, 220–242.

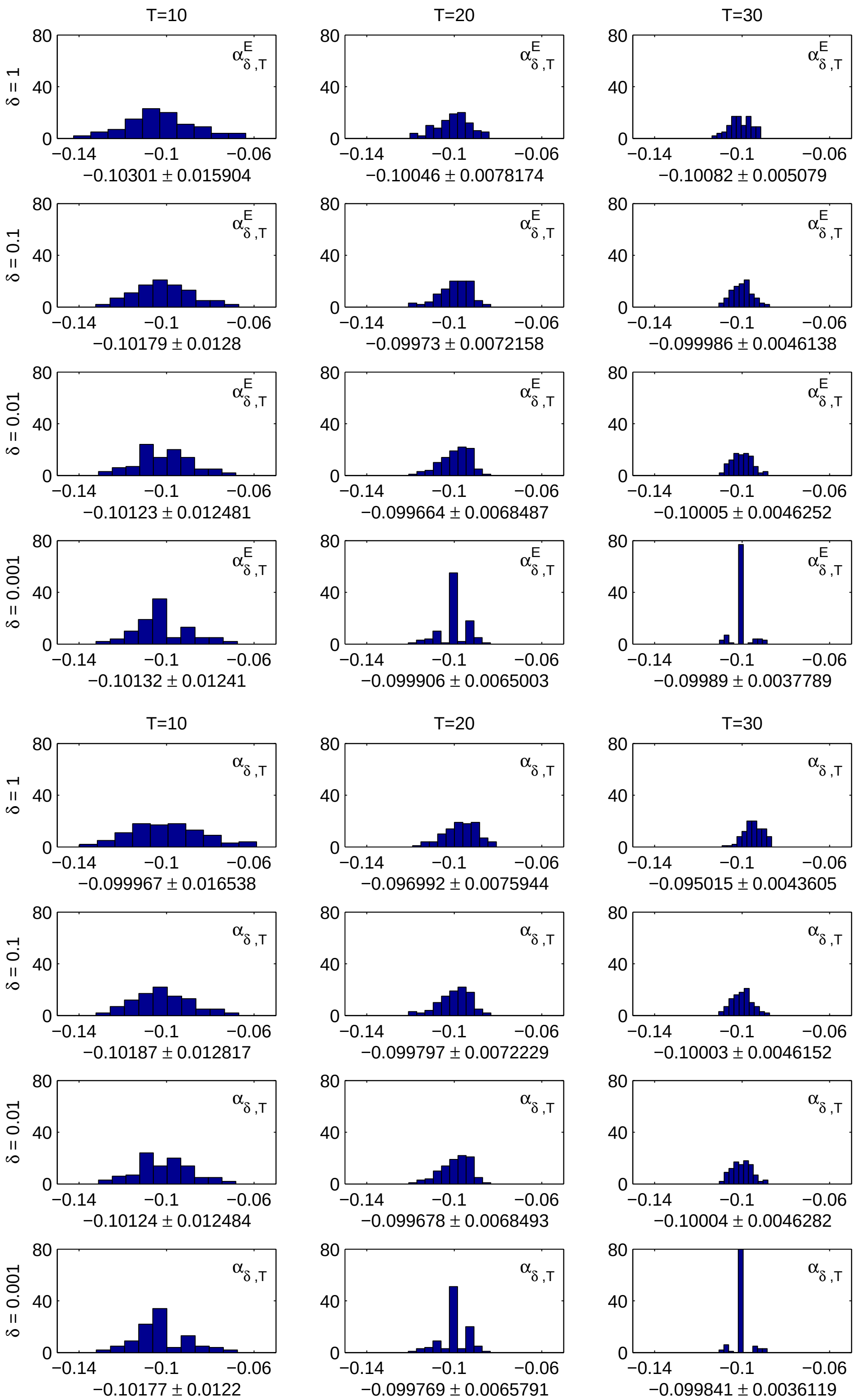

Fig. 1 Histograms and confidence limits for the exact ($\widehat{\alpha}^{E}_{\delta,T}$) and the conventional ($\widehat{\alpha}_{\delta,T}$) QML estimators of α computed from the Example 1 data with sampling period δ and time interval of length T.

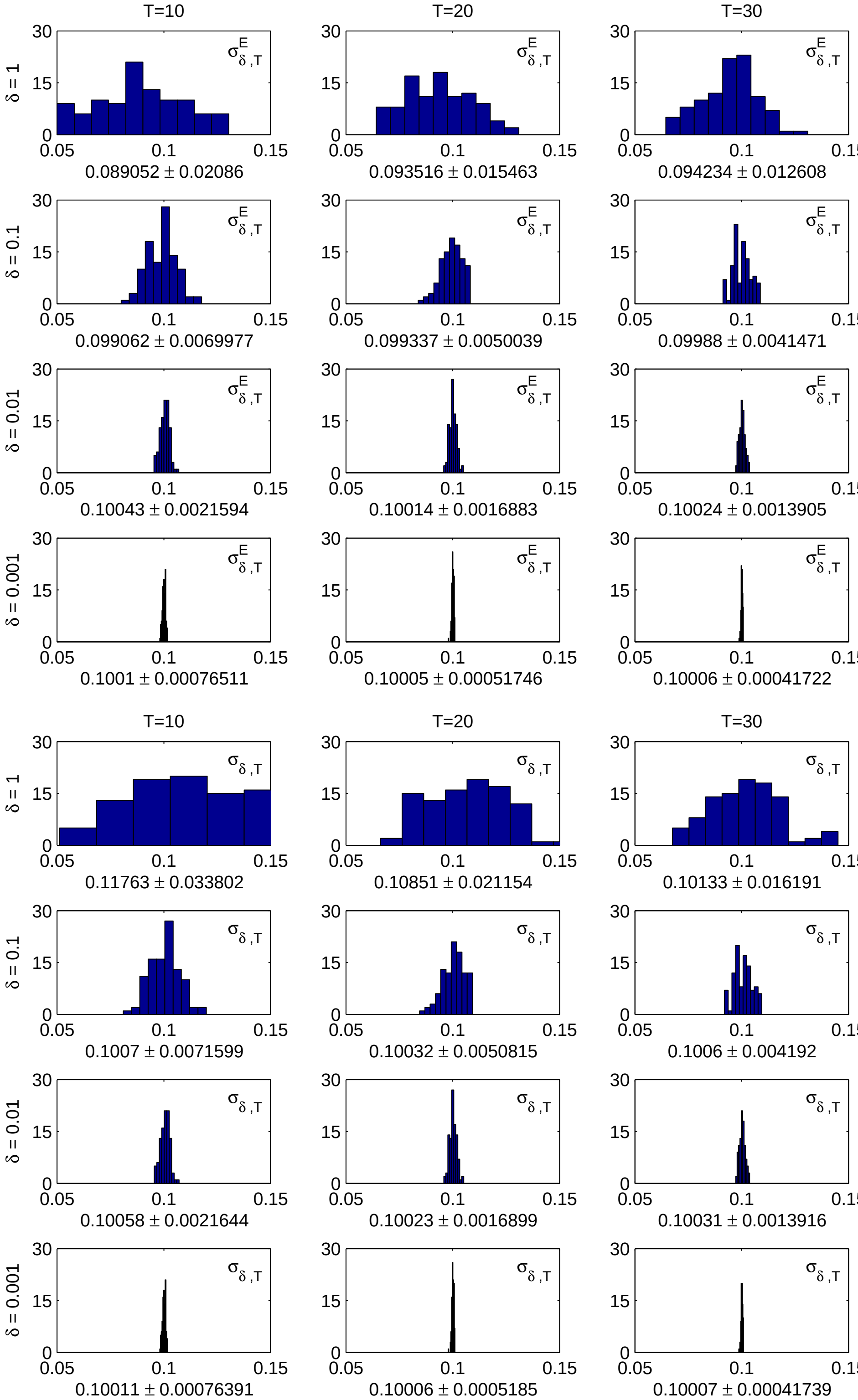

Fig. 2 Histograms and confidence limits for the exact $(\widehat{\sigma}^E_{\delta,T})$ and the conventional $(\widehat{\sigma}_{\delta,T})$ QML estimators of σ computed from the Example 1 data with sampling period δ and time interval of length T.

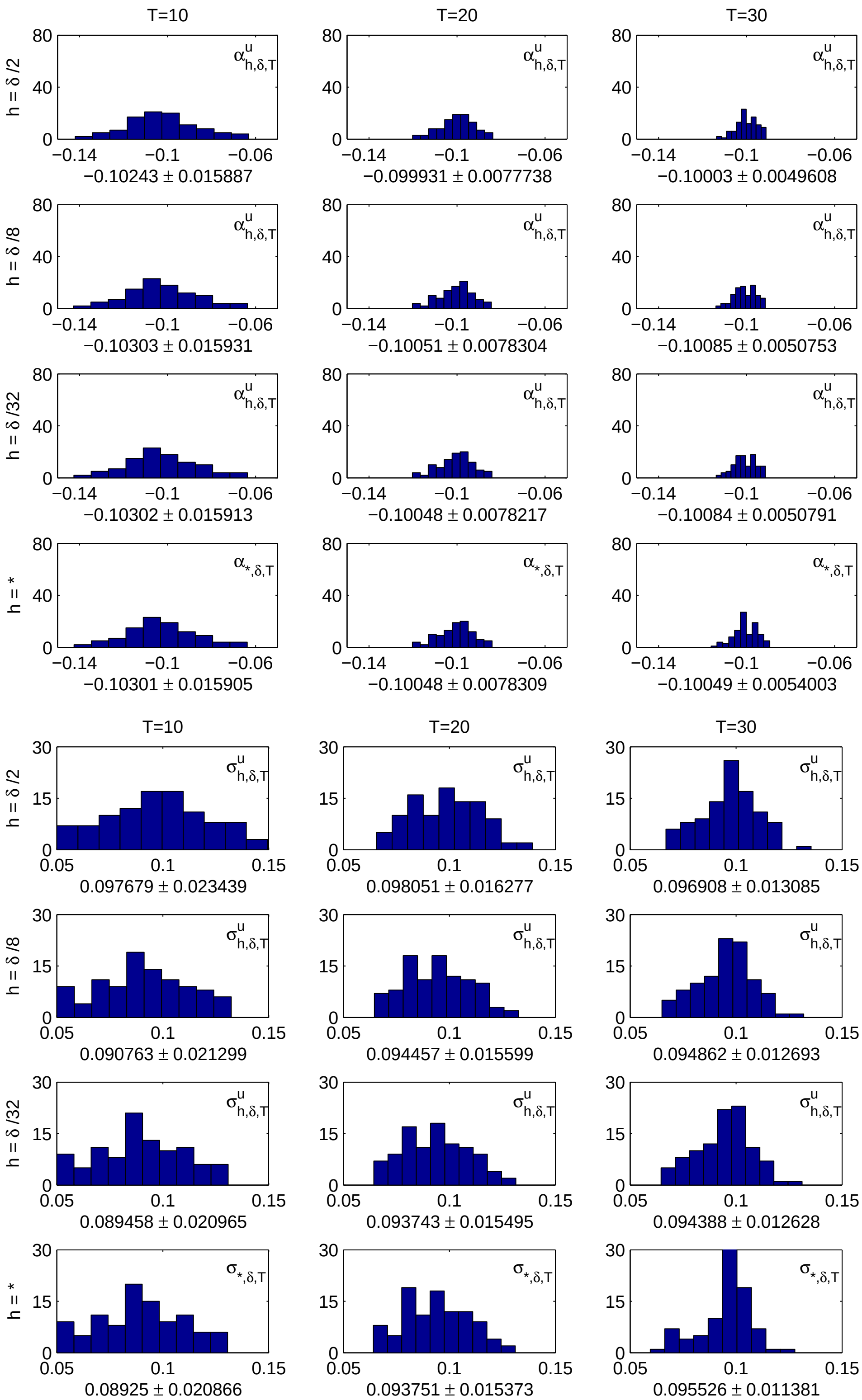

Fig. 3 Histograms and confidence limits for the oder-1 QML estimators of α and σ computed on uniform $(\tau)^u_{h,T}$ and adaptive $(\tau)_{\cdot,T}$ time discretizations from the Example 1 data with sampling period $\delta = 1$ and time interval of length T.

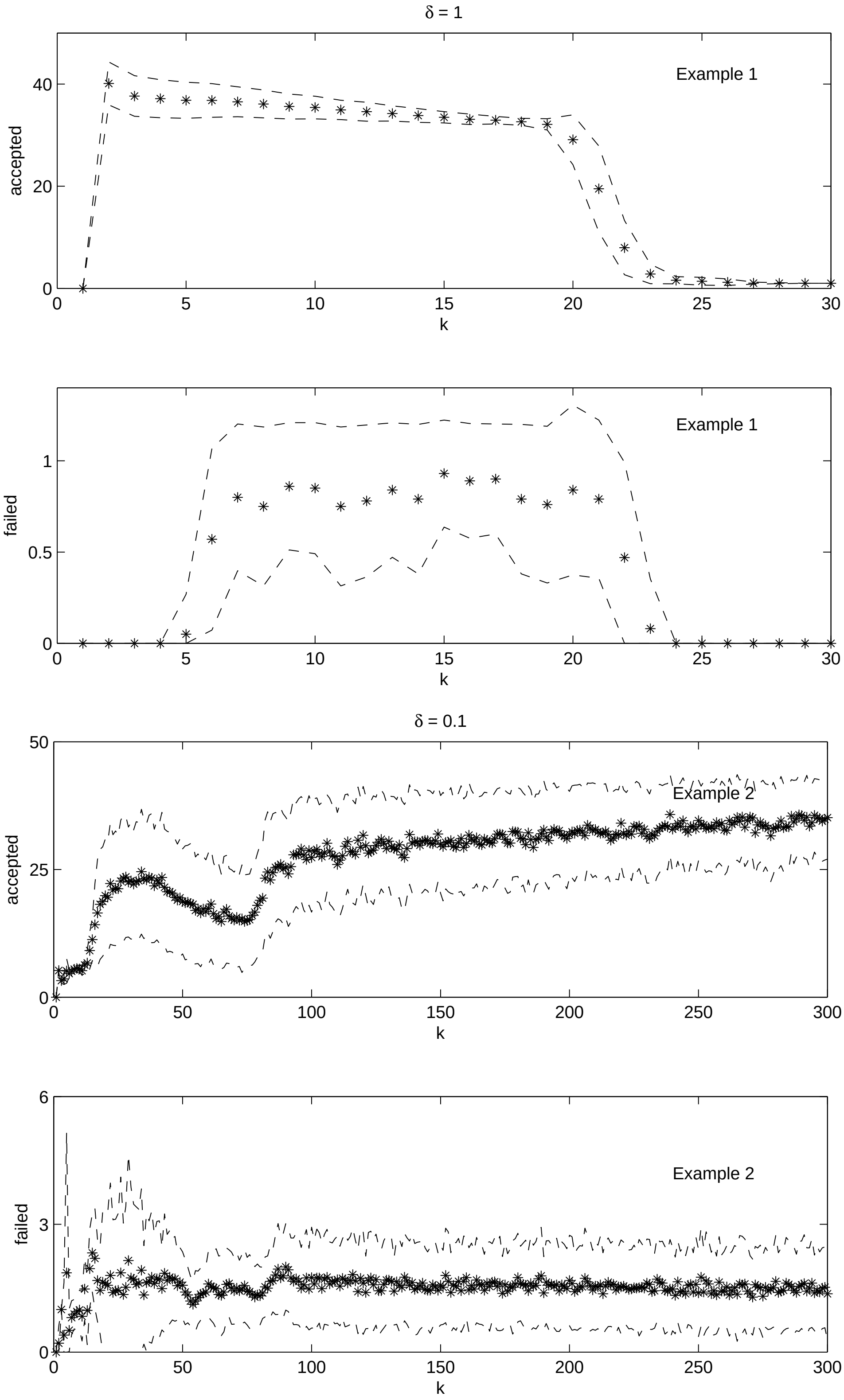

Fig. 4 Average (*) and 90% confidence limits (-) of accepted and failed steps of the adaptive QML estimator at each $t_k \in \{t\}_N$ in the Examples 1 and 2.

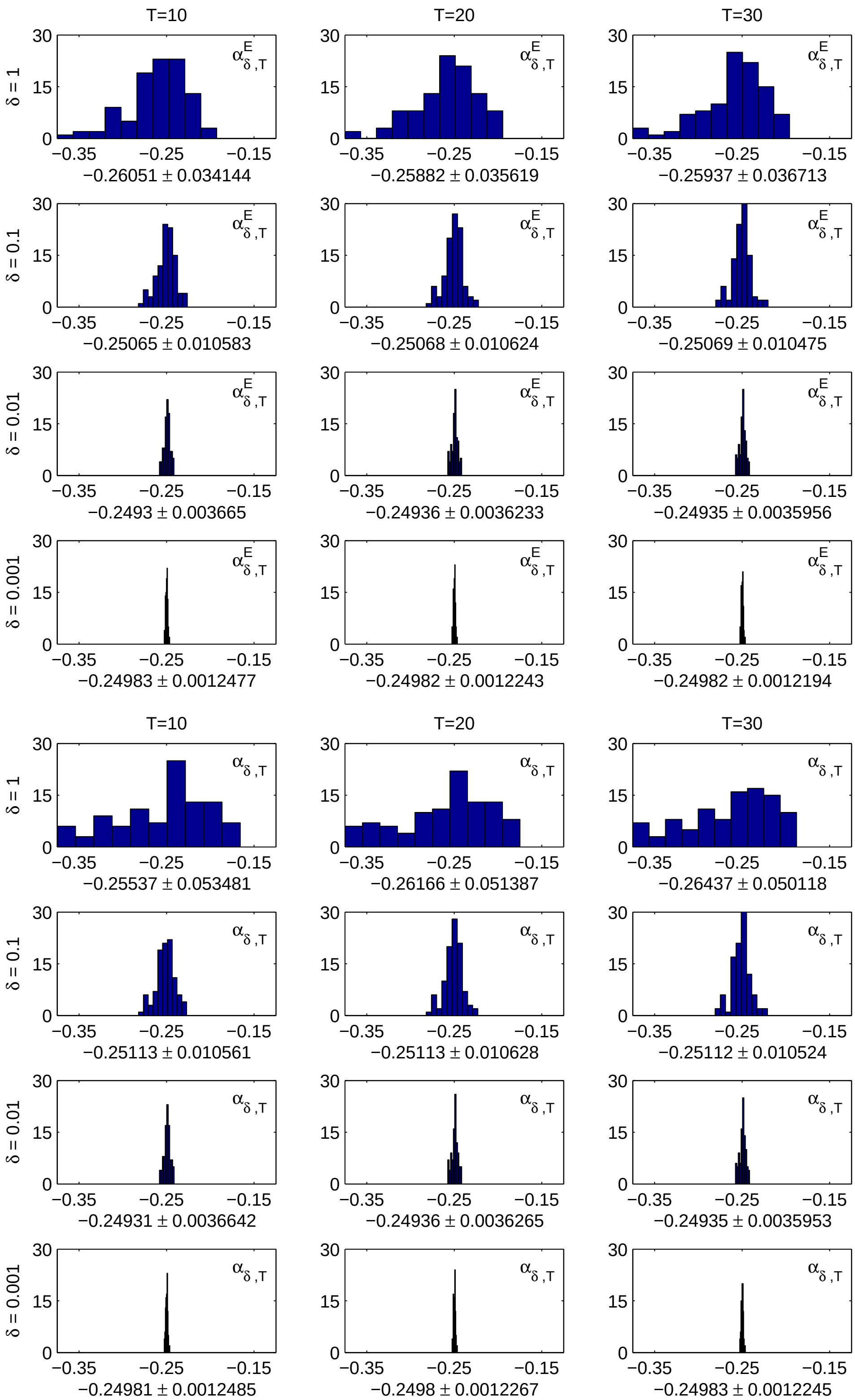

Fig. 5 Histograms and confidence limits for the exact ($\widehat{\alpha}^E_{\delta,T}$) and the conventional ($\widehat{\alpha}_{\delta,T}$) QML estimators of α computed from the Example 2 data with sampling period δ and time interval of length T.

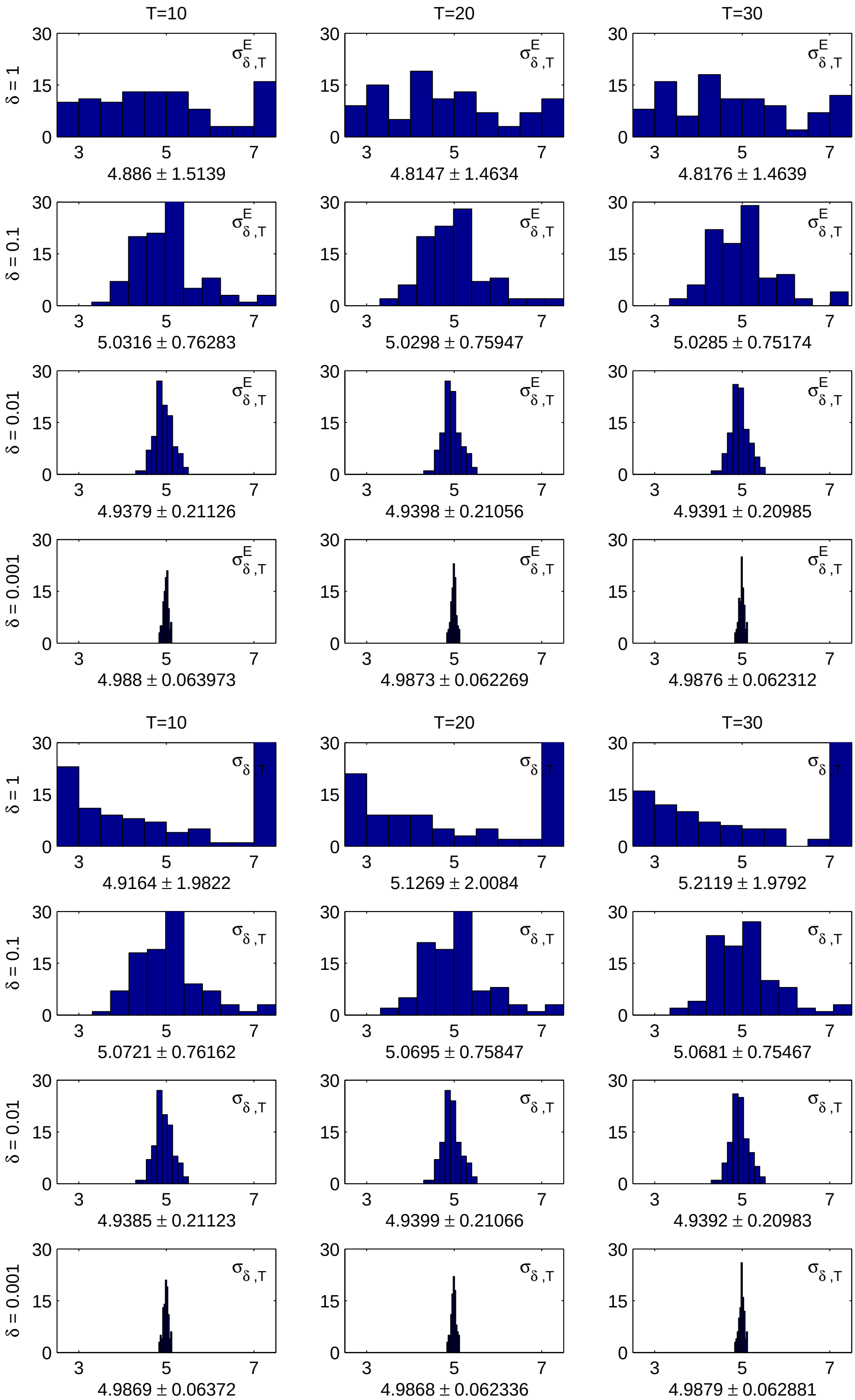

Fig. 6 Histograms and confidence limits for the exact $(\widehat{\sigma}^{E}_{\delta,T})$ and the conventional $(\widehat{\sigma}_{\delta,T})$ QML estimators of σ computed from the Example 2 data with sampling period δ and time interval of length T.

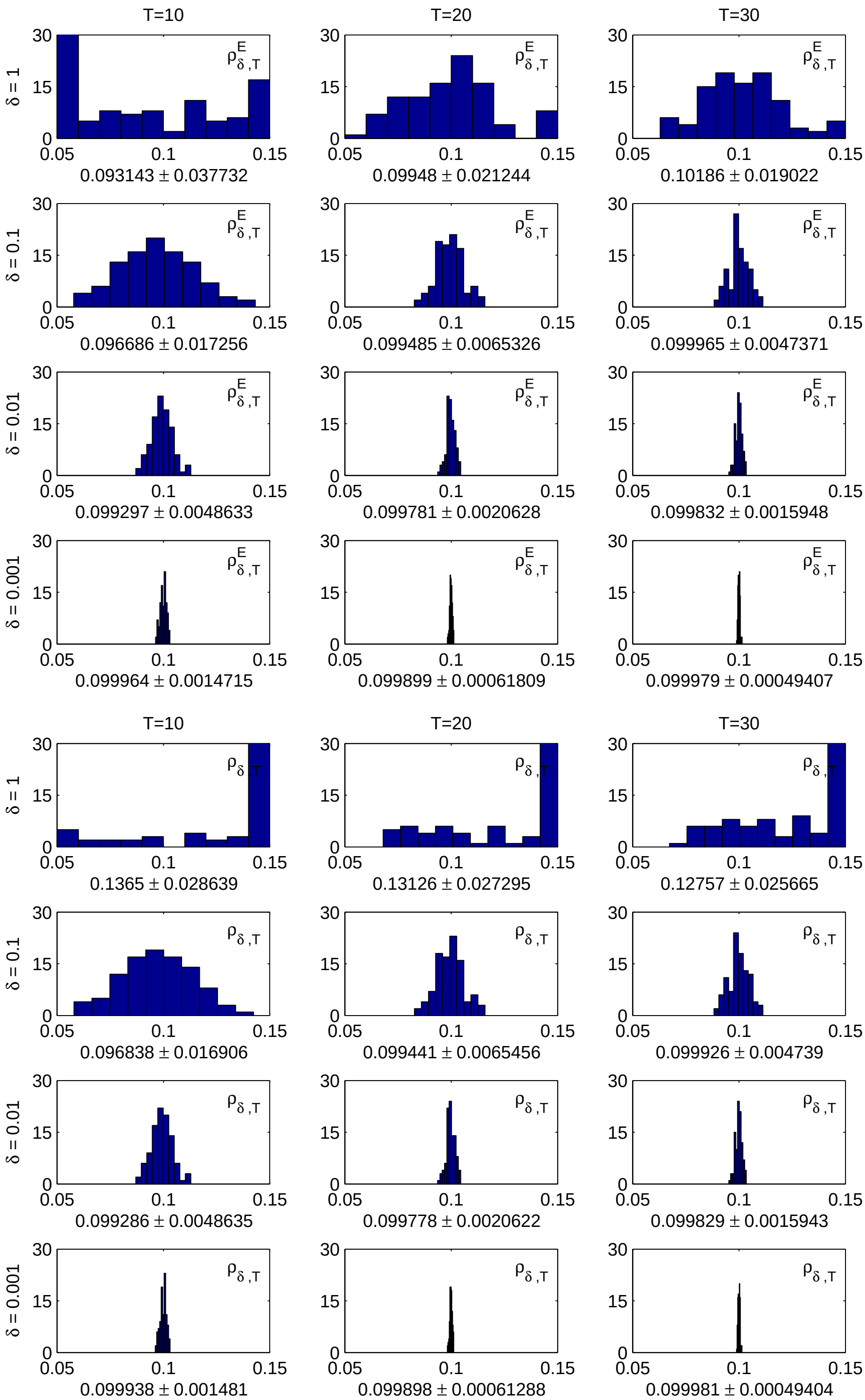

Fig. 7 Histograms and confidence limits for the exact ($\widehat{\rho}^E_{\delta,T}$) and the conventional ($\widehat{\rho}_{\delta,T}$) QML estimators of ρ computed from the Example 2 data with sampling period δ and time interval of length T.

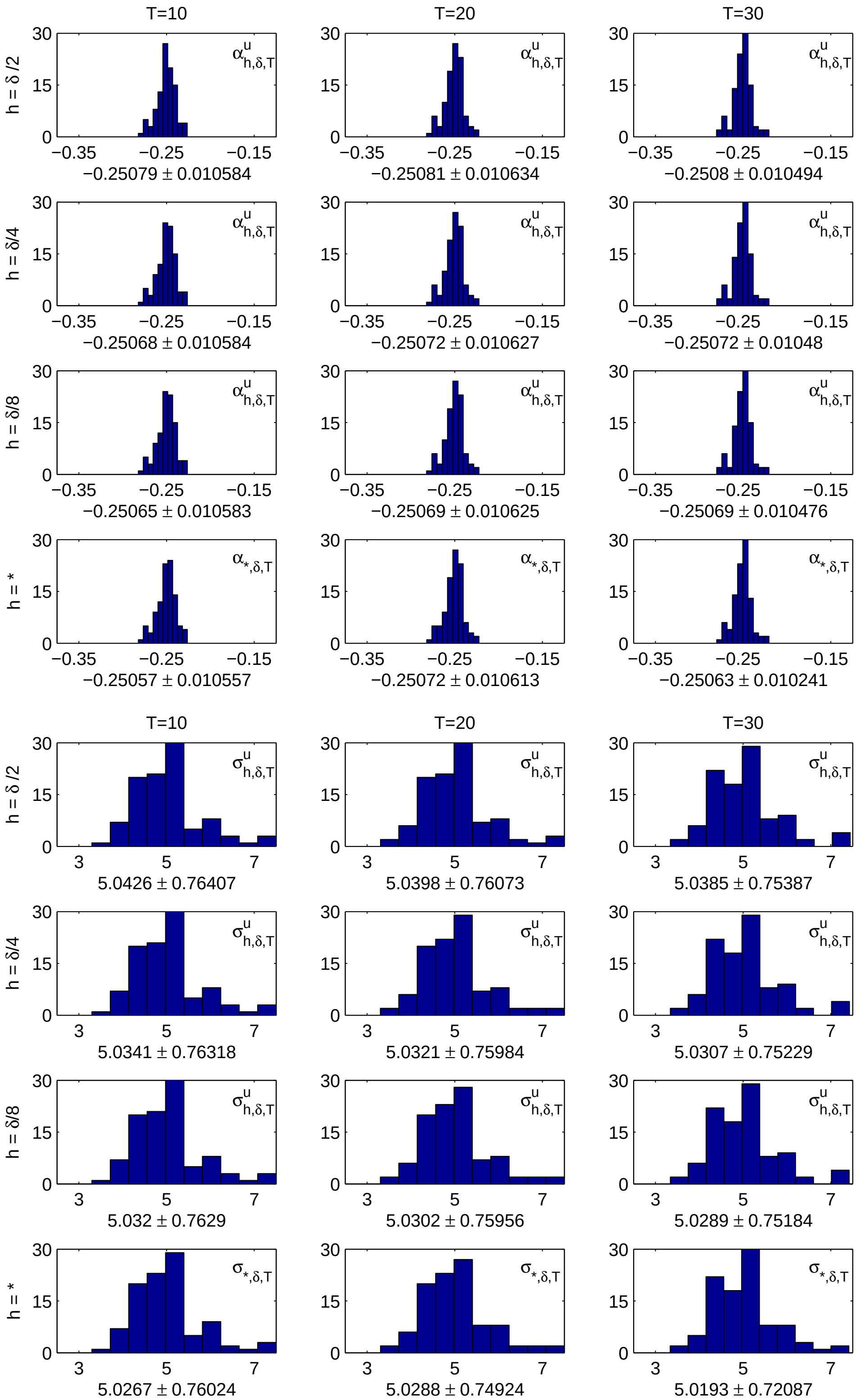

Fig. 8a Histograms and confidence limits for the oder-1 QML estimators of α and σ computed on uniform $(\tau)_{h,T}^{u}$ and adaptive $(\tau)_{\cdot,T}$ time discretizations from the Example 2 data with sampling period $\delta = 0.1$ and time interval of length T.

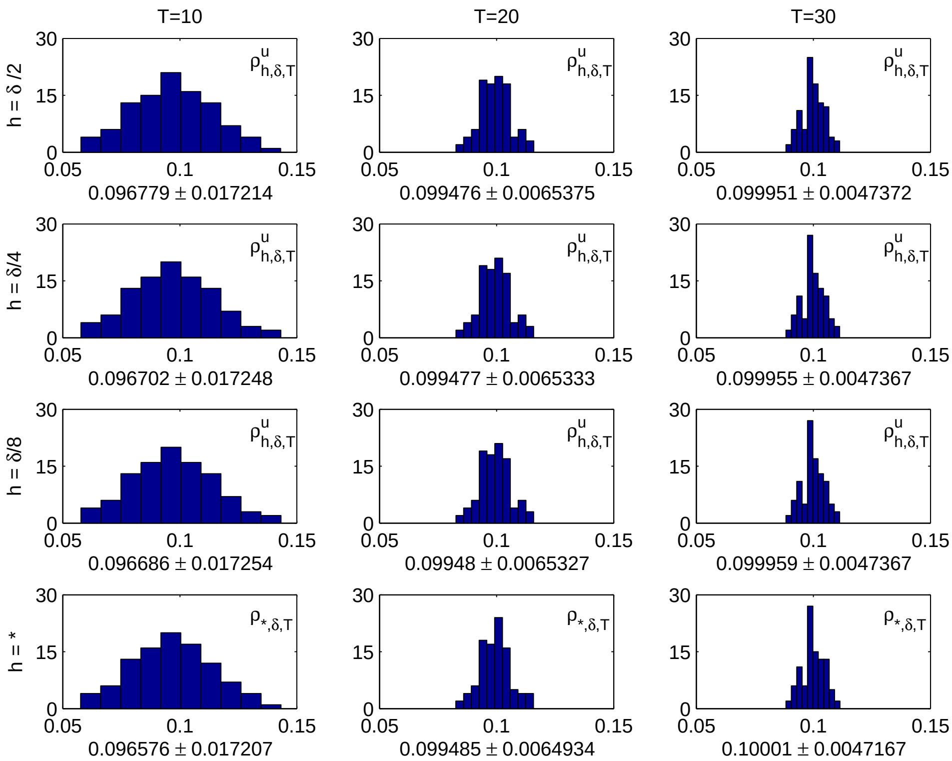

Fig. 8b Histograms and confidence limits for the oder-1 QML estimators of ρ computed on uniform $(\tau)^u_{h,T}$ and adaptive $(\tau)_{\cdot,T}$ time discretizations from the Example 2 data with sampling period $\delta = 0.1$ and time interval of length T.

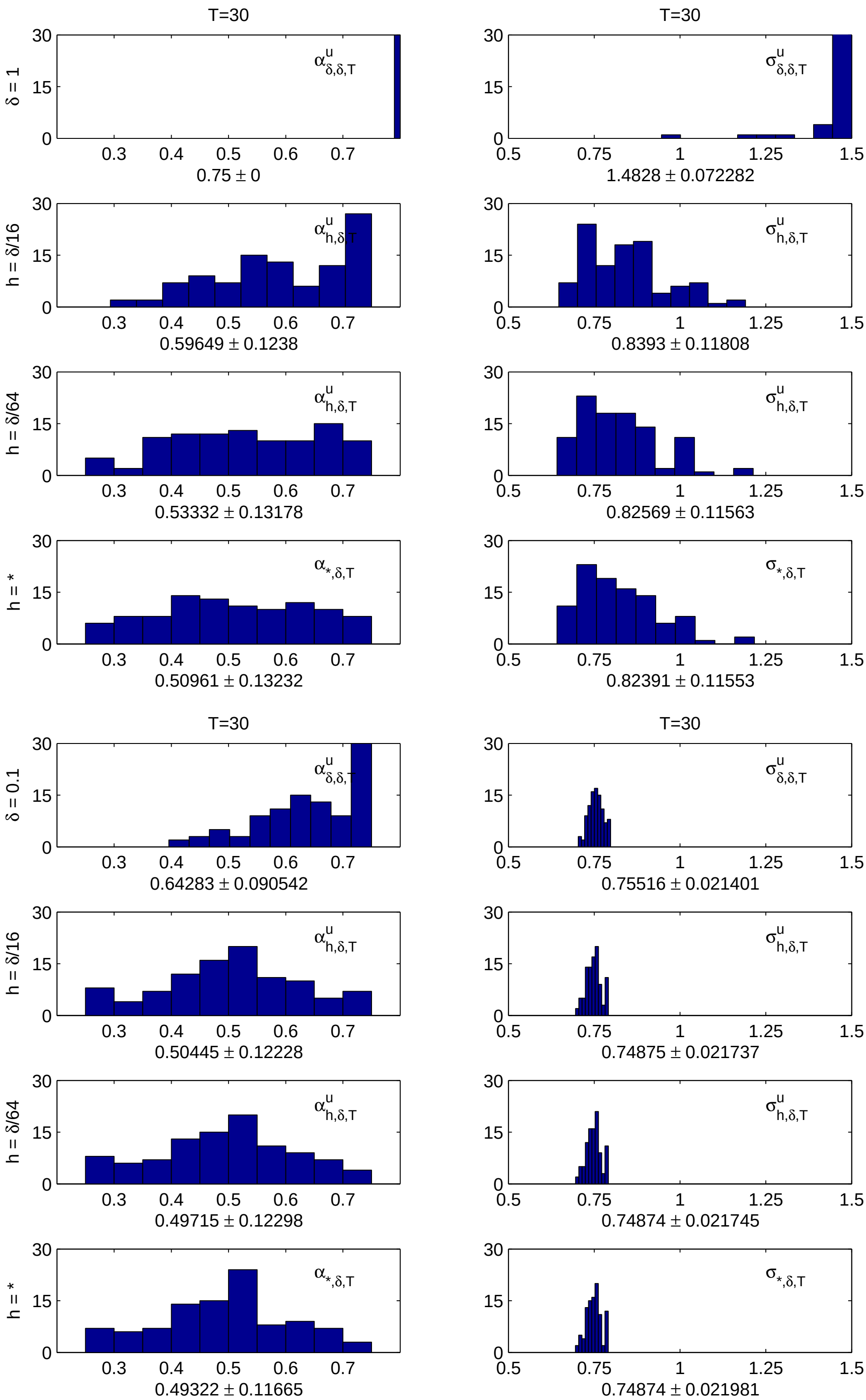

Fig. 9 Histograms and confidence limits for the oder-1 QML estimators of α and σ computed on uniform $(\tau)^{u}_{h,T}$ and adaptive $(\tau)_{\cdot,T}$ time discretizations from the Example 3 data with sampling period δ and time interval of length $T = 30$.

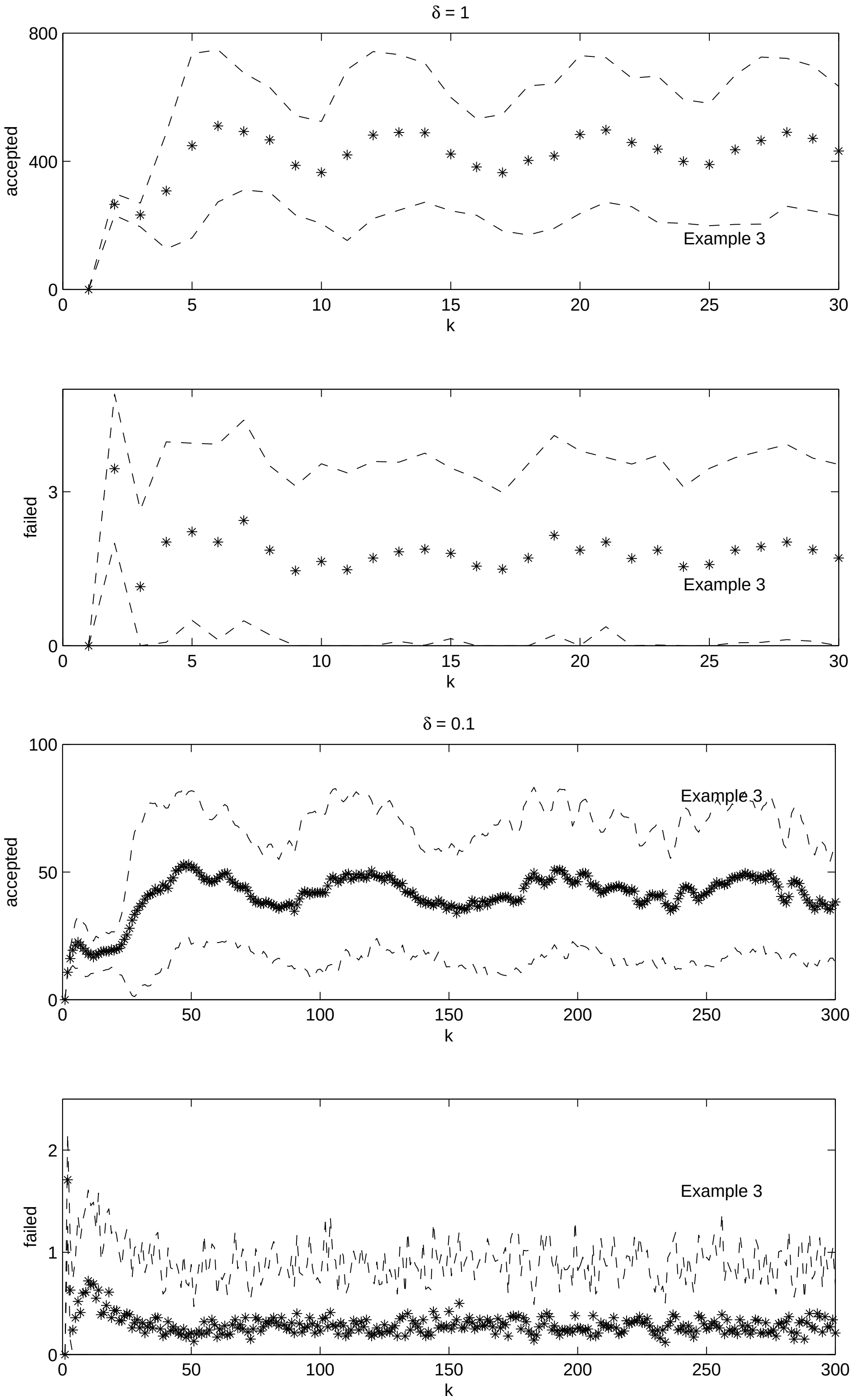

Fig. 10 Average (*) and 90% confidence limits (-) of accepted and failed steps of the adaptive QML estimator at each $t_k \in \{t\}_N$ in the Example 3.

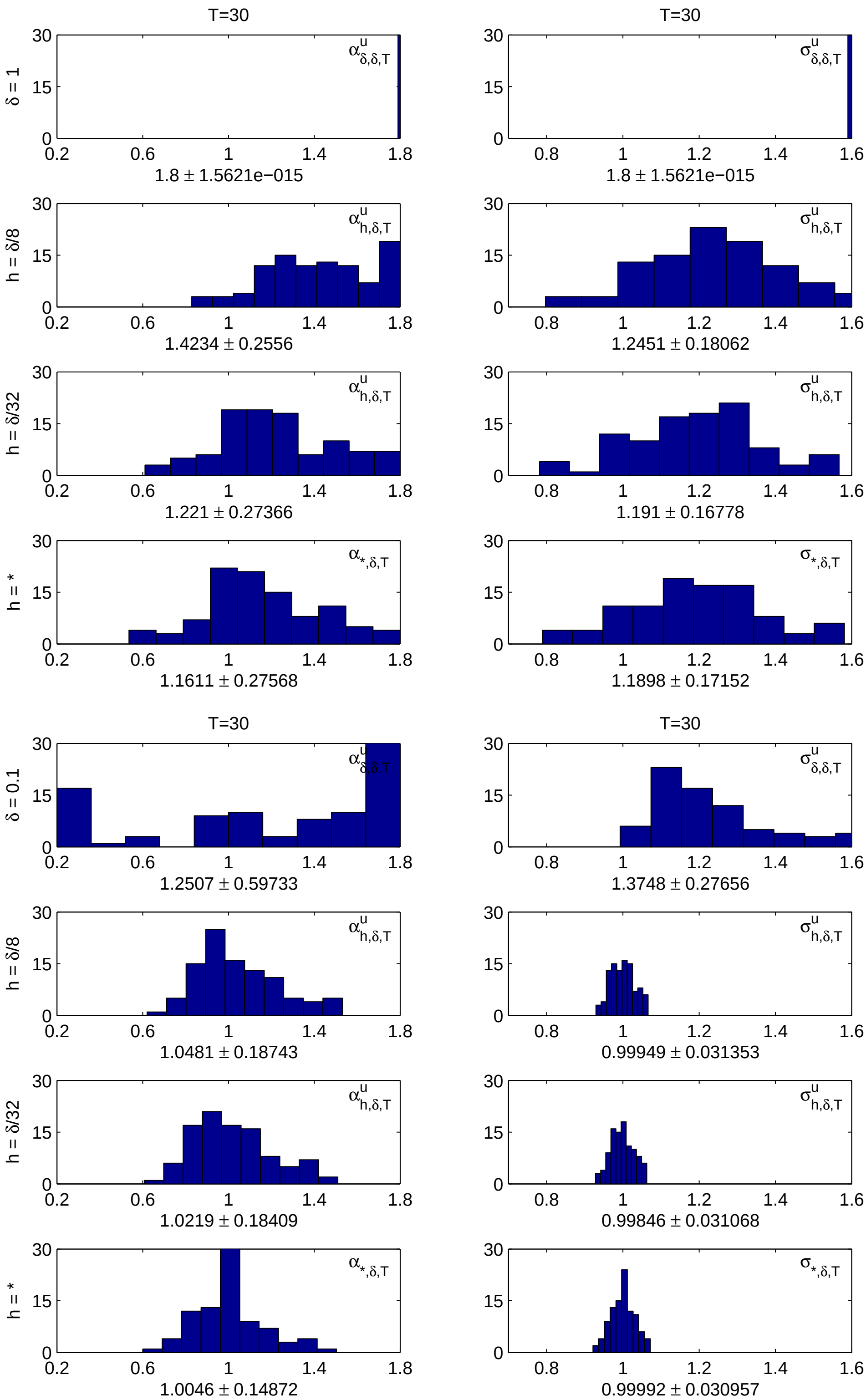

Fig. 11 Histograms and confidence limits for the oder-1 QML estimators of α and σ computed on uniform $(\tau)^u_{h,T}$ and adaptive $(\tau)_{\cdot,T}$ time discretizations from the Example 4 data with sampling period δ and time interval of length $T = 30$.

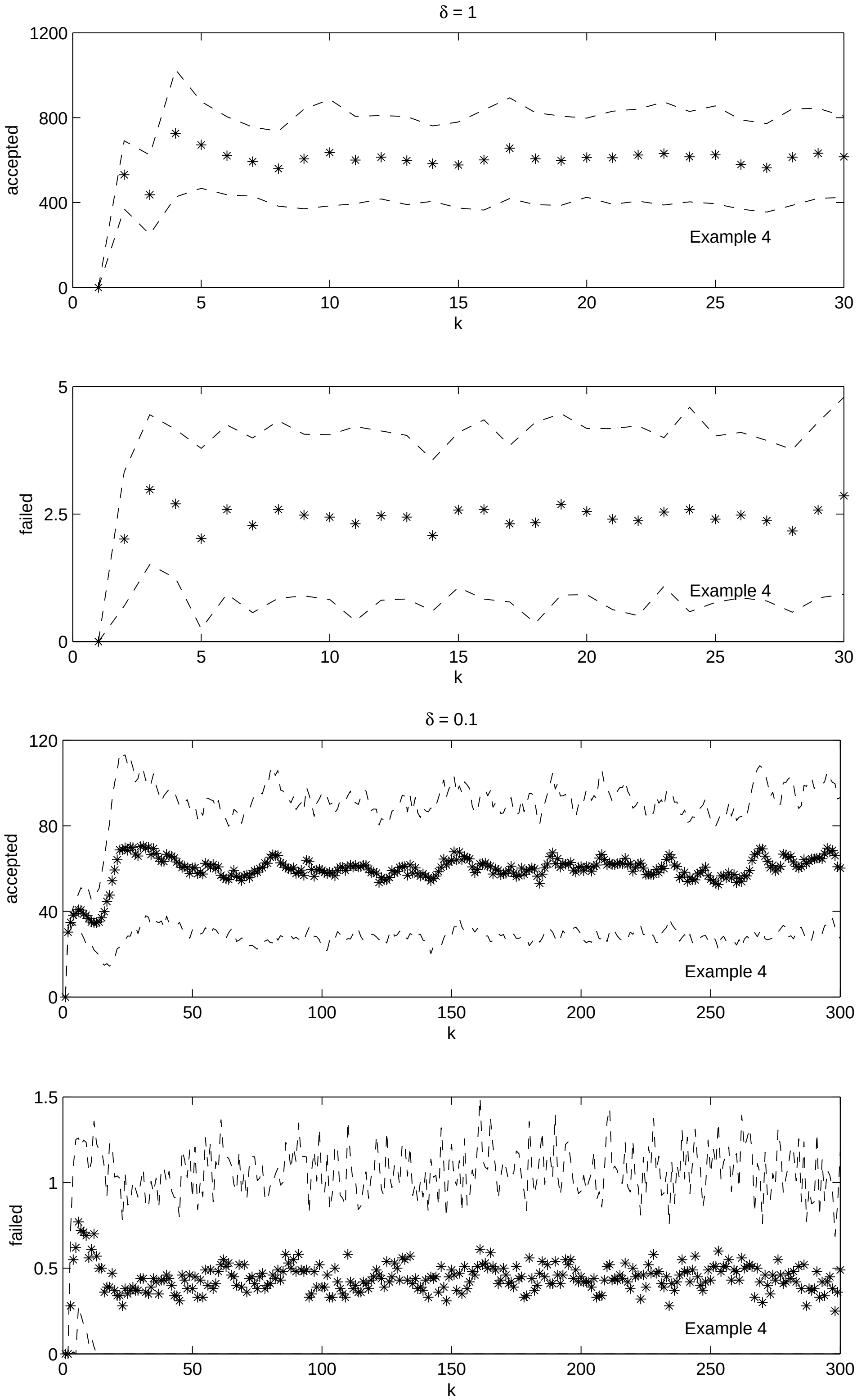

Fig. 12 Average (*) and 90% confidence limits (-) of accepted and failed steps of the adaptive QML estimator at each $t_k \in \{t\}_N$ in the Example 4.