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

Approximate continuous-discrete filters for the estimation of diffusion processes from partial and noisy observations

J.C. Jimenez

Received: date / Accepted: date

Abstract In this paper, an alternative approximation to the innovation method is introduced for the parameter estimation of diffusion processes from partial and noisy observations. This is based on a convergent approximation to the first two conditional moments of the innovation process through approximate continuous-discrete filters of minimum variance. It is shown that, for finite samples, the resulting approximate estimators converge to the exact one when the error of the approximate filters decreases. For an increasing number of observations, the estimators are asymptotically normal distributed and their bias decreases when the above 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 partial and noisy observations distant in time.

Keywords system identification · innovation estimator · diffusion process · stochastic differential equation · linear minimum variance filter · local linearization filter

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. In addition, if only some components of the diffusion process contaminated with noise are observed, then an extra complication arises. Typically, in this situation, the statistical problem under consideration is reformulated in the framework of continuous-discrete state space models, where the SDE to be estimated defines the continuous state equation and the given observations are described in terms of an discrete observation equation. For such class of models, a number of estimators based on analytical and simulated approximations have been developed in the last four decades. See, for instance, Nielsen et. al (2000a) and Jimenez et al. (2006) for a review.

In particular, the present paper deals with the class of innovation estimators for the parameter estimation of SDEs given a time series of partial and noisy observations. These are the estimators obtained by maximizing a normal log-likelihood function of the discrete-time innovations associated with the underlying continuous-discrete state space model. Approximations to this class of estimators have been derived by approximating the discrete-time innovations by means of inexact filters. With this purpose, approximate continuous-discrete filters like the Local Linearization (Ozaki 1994, Shoji 1998, Jimenez & Ozaki 2006), the extended Kalman (Nielsen & Madsen 2001, Singer 2002), and the second order (Nielsen et al. 2000b, Singer

J.C. Jimenez
Departamento de Matemática Interdisciplinaria,
Instituto de Cibernética, Matemática y Física,
Calle 15, no. 551, Vedado, La Habana, Cuba.
email: jcarlos@icimaf.cu

2002) filters have been used, as well as, discrete-discrete filters after the discretization of the SDE by means of a numerical scheme (Ozaki & Iino 2001, and Peng et al. 2002). The approximate innovation estimators obtained in this way have been useful for the identification, from actual data, of a variety of neurophysiological, financial and molecular models among others (see, e.g., Calderon 2009, Chiarella et al. 2009, Jimenez et al. 2006, Riera et al. 2004, Valdes et al. 1999). However, a common feature of the approximate innovation estimators mentioned above is that, once the observations are given, the error between the approximate and the exact innovations is fixed and completely determined by the distance between observations. Clearly, this fixes the bias of the approximate estimators for finite samples and obstructs its asymptotic correction when the number of observations increases.

In this paper, an alternative approximation to the innovation estimator for diffusion processes is introduced, which is oriented to reduce and control the estimation bias. This is based on a recursive approximation of the first two conditional moments of the innovation process through approximate filters that converge to the linear one of minimum variance. It is shown that, for finite samples, the resulting approximate estimators converge to the exact one when the error of the approximate filters decreases. For an increasing number of observations, they are asymptotically normal distributed and their bias decreases when the above mentioned error does it. As a particular instance, the approximate innovation estimators designed with the order-β Local Linearization filters are presented. Their convergence, practical algorithms and performance in simulations are also considered in detail. The simulations show that, with respect to the conventional approximations to the innovation estimators, the new approximate estimators significantly enhance the parameter estimation of the test equations given a reduced number of partial and noisy 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 studied. As a particular instance, the order-β innovation estimator based on convergent Local Linearization filters 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 $\mathcal{F}$, and $\mathbf{x}$ be a d-dimensional diffusion process defined by the 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 $\boldsymbol{\theta} \in \mathcal{D}_\theta$.

Let us consider the state space model defined by the continuous state equation (1) and the discrete observation equation

$$\mathbf{z}(t_k) = \mathbf{C}\mathbf{x}(t_k) + \mathbf{e}_{t_k}, \text{ for } k = 0, 1, .., M-1, \tag{2}$$

where $\{\mathbf{e}_{t_k} : \mathbf{e}_{t_k} \sim \mathcal{N}(0, \boldsymbol{\Pi}_{t_k}),\ k = 0, .., M-1\}$ is a sequence of r-dimensional i.i.d. Gaussian random vectors independent of $\mathbf{w}$, $\boldsymbol{\Pi}_{t_k}$ an $r \times r$ positive semi-definite matrix, and $\mathbf{C}$ an $r \times d$ matrix. Here, it is assumed that the M time instants t_k define an increasing sequence $\{t\}_M = \{t_k : t_k < t_{k+1},\ k = 0, 1, .., M-1\}$.

Suppose that, through (2), M partial and noisy observations of the diffusion process $\mathbf{x}$ defined by (1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0 \in \mathcal{D}_\theta$ are given on $\{t\}_M$. In particular, denote by $Z = \{\mathbf{z}_{t_0}, .., \mathbf{z}_{t_{M-1}}\}$ the sequence of these observations, where $\mathbf{z}_{t_k}$ denotes the observation at t_k for all $t_k \in \{t\}_M$.

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. Specifically, let us consider the innovation estimator defined as follows.

Definition 1 (Ozaki, 1994) Given M observations Z of the continuous-discrete state space model (1)-(2) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0 \in \mathcal{D}_\theta$ on $\{t\}_M$,

$$\widehat{\boldsymbol{\theta}}_M = \arg\{\min_{\boldsymbol{\theta}} U_M(\boldsymbol{\theta}, Z)\} \tag{3}$$

defines the innovation estimator of $\boldsymbol{\theta}_0$, where

$$U_M(\boldsymbol{\theta}, Z) = (M-1)\ln(2\pi) + \sum_{k=1}^{M-1} \ln(\det(\boldsymbol{\Sigma}_{t_k})) + \boldsymbol{\nu}_{t_k}^{\top} (\boldsymbol{\Sigma}_{t_k})^{-1} \boldsymbol{\nu}_{t_k},$$

$\boldsymbol{\nu}_{t_k}$ is the discrete innovations of the model (1)-(2) and $\boldsymbol{\Sigma}_{t_k}$ the innovation variance for all $t_k \in \{t\}_M$.

In the above definition,

$$\boldsymbol{\nu}_{t_k} = \mathbf{z}_{t_k} - \mathbf{C}\mathbf{x}_{t_k/t_{k-1}} \qquad \text{and} \qquad \boldsymbol{\Sigma}_{t_k} = \mathbf{C}\mathbf{U}_{t_k/t_{k-1}}\mathbf{C}^{\intercal} + \boldsymbol{\Pi}_{t_k},$$

where $\mathbf{x}_{t_k/t_{k-1}} = E(\mathbf{x}(t_k)|Z_{t_{k-1}})$ and $\mathbf{U}_{t_k/t_{k-1}} = E(\mathbf{x}(t_k)\mathbf{x}^{\intercal}(t_k)|Z_{t_{k-1}}) - \mathbf{x}_{t_k/t_{k-1}}\mathbf{x}^{\intercal}_{t_k/t_{k-1}}$ denote the conditional mean and variance of the diffusion process $\mathbf{x}$ at t_k given the observations $Z_{t_{k-1}} = \{\mathbf{z}_{t_0}, .., \mathbf{z}_{t_{k-1}}\}$ for all $t_{k-1}, t_k \in \{t\}_M$ and $\boldsymbol{\theta} \in \mathcal{D}_\theta$. Here, the predictions $E(\mathbf{x}(t_k)|Z_{t_{k-1}})$ and $E(\mathbf{x}(t_k)\mathbf{x}^{\intercal}(t_k)|Z_{t_{k-1}})$ are recursively computed through the Linear Minimum Variance filter for the model (1)-(2). Because the first two conditional moments of $\mathbf{x}$ are correctly specified, Theorem 1 in Ljung & Caines (1979) for prediction error estimators implies the consistent and the asymptotic normality of the innovation estimator (3) under conventional regularity conditions (Ozaki 1994, Nolsoe et al. 2000).

In general, since the conditional mean and variance of equation (1) have not explicit formulas, approximations to them are needed. If $\widetilde{\mathbf{x}}_{t_k/t_{k-1}}$ and $\widetilde{\mathbf{U}}_{t_k/t_{k-1}}$ are approximations to $\mathbf{x}_{t_k/t_{k-1}}$ and $\mathbf{U}_{t_k/t_{k-1}}$, then the estimator

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

with

$$\widetilde{U}_M(\boldsymbol{\theta}, Z) = (M-1)\ln(2\pi) + \sum_{k=1}^{M-1} \ln(\det(\widetilde{\boldsymbol{\Sigma}}_{t_k})) + \widetilde{\boldsymbol{\nu}}_{t_k}^{\top} (\widetilde{\boldsymbol{\Sigma}}_{t_k})_{t_k}^{-1} \widetilde{\boldsymbol{\nu}}_{t_k}$$

provides an approximation to the innovation estimator (3), where

$$\widetilde{\boldsymbol{\nu}}_{t_k} = \mathbf{z}_{t_k} - \mathbf{C}\widetilde{\mathbf{x}}_{t_k/t_{k-1}} \qquad \text{and} \qquad \widetilde{\boldsymbol{\Sigma}}_{t_k} = \mathbf{C}\widetilde{\mathbf{U}}_{t_k/t_{k-1}}\mathbf{C}^{\intercal} + \boldsymbol{\Pi}_{t_k}$$

are approximations to $\boldsymbol{\nu}_{t_k}$ and $\boldsymbol{\Sigma}_{t_k}$.

Approximate estimators of this type have early been considered in a number of papers. Approximate continuous-discrete filters like Local Linearization filters (Ozaki 1994, Shoji 1998, Jimenez & Ozaki 2006), extended Kalman filter (Nielsen & Madsen 2001, Singer 2002), and second order filters (Nielsen et al. 2000b, Singer 2002) have been used for approximating the values of $\boldsymbol{\nu}_{t_k}$ and $\boldsymbol{\Sigma}_{t_k}$. On the other hand, in Ozaki & Iino (2001) and Peng et. al (2002), discrete-discrete filters have also been used after the discretization of the equation (1) by means of a numerical scheme. In all these approximations, once the data Z are given (and so the time partition $\{t\}_M$ is specified), the error between $\widetilde{\boldsymbol{\nu}}_{t_k}$ and $\boldsymbol{\nu}_{t_k}$ is completely settled by $t_k - t_{k-1}$ and can not be reduced. In this way, the difference between the approximate innovation estimator $\widehat{\boldsymbol{\vartheta}}_M$ and the exact one $\widehat{\boldsymbol{\theta}}_M$ can not be reduced neither. Clearly, this is a important limitation of these approximate estimators. Nevertheless, in a number of practical situations (see Jimenez & Ozaki 2006, Jimenez et al. (2006), and references therein) the bias the approximate innovation estimators is negligible. Therefore, these estimators has been useful for the identification, from actual data, of a variety of neurophysiological, financial and molecular models among others as it was mentioned above. Further, in a simulation study with the Cox-Ingersoll-Ross model of short-term interest rate, approximate innovation methods have provided similar or better results than those obtained by prediction-based estimating functions but with much lower computational cost (Nolsoe et al., 2000). Similar results have been reported in a comparative study with the approximate likelihood via simulation method (Singer, 2002).

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 innovation 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

$$E\left(\mathbf{y}(t_0)\middle|\mathcal{F}_{t_0}\right) = E\left(\mathbf{x}(t_0)\middle|\mathcal{F}_{t_0}\right) \quad \text{and} \quad E\left(\mathbf{y}(t_0)\mathbf{y}^{\intercal}(t_0)\middle|\mathcal{F}_{t_0}\right) = E\left(\mathbf{x}(t_0)\mathbf{x}^{\intercal}(t_0)\middle|\mathcal{F}_{t_0}\right); \tag{4}$$

satisfying the bound condition

$$E\left(\left|\mathbf{y}(t)\right|^{2q}\middle|\mathcal{F}_{t_0}\right) \leq L \tag{5}$$

for all $t \in [t_0, t_{M-1}]$; and the weak convergence criteria

$$\sup_{t_k \leq t \leq t_{k+1}} \left| E\left(g(\mathbf{x}(t))\middle|\mathcal{F}_{t_k}\right) - E\left(g(\mathbf{y}(t))\middle|\mathcal{F}_{t_k}\right)\right| \leq L_k h^{\beta} \tag{6}$$

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). The second conditional moment of $\mathbf{y}$ is also assumed to be positive definite and continuous for all $\boldsymbol{\theta} \in \mathcal{D}_\theta$.

In addition, let us consider the following approximation to the Linear Minimum Variance (LMV) filter of the model (1)-(2).

Definition 2 (Jimenez 2012b) Given a time discretization $(\tau)_h \supset \{t\}_M$, the order-β Linear Minimum Variance filter for the state space model (1)-(2) is defined, between observations, by

$$\mathbf{y}_{t/t} = E(\mathbf{y}(t)/Z_t) \qquad \text{and} \qquad \mathbf{V}_{t/t} = E(\mathbf{y}(t)\mathbf{y}^{\intercal}(t)/Z_t) - \mathbf{y}_{t/t}\mathbf{y}_{t/t}^{\intercal} \tag{7}$$

for all $t \in (t_k, t_{k+1})$, and by

$$\mathbf{y}_{t_{k+1}/t_{k+1}} = \mathbf{y}_{t_{k+1}/t_k} + \mathbf{K}_{t_{k+1}}(\mathbf{z}_{t_{k+1}} - \mathbf{C}\mathbf{y}_{t_{k+1}/t_k}), \tag{8}$$

$$\mathbf{V}_{t_{k+1}/t_{k+1}} = \mathbf{V}_{t_{k+1}/t_k} - \mathbf{K}_{t_{k+1}}\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}, \tag{9}$$

for each observation at t_{k+1}, with filter gain

$$\mathbf{K}_{t_{k+1}} = \mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal}(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal} + \boldsymbol{\Pi}_{t_{k+1}})^{-1} \tag{10}$$

for all $t_k, t_{k+1} \in \{t\}_M$, where $\mathbf{y}$ is an order-β approximation to the solution of (1) in weak sense, and $Z_t = \{\mathbf{z}(t_k) : t_k \leq t,\ t_k \in \{t\}_M\}$ are given observations of (1)-(2) until the time instant t. The predictions $\mathbf{y}_{t/t_k} = E(\mathbf{y}(t)/Z_{t_k})$ and $\mathbf{V}_{t/t_k} = E(\mathbf{y}(t)\mathbf{y}^{\intercal}(t)/Z_{t_k}) - \mathbf{y}_{t/t_k}\mathbf{y}_{t/t_k}^{\intercal}$, with initial conditions $\mathbf{y}_{t_k/t_k}$ and $\mathbf{V}_{t_k/t_k}$, are defined for all $t \in (t_k, t_{k+1}]$ and $t_k, t_{k+1} \in \{t\}_M$.

Once an order-β approximation to the solution of equation (1) is chosen, and so an order-β LMV filter is specified, the following approximate innovation estimator can naturally be defined.

Definition 3 Given M observations Z of the continuous-discrete state space model (1)-(2) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0 \in \mathcal{D}_\theta$ on $\{t\}_M$, the order-β innovation estimator for the parameters of (1) is defined by

$$\widehat{\boldsymbol{\theta}}_M(h) = \arg\{\min_{\boldsymbol{\theta}} U_{M,h}(\boldsymbol{\theta}, Z)\}, \tag{11}$$

where

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

with $\boldsymbol{\nu}_{h,t_k} = \mathbf{z}_{t_k} - \mathbf{C}\mathbf{y}_{t_k/t_{k-1}}$ and $\boldsymbol{\Sigma}_{h,t_k} = \mathbf{C}\mathbf{V}_{t_k/t_{k-1}}\mathbf{C}^{\intercal} + \boldsymbol{\Pi}_{t_k}$, being $\mathbf{y}_{t_k/t_{k-1}}$ and $\mathbf{V}_{t_k/t_{k-1}}$ the prediction mean and variance of an order-β LMV filter for the model (1)-(2), and h the maximum stepsize of the time discretization $(\tau)_h \supset \{t\}_M$ associated to the filter.

In principle, according to the above definitions, any kind of approximation $\mathbf{y}$ converging to $\mathbf{x}$ in a weak sense can be used to construct an approximate order-β LMV filter and so an approximate order-β innovation estimator. In this way, the Euler-Maruyama, the Local Linearization and any high order numerical scheme for SDEs as those considered in Kloeden & Platen (1999) might be used as well. However, the approximations $\boldsymbol{\nu}_{h,t_k}$ and $\boldsymbol{\Sigma}_{h,t_k}$ to $\boldsymbol{\nu}_{t_k}$ and $\boldsymbol{\Sigma}_{t_k}$ in (3) at each t_k will be now derived from the predictions of approximate LMV filter after various iterations with stepsizes lower than $t_k - t_{k-1}$. Note that, when $(\tau)_h \equiv \{t\}_M$, an order-β LMV filter might reduce to some one of the conventional approximation to the exact LMV filter. In this situation, the corresponding order-β innovation estimator reduces to some one of the approximate innovation estimator mentioned in Section 2. In particular, to those considered in Ozaki (1994), Shoji (1998), Jimenez & Ozaki (2006) when Local Linearization schemes are used to define order-β LMV filters.

Note that the goodness of the approximation $\mathbf{y}$ to $\mathbf{x}$ is measured (in weak sense) by the left hand side of (6). Thus, the inequality (6) 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 $\mathbf{x}$, which implies the convergence of order-β LMV filter (7)-(10) to the exact LMV filter stated by Theorem 3.2 in Jimenez (2012). Since the approximate innovation estimator (11) is designed in terms of the order-β LMV filter (7)-(10), the weak convergence of $\mathbf{y}$ to $\mathbf{x}$ should then imply the convergence of the approximate innovation estimator (11) to the exact one (3) 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 the state space model (1)-(2), Theorem 5 in Jimenez (2012b) states the convergence of the order-β LMV filters to the exact LMV one when h decreases. Therefore, the convergence of the order-β innovation estimator to the exact innovation estimator is predictable when h goes to zero.

Theorem 1 *Let Z be a time series of M observations of the state space model (1)-(2) 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 innovation and an order-β innovation 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 model (1)-(2) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

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

$$\begin{aligned}\det(\boldsymbol{\Sigma}_{h,t_k}) &= \det(\boldsymbol{\Sigma}_{t_k} - \Delta\boldsymbol{\Sigma}_{h,t_k}) \\ &= \det(\boldsymbol{\Sigma}_{t_k}) \det(\mathbf{I} - \boldsymbol{\Sigma}_{t_k}^{-1}\Delta\boldsymbol{\Sigma}_{h,t_k}) \end{aligned} \tag{12}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{h,t_k}^{-1} &= (\boldsymbol{\Sigma}_{t_k} - \Delta\boldsymbol{\Sigma}_{h,t_k})^{-1} \\ &= \boldsymbol{\Sigma}_{t_k}^{-1} + \boldsymbol{\Sigma}_{t_k}^{-1}\Delta\boldsymbol{\Sigma}_{h,t_k}(\mathbf{I} - \boldsymbol{\Sigma}_{t_k}^{-1}\Delta\boldsymbol{\Sigma}_{h,t_k})^{-1}\boldsymbol{\Sigma}_{t_k}^{-1}. \end{aligned} \tag{13}$$

By using these two identities and the identity

$$\begin{aligned}(\mathbf{z}_{t_k} - \boldsymbol{\mu}_{h,t_k})^{\intercal}(\boldsymbol{\Sigma}_{h,t_k})^{-1}(\mathbf{z}_{t_k} - \boldsymbol{\mu}_{h,t_k}) &= (\mathbf{z}_{t_k} - \boldsymbol{\mu}_{t_k})^{\intercal}(\boldsymbol{\Sigma}_{h,t_k})^{-1}(\mathbf{z}_{t_k} - \boldsymbol{\mu}_{t_k}) \\ &\quad + (\mathbf{z}_{t_k} - \boldsymbol{\mu}_{t_k})^{\intercal}(\boldsymbol{\Sigma}_{h,t_k})^{-1}(\boldsymbol{\mu}_{t_k} - \boldsymbol{\mu}_{h,t_k}) \\ &\quad + (\boldsymbol{\mu}_{t_k} - \boldsymbol{\mu}_{h,t_k})^{\intercal}(\boldsymbol{\Sigma}_{h,t_k})^{-1}(\mathbf{z}_{t_k} - \boldsymbol{\mu}_{t_k}) \\ &\quad + (\boldsymbol{\mu}_{t_k} - \boldsymbol{\mu}_{h,t_k})^{\intercal}(\boldsymbol{\Sigma}_{h,t_k})^{-1}(\boldsymbol{\mu}_{t_k} - \boldsymbol{\mu}_{h,t_k}), \end{aligned} \tag{14}$$

with $\boldsymbol{\mu}_{t_k} = \mathbf{C}\mathbf{x}_{t_k/t_{k-1}}$ and $\boldsymbol{\mu}_{h,t_k} = \mathbf{C}\mathbf{y}_{t_k/t_{k-1}}$, it is obtained that

$$U_{M,h}(\boldsymbol{\theta}, Z) = U_M(\boldsymbol{\theta}, Z) + R_{M,h}(\boldsymbol{\theta}), \tag{15}$$

where U_M and $U_{M,h}$ are defined in (3) and (11), respectively, and

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

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

Theorem 5 in Jimenez (2012b) deals with the convergence of the order-β filters to the exact LMV one. In particular, for the predictions, it states that

$$\left|\mathbf{x}_{t_k/t_{k-1}} - \mathbf{y}_{t_k/t_{k-1}}\right| \leq Kh^{\beta} \quad \text{and} \quad \left|\mathbf{U}_{t_k/t_{k-1}} - \mathbf{V}_{t_k/t_{k-1}}\right| \leq Kh^{\beta} \tag{16}$$

for all $t_k, t_{k+1} \in \{t\}_M$, where K is a positive constant. Here, we recall that $\mathbf{x}_{t_k/t_{k-1}}$ and $\mathbf{U}_{t_k/t_{k-1}}$ are the predictions of the exact LMV filter for the model (1)-(2), whereas $\mathbf{y}_{t_k/t_{k-1}}$ and $\mathbf{V}_{t_k/t_{k-1}}$ are the predictions of the order-β filter. From this and taking into account that $\boldsymbol{\mu}_{t_k} - \boldsymbol{\mu}_{h,t_k} = \mathbf{C}(\mathbf{x}_{t_k/t_{k-1}} - \mathbf{y}_{t_k/t_{k-1}})$ and $\Delta\boldsymbol{\Sigma}_{h,t_k} = \mathbf{C}(\mathbf{U}_{t_k/t_{k-1}} - \mathbf{V}_{t_k/t_{k-1}})\mathbf{C}^{\intercal}$ follows that

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

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 (15),

$$\left|\widehat{\boldsymbol{\theta}}_M(h) - \widehat{\boldsymbol{\theta}}_M\right| = \left|\arg\{\min_{\boldsymbol{\theta}}\{U_M(\boldsymbol{\theta}, Z) + R_{M,h}(\boldsymbol{\theta})\}\} - \arg\{\min_{\boldsymbol{\theta}} U_M(\boldsymbol{\theta}, Z)\}\right| \to 0 \tag{17}$$

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

On the other hand, since the constant K in (16) does not depends of a specific realizations of the model (1)-(2), from these inequalities follows that

$$E(\left|\mathbf{x}_{t_k/t_{k-1}} - \mathbf{y}_{t_k/t_{k-1}}\right|) \leq Kh^{\beta} \quad \text{and} \quad E(\left|\mathbf{U}_{t_k/t_{k-1}} - \mathbf{V}_{t_k/t_{k-1}}\right|) \leq Kh^{\beta},$$

where the new expectation here is with respect to the measure on the underlying probability space generating the realizations of the model (1)-(2) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. From this and (17) 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 first assertion of this theorem states that, for each given data Z, the order-β innovation 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 (6) is then clear that the order-β innovation estimator (11) converges to the exact one (3) when the error (in weak sense) of the order-β approximation $\mathbf{y}$ to $\mathbf{x}$ decreases or, equivalently, when the error between the order-β and the LMV filter 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 the model (1)-(2) 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 state space model.

Theorem 2 *Let Z be a time series of M observations of the state space model (1)-(2) 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 innovation and an order-β innovation 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 model (1)-(2) 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 model (1)-(2) 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 innovation estimators mentioned in Section 2 do not have the desired convergence properties stated in the theorems above for the order-β innovation estimator. Further note that, either in Definition 3 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 innovation 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 innovation 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 innovation 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

$$W_M(\boldsymbol{\theta}) = E(U_M(\boldsymbol{\theta}, Z)) \text{ with } \boldsymbol{\theta} \in \mathcal{D}_\theta \tag{18}$$

should be considered, where U_M is defined in (3) and the expectation is taken with respect to the measure on the underlying probability space generating the realizations of the state space model (1)-(2). 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 state space model (1)-(2) are assumed, which are typically satisfied by stationary and ergodic diffusion processes (see, e.g., Ljung & Caines, 1979).

Lemma 1 *If $\boldsymbol{\Sigma}_{t_k}$ is positive definite for all $k = 1, .., M-1$, then the function $W_M(\boldsymbol{\theta})$ defined in (18) has an unique minimum and*

$$\arg\{\min_{\boldsymbol{\theta}\in\mathcal{D}_\theta} W_M(\boldsymbol{\theta})\} = \boldsymbol{\theta}_0. \tag{19}$$

Proof Since $\boldsymbol{\Sigma}_{t_k}$ 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

$$l_k(\boldsymbol{\theta}) = E(\ln(\det(\boldsymbol{\Sigma}_{t_k})) + \boldsymbol{\nu}_{t_k}^\top (\boldsymbol{\Sigma}_{t_k})^{-1} \boldsymbol{\nu}_{t_k} | Z_{t_{k-1}})$$

on $\mathcal{D}_\theta$ for all k, where $\boldsymbol{\nu}_{t_k} = \mathbf{z}_{t_k} - \mathbf{C}\mathbf{x}_{t_k/t_{k-1}}$ and $\boldsymbol{\Sigma}_{t_k} = \mathbf{C}\mathbf{U}_{t_k/t_{k-1}}\mathbf{C}^\intercal + \boldsymbol{\Pi}_{t_k}$. Consequently and under the assumed unique identifiability of the model (1)-(2), $\boldsymbol{\theta}_0$ 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 $\mathcal{D}_\theta$.

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 state space model (1)-(2) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ on the time partition $\{t\}_M$. Let $\widehat{\boldsymbol{\theta}}_M(h)$ be an order-β innovation estimator for the parameters of (1) given Z. Then*

$$\widehat{\boldsymbol{\theta}}_M(h) - \boldsymbol{\theta}_0 \to \Delta\boldsymbol{\theta}_M(h) \tag{20}$$

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

$$W_M^{''}(\boldsymbol{\theta}) > \epsilon\mathbf{I} \quad \text{ and } \quad \mathbf{H}_{M,h}(\boldsymbol{\theta}) = ME(U_{M,h}'(\boldsymbol{\theta}, Z)(U_{M,h}'(\boldsymbol{\theta}, Z))^\intercal) > \epsilon\mathbf{I} \tag{21}$$

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

$$\sqrt{M}\mathbf{P}_{M,h}^{-1/2}(\widehat{\boldsymbol{\theta}}_M(h) - \boldsymbol{\theta}_0) \sim \mathcal{N}(\Delta\boldsymbol{\theta}_M(h), \mathbf{I}) \tag{22}$$

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 (11).

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

$$\widehat{\boldsymbol{\theta}}_M(h) - \boldsymbol{\alpha}_M(h) \to 0 \tag{23}$$

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

$$\sqrt{M}\mathbf{P}_{M,h}^{-1/2}(\boldsymbol{\alpha}_M(h))(\widehat{\boldsymbol{\theta}}_M(h) - \boldsymbol{\alpha}_M(h)) \sim \mathcal{N}(0, \mathbf{I}) \tag{24}$$

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 (12)-(14), the function

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

with $\boldsymbol{\mu}_{h,t_k} = \mathbf{C}\mathbf{y}_{t_k/t_{k-1}}$, can be written as

$$W_{M,h}(\boldsymbol{\theta}) = W_M(\boldsymbol{\theta}) + E(R_{M,h}(\boldsymbol{\theta})), \tag{25}$$

where W_M is defined in (18) and

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

with $\mathbf{M}_{h,t_k} = \boldsymbol{\Sigma}_{t_k}^{-1}\Delta\boldsymbol{\Sigma}_{h,t_k}(\mathbf{I} - \boldsymbol{\Sigma}_{t_k}^{-1}\Delta\boldsymbol{\Sigma}_{h,t_k})^{-1}\boldsymbol{\Sigma}_{t_k}^{-1}$, $\boldsymbol{\mu}_{t_k} = \mathbf{C}\mathbf{x}_{t_k/t_{k-1}}$ and $\Delta\boldsymbol{\Sigma}_{h,t_k} = \boldsymbol{\Sigma}_{t_k} - \boldsymbol{\Sigma}_{h,t_k}$.

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

$$\mathbf{P}_{M,h}(\boldsymbol{\theta}) = (W_M^{''}(\boldsymbol{\theta}))^{-1}\mathbf{H}_{M,h}(\boldsymbol{\theta})(W_M^{''}(\boldsymbol{\theta}))^{-1} + \Delta\mathbf{P}_{M,h}(\boldsymbol{\theta}), \tag{26}$$

where

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

Theorem 5 in Jimenez (2012b) deals with the convergence of the order-β filters to the exact LMV one. In particular, for the predictions, it states that

$$\left|\mathbf{x}_{t_k/t_{k-1}} - \mathbf{y}_{t_k/t_{k-1}}\right| \leq Kh^{\beta} \qquad \text{and} \qquad \left|\mathbf{U}_{t_k/t_{k-1}} - \mathbf{V}_{t_k/t_{k-1}}\right| \leq Kh^{\beta}$$

for all $t_k, t_{k+1} \in \{t\}_M$, where K is a positive constant. Here, we recall that $\mathbf{x}_{t_k/t_{k-1}}$ and $\mathbf{U}_{t_k/t_{k-1}}$ are the predictions of the exact LMV filter for the model (1)-(2), whereas $\mathbf{y}_{t_k/t_{k-1}}$ and $\mathbf{V}_{t_k/t_{k-1}}$ are the predictions of the order-β filter. From this and taking into account that $\boldsymbol{\mu}_{t_k} - \boldsymbol{\mu}_{h,t_k} = \mathbf{C}(\mathbf{x}_{t_k/t_{k-1}} - \mathbf{y}_{t_k/t_{k-1}})$ and $\Delta\boldsymbol{\Sigma}_{h,t_k} = \mathbf{C}(\mathbf{U}_{t_k/t_{k-1}} - \mathbf{V}_{t_k/t_{k-1}})\mathbf{C}^{\intercal}$ follows that

$$\left|\boldsymbol{\mu}_{t_k} - \boldsymbol{\mu}_{h,t_k}\right| \rightarrow \mathbf{0} \qquad \text{and} \qquad \left|\boldsymbol{\Sigma}_{t_k} - \boldsymbol{\Sigma}_{h,t_k}\right| \rightarrow \mathbf{0}$$

as $h \rightarrow 0$ 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 $|R_{M,h}(\boldsymbol{\theta}, Z)| \rightarrow 0$ and $\left|R''_{M,h}(\boldsymbol{\theta}, Z)\right| \rightarrow 0$ as well with h. From this and (25), it is obtained that

$$W_{M,h}(\boldsymbol{\theta}) \rightarrow W_M(\boldsymbol{\theta}) \qquad \text{and} \qquad W''_{M,h}(\boldsymbol{\theta}) \rightarrow W''_M(\boldsymbol{\theta}) \quad \text{as} \qquad h \rightarrow 0. \tag{27}$$

In addition, left (27) and Lemma 1 imply that

$$\Delta\boldsymbol{\theta}_M(h) = \boldsymbol{\alpha}_M(h) - \boldsymbol{\theta}_0 = \arg\{\min_{\boldsymbol{\theta}\in\mathcal{D}_\theta} W_{M,h}(\boldsymbol{\theta})\} - \arg\{\min_{\boldsymbol{\theta}\in\mathcal{D}_\theta} W_M(\boldsymbol{\theta})\} \rightarrow 0 \quad \text{as} \qquad h \rightarrow 0, \tag{28}$$

whereas from right (27) follows that

$$\Delta\mathbf{P}_{M,h}(\boldsymbol{\theta}) \rightarrow 0 \quad \text{as} \qquad h \rightarrow 0. \tag{29}$$

Finally, (28)-(29) together (23), (24) and (26) imply that (20) and (22) hold, which completes the proof.

Theorem 3 states that, for an increasing number of observations, the order-β innovation 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 asymptotic properties of the exact innovation estimator $\widehat{\boldsymbol{\theta}}_M$ derived from Theorem 1 in Ljung & Caines (1979) and the convergence of the approximate estimator $\widehat{\boldsymbol{\theta}}_M(h)$ to $\widehat{\boldsymbol{\theta}}_M$ given by Theorem 1 when h goes to zero. Further note that, when $h = 0$, the Theorem 3 reduces to Theorem 1 in Ljung & Caines (1979) for the exact innovation estimator $\widehat{\boldsymbol{\theta}}_M$. This is other expected result since the order-β innovation 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.

3.3 Models with nonlinear observation equation

Previous definitions and results have been stated for models with linear observation equation. However, by following the procedure proposed in Jimenez and Ozaki (2006), they can be easily applied as well to state space models with nonlinear observation equation.

For illustrate this, let us consider the state space model defined by the continuous state equation (1) and the discrete observation equation

$$\mathbf{z}(t_k) = \mathbf{h}(t_k, \mathbf{x}(t_k)) + \mathbf{e}_{t_k}, \text{ for } k = 0, 1, .., M-1, \tag{30}$$

where $\mathbf{e}_{t_k}$ is defined as in (2) and $\mathbf{h} : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}^r$ is a twice differentiable function. By using the Ito formula,

$$\begin{aligned} d\mathbf{h}^j &= \{\frac{\partial \mathbf{h}^j}{\partial t} + \sum_{k=1}^{d} f^k \frac{\partial \mathbf{h}^j}{\partial \mathbf{x}^k} + \frac{1}{2}\sum_{s=1}^{m}\sum_{k,l=1}^{d} \mathbf{g}_s^l \mathbf{g}_s^k \frac{\partial^2 \mathbf{h}^j}{\partial \mathbf{x}^l \partial \mathbf{x}^k}\}dt + \sum_{s=1}^{m}\sum_{l=1}^{d} \mathbf{g}_s^l \frac{\partial \mathbf{h}^j}{\partial \mathbf{x}^l} d\mathbf{w}^s \\ &= \boldsymbol{\rho}^j dt + \sum_{s=1}^{m} \boldsymbol{\sigma}_s^j d\mathbf{w}^s \end{aligned}$$

with $j = 1, .., r$. Hence, the state space model (1)+(30) is transformed to the following higher-dimensional state space model with linear observation

$$d\mathbf{v}(t) = \mathbf{a}(t, \mathbf{v}(t))dt + \sum_{i=1}^{m} \mathbf{b}_i(t, \mathbf{v}(t))d\mathbf{w}^i(t),$$

$$\mathbf{z}(t_k) = \mathbf{C}\mathbf{v}(t_k) + \mathbf{e}_{t_k}, \text{ for } k = 0, 1, .., M-1,$$

where

$$\mathbf{v} = \begin{bmatrix} \mathbf{x} \\ \mathbf{h} \end{bmatrix}, \ \mathbf{a} = \begin{bmatrix} \mathbf{f} \\ \boldsymbol{\rho} \end{bmatrix}, \ \mathbf{b}_i = \begin{bmatrix} \mathbf{g}_i \\ \boldsymbol{\sigma}_i \end{bmatrix}$$

and the matrix $\mathbf{C}$ is such that $\mathbf{h}(t_k, \mathbf{x}(t_k)) = \mathbf{C}\mathbf{v}(t_k)$.

In this way, the state space model (1)+(30) is transformed to the form of the model (1)-(2), and so the previous definition and results related to the order-β innovation estimator can be applied. Further, note that if the nonlinear function h depends of unknown parameters, they can be estimated as well by the approximate innovation method.

3.4 Models with noise free complete observations

This section deals with the particular case that the observation noise is zero and all components of the diffusion process defined in (1) are discretely observed. That is, when $\mathbf{C} \equiv \mathbf{I}$ and $\mathbf{\Pi}_{t_k} = \mathbf{0}$ in (2) for all k, where $\mathbf{I}$ denotes the d-dimensional identity matrix. Hence, the inference problem under consideration in this paper reduces then to the well known problem of parameter estimation of diffusion processes from complete observations. In this situation, it is easy to realize that the innovation estimator (3) reduces to the well known quasi-maximum likelihood (QML) estimator for SDEs, and that the approximate order-β innovation estimator (11) reduces to the approximate order-β QML estimator introduced in Jimenez (2012c) for the estimation of SDEs from complete observations. For the same reason, Theorems 1, 2 and 3 reduce to those corresponding in Jimenez (2012c) concerning the convergence and asymptotic properties of the approximate order-β QML estimator.

4 Order-β innovation estimator based on LL filters

Since, in principle, any approximate filter converging to LMV filter of the model (1)-(2) can be used to construct an order-β innovation estimator, some additional criterions could be considered for the selection of one of them. For instance, high order of convergence, efficiency of the algorithm from computational viewpoint, and so on. In this paper, we elected the order-β Local Linearization (LL) filters proposed in Jimenez (2012b) for the following reasons: 1) their predictions have simple explicit formulas that can be computed by means of efficient algorithm (including high dimensional equations); 2) their predictions are exact for linear SDEs in all the possible variants (with additive and/or multiplicative noise, autonomous or not); 3) they have an adequate order $\beta = 1, 2$ of convergence; and 4) the better performance of the approximate innovation estimators based on conventional LL filters (see, e.g., Ozaki, 1994; Shoji, 1998; Singer, 2002).

According to Jimenez (2012b), the order-β LL filter is defined on $(\tau)_h \supset \{t\}_M$ in terms of the order-β Local Linear approximation $\mathbf{y}$ that satisfies the conditions (4)-(6). Denote by $\mathbf{y}_{\tau_n/t_k} = E(\mathbf{y}(\tau_n)|Z_{t_k})$ and $\mathbf{P}_{\tau_n/t_k} = E(\mathbf{y}(\tau_n)\mathbf{y}^\intercal(\tau_n)|Z_{t_k})$ the first two conditional moment of $\mathbf{y}$ at τ_n given the observations Z_{t_k}, for all $\tau_n \in \{(\tau)_h \cap [t_k, t_{k+1}]\}$ and $k = 0, .., M-2$.

Starting with the initial filter values $\mathbf{y}_{t_0/t_0} = \mathbf{x}_{t_0/t_0}$ and $\mathbf{P}_{t_0/t_0} = \mathbf{Q}_{t_0/t_0}$, the LL filter algorithm performs the recursive computation of :

1. the predictions $\mathbf{y}_{\tau_n/t_k}$and $\mathbf{P}_{\tau_n/t_k}$ for all $\tau_n \in \{(\tau)_h \cap (t_k, t_{k+1}]\}$ by means of the recursive formulas (42)-(43) given in the Appendix, and the prediction variance

$$\mathbf{V}_{t_{k+1}/t_k} = \mathbf{P}_{t_{k+1}/t_k} - \mathbf{y}_{t_{k+1}/t_k}\mathbf{y}^\intercal_{t_{k+1}/t_k};$$

2. the filters

$$\mathbf{y}_{t_{k+1}/t_{k+1}} = \mathbf{y}_{t_{k+1}/t_k} + \mathbf{K}_{t_{k+1}}(\mathbf{z}_{t_{k+1}} - \mathbf{C}\mathbf{y}_{t_{k+1}/t_k}),$$
$$\mathbf{V}_{t_{k+1}/t_{k+1}} = \mathbf{V}_{t_{k+1}/t_k} - \mathbf{K}_{t_{k+1}}\mathbf{C}\mathbf{V}_{t_{k+1}/t_k},$$
$$\mathbf{P}_{t_{k+1}/t_{k+1}} = \mathbf{V}_{t_{k+1}/t_{k+1}} + \mathbf{y}_{t_{k+1}/t_{k+1}}\mathbf{y}^{\intercal}_{t_{k+1}/t_{k+1}},$$

with filter gain

$$\mathbf{K}_{t_{k+1}} = \mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal}(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal} + \mathbf{\Pi}_{t_{k+1}})^{-1};$$

for each k, with $k = 0, 1, \ldots, M-2$.

Under general conditions, the convergence of the order-β LL filter to exact LMV filter when h goes to zero has been stated by Theorem 10 in Jimenez (2012b). Hence, Theorem 1 implies that the LL-based innovation estimator

$$\widehat{\boldsymbol{\theta}}_M(h) = \arg\{\min_{\boldsymbol{\theta}} U_{M,h}(\boldsymbol{\theta}, Z)\}, \tag{31}$$

with

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

converges to the exact one (3) as h goes to zero for all given Z, where $\mathbf{\Sigma}_{t_k/t_{k-1}} = \mathbf{C}\mathbf{V}_{t_k/t_{k-1}}\mathbf{C}^{\intercal} + \mathbf{\Pi}_{t_k}$. For the same reason, this order-β innovation estimator 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.

Note that, when $(\tau)_h \equiv \{t\}_M$, the order-β LL filter reduces to the conventional LL filter. In this situation, the order-β innovation estimator (31) reduces to the conventional innovation estimators of Ozaki (1994) or Shoji (1998) for SDEs with additive noise, and to that of Jimenez and Ozaki (2006) for SDEs with multiplicative noise. It is worth to emphasize here that, for each data $\mathbf{z}_k$, the formulas (42)-(43) for the predictions are recursively evaluated at all the time instants $\tau_n \in \{(\tau)_h \cap (t_k, t_{k+1}]\}$ for the order-β estimator, whereas they are evaluated only at $t_{k+1} = \{(\tau)_h \cap (t_k, t_{k+1}]\}$ for the conventional ones. In addition, since the predictions of the order-β LL filter are exact for linear SDEs, the order-β innovation estimator (31) reduces to the maximum likelihood estimator of Schweppe (1965) for linear equations with additive noise.

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 innovation estimators are computed and compared: 1) the exact one (3), when it is possible; 2) the conventional one based on the LL filter. That is, the estimator defined by (31) with $(\tau)_h \equiv \{t\}_M$ and $\beta = 1$; 3) the order-1 innovation estimator (31) with various uniform time discretizations $(\tau)^u_{h,T}$; and 4) the adaptive order-1 innovation estimator (31) 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 and noisy observations taken with different time distances (sampling periods) on time intervals with distinct lengths.

5.1 Test models

Example 1. State equation with multiplicative noise

$$dx = \alpha t x dt + \sigma\sqrt{t} x dw_1 \tag{32}$$

and observation equation

$$z_{t_k} = x(t_k) + e_{t_k}, \text{ for } k = 0, 1, .., M-1, \tag{33}$$

with $\alpha = -0.1$, $\sigma = 0.1$ and observation noise variance $\Pi = 0.0001$. For this state equation, the predictions for the first two conditional moments are

$$x_{t_{k+1}/t_k} = x_{t_k/t_k} e^{\alpha(t_{k+1}^2 - t_k^2)/2} \quad \text{and} \quad Q_{t_{k+1}/t_k} = Q_{t_k/t_k} e^{(\alpha+\sigma^2/2)(t_{k+1}^2 - t_k^2)},$$

where the filters x_{t_k/t_k} and Q_{t_k/t_k} are obtained from the well-known formulas of the exact LMV filter for all $k = 0, 1, .., M-2$, with initial values $x_{t_0/t_0} = 1$ and $Q_{t_0/t_0} = 1$ at $t_0 = 0.5$.

Example 2. State equation with two additive noise

$$dx = \alpha t x dt + \sigma t^2 e^{\alpha t^2/2} dw_1 + \rho\sqrt{t} dw_2 \tag{34}$$

and observation equation

$$z_{t_k} = x(t_k) + e_{t_k}, \text{ for } k = 0, 1, .., M-1, \tag{35}$$

with $\alpha = -0.25$, $\sigma = 5$, $\rho = 0.1$ and observation noise variance $\Pi = 0.0001$. For this state equation, the predictions for the first two conditional moments are

$$x_{t_{k+1}/t_k} = x_{t_k/t_k} e^{\alpha(t_{k+1}^2 - t_k^2)/2}$$

and

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

where the filters x_{t_k/t_k} and Q_{t_k/t_k} are obtained from the formulas of the exact LMV filter for all $k = 0, 1, .., M-2$, with initial values $x_{t_0/t_0} = 10$ and $Q_{t_0/t_0} = 100$ at $t_0 = 0.01$.

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

$$dx_1 = x_2 dt \tag{36}$$
$$dx_2 = (-(x_1^2 - 1)x_2 - x_1 + \alpha)dt + \sigma dw \tag{37}$$

and observation equation

$$z_{t_k} = x_1(t_k) + e_{t_k}, \text{ for } k = 0, 1, .., M-1, \tag{38}$$

where $\alpha = 0.5$ and $\sigma^2 = (0.75)^2$ are the intensity and the variance of the random input, respectively. In addition, $\Pi = 0.001$ is the observation noise variance, and $\mathbf{x}_{t_0/t_0}^{\intercal} = [1\ 1]$ and $\mathbf{Q}_{t_0/t_0} = \mathbf{x}_{t_0/t_0}\mathbf{x}_{t_0/t_0}^{\intercal}$ are the initial filter values at $t_0 = 0$.

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

$$dx_1 = x_2 dt \tag{39}$$
$$dx_2 = (-(x_1^2 - 1)x_2 - \alpha x_1)dt + \sigma x_1 dw \tag{40}$$

and observation equation

$$z_{t_k} = x_1(t_k) + e_{t_k}, \text{ for } k = 0, 1, .., M-1, \tag{41}$$

where $\alpha = 1$ and $\sigma^2 = 1$ are the frequency mean value and variance, respectively. $\Pi = 0.001$ is the observation noise variance, and $\mathbf{x}_{t_0/t_0}^{\intercal} = [1\ 1]$ and $\mathbf{Q}_{t_0/t_0} = \mathbf{x}_{t_0/t_0}\mathbf{x}_{t_0/t_0}^{\intercal}$ are the initial filter values at $t_0 = 0$.

In these examples, autonomous or non autonomous, linear or nonlinear, one or two dimensional SDEs 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 SDEs in Examples 1 and 2 have explicit expressions, the exact innovation estimator (3) can be computed.

These four state space models have previously been used in Jimenez (2012b) to illustrate the convergence of the order-β LL filter 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 state equations

For the first two examples, 100 realizations of the state equation 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 for evaluating the corresponding observation equation 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 $Z^i_{\delta,T} = \{z^i_{t_k} : t_k \in \{t\}_{M,T}, M = T/\delta\}$, with $i = 1,..,100$, of M observations $z^i_{t_k}$ each one were finally available for both state space models to make inference. 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 ($\widehat{\alpha}_{\delta,T}$) innovation 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 ($\widehat{\sigma}_{\delta,T}$) innovation 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 (32) is better estimated by $\widehat{\alpha}^E_{\delta,T}$. Contrarily, from these samples, the conventional innovation estimators $\widehat{\alpha}_{\delta,T}$ can not provided a good approximation to α, and so the whole unexplained component of the drift coefficient of (32) included in the samples is interpreted as noise by the conventional 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, the difference between the exact ($\widehat{\alpha}^E_{\delta,T}, \widehat{\sigma}^E_{\delta,T}$) and the conventional ($\widehat{\alpha}_{\delta,T}, \widehat{\sigma}_{\delta,T}$) innovation 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.

For the data of (32) with largest sampling period $\delta = 1$, the order-1 innovation 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 ($\widehat{\alpha}^u_{h,\delta,T}, \widehat{\sigma}^u_{h,\delta,T}$) innovation 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 innovation 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 innovation 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 innovation 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. These findings are more precisely summarized in Table II, which shows the difference between the averages of the exact and the approximate innovation estimators. Further, note the small difference between the adaptive estimators ($\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T}$) and the exact ones ($\widehat{\alpha}^E_{\delta,T}, \widehat{\sigma}^E_{\delta,T}$), which illustrates the usefulness of the adaptive strategy for improving the innovation parameter estimation for finite samples with large sampling periods. The number of accepted and fail steps of the adaptive innovation estimators at each $t_k \in \{t\}_{T/\delta,T}$ are shown in Figure 4. Further, note that the results of Table II illustrate the convergence findings of Theorem 2.

$\delta = 1$		$h = \delta$	$h = \delta/2$	$h = \delta/8$	$h = \delta/32$
α	$T = 10$	$7.5 \pm 5.5 \times 10^{-3}$	$1.8 \pm 1.2 \times 10^{-3}$	$2.9 \pm 2.3 \times 10^{-4}$	$6.8 \pm 5.6 \times 10^{-5}$
	$T = 20$	$7.7 \pm 8.0 \times 10^{-3}$	$1.7 \pm 1.2 \times 10^{-3}$	$2.7 \pm 2.2 \times 10^{-4}$	$6.4 \pm 5.3 \times 10^{-5}$
	$T = 30$	$7.1 \pm 5.2 \times 10^{-3}$	$1.7 \pm 1.2 \times 10^{-3}$	$2.7 \pm 2.2 \times 10^{-4}$	$6.3 \pm 5.3 \times 10^{-5}$
σ	$T = 10$	$3.2 \pm 1.9 \times 10^{-2}$	$1.0 \pm 0.6 \times 10^{-2}$	$2.1 \pm 1.1 \times 10^{-3}$	$5.1 \pm 2.6 \times 10^{-4}$
	$T = 20$	$3.2 \pm 1.9 \times 10^{-2}$	$1.0 \pm 0.6 \times 10^{-2}$	$2.1 \pm 1.1 \times 10^{-3}$	$5.1 \pm 2.6 \times 10^{-4}$
	$T = 30$	$3.2 \pm 1.9 \times 10^{-2}$	$1.0 \pm 0.6 \times 10^{-2}$	$2.1 \pm 1.1 \times 10^{-3}$	$5.1 \pm 2.6 \times 10^{-4}$

Table I. Confidence limits for the error between the exact and the approximate innovation estimators of the equation (32). $h = \delta$, 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	-0.00403	-0.00433	-0.00373	-0.0321	-0.0321	-0.0321
$\delta/2$	-0.00083	-0.00077	-0.00077	-0.0107	-0.0106	-0.0106
$\delta/8$	-0.00004	-0.00002	-0.00002	-0.0021	-0.0021	-0.0021
$\delta/32$	0	0.00001	0	-0.0005	-0.0005	-0.0005
$\cdot$	-0.00010	-0.00014	-0.00010	-0.0003	-0.0002	-0.0003

Table II: Difference between the averages of the exact and the approximate innovation estimators for the equation (32). $h = \delta$, 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 $(\tau)_{\cdot,T}$.

$\delta = 0.1$	α			σ			ρ		
h	$T = 10$	$T = 20$	$T = 30$	$T = 10$	$T = 20$	$T = 30$	$T = 10$	$T = 20$	$T = 30$
δ	0.00039	0.00031	0.00029	-0.0311	-0.0291	-0.0287	-2.13×10^{-4}	2.4×10^{-5}	3.3×10^{-5}
$\delta/2$	0.00010	0.00007	0.00007	-0.0067	-0.0059	-0.0060	-0.54×10^{-4}	1.4×10^{-5}	1.8×10^{-5}
$\delta/4$	0.00003	0.00002	0.00001	-0.0013	-0.0012	-0.0012	-0.05×10^{-4}	1.0×10^{-5}	1.1×10^{-5}
$\delta/8$	0.00001	0	0	-0.0002	-0.0001	-0.0001	0.03×10^{-4}	0.6×10^{-5}	0.6×10^{-5}
$\cdot$	-0.00005	0.00002	0.00009	0	-0.0023	-0.0106	2.14×10^{-4}	2.8×10^{-5}	9.4×10^{-5}

Table III: Difference between the averages of the exact and the approximate innovation estimators for the equation (34). $h = \delta$, 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	$T = 10$	$5.2 \pm 4.0 \times 10^{-4}$	$1.1 \pm 0.9 \times 10^{-4}$	$2.7 \pm 2.0 \times 10^{-5}$	$7.4 \pm 5.6 \times 10^{-6}$
	$T = 20$	$5.5 \pm 4.0 \times 10^{-4}$	$1.2 \pm 0.8 \times 10^{-4}$	$2.6 \pm 1.8 \times 10^{-5}$	$7.1 \pm 5.9 \times 10^{-6}$
	$T = 30$	$5.4 \pm 3.9 \times 10^{-4}$	$1.1 \pm 0.8 \times 10^{-4}$	$2.6 \pm 1.8 \times 10^{-5}$	$6.7 \pm 5.4 \times 10^{-6}$
σ	$T = 10$	$4.8 \pm 3.5 \times 10^{-2}$	$9.7 \pm 6.6 \times 10^{-3}$	$1.8 \pm 1.4 \times 10^{-3}$	$3.6 \pm 3.6 \times 10^{-4}$
	$T = 20$	$4.9 \pm 3.5 \times 10^{-2}$	$1.0 \pm 0.7 \times 10^{-2}$	$1.9 \pm 1.5 \times 10^{-3}$	$3.9 \pm 4.7 \times 10^{-4}$
	$T = 30$	$4.9 \pm 3.4 \times 10^{-2}$	$1.0 \pm 0.7 \times 10^{-2}$	$1.9 \pm 1.4 \times 10^{-3}$	$3.7 \pm 3.8 \times 10^{-4}$
ρ	$T = 10$	$0.8 \pm 1.2 \times 10^{-3}$	$1.9 \pm 2.6 \times 10^{-4}$	$3.9 \pm 5.0 \times 10^{-5}$	$9.9 \pm 8.9 \times 10^{-6}$
	$T = 20$	$1.3 \pm 1.2 \times 10^{-4}$	$3.7 \pm 3.0 \times 10^{-5}$	$1.3 \pm 0.5 \times 10^{-5}$	$6.6 \pm 2.0 \times 10^{-6}$
	$T = 30$	$7.5 \pm 5.7 \times 10^{-5}$	$2.5 \pm 1.2 \times 10^{-5}$	$1.1 \pm 0.3 \times 10^{-5}$	$6.0 \pm 1.3 \times 10^{-6}$

Table IV: Confidence limits for the error between the exact and the approximate innovation estimators of the equation (34). $h = \delta$, 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 ($\widehat{\alpha}_{\delta,T}$) innovation 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 ($\widehat{\sigma}_{\delta,T}$) innovation 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 innovation estimators is not so large as in the previous example. Nevertheless, in this extreme situation of low information in the data, the order-1 innovation 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 innovation 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 $rtol_{\mathbf{y}} = rtol_{\mathbf{P}} = 5 \times 10^{-7}$ and $atol_{\mathbf{y}} = 5 \times 10^{-10}$, $atol_{\mathbf{P}} = 5 \times 10^{-13}$. The average of accepted and fail steps of the adaptive innovation estimators at each $t_k \in \{t\}_{T/\delta,T}$ are shown in Figure 4. Observe in Table III the higher difference between the averages of the exact and the adaptive estimators for the three parameters when $T = 30$. The reason is that, for $t_k > 200$, the mean and variance of the diffusion process (34) becomes almost indistinguishable of zero in such a way that the signal noise ratio is very small. This is so small that the adaptive strategy computes inaccurate estimates of the integration errors for the predictions and so less accurate estimators for the parameters of the SDE (34). For this example, the convergence of the order-1 innovation estimators to the exact one is shown in Table IV, which gives the confidence limits for the error between theses 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 state equations

For the examples 3 and 4, 100 realizations of the state equation solution 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 for evaluating the corresponding observation equation, 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 $Z^{i}_{\delta,T} = \{z^{i}_{t_k} : t_k \in \{t\}_{M,T},\ M = T/\delta\}$, with $i = 1, .., 100$, of M observations $z^{i}_{t_k}$ each one were available for both state space models with the two values of (δ, T) mentioned above.

For both examples, the order-1 innovation 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 innovation 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 innovation estimators is negligible for $(\tau)_{\cdot,T}$ thin enough. In addition, Table V and VI show the bias of the approximate innovation estimators for these examples. Observe as the adaptive $(\widehat{\alpha}_{\cdot,\delta,T}, \widehat{\sigma}_{\cdot,\delta,T})$ and the order-1 innovation 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 innovation 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 innovation 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 innovation 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 innovation 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 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.

$T = 30$	α		σ	
h	$\delta = 1$	$\delta = 0.1$	$\delta = 1$	$\delta = 0.1$
δ	-0.4588	-0.1403	-0.7240	-0.0140
$\delta/16$	-0.1244	-0.0026	-0.2180	0.0103
$\delta/64$	-0.0336	0.0041	-0.1883	0.0104
$\cdot$	-0.0108	0.0064	-0.1803	0.0099

Table V: Bias of the approximate innovation estimators for the equation (36)-(37). $h = \delta$, 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	-0.8511	-0.2740	-1.0347	-0.0239
$\delta/8$	-0.2488	-0.0662	-0.3107	0.0071
$\delta/32$	-0.1887	-0.0472	-0.2857	0.0072
$\cdot$	-0.1550	-0.0373	-0.2805	0.0084

Table VI: Bias of the approximate innovation estimators for the equation (39)-(40). $h = \delta$, 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 $(\tau)_{\cdot,T}$.

5.4 Simulations with noise free observation equations

In section 3.4, the connection among the innovation and quasi-maximum likelihood estimators was early mentioned for the identification of models with noise free complete observations. In this situation, it is easy to verify that the LL-based innovation estimator (31) reduces to the LL-based quasi-maximum likelihood estimator introduced in Jimenez (2012b). In that paper, the state equations of the four models considered in Section 5.1 were also used as test examples in simulations. The reader interested in this identification problem is encouraged to consider these simulations.

6 Conclusions

An alternative approximation to the innovation method was introduced for the parameter estimation of diffusion processes given a time series of partial and noisy observations. This is based on a convergent approximation to the first two conditional moments of the innovation process through approximate continuous-discrete filters of minimum variance. For finite samples, the convergence of the approximate innovation estimators to the exact one was proved when the error between the approximate and the exact linear minimum variance filters decreases. It was also demonstrated that, 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 particular instance, the order-β innovation estimators based on Local Linearization filters 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 innovation estimator provides satisfactory approximations to the exact innovation estimator; 2) the convergence of the order-1 innovation estimator to the exact one when the maximum stepsize of the time discretization between observations decreases; 3) with respect to the conventional innovation estimator, the order-1 innovation estimator gives much better approximation to the exact innovation estimator, and has less bias and higher efficiency; 4) with an adequate tolerance, the adaptive order-1 innovation estimator provides an automatic, suitable and computational efficient approximation to the exact innovation estimator; and 5) the effectiveness of the order-1 innovation estimator for the identification of SDEs from a reduced number of partial and noisy 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.

Acknowledgement. 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.

7 Appendix

According to Jimenez (2012b), given the filters values $\mathbf{y}_{t_k/t_k}$ and $\mathbf{P}_{t_k/t_k}$, the predictions $\mathbf{y}_{t/t_k}$ and $\mathbf{P}_{t/t_k}$ of the order-β LL filter are computed by the recursive formulas

$$\mathbf{y}_{t/t_k} = \mathbf{y}_{\tau_{n_t}/t_k} + \mathbf{L}_2 e^{\mathbf{M}(\tau_{n_t})(t-\tau_{n_t})}\mathbf{u}_{\tau_{n_t},t_k} \tag{42}$$

and

$$vec(\mathbf{P}_{t/t_k}) = \mathbf{L}_1 e^{\mathbf{M}(\tau_{n_t})(t-\tau_{n_t})}\mathbf{u}_{\tau_{n_t},t_k} \tag{43}$$

for all $t \in (t_k, t_{k+1}]$ and $t_k, t_{k+1} \in \{t\}_M$, where

$$n_t = \max\{n = 0, 1, \ldots : \tau_n \leq t \text{ and } \tau_n \in (\tau)_h\},$$

and 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

$$\boldsymbol{\beta}_1(\tau) = \sum_{i=1}^{m} \mathbf{b}_{i,0}(\tau)\mathbf{b}_{i,0}^{\intercal}(\tau)$$

$$\boldsymbol{\beta}_2(\tau) = \sum_{i=1}^{m} \mathbf{b}_{i,0}(\tau)\mathbf{b}_{i,1}^{\intercal}(\tau) + \mathbf{b}_{i,1}(\tau)\mathbf{b}_{i,0}^{\intercal}(\tau)$$

$$\boldsymbol{\beta}_3(\tau) = \sum_{i=1}^{m} \mathbf{b}_{i,1}(\tau)\mathbf{b}_{i,1}^{\intercal}(\tau)$$

$$\boldsymbol{\beta}_4(\tau) = \mathbf{a}_0(\tau) \oplus \mathbf{a}_0(\tau) + \sum_{i=1}^{m} \mathbf{b}_{i,0}(\tau) \otimes \mathbf{B}_i(\tau) + \mathbf{B}_i(\tau) \otimes \mathbf{b}_{i,0}(\tau)$$

$$\boldsymbol{\beta}_5(\tau) = \mathbf{a}_1(\tau) \oplus \mathbf{a}_1(\tau) + \sum_{i=1}^{m} \mathbf{b}_{i,1}(\tau) \otimes \mathbf{B}_i(\tau) + \mathbf{B}_i(\tau) \otimes \mathbf{b}_{i,1}(\tau),$$

$\mathbf{L} = \left[\mathbf{I}_d \ \mathbf{0}_{d\times 2} \right]$, 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.

From computational viewpoint, each evaluation of the formulas (42)-(43) 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 moments $\mathbf{y}_{t_{k+1}/t_k}$ and $\mathbf{V}_{t_{k+1}/t_k}$ required by the innovation estimator (31).

8 References

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.

Calderon C.P., Harris N.C. and Kiang C.H. and Cox D.D. (2009). Analyzing single-molecule manipulation experiments, Journal of Molecular Recognition, 22, 356-362.

Chiarella C., Hung H. and To T.D. (2009). The volatility structure of the fixed income market under the HJM framework: A nonlinear filtering approach. Comput. Stat. Data Anal., 53, 2075-2088.

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

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. (2012c) Approximate discrete-time schemes for the estimation of diffusion processes from complete observations. Submitted. http://arxiv.org/pdf/1212.1788v1.pdf.

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., 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.

Jimenez, J.C. and Ozaki, T. (2006) An approximate innovation method for the estimation of diffusion processes from discrete data, J. Time Series Analysis, 27, 77-97.

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. and Madsen H. (2001) Applying the EKF to stochastic differential equations with level effects, Automatica, 37, 107-112.

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

Nielsen J.N., Vestergaard M. and Madsen H. (2000b) Estimation in continuous-time stochastic volatility models using nonlinear filters, Int. J. Theor. Appl. Finance, 3, 279-308.

Nolsoe K., Nielsen J.N. and Madsen H. (2000) Prediction-based estimating function for diffusion processes with measurement noise. Technical Reports 2000, No.10, Informatics and Mathematical Modelling, Technical University of Denmark.

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

Ozaki T. (1994) The local linearization filter with application to nonlinear system identification. In Bozdogan H. (ed.) Proceedings of the first US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach, 217-240. Kluwer Academic Publishers.

Ozaki T. and Iino M. (2001) An innovation approach to non-Gaussian time series analysis, J. Appl. Prob., 38A, 78-92.

Peng H., Ozaki T. and Jimenez J.C. (2002) Modeling and control for foreign exchange based on a continuous time stochastic microstructure model, in Proceedings of the 41st IEEE Conference on Decision and Control, Las Vegas, Nevada USA, December 2002, 4440-4445.

Riera J.J., Watanabe J., Iwata K., Miura N., Aubert E., Ozaki T. and Kawashima R. (2004) A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. Neuroimage, 21, 547-567.

Schweppe F. (1965) Evaluation of likelihood function for Gaussian signals, IEEE Trans. Inf. Theory, 11, 61-70.

Shoji I. (1998) A comparative study of maximum likelihood estimators for nonlinear dynamical systems, Int. J. Control, 71, 391-404.

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.

Valdes P.A., Jimenez J.C., Riera J., Biscay R. and Ozaki T. (1999) Nonlinear EEG analysis based on a neural mass model. Biol. Cyb., 81, 415-424.

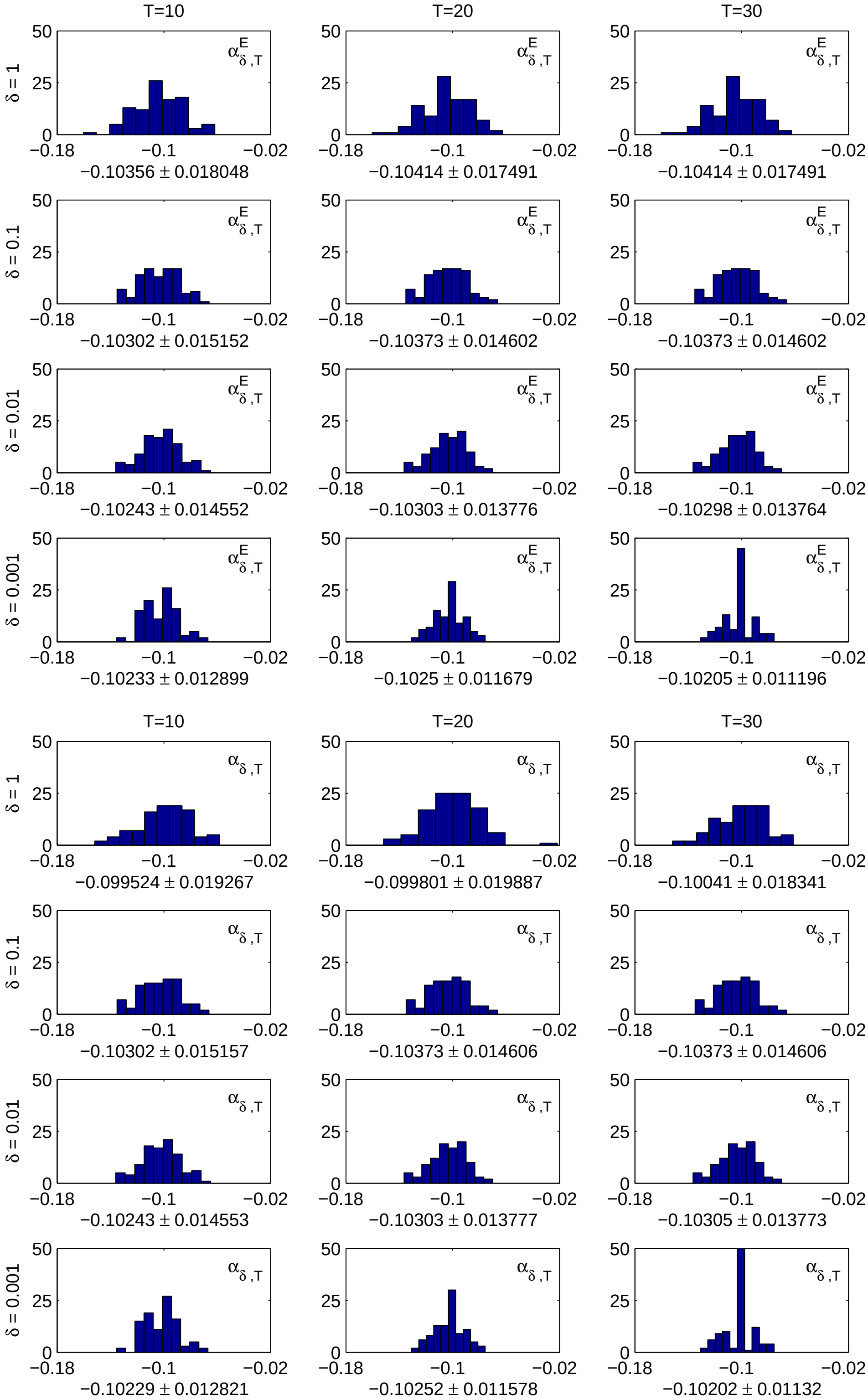

Fig. 1 Histograms and confidence limits for the exact ($\widehat{\alpha}^E_{\delta,T}$) and the conventional ($\widehat{\alpha}_{\delta,T}$) innovation estimators of α computed from the Example 1 data with sampling period δ and time interval of length T.

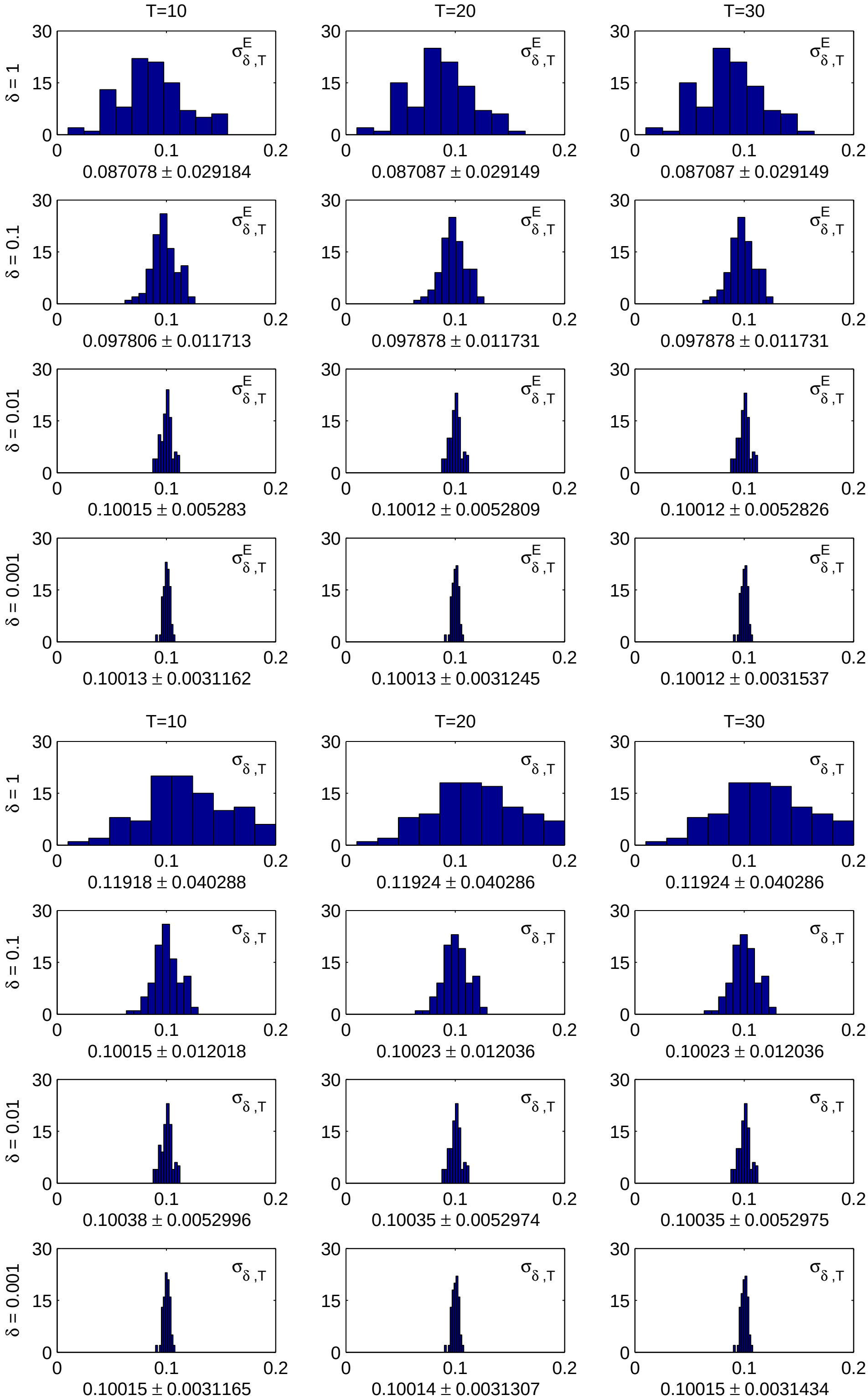

Fig. 2 Histograms and confidence limits for the exact ($\widehat{\sigma}^{E}_{\delta,T}$) and the conventional ($\widehat{\sigma}_{\delta,T}$) innovation estimators of σ computed from the Example 1 data with sampling period δ and time interval of length T.

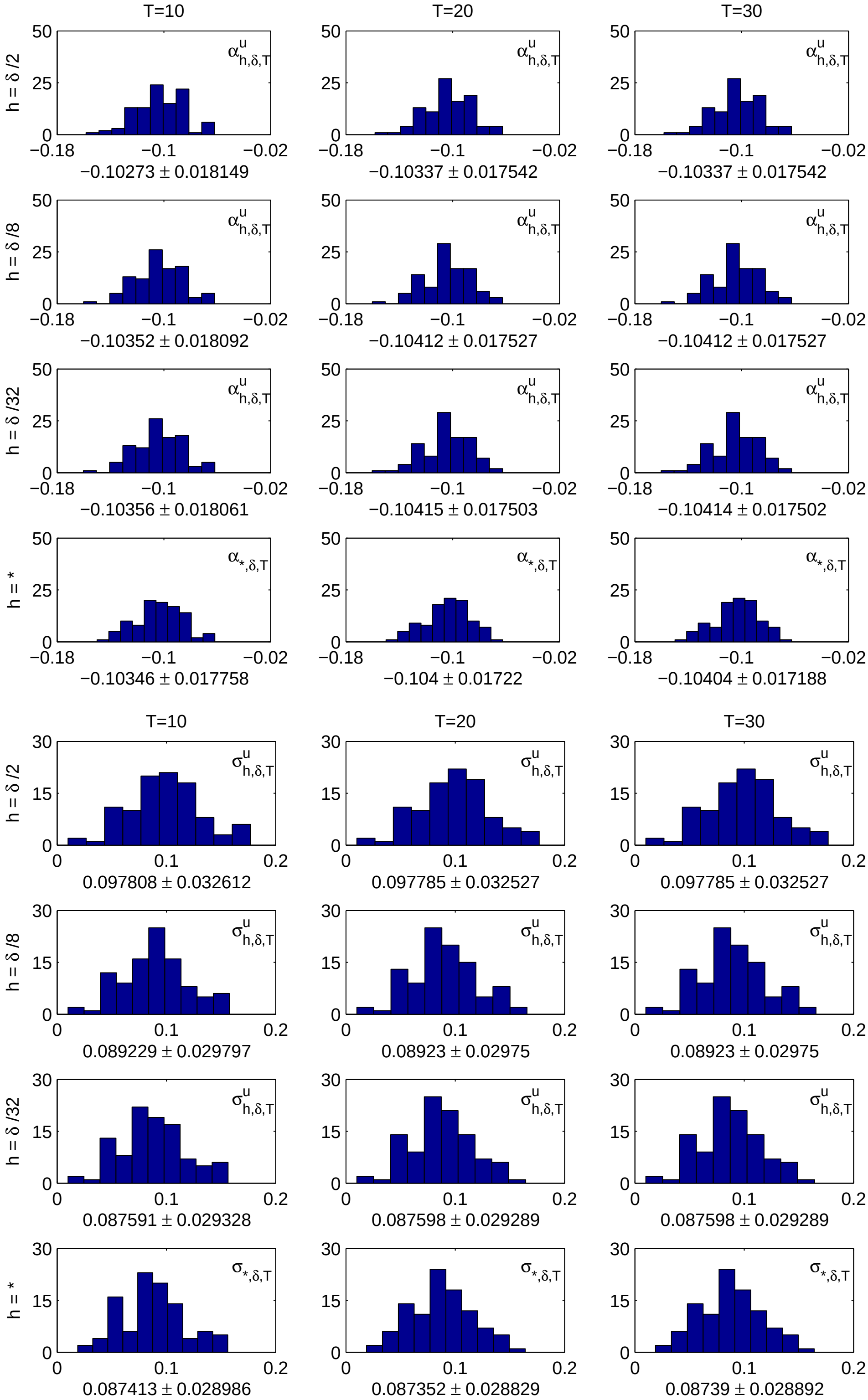

Fig. 3 Histograms and confidence limits for the oder-1 innovation estimators of α and σ computed on uniform $(\tau)^u_{h,T}$ and adaptive $(\tau)_{\cdot,T}$ time discretizations from the Example 1 data with sampling period $\delta = 1$ and time interval of length T.

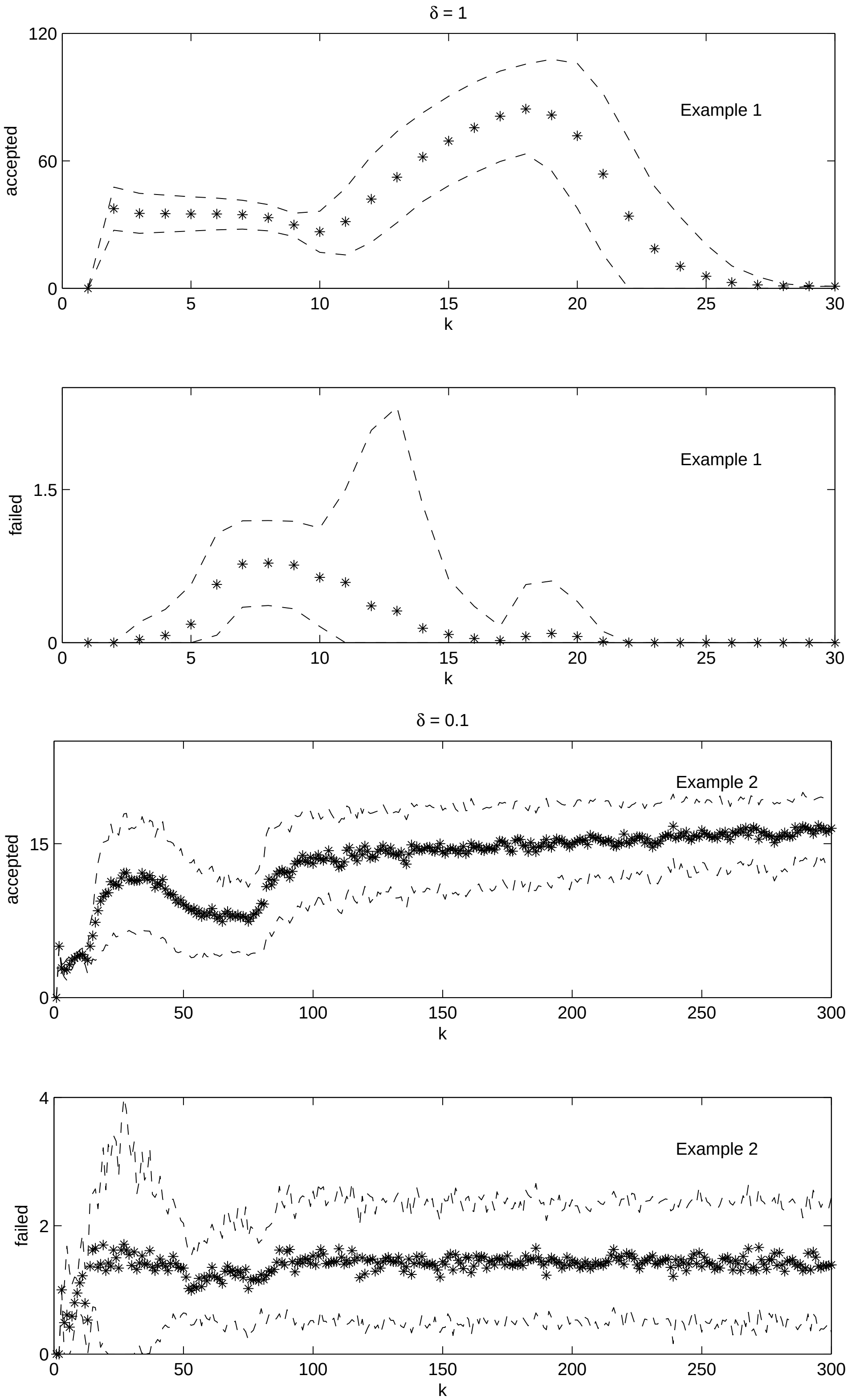

Fig. 4 Average (*) and 90% confidence limits (-) of accepted and failed steps of the adaptive innovation estimator at each $t_k \in \{t\}_N$ in the Examples 1 and 2.

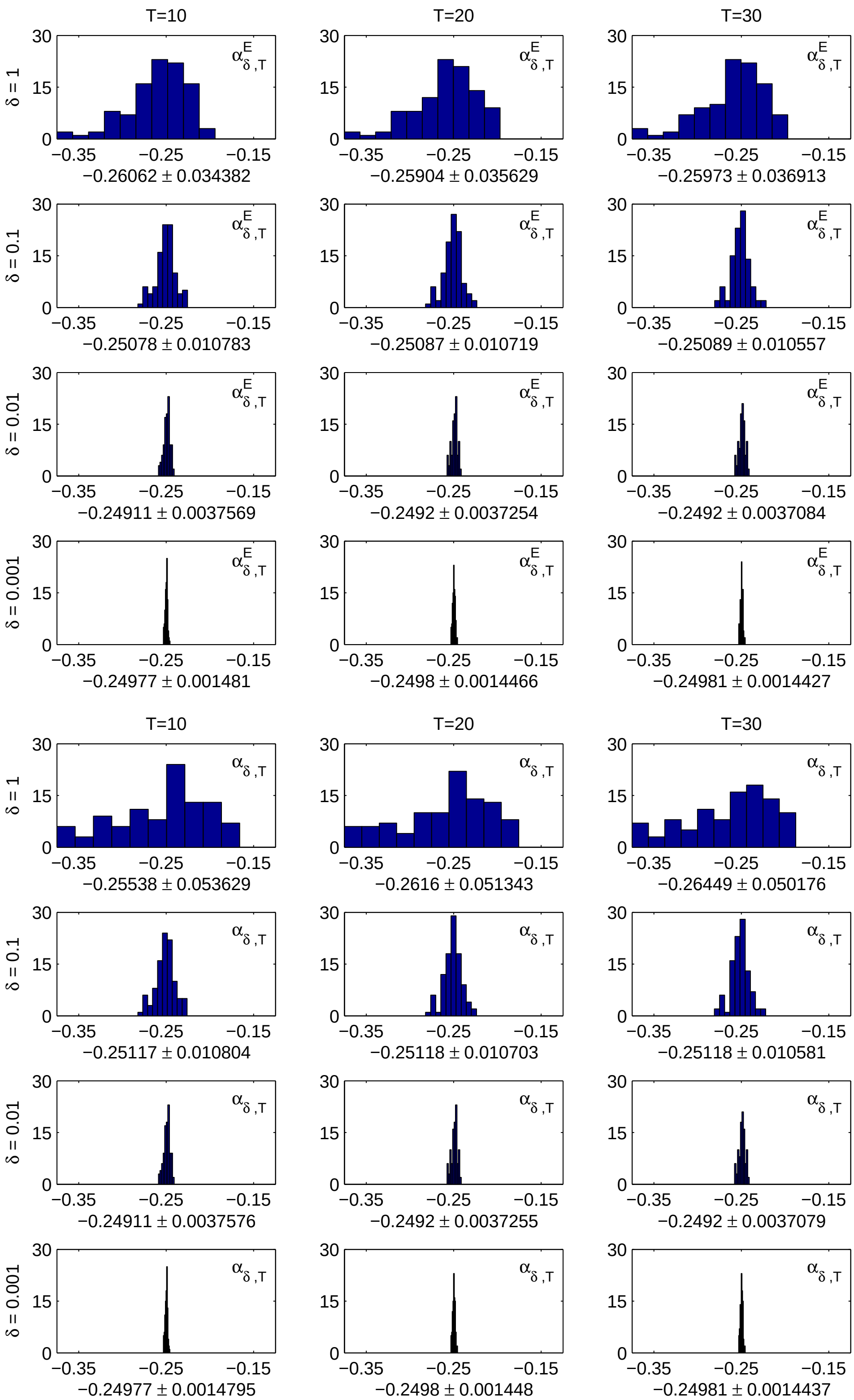

Fig. 5 Histograms and confidence limits for the exact ($\widehat{\alpha}^E_{\delta,T}$) and the conventional ($\widehat{\alpha}_{\delta,T}$) innovation estimators of α computed from the Example 2 data with sampling period δ and time interval of length T.

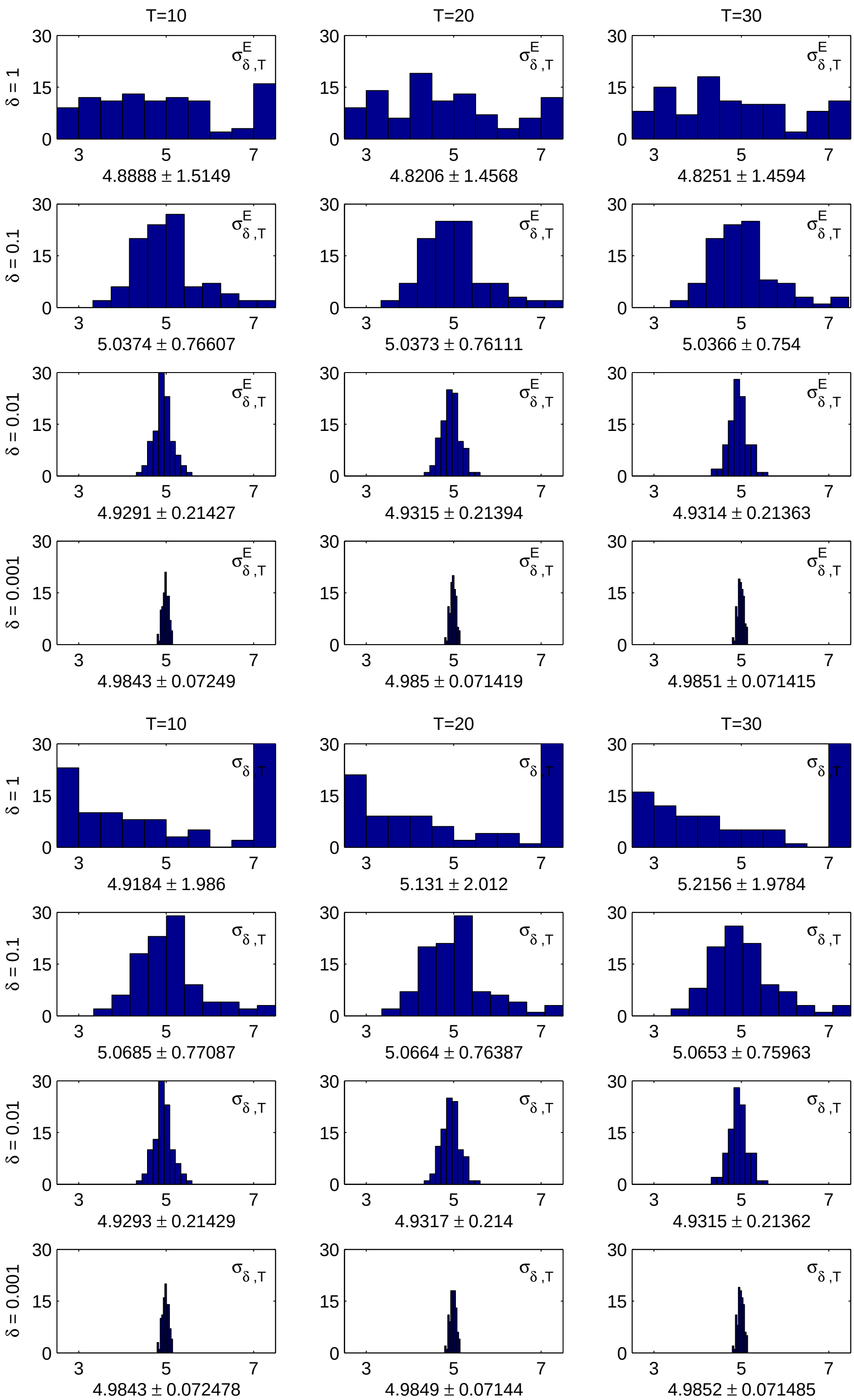

Fig. 6 Histograms and confidence limits for the exact ($\widehat{\sigma}^E_{\delta,T}$) and the conventional ($\widehat{\sigma}_{\delta,T}$) innovation estimators of σ computed from the Example 2 data with sampling period δ and time interval of length T.

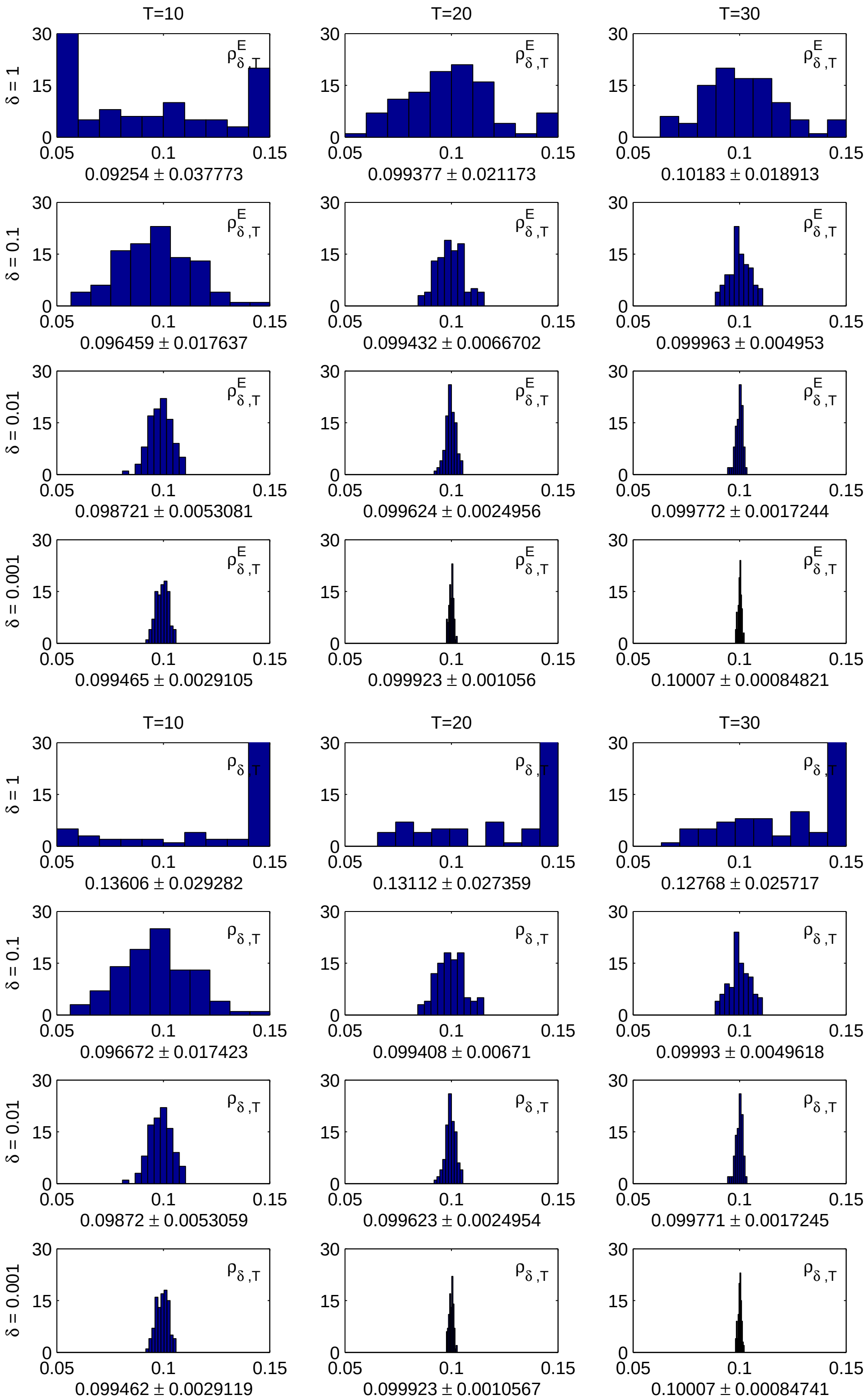

Fig. 7 Histograms and confidence limits for the exact $(\widehat{\rho}^{E}_{\delta,T})$ and the conventional $(\widehat{\rho}_{\delta,T})$ innovation estimators of ρ computed from the Example 2 data with sampling period δ and time interval of length T.

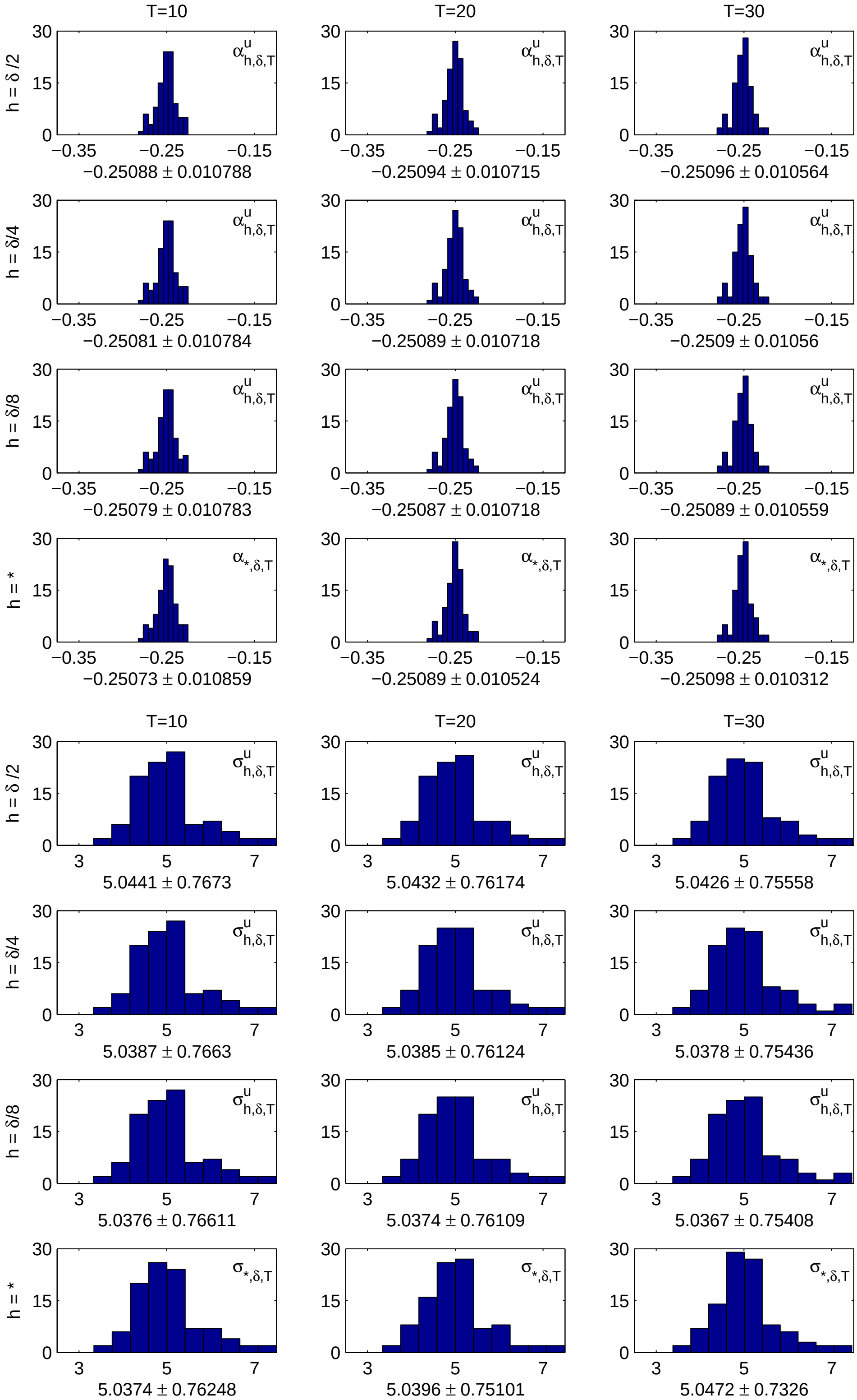

Fig. 8a Histograms and confidence limits for the oder-1 innovation estimators of α and σ computed on uniform $(\tau)^u_{h,T}$ and adaptive $(\tau)_{\cdot,T}$ time discretizations from the Example 2 data with sampling period $\delta = 0.1$ and time interval of length T.

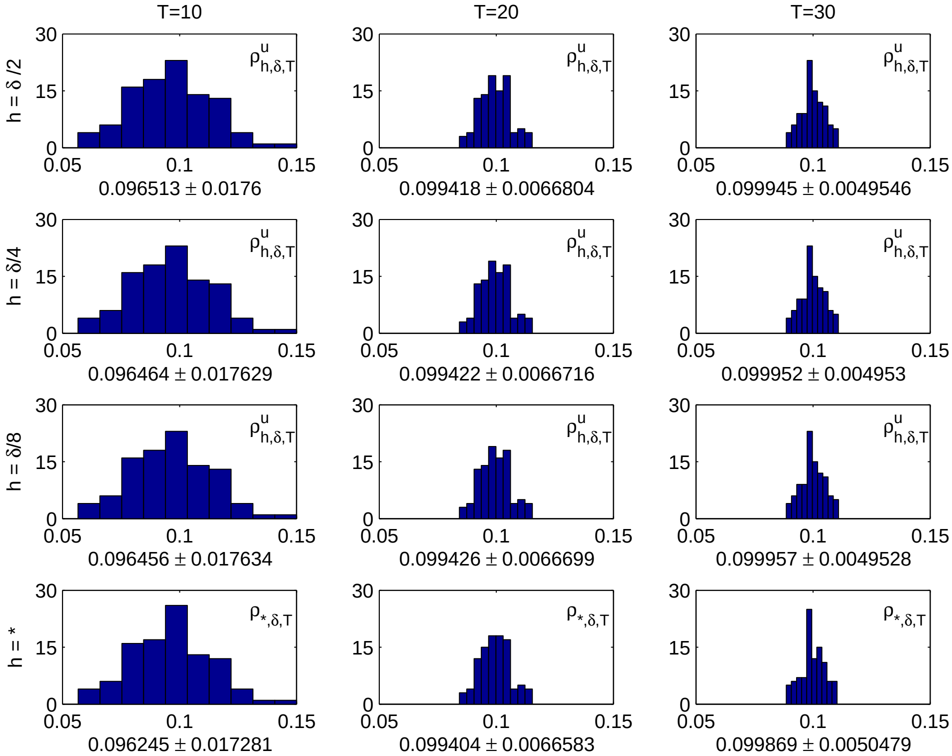

Fig. 8b Histograms and confidence limits for the oder-1 innovation estimators of ρ computed on uniform $(\tau)^u_{h,T}$ and adaptive $(\tau)_{\cdot,T}$ time discretizations from the Example 2 data with sampling period $\delta = 0.1$ and time interval of length T.

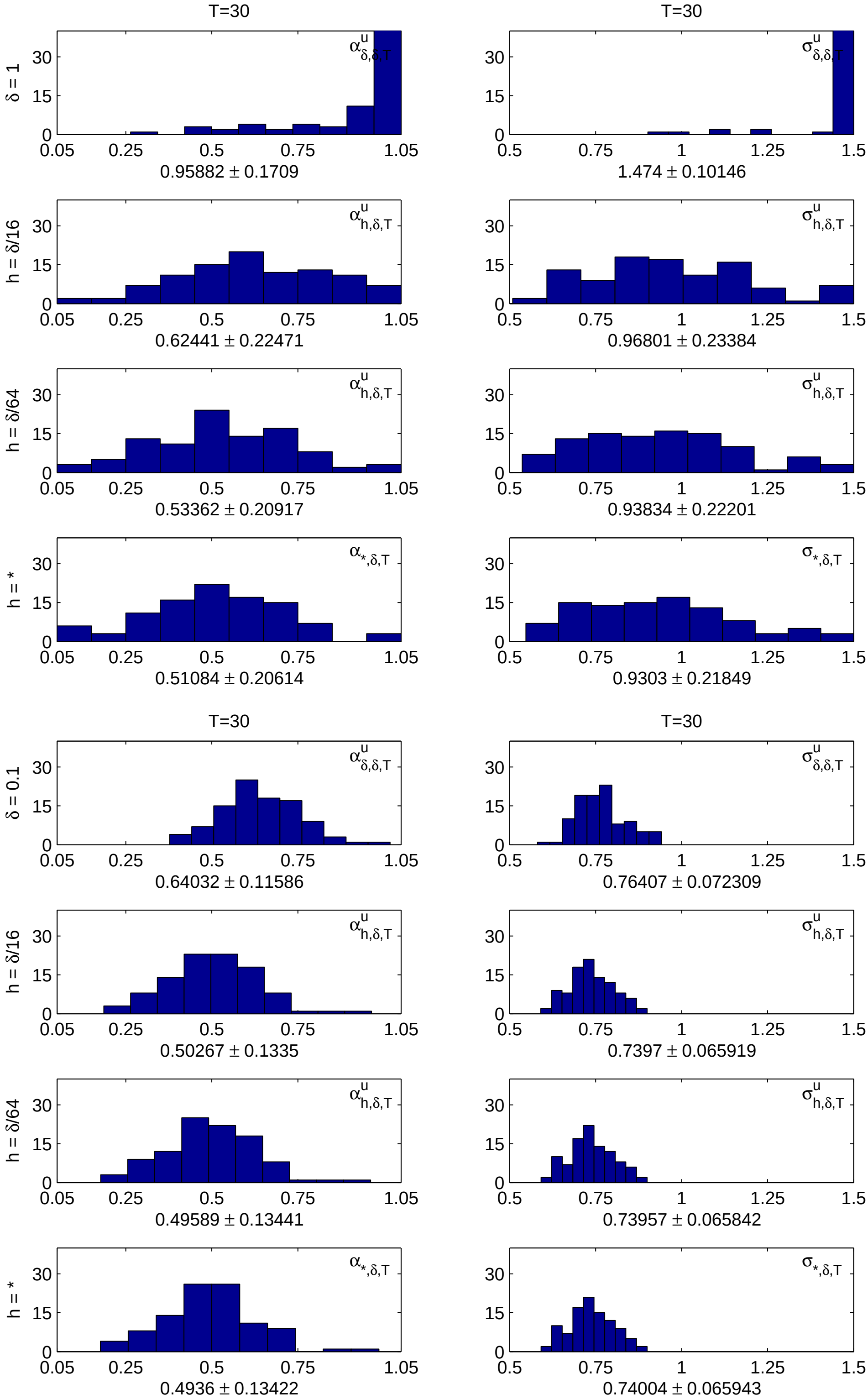

Fig. 9 Histograms and confidence limits for the oder-1 innovation estimators of α and σ computed on uniform $(\tau)^u_{h,T}$ and adaptive $(\tau)_{\cdot,T}$ time discretizations from the Example 3 data with sampling period δ and time interval of length $T = 30$.

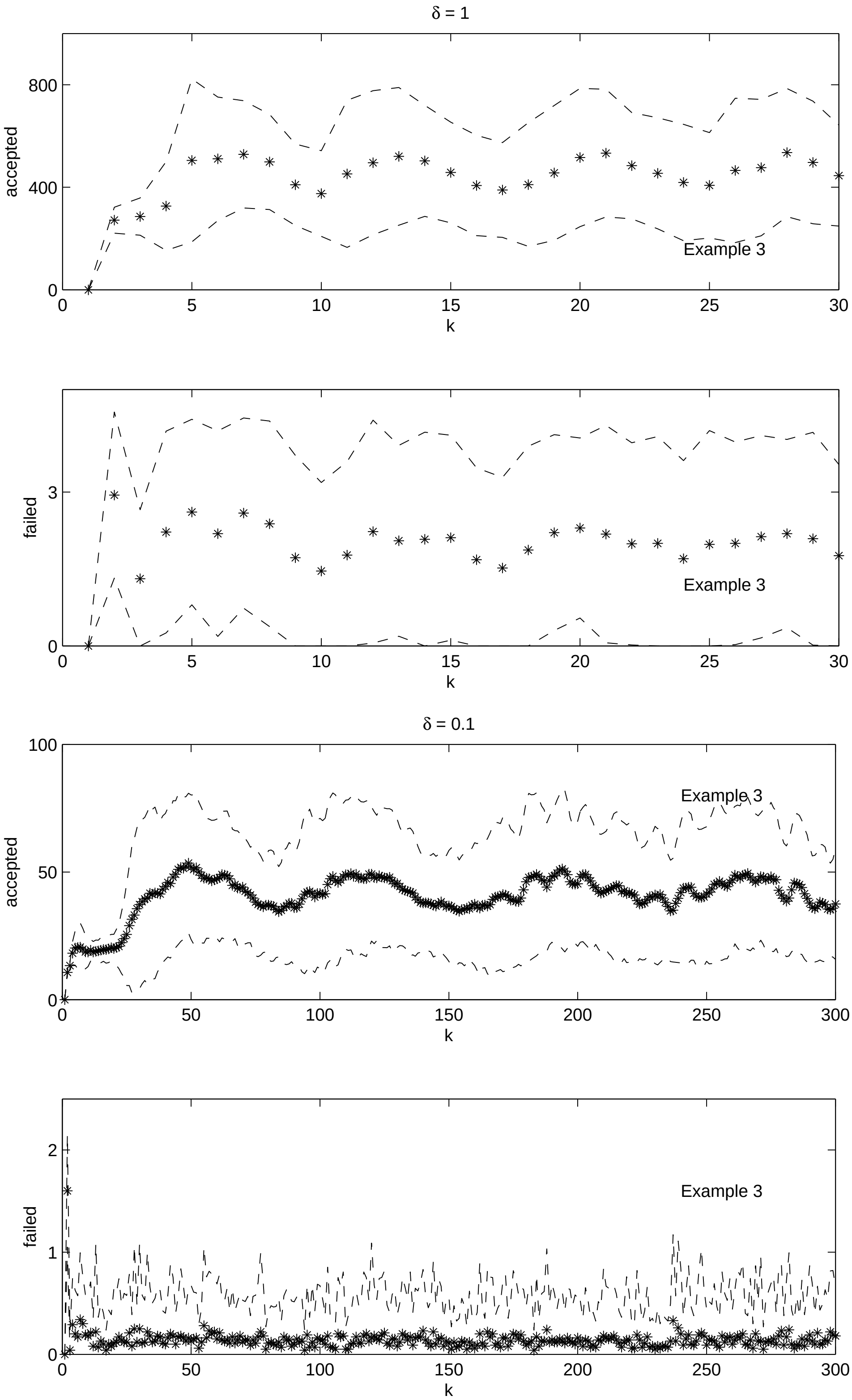

Fig. 10 Average (*) and 90% confidence limits (-) of accepted and failed steps of the adaptive innovation estimator at each $t_k \in \{t\}_N$ in the Example 3.

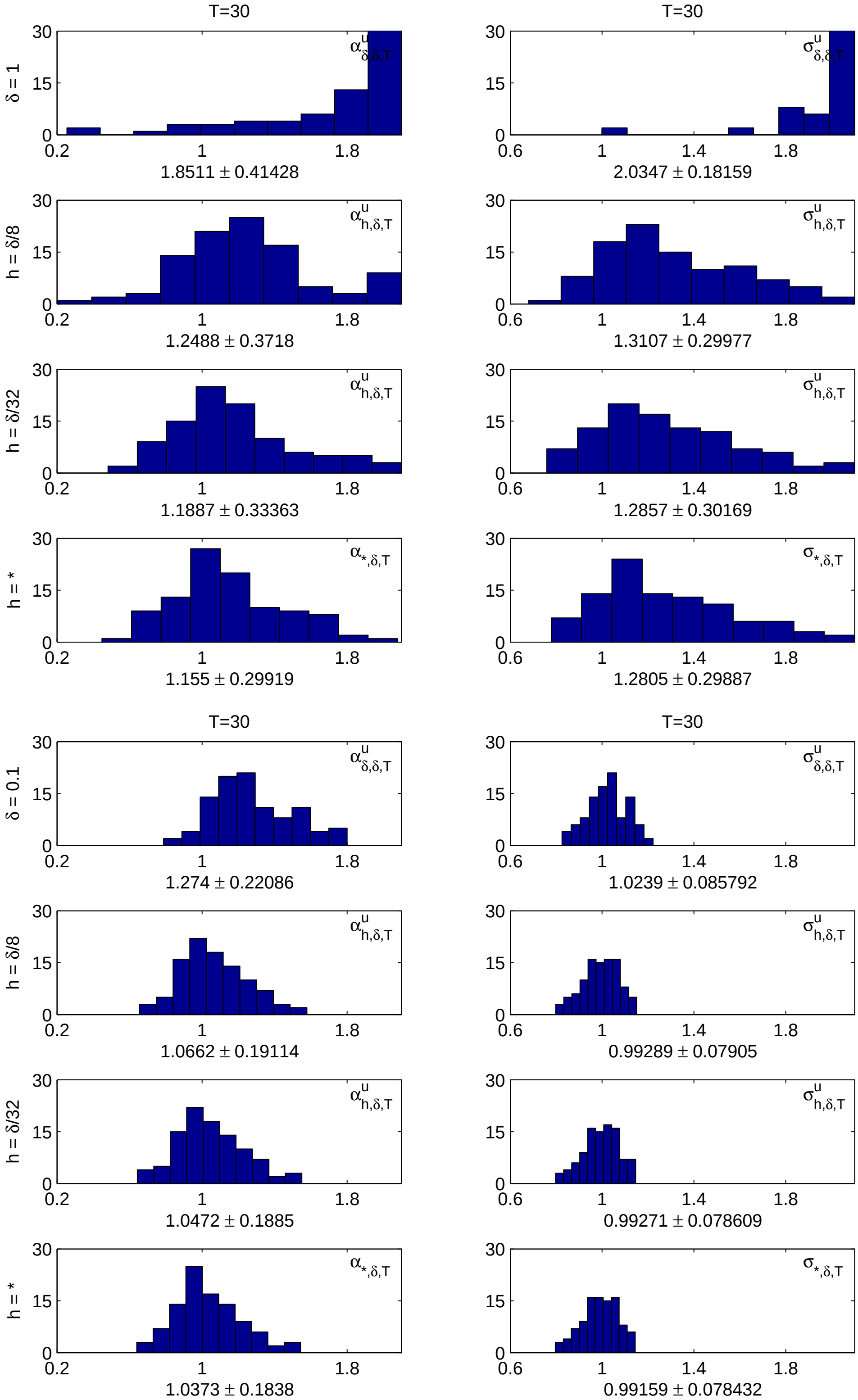

Fig. 11 Histograms and confidence limits for the oder-1 innovation estimators of α and σ computed on uniform $(\tau)^u_{h,T}$ and adaptive $(\tau)_{\cdot,T}$ time discretizations from the Example 4 data with sampling period δ and time interval of length $T = 30$

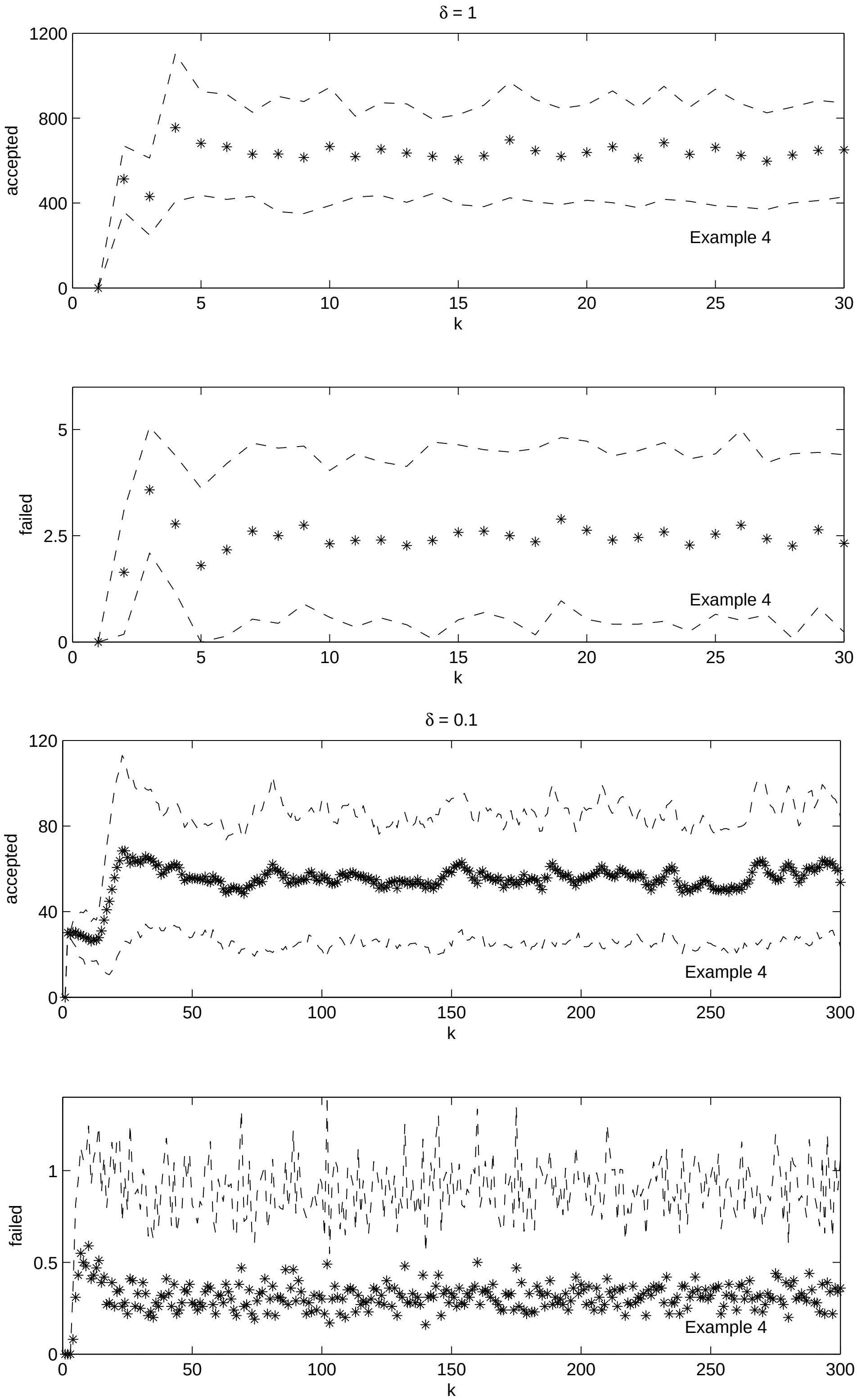

Fig. 12 Average (*) and 90% confidence limits (-) of accepted and failed steps of the adaptive innovation estimator at each $t_k \in \{t\}_N$ in the Example 4.