

SPECIAL ISSUE

## INDIRECT INFERENCE FOR TIME SERIES USING THE EMPIRICAL CHARACTERISTIC FUNCTION AND CONTROL VARIATES

RICHARD A. DAVIS,<sup>a</sup> THIAGO DO RÊGO SOUSA<sup>b</sup> AND CLAUDIA KLÜPPELBERG<sup>b</sup>

<sup>a</sup> *Department of Statistics, Columbia University, New York, 10027, NY, USA*

<sup>b</sup> *Center for Mathematical Sciences, Technical University of Munich, Boltzmannstr. 3 Garching, 85748, Germany*

We estimate the parameter of a stationary time series process by minimizing the integrated weighted mean squared error between the empirical and simulated characteristic function, when the true characteristic functions cannot be explicitly computed. Motivated by Indirect Inference, we use a Monte Carlo approximation of the characteristic function based on i.i.d. simulated blocks. As a classical variance reduction technique, we propose the use of control variates for reducing the variance of this Monte Carlo approximation. These two approximations yield two new estimators that are applicable to a large class of time series processes. We show consistency and asymptotic normality of the parameter estimators under strong mixing, moment conditions, and smoothness of the simulated blocks with respect to its parameter. In a simulation study we show the good performance of these new simulation based estimators, and the superiority of the control variates based estimator for Poisson driven time series of counts.

*Received 13 August 2020; Revised 15 December 2020; Accepted 27 December 2020*

**Keywords:** Asymptotic normality; characteristic function; control variates; indirect inference estimation; time series of counts; SLLN; variance reduction.

**JEL.** 62F12; 62G20; 62M10; 65C05; 91G70.

### 1. INTRODUCTION

Let  $(X_j)_{j \in \mathbb{Z}}$  be a stationary time series, whose distribution depends on  $\theta \in \Theta \subset \mathbb{R}^q$  for some  $q \in \mathbb{N}$ . Denote by  $\theta_0 \in \Theta$  the true parameter, which we want to estimate from observations  $X_1, \dots, X_T$  of the time series. Maximum likelihood estimation (MLE) has been extensively used for parameter estimation, since under weak regularity conditions it is known to be asymptotically efficient. For many models, however, MLE is not always feasible to carry out, due to a likelihood that may be intractable to compute, or maximization of the likelihood is difficult, or because the likelihood function is unbounded on  $\Theta$ . To overcome such problems, alternative methods have been developed, for instance, the generalized method of moments (GMM) in Hansen (1982), the quasi-maximum likelihood estimation (QMLE) in White (1982), and composite likelihood methods in Lindsay (1988).

In a similar vein, Feuerverger (1990) proposed an estimator based on matching the empirical characteristic function (chf) computed from blocks of the observed time series and the true chf. More specifically, given a fixed  $p \in \mathbb{N}$ , the observed blocks of  $X_1, \dots, X_T$  are

$$\mathbf{X}_j = (X_j, \dots, X_{j+p-1}), \quad j = 1, \dots, n, \quad (1.1)$$

where  $n = T - p + 1$ . In that article, a finite set of points in  $\mathbb{R}^p$  needs to be chosen as arguments for which the true and the empirical chf are compared. However, the practical choice of this set depends on the problem at hand and the asymptotic results derived in Feuerverger (1990) do not offer practical guidance for choosing these points. To

\*Correspondence to: Thiago do Rêgo Sousa, Center for Mathematical Sciences, Technical University of Munich, 85748 Garching, Boltzmannstr. 3, Germany. Email: thiago.sousa@tum.de

overcome this limitation Yu (1998) and Knight and Yu (2002) considered an integrated weighted squared distance between the empirical and the true chfs.

This method has been used in a variety of applications; an interesting review article, Yu (2004) contains a wealth of examples and references. More recent publications, where the method has been successfully applied to discrete-time models include Knight *et al.* (2002), Meintanis and Taufer (2012), Kotchoni (2012), Milovanovic *et al.* (2014), Francq and Meintanis (2016), and Ndongo *et al.* (2016). The method also applies to continuous-time processes after discretization and has been used prominently for Lévy-driven models. The book Belomestny *et al.* (2015) provides additional insight and references in this field.

The principal goal of this article is to extend the ideas of these articles to a more general setting. For example, we do not assume the idealized situation for which the chf has an explicit expression as a function of  $\theta \in \Theta$ . We propose two new estimators of  $\theta$ , which are based on replacing the true chf with estimates that are constructed from a functional approximation of the chf constructed from simulated sample paths of  $(X_j(\theta))_{j \in \mathbb{Z}}$ .

While much attention has been given to the choice of the integrated distance used when computing such estimators, which under some regularity conditions can achieve the Cramér-Rao efficiency bound (see eq. (2.3) of Knight and Yu (2002) and Proposition 4.2 of Carrasco *et al.* (2007)), the focus of our article is on the practical and theoretical aspects that emerge when it is required to approximate the theoretical chf for parameter estimation. For more details on the search for efficient estimators we refer to (Carrasco *et al.*, 2007; Carrasco and Florens, 2014; Carrasco and Kotchoni, 2017).

Our first estimator is computed from a simple Monte Carlo approximation to replace the true, but unknown chf. This is similar to the simulated method of moments of McFadden (1989) and of the indirect inference method (Smith (1993) and Gouriéroux *et al.* (1993)). In particular, indirect inference has been successfully applied in a variety of situations: parameter estimation of continuous time models with stochastic volatility (Bianchi and Cleur (1996), Jiang (1998), Raknerud and Skare (2012), Laurini and Hotta (2013) and Wahlberg *et al.* (2015)), robust estimation (de Luna and Genton (2001) and Fasen-Hartmann and Kimmig (2020)), and finite sample bias reduction (Gouriéroux *et al.* (2000); Gouriéroux *et al.* (2010) and Do Rêgo Sousa *et al.* (2019)).

More precisely, for many different  $\theta \in \Theta$ , we simulate an i.i.d. sample of blocks denoted by

$$\tilde{X}_j(\theta) = (\tilde{X}_1^{(j)}(\theta), \dots, \tilde{X}_p^{(j)}(\theta)), \quad j = 1, \dots, H, \quad (1.2)$$

for  $H \in \mathbb{N}$ , and define a *simulation based parameter estimator*, which minimizes the integrated weighted mean squared error, which is the integrated distance we use, between the empirical chf computed from the blocks (1.2) of the observed time series and its simulated version computed from a large number of simulated paths of the time series.

This is in contrast to the simulation based estimator defined in Section 5.2 of Carrasco *et al.* (2007), which is computed from one long time series path instead of the i.i.d. sample of blocks in (1.2) (a similar method has been applied by Forneron (2018) to estimate the structural parameters and the distribution of shocks in dynamic models). Since we compute the Monte Carlo approximation of the chf from independent blocks, it should have smaller variance than the corresponding one for dependent blocks. Our method gives a chf approximation which yields strongly consistent and asymptotically normal parameter estimators. We also report their small sample properties for different models.

Furthermore, as the Monte Carlo approximation of the chf is computed from i.i.d. blocks of a time series, control variates techniques (see Glynn and Szechtman (2002) and Robert and Casella (2004)) provide an even more accurate approximation for the chf. Control variates techniques are classical variance reduction methods in simulation. The idea is to use a set of control variates, which are correlated with the chf. The method then approximates the joint covariance matrix of the control variates and the chf, and uses it to construct a new Monte Carlo approximation of the chf. We choose the first two terms in the Taylor expansion of the complex exponential  $e^{i\langle t, X_1(\theta) \rangle}$ ,  $\langle t, X_1(\theta) \rangle$  and  $\langle t, X_1(\theta) \rangle^2$  for  $\theta \in \Theta$  as control variates, where  $\langle \cdot, \cdot \rangle$  denotes the usual Euclidean inner product in  $\mathbb{R}^d$ . This requires knowing the mean and covariance matrix of  $X_1(\theta)$  for  $\theta \in \Theta$ .

In assessing the performance of both the Monte Carlo approximation and the control variates approximation of the chf, two trends emerge. First, both the Monte Carlo and the control variates approximations work better for small values of the argument. Second, the control variates approximation performs much better than the Monte Carlo approximation, in particular, for small values of the argument. As a consequence, we propose a *control variates based parameter estimator* whose integrated mean squared error distinguishes between small and large values of the argument.

Under regularity conditions we prove strong consistency of the proposed parameter estimators and asymptotic normality of the simulation based parameter estimator. We find that the simulation based parameter estimator is asymptotically normal with asymptotic covariance matrix equal to the one of the oracle estimator as derived in Knight and Yu (2002). From this we conclude that there cannot be any improvement in the limit law for the asymptotic normality of the control variates based estimator. However, we prove that it is computed from a better approximation of the chf. Thus, the control variates estimator improves the finite sample performance compared to the simulation based parameter estimator.

It is assumed throughout that  $(X_j)_{j \in \mathbb{Z}}$  is a stationary time series. This ensures that the blocks of random variables in (1.1) are stationary, from which we obtain convergence of the empirical chf to the joint chf. Now in some restricted cases, our method can be adapted to special types of non-stationarity. For example, if  $(X_j)_{j \in \mathbb{Z}}$  is non-stationary, but the differenced process  $\nabla X_j = X_j - X_{j-1}$  is stationary, then our methodology can be applied directly to  $\nabla X_j$ . Similarly, if  $X_j = Y_j + \mu_j$ , where  $Y_j$  is stationary and  $\mu_j$  is a mean function that can be estimated consistently say by  $\hat{\mu}_j$ , then the methodology can be applied to  $X_j - \hat{\mu}_j$ . We do not pursue this line of investigation here.

The finite sample performance of the estimators is investigated for two important models. We begin with a stationary Gaussian ARFIMA model, whose chf is explicitly known so that we can use the oracle estimator and compare its performance with the simulated based estimator. Their performance is comparable and also very close to the MLE, so in this model there is no need to use control variates. The second example is a nonlinear model for time series of counts, which has been proposed originally in Zeger (1988) and applied, for instance, for modeling disease counts (see also Campbell (1994), Chan and Ledolter (1995) and Davis *et al.* (1999)).

In the second example, the oracle estimator does not apply, since the chf of a Poisson-AR process cannot be computed in closed form. For this model and different parameter sets, both the simulation based and the control variates based estimators perform satisfactorily, and the control variates based estimator improves the performance of the simulation based estimator considerably. When compared with the composite pairwise likelihood estimator in Davis and Yau (2011), the control variates based estimator has comparable or even smaller bias.

Our article is organized as follows. In Section 2 we present the oracle estimator, and the estimators computed from a Monte Carlo approximation and from a control variates approximation of the chf in detail. Here we also motivate the choice of the control variates used. The asymptotic properties of the two new estimators are established in Section 3. As all estimators are computed from true or approximated chf's we assess their performance in Section 4, first for a Gaussian AR(1) process and then for the Poisson-AR process. Practical aspects of calculating the weighted least squares function are discussed in Section 5, as well as the estimation results for finite samples. In Section 5.1 we compare the oracle estimator, the simulation based parameter estimator and the MLE for a Gaussian ARFIMA model, whereas in Section 5.2 we compare the simulation based parameter estimator and the control variates based estimator for the Poisson-AR process. The proofs of the main results in Section 3, of Lemma 1 of Section 5, and the Tables discussed in Sections 5.1 and 5.2 are provided in the Appendix.

## 2. PARAMETER ESTIMATION BASED ON THE EMPIRICAL CHARACTERISTIC FUNCTION

Throughout we use the following notation. For  $z \in \mathbb{C}$  we use the  $L^2$ -norm:  $|z| = \sqrt{z \bar{z}}$ , where  $\bar{z}$  is the complex conjugate of  $z$ . For  $x \in \mathbb{R}^d$  and  $d \in \mathbb{N}$  we denote by  $|x|$  the  $L^2$ -norm, but recall that in  $\mathbb{R}^d$  all norms are equivalent. For  $z \in \mathbb{C}$  the symbols  $\Re(z)$  and  $\Im(z)$  denote its real and imaginary part. For a function  $f : \mathbb{R}^q \rightarrow \mathbb{R}^p$  its Jacobi matrix is given by  $\nabla_{\theta} f(\theta) = \frac{\partial f(\theta)}{\partial \theta^T} \in \mathbb{R}^{p \times q}$  and  $\nabla_{\theta}^2 f(\theta) = \frac{\partial \text{vec}(\nabla_{\theta} f(\theta))}{\partial \theta^T} \in \mathbb{R}^{pq \times q}$ .

## 2.1. The Oracle Estimator

Let  $(X_j(\theta))_{j \in \mathbb{Z}}$  be a stationary time series process, whose distribution depends on  $\theta \in \Theta \subset \mathbb{R}^q$  for some  $q \in \mathbb{N}$ . Denote by  $\theta_0 \in \Theta$  the true parameter, which we want to estimate, and suppose that we observe  $X_1, \dots, X_T$ . Given a fixed  $p \in \mathbb{N}$ , define for  $\theta \in \Theta$  the  $p$ -dimensional blocks

$$X_j(\theta) = (X_j(\theta), \dots, X_{j+p-1}(\theta)), \quad j = 1, \dots, n, \quad (2.1)$$

where  $n = T - p + 1$ . For  $j = 1, \dots, n$ , the observed blocks correspond to  $X_j = (X_j, \dots, X_{j+p-1})$ , which can be used to calculate the *empirical characteristic function (chf)*, defined as

$$\varphi_n(t) = \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle}, \quad t \in \mathbb{R}^p. \quad (2.2)$$

Under mild conditions such as ergodicity,  $\varphi_n(t)$  converges a.s. pointwise to the true chf  $\varphi(t) = \mathbb{E}e^{i\langle t, X_1 \rangle}$  for all  $t \in \mathbb{R}^p$ . We assume that  $p$  is chosen in such a way that  $\varphi(\cdot)$  uniquely identifies the parameter of interest  $\theta$ . The idea of estimating  $\theta_0$  from a single time series observation by matching the empirical chf of blocks of the observed time series and the true one has been proposed in Yu (1998) and Knight and Yu (2002), and we use the one in Knight and Yu (2002), where the *oracle estimator* of  $\theta_0$  is defined as

$$\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} Q_n(\theta), \quad (2.3)$$

where

$$Q_n(\theta) = \int_{\mathbb{R}^p} |\varphi_n(t) - \varphi(t, \theta)|^2 w(t) dt, \quad \theta \in \Theta, \quad (2.4)$$

with suitable weight function  $w$  such that the integral is well-defined, and chf

$$\varphi(t, \theta) = \mathbb{E}e^{i\langle t, X_1(\theta) \rangle}, \quad t \in \mathbb{R}^p. \quad (2.5)$$

In an ideal situation,  $\varphi(\cdot, \theta)$  has an explicit expression, which is known for all  $\theta \in \Theta$ .

## 2.2. Estimator Based on a Monte Carlo Approximation of $\varphi(\cdot, \theta)$

Unfortunately, a closed form expression of the chf  $\varphi(\cdot, \theta)$  is for many time series processes not available. However, it can be approximated by a Monte Carlo simulation, and an idea borrowed from the simulated method of moments (McFadden (1989), see also Smith (1993) and Gouriéroux *et al.* (1993) for a similar idea in the context of indirect inference) is to replace  $\varphi(\cdot, \theta)$  by its functional approximation constructed from simulated sample paths of  $(X_j(\theta))_{j \in \mathbb{Z}}$ . For many different  $\theta \in \Theta$ , we simulate, independent of the observed time series, an i.i.d. sample of the blocks in (2.1) denoted by

$$\tilde{X}_j(\theta) = (\tilde{X}_1^{(j)}(\theta), \dots, \tilde{X}_p^{(j)}(\theta)), \quad j = 1, \dots, H, \quad (2.6)$$

for  $H \in \mathbb{N}$ , and define the *Monte Carlo approximation* of  $\varphi(\cdot, \theta)$  based on these simulations as

$$\varphi_H(t, \theta) = \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \tilde{X}_j(\theta) \rangle}, \quad t \in \mathbb{R}^p. \quad (2.7)$$

If we replace  $\varphi(\cdot, \theta)$  in (2.4) by  $\varphi_H(\cdot, \theta)$ , we obtain the *simulation based parameter estimator*

$$\hat{\theta}_{n,H} = \arg \min_{\theta \in \Theta} Q_{n,H}(\theta), \tag{2.8}$$

where

$$Q_{n,H}(\theta) = \int_{\mathbb{R}^p} |\varphi_n(t) - \varphi_H(t, \theta)|^2 w(t) dt, \tag{2.9}$$

with suitable weight function  $w$  such that the integral is well-defined.

**Remark 2.1.** An alternative approximation to (2.7) of the chf is based on generating one long time series path and use the empirical chf of the consecutive blocks of  $p$ -dimensional random variables constructed as in (2.1) (see Carrasco *et al.* (2007)). While being unbiased, the approximation will generally have larger variance than the approximation (2.7). Nevertheless, when it is expensive to generate realizations even of dimension  $p$ , for instance, when a long burn-in time is required to achieve stationarity, it may be computationally more efficient to generate one long time series. While we do not pursue this approach here, the technical aspects of working with one long time series are not much different than the estimate based on independent replicates as in (2.7), but might require a much larger sample size than desired to control the variance of the estimate. This is especially true for long-memory time series.

Since  $\varphi_H(\cdot, \theta)$  is based on  $H$  i.i.d. time series blocks, we can reduce its variance further using control variates to produce an even more accurate approximation for the chf. This will result in an improved version of  $\hat{\theta}_{n,H}$ .

### 2.3. Estimator Based on a Control Variates Approximation of $\varphi(\cdot, \theta)$

The estimator  $\hat{\theta}_{n,H}$  in (2.8) requires only that the stationary time series process can be simulated, and is therefore easily applicable to a large class of models. When computing  $Q_{n,H}(\theta)$  of (2.9), it is very important that the error

$$\xi_H(t, \theta) = |\varphi_H(t, \theta) - \varphi(t, \theta)|, \quad t \in \mathbb{R}^p, \theta \in \Theta, \tag{2.10}$$

in approximating the true chf is small, since it propagates to  $\hat{\theta}_{n,H}$ . To reduce the variance of the empirical chf  $\varphi_H(\cdot, \theta)$ , we use the method of control variates, an often used variance reduction technique in the context of Monte Carlo integration (Glynn and Szechtman (2002), Oates *et al.* (2017), Portier and Segers (2019)).

We construct a control variates approximation of  $\varphi(\cdot, \theta)$  from the i.i.d. sample  $\tilde{X}_j(\theta), j = 1, \dots, H$ , as in (2.6). We also require explicit expressions for the moments  $\mathbb{E}\langle t, X_1(\theta) \rangle^\nu$  for  $\nu = 1, 2$  and  $\theta \in \Theta$ .

Recall that  $\tilde{X}_1(\theta) \stackrel{d}{=} X_1(\theta)$  for all  $\theta \in \Theta$ , so that both random variables have the same moments. As in Portier and Segers (2019), we denote by  $P_\theta$  the distribution of the block  $X_1(\theta)$  and by  $P_{H,\theta}$  its empirical version. For example, if  $f_t(x) = e^{i\langle t, x \rangle}$  for  $t, x \in \mathbb{R}^p$ , we want to provide a good approximation for  $\varphi(t, \theta) = \mathbb{E}f_t(X_1(\theta)) =: P_\theta(f_t)$  for  $\theta \in \Theta$ . To apply the control variates technique, we need control functions, which are correlated with  $f_t(X_1(\theta))$  and whose expectations are known. In the time series context, it is often that we know the first and second order structure of the process in closed form. Even for complicated models, for example, models defined in terms of stochastic integrals (see e.g., (Brockwell, 2001; Klüppelberg *et al.*, 2004; Brockwell *et al.*, 2006; Stelzer, 2010)) these expressions are available. The first and second order of  $X_1(\theta)$  appear in the Taylor series of  $f_t(X_1(\theta))$  and therefore they are natural choices of control functions. We also remark that if the time series process also allows for the computation of additional moments expressions in closed form, which are correlated with  $f_t(X_1(\theta))$ , then we encourage using them as control functions while approximating the chf. We describe now the construction of the control variates approximation in detail.

We use the first two terms in the Taylor series of the complex function  $f_i(x)$ , which suggests the vector of control functions  $h_{i,\theta} = (h_{1,i,\theta}, h_{2,i,\theta})^T$ , where for  $v = 1, 2$ ,

$$h_{v,i,\theta}(x) = \langle t, x \rangle^v - \mathbb{E} \langle t, X_1(\theta) \rangle^v, \quad t \in \mathbb{R}^p,$$

so that  $P_\theta(h_{i,\theta}) = 0$ , the zero vector in  $\mathbb{R}^2$ . The Monte Carlo approximation of  $\varphi(\cdot, \theta)$  based on the i.i.d. sample  $\tilde{X}_j(\theta), j = 1, \dots, H$ , is then

$$P_{H,\theta}(f_i) = \frac{1}{H} \sum_{j=1}^H f_i(\tilde{X}_j(\theta)) = \frac{1}{H} \sum_{j=1}^H e^{i \langle t, \tilde{X}_j(\theta) \rangle} = \varphi_H(t, \theta). \quad (2.11)$$

Since  $\mathbb{E}P_{H,\theta}(f_i) = \mathbb{E}f_i(X_1(\theta))$ , the Monte Carlo approximation  $\varphi_H(t, \theta)$  is unbiased and has variance

$$\text{Var}[P_{H,\theta}(f_i)] = H^{-1} \sigma_\theta^2(f_i) \quad \text{with} \quad \sigma_\theta^2(f_i) = P_\theta(\{f_i - P_\theta(f_i)\}^2). \quad (2.12)$$

Then for every vector  $\beta \in \mathbb{C}^2$ , we have that  $P_{H,\theta}(f_i) - \beta^T P_{H,\theta}(h_{i,\theta})$  is also an unbiased estimator of  $\varphi(t, \theta)$ . Since  $\tilde{X}_j(\theta), j = 1, \dots, H$ , is an independent sample,  $\text{Var}[P_{H,\theta}(f_i) - \beta^T P_{H,\theta}(h_{i,\theta})] = H^{-1} \sigma_\theta^2(f_i - \beta^T h_{i,\theta})$  and, if we differentiate the map  $\beta \mapsto \sigma_\theta^2(f_i - \beta^T h_{i,\theta})$  with respect to  $\beta$  and set it equal to zero, we obtain (cf. Approach 1 in Glynn and Szechtmann (2002)) the theoretical optimum

$$\beta_{\theta, f_i}^{(\text{opt})}(h_{i,\theta}) = \{P_\theta(h_{i,\theta} h_{i,\theta}^T)\}^{-1} P_\theta(h_{i,\theta} f_i), \quad (2.13)$$

provided the inverse exists. In this case, the estimator

$$\varphi_H^{(\text{cvopt})}(t, \theta) = P_{H,\theta}(f_i) - \left(\beta_{\theta, f_i}^{(\text{opt})}(h_{i,\theta})\right)^T P_{H,\theta}(h_{i,\theta}) \quad (2.14)$$

has minimal asymptotic variance. To investigate the existence of the above inverse note that for each fixed  $t \in \mathbb{R}^p$  and  $\theta \in \Theta$ , the determinant of  $P_\theta(h_{i,\theta} h_{i,\theta}^T)$  is

$$\text{Var}[\langle t, \tilde{X}_1(\theta) \rangle] \text{Var}[\langle t, \tilde{X}_1(\theta) \rangle^2] - \left\{ \text{Cov}[\langle t, \tilde{X}_1(\theta) \rangle, \langle t, \tilde{X}_1(\theta) \rangle^2] \right\}^2.$$

Since by the Cauchy–Schwarz inequality,

$$\left\{ \text{Cov}[\langle t, \tilde{X}_1(\theta) \rangle, \langle t, \tilde{X}_1(\theta) \rangle^2] \right\}^2 \leq \text{Var}[\langle t, \tilde{X}_1(\theta) \rangle] \text{Var}[\langle t, \tilde{X}_1(\theta) \rangle^2],$$

it follows (see e.g., Klenke (2013), Theorem 5.8) that

$$\det(P_\theta(h_{i,\theta} h_{i,\theta}^T)) = 0 \Leftrightarrow a \langle t, \tilde{X}_1(\theta) \rangle + b \langle t, \tilde{X}_1(\theta) \rangle^2 + c \stackrel{\text{a.s.}}{=} 0, \quad (2.15)$$

for some  $a, b, c \in \mathbb{R}$  with  $|a| + |b| + |c| > 0$ . As the scalar product is random, universal coefficients to satisfy the right-hand side of (2.15) exist only in degenerate cases, which we do not consider.

Since  $\beta_{\theta, f_i}^{(\text{opt})}(h_{i,\theta})$  is unknown, it needs to be estimated (e.g., by one of the methods in Glynn and Szechtmann (2002), and we use the one described in eqs. (6) and (7) in Portier and Segers (2019)):

$$\hat{\beta}_{H,\theta, f_i}(h_{i,\theta}) = \left\{ P_{H,\theta}(h_{i,\theta} h_{i,\theta}^T) - P_{H,\theta}(h_{i,\theta}) P_{H,\theta}(h_{i,\theta}^T) \right\}^{-1} \left\{ P_{H,\theta}(h_{i,\theta} f_i) - P_{H,\theta}(h_{i,\theta}) P_{H,\theta}(f_i) \right\}. \quad (2.16)$$

For the i.i.d. sample  $\tilde{X}_j(\theta), j = 1, \dots, H$ , as in (2.6) we obtain the *control variates approximation* of  $\varphi(\cdot, \theta)$  given by

$$\varphi_H^{(cv)}(t, \theta) = P_{H,\theta}(f_t) - \kappa_H(t, \theta), \quad t \in \mathbb{R}^p, \tag{2.17}$$

where

$$\kappa_H(t, \theta) = \left( \hat{\beta}_{H,\theta,f_t}(h_{t,\theta}) \right)^T P_{H,\theta}(h_{t,\theta}). \tag{2.18}$$

Recall from (2.11) that  $P_{H,\theta}(f_t) = \varphi_H(t, \theta)$ , so we could simply replace  $\varphi_H(t, \theta)$  in (2.9) by  $\varphi_H^{(cv)}(t, \theta)$  as given in (2.17). However, as we shall see in Section 4, the control variates approximation  $\varphi_H^{(cv)}(t, \theta)$  provides superior approximations of  $\varphi(t, \theta)$  only for values of  $t$ , for which  $\widehat{\text{Var}}(\langle t, \tilde{X}_1(\theta) \rangle)$  is small. Thus, we replace  $\varphi_H(t, \theta)$  in (2.9) by a combination of  $\varphi_H(t, \theta)$  and  $\varphi_H^{(cv)}(t, \theta)$ . More precisely, we propose the following *control variates based estimator*:

$$\hat{\theta}_{n,H,k}^{(cv)} = \underset{\theta \in \Theta}{\operatorname{argmin}} Q_{n,H,k}^{(cv)}(\theta), \tag{2.19}$$

where for appropriate  $k > 0$ ,

$$Q_{n,H,k}^{(cv)}(\theta) = \int_{\mathbb{R}^p} \left| \varphi_n(t) - \left( \varphi_H^{(cv)}(t, \theta) 1_{\{\widehat{\text{Var}}(\langle t, X_1 \rangle) < k\}} + \varphi_H(t, \theta) 1_{\{\widehat{\text{Var}}(\langle t, X_1 \rangle) \geq k\}} \right) \right|^2 \bar{w}(t) dt, \tag{2.20}$$

$\bar{w}(t) = \frac{w(t)}{\widehat{\text{Var}}(\langle t, X_1 \rangle)}$ , with suitable weight function  $w$  such that the integral is well-defined.

It is worth mentioning that, for a fixed weight function  $w(\cdot)$ , the weight function  $\bar{w}(\cdot)$  can always be computed since  $\widehat{\text{Var}}(\langle t, X_1 \rangle)$  depends only on the time series data. The downside of using the control variates based estimator (2.19) is that one needs to resort to numerical integration. However, the procedure is feasible for moderate dimension  $p$ . As illustrated in the Poisson-AR example of Section 4.2, the control variates based estimator has improved the performance over the simulation based estimator (2.8) considerably.

Note that  $\widehat{\text{Var}}(\langle t, X_1 \rangle) = t^T \hat{\Gamma}_p t$  where  $\hat{\Gamma}_p = (\hat{\gamma}_p(i-j))_{i,j=1}^p$  with

$$\hat{\gamma}_p(h) = \frac{1}{n-h} \sum_{j=1}^{n-h} (X_j - \hat{\mu}_n)(X_{j+h} - \hat{\mu}_n), \quad h = 1, \dots, p, \tag{2.21}$$

and  $\hat{\mu}_n = \frac{1}{n} \sum_{j=1}^n X_j$ . The choice of the indicator function  $1_{\{\widehat{\text{Var}}(\langle t, X_1 \rangle) < k\}}$  is justified by the fact that, when estimating the parameter  $\theta_0$ , we focus on approximations of  $\varphi(t, \theta)$  for  $\theta$  close to  $\theta_0$ .

### 3. ASYMPTOTIC BEHAVIOR OF THE PARAMETER ESTIMATORS

Before performing the parameter estimation we need to make sure that the parameters are identifiable from the model.

In the following we assume that the model parameters are identifiable from the chf. In our examples, the dimension  $p$  must be at least 2. For a specific choice of  $p$ , the minimum in (2.19) may not be unique giving an identifiability problem of the estimated model. This may be remedied by increasing the dimension  $p$ .

In the sequel, we will make various assumptions on different aspects of the underlying process, smoothness of the model, moments of the process, and properties of the weight function. We group these assumptions into the following categories.

**Assumptions A.** (Parameter space and time series process).

- (a.1)  $\Theta$  is a compact subset of  $\mathbb{R}^q$  and  $\theta_0 \in \Theta^\circ$ , the interior of  $\Theta$ .
- (a.2)  $(X_j)_{j \in \mathbb{Z}}$  is a stationary and ergodic sequence.
- (a.3)  $(X_j)_{j \in \mathbb{Z}}$  is  $\alpha$ -mixing with rate function  $(\alpha_j)_{j \in \mathbb{N}}$  satisfying  $\sum_{j=1}^{\infty} (\alpha_j)^{1/r} < \infty$  for some  $r > 1$ .

**Assumptions B.** (Continuity and differentiability in  $\theta$ ).

- (b.1) For each  $j \in \mathbb{N}$ , the map  $\theta \mapsto \tilde{X}_j(\theta)$  is continuous on  $\Theta$ .
- (b.2) For each  $j \in \mathbb{N}$ , the map  $\theta \mapsto \tilde{X}_j(\theta)$  is twice continuously differentiable in an open neighborhood around  $\theta_0$ .

**Assumptions C.** (Moments).

- (c.1)  $\mathbb{E}|X_1|^u < \infty$ , where  $u = 2r/(r-1)$  with  $r > 1$  being such that (a.3) holds.
- (c.2)  $\mathbb{E} \prod_{j=1}^p |X_j|^\alpha < \infty$  for some  $\alpha \in (u/2, u]$  where  $u = 2r/(r-1)$  with  $r > 1$  being such that (a.3) holds.
- (c.3)  $\mathbb{E} \sup_{\theta \in \Theta} |X_1(\theta)|^4 < \infty$ .
- (c.4) For each  $\theta \in \Theta$ ,  $\mathbb{E} |\nabla_\theta X_1(\theta)| < \infty$ .
- (c.5)  $\mathbb{E} \sup_{\theta \in \Theta} |\nabla_\theta X_1(\theta)|^{2(1+\varepsilon)} < \infty$  and  $\mathbb{E} \sup_{\theta \in \Theta} |\nabla_\theta^2 X_1(\theta)|^{1+\varepsilon} < \infty$  for some  $\varepsilon > 0$ .

**Assumptions D.** (Weight function).

- (d.1)  $\int_{\mathbb{R}^p} w(t) dt < \infty$ .
- (d.2)  $\int_{\mathbb{R}^p} |t| w(t) dt < \infty$ .
- (d.3)  $\int_{\mathbb{R}^p} |t|^{2(1+\varepsilon)} w(t) dt < \infty$  for some  $\varepsilon > 0$ .
- (d.4)  $\int_{\mathbb{R}^p} \frac{w(t)}{|t|^2} dt < \infty$ .

Assumption B is indeed satisfied by many linear and nonlinear time series processes, in particular, when they have a representation  $X_j(\theta) = f(Z_j, Z_{j-1}, \dots; \theta)$  or  $X_j(\theta) = f(Z_j, X_{j-1}(\theta), X_{j-2}(\theta), \dots; \theta)$  for i.i.d. noise variables  $(Z_j)_{j \in \mathbb{Z}}$ , and  $f : \mathbb{R}^\infty \times \Theta \mapsto \mathbb{R}$  is a measurable function. Prominent examples are the MA( $\infty$ ) and AR( $\infty$ ) representations of a causal or invertible ARMA( $p, q$ ) model (see e.g., eqs. (3.1.15) and (3.1.18) in Brockwell and Davis (2013)) or the ARCH( $\infty$ ) representation of a GARCH( $p, q$ ) model (see e.g., Francq and Zakoian (2011), Theorem 2.8). In this case, assumptions (b.1) and (b.2) will hold whenever the map  $f$  is continuously differentiable for  $\theta \in \Theta$ . For example, if  $f$  is Lipschitz-continuous for  $\theta \in \Theta$ , then the continuity assumption (b.1) holds.

The key asymptotic properties, consistency and asymptotic normality of our estimates are stated in the following theorems. The proofs of these results are presented in the Appendix.

We formulate first the strong consistency results of the parameters.

**Theorem 3.1.** (Consistency of  $\hat{\theta}_{n,H}$ ). Assume that (a.1), (a.2), (b.1), and (d.1) hold. Let  $H = H(n) \rightarrow \infty$  as  $n \rightarrow \infty$ . Then  $\hat{\theta}_{n,H} \xrightarrow{\text{a.s.}} \theta_0$  as  $n \rightarrow \infty$ .

**Theorem 3.2.** (Consistency of  $\hat{\theta}_{n,H,k}^{(\text{cv})}$ ). Assume that the conditions of Theorem 3.1 hold, and additionally (c.1), (c.3), and (d.4). Let  $H = H(n) \rightarrow \infty$  as  $n \rightarrow \infty$ . Then  $\hat{\theta}_{n,H,k}^{(\text{cv})} \xrightarrow{\text{a.s.}} \theta_0$  as  $n \rightarrow \infty$ .

The asymptotic normality of the simulation based parameter estimator reads as follows.

**Theorem 3.3.** (Asymptotic normality of  $\hat{\theta}_{n,H}$ ). Assume that Assumptions A and B, and the moment conditions (c.2), (c.4), and (c.5) hold. Furthermore, assume that the weight function satisfies (d.1), (d.2) and (d.3). Set  $H = H(n) := \bar{H}(n)n$  and  $\bar{H}(n) \rightarrow \infty$  as  $n \rightarrow \infty$  and define

$$K_j(\theta) = \int_{\mathbb{R}^p} \left( \frac{\partial}{\partial \theta} \Re(\varphi(t, \theta)), \frac{\partial}{\partial \theta} \Im(\varphi(t, \theta)) \right) \begin{pmatrix} \cos(\langle t, X_j \rangle) - \Re(\varphi(t, \theta)) \\ \sin(\langle t, X_j \rangle) - \Im(\varphi(t, \theta)) \end{pmatrix} w(t) dt, \quad j \in \mathbb{N} \quad (3.1)$$

and

$$Q = \int_{\mathbb{R}^p} \left( \frac{\partial}{\partial \theta} \mathfrak{R}(\varphi(t, \theta_0)) \right), \frac{\partial}{\partial \theta} \mathfrak{S}(\varphi(t, \theta_0)) \left( \frac{\partial}{\partial \theta} \mathfrak{R}(\varphi(t, \theta_0)) \right), \frac{\partial}{\partial \theta} \mathfrak{S}(\varphi(t, \theta_0)) \right)^T w(t) dt. \tag{3.2}$$

If  $Q$  is a non-singular matrix, then

$$\sqrt{n}(\hat{\theta}_{n,H} - \theta_0) \xrightarrow{d} N(0, Q^{-1} W Q^{-1}), \quad n \rightarrow \infty, \tag{3.3}$$

where

$$W = \mathbb{V}\text{ar}[K_1(\theta_0)] + 2 \sum_{j=2}^{\infty} \text{Cov}[K_1(\theta_0), K_j(\theta_0)]. \tag{3.4}$$

Theorem 3.3 shows that  $\hat{\theta}_{n,H}$  is asymptotically normal and achieves the same asymptotic efficiency as the oracle estimator from (2.3) (see Theorem 2.1 in Knight and Yu (2002)). Therefore, there cannot be any improvement in the limit law for the asymptotic normality of  $\hat{\theta}_{n,H,k}^{(cv)}$ . However, as we show in Section 4,  $\hat{\theta}_{n,H,k}^{(cv)}$  is based on a better approximation of the chf  $\varphi(\cdot, \theta)$  than that used for  $\hat{\theta}_{n,H}$ . Thus, the control variates estimator  $\hat{\theta}_{n,H,k}^{(cv)}$  improves the finite sample performance compared to the simulation based estimator  $\hat{\theta}_{n,H}$ .

**Remark 3.4.** As pointed out in (Knight and Yu, 2002, Remark 2.3), the asymptotic variance of  $\hat{\theta}_{n,H}$  in (3.3) can be approximated by replacing  $\theta_0$  by  $\hat{\theta}_{n,H}$  in (3.2) and (3.4) and by replacing the infinite sum in (3.4) by an approximating sum with a kernel and a convenient bandwidth using the methods suggested in Andrews (1991) and Newey and West (1994).

#### 4. ASSESSING THE QUALITY OF THE ESTIMATED CHF

We compare the performance of both the Monte Carlo approximation  $\varphi_H(\cdot, \theta)$  and the control variates approximation  $\varphi_H^{(cv)}(\cdot, \theta)$  of the chf as defined in (2.7) and (2.17), respectively. We start with the following comparison of the two chf approximations.

**Remark 4.1.** (Comparison of  $\varphi_H^{(cv)}(\cdot, \theta)$  and  $\varphi_H(\cdot, \theta)$ ). Assume that (c.3) holds, and let  $\varphi_H^{(cvopt)}$  and  $\varphi_H^{(cv)}$  be as defined in (2.14) and (2.17), respectively. We use that  $\hat{\beta}_{H,\theta,f_i}(h_{t,\theta}) \xrightarrow{\text{a.s.}} \beta_{\theta,f_i}^{(opt)}(h_{t,\theta})$  as  $n \rightarrow \infty$  with limit given in (2.13). This follows from the representation of  $\hat{\beta}_{H,\theta,f_i}(h_{t,\theta})$  as

$$\hat{\beta}_{H,\theta,f_i}(h_{t,\theta}) = \hat{\beta}_{H,\theta,\mathfrak{R}(f_i)}(h_{t,\theta}) + i \hat{\beta}_{H,\theta,\mathfrak{S}(f_i)}(h_{t,\theta})$$

and the almost sure convergence of both terms. The quantities needed to compute the estimator in (2.16) are, for each  $v, \kappa = 1, 2$ :

$$P_{H,\theta}(f_t) = \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \tilde{X}_j(\theta) \rangle}, \tag{4.1}$$

$$P_{H,\theta}(h_{v,t,\theta}) = \frac{1}{H} \sum_{j=1}^H (\langle t, \tilde{X}_j(\theta) \rangle^v - \mathbb{E} \langle t, X_1(\theta) \rangle^v),$$

$$P_{H,\theta}(f_t h_{v,t,\theta}) = \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \tilde{X}_j(\theta) \rangle} (\langle t, \tilde{X}_j(\theta) \rangle^v - \mathbb{E} \langle t, X_1(\theta) \rangle^v),$$

$$P_{H,\theta}(h_{\nu,t,\theta}h_{\kappa,t,\theta}) = \frac{1}{H} \sum_{j=1}^H \left( \langle t, \tilde{X}_j(\theta) \rangle^\nu - \mathbb{E} \langle t, X_1(\theta) \rangle^\nu \right) \left( \langle t, \tilde{X}_j(\theta) \rangle^\kappa - \mathbb{E} \langle t, X_1(\theta) \rangle^\kappa \right). \tag{4.2}$$

Hence, strong consistency of  $\hat{\beta}_{H,\theta,f_i}(h_{t,\theta})$  follows from the SLLN. This together with  $P_\theta(h_{t,\theta}) = 0$  implies by Theorem 1 in Glynn and Szechtman (2002) that, as  $H \rightarrow \infty$ ,

$$H^{1/2} \left( \mathfrak{R} \left( \varphi_H^{(cv)}(t, \theta) - \varphi(t, \theta) \right) \right) \xrightarrow{d} N \left( 0, \sigma_\theta^2 \left( \mathfrak{R}(f_t) - [\beta_{\theta,\mathfrak{R}(f_t)}^{(opt)}(h_{t,\theta})]^T h_{t,\theta} \right) \right),$$

$$H^{1/2} \left( \mathfrak{I} \left( \varphi_H^{(cv)}(t, \theta) - \varphi(t, \theta) \right) \right) \xrightarrow{d} N \left( 0, \sigma_\theta^2 \left( \mathfrak{I}(f_t) - [\beta_{\theta,\mathfrak{I}(f_t)}^{(opt)}(h_{t,\theta})]^T h_{t,\theta} \right) \right),$$

with

$$\sigma_\theta^2 \left( \mathfrak{R}(f_t) - [\beta_{\theta,\mathfrak{R}(f_t)}^{(opt)}(h_{t,\theta})]^T h_{t,\theta} \right) \leq \sigma_\theta^2 \left( \mathfrak{R}(f_t) \right) \quad \text{and}$$

$$\sigma_\theta^2 \left( \mathfrak{I}(f_t) - [\beta_{\theta,\mathfrak{I}(f_t)}^{(opt)}(h_{t,\theta})]^T h_{t,\theta} \right) \leq \sigma_\theta^2 \left( \mathfrak{I}(f_t) \right),$$

with  $\sigma_\theta^2(\cdot)$  as defined in (2.12). Therefore,  $\varphi_H^{(cv)}(\cdot, \theta)$  provides an approximation of the integral  $Q_n(\theta)$  in (2.4) with smaller variance than  $\varphi_H(\cdot, \theta)$ . As a consequence, this favors the control variates estimator  $\hat{\theta}_{n,H,k}^{(cv)}$  over the simulation based estimator  $\hat{\theta}_{n,H}$  for large sample sizes  $n \in \mathbb{N}$ .

For all forthcoming examples we choose  $p = 3$  and  $H = 3000$ . We begin with a stationary Gaussian AR(1) process, where we know the chf  $\varphi(\cdot)$  explicitly, and then proceed to the Poisson-AR process, where we approximate the true unknown chf by a precise simulated version.

### 4.1. The Gaussian AR(1) Process

We start with a stationary Gaussian AR(1) process to show how the method of control variates improves the Monte Carlo approximation of its chf. Let  $(X_j(\theta))_{j \in \mathbb{Z}}$  be the AR(1) process

$$X_j(\theta) = \phi X_{j-1}(\theta) + Z_j(\theta), \quad j \in \mathbb{Z}, \quad (Z_j(\theta))_{j \in \mathbb{Z}} \stackrel{iid}{\sim} N(0, \sigma^2), \tag{4.3}$$

with parameter space  $\Theta$  being a compact subset of  $\{\theta = (\phi, \sigma) : |\phi| < 1, \sigma > 0\}$ . Then the true chf of  $X_1(\theta) = (X_1(\theta), X_2(\theta), X_3(\theta))$  is given by  $\varphi(t, \theta) = e^{-\frac{1}{2}t^T \Gamma_3(\theta) t}$  for  $t \in \mathbb{R}^3$ , where the covariance matrix  $\Gamma_3(\theta)$  is explicitly known and identifies the parameter  $\theta$  uniquely; see for example, Brockwell and Davis (2013), Example 3.1.2. For a fixed  $\theta \in \Theta$  and many  $t \in \mathbb{R}^3$  we compute the absolute errors

$$\xi_H(t, \theta) = |\varphi_H(t, \theta) - \varphi(t, \theta)| \quad \text{and} \quad \xi_H^{(cv)}(t, \theta) = |\varphi_H^{(cv)}(t, \theta) - \varphi(t, \theta)| \tag{4.4}$$

where  $\varphi_H(\cdot, \theta)$  is the Monte Carlo approximation of the chf of  $X_1(\theta) = (X_1(\theta), X_2(\theta), X_3(\theta))$  and  $\varphi_H^{(cv)}(\cdot, \theta)$  its control variates approximation. To understand how well we can approximate  $\varphi(\cdot, \theta)$ , we plot in Figure 1,  $\xi_H(t, \theta)$  and  $\xi_H^{(cv)}(t, \theta)$  against  $\sqrt{\text{Var}[\langle t, X_1(\theta) \rangle]}$  for different parameters  $\theta$ . These quantities are computed from an i.i.d. sample  $X_j(\theta), j = 1, \dots, H$  as in (2.6). To simulate i.i.d. observations from the model (4.3), we use the fact that the one-dimensional stationary distribution is  $X_1(\theta) \sim N(0, \sigma^2/(1-\phi^2))$ , and then use the recursion in (4.3) to simulate  $X_2(\theta)$  and  $X_3(\theta)$ . We chose 500 randomly generated values of  $t$  from the 3-dimensional Laplace distribution with chf given in (5.2).

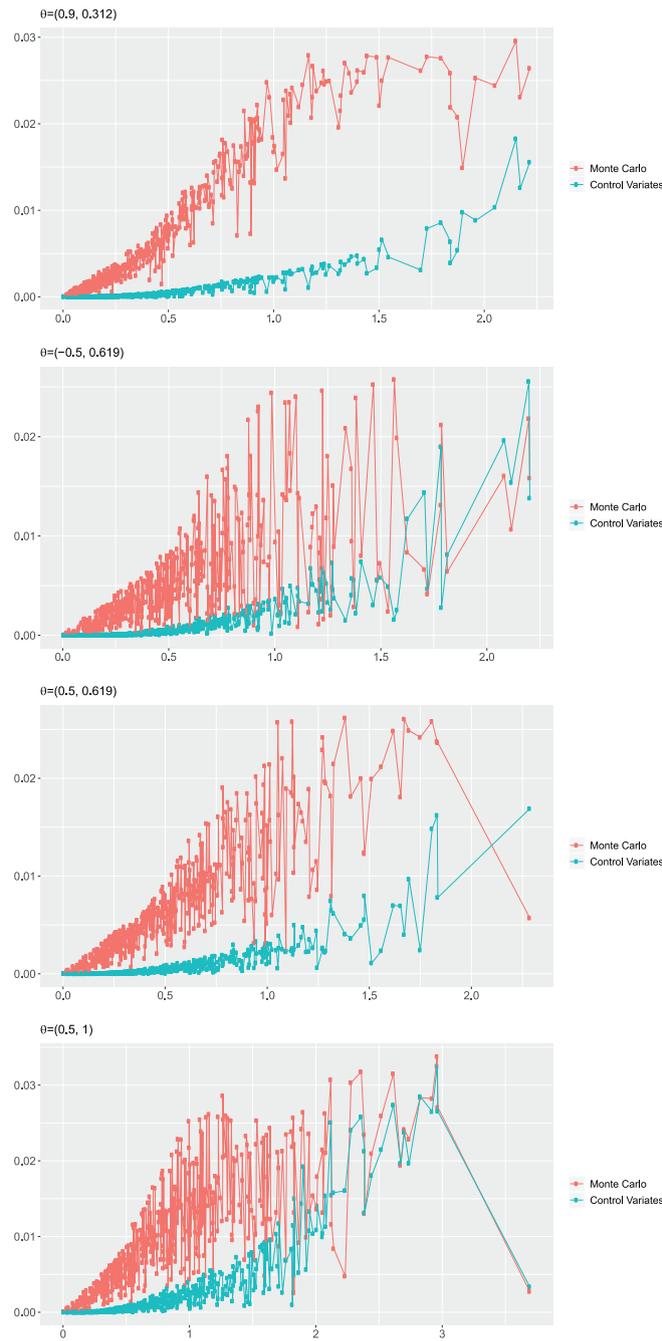


FIGURE 1. Gaussian AR(1) model: absolute error  $\xi_H(t, \theta)$  (red) and  $\xi_H^{(cv)}(t, \theta)$  (green) for  $p = 3$  and  $H = 3000$  as in (4.4). We use 500 randomly generated values of  $t \in \mathbb{R}^3$  from the Laplace distribution (with chf as in (5.2) below), which are plotted against  $\sqrt{\text{Var}[\langle t, X_1(\theta) \rangle]}$

It is clear from Figure 1 that both the Monte Carlo and the control variates approximations work better when  $\sqrt{\text{Var}[\langle t, X_1(\theta) \rangle]}$  is small, and also that the control variates approximations are best for small values of  $\sqrt{\text{Var}[\langle t, X_1(\theta) \rangle]}$ . The superiority of the control variates approximation for all  $t$  and all parameter settings is clearly visible, and already expected from Remark 4.1.

## 4.2. The Poisson-AR Model

We consider a nonlinear time series process for time series of counts, which has been proposed originally in Zeger (1988). A prototypical Poisson-AR(1) model suggested in Davis and Rodriguez-Yam (2005) assumes that the observations  $(X_j(\theta))_{j \in \mathbb{Z}}$  are independent and Poisson-distributed with means  $e^{\beta + \alpha_j(\theta)}$  where the process  $(\alpha_j(\theta))_{j \in \mathbb{Z}}$  is a latent stationary Gaussian AR(1) process, given by the equations

$$\alpha_j(\theta) = \phi \alpha_{j-1}(\theta) + \eta_j(\theta), \quad j \in \mathbb{Z}, \quad (\eta_j(\theta))_{j \in \mathbb{Z}} \stackrel{\text{iid}}{\sim} N(0, \sigma^2),$$

with parameter space  $\Theta$  being a compact subset of  $\{\theta = (\beta, \phi, \sigma) : |\phi| < 1, \beta \in \mathbb{R}, \sigma > 0\}$ . The parameter  $\theta$  is uniquely identifiable from the second order structure, which has been computed in Section 2.1 of Davis *et al.* (2000).

For this model, the true chf of  $X_1(\theta) = (X_1(\theta), X_2(\theta), X_3(\theta))$  cannot be computed in closed form. To mimic the assessment of the errors in (4.4), we simulate 1,000,000 i.i.d. observations from  $X_1(\theta)$  by first simulating a Gaussian AR(1) process  $(\alpha_1(\theta), \alpha_2(\theta), \alpha_3(\theta))$  (as described in Section 4.1) and then simulating independent Poisson random variables with means  $e^{\beta + \alpha_1(\theta)}$ ,  $e^{\beta + \alpha_2(\theta)}$  and  $e^{\beta + \alpha_3(\theta)}$ , respectively. From this we compute the empirical characteristic function and take it as  $\varphi(\cdot, \theta)$  in the absolute error terms (4.4).

We compare the performance of both the Monte Carlo approximation and the control variates approximation of the chf. Figure 2 presents the results. The plots in Figure 2 are also in favor of the control variates approximation, when compared to the Monte Carlo approximation.

## 5. PRACTICAL ASPECTS AND SIMULATION RESULTS

Our objective is to obtain a simple expression of the integrated mean squared error  $Q_{n,H}(\theta)$  in (2.9), which is needed to compute the estimator in (2.8). For a weight function  $w$  in (2.9), we write

$$\tilde{w}(x) = \int_{\mathbb{R}^p} e^{i\langle t, x \rangle} w(t) dt, \quad x \in \mathbb{R}^p, \quad (5.1)$$

for its Fourier transform. Our preference is on weight functions such that (5.1) is known explicitly.

**Example 5.1.** (Weight functions and their characteristic functions).

(i) Laplace:  $w$  is a multivariate Laplace density with chf

$$\tilde{w}(t) = \frac{1}{(1 + (2\pi^2)^{-1} t^T t)}, \quad t \in \mathbb{R}^p. \quad (5.2)$$

(ii) Cauchy:  $w$  is a multivariate Cauchy density with chf

$$\tilde{w}(t) = e^{-\sqrt{t^T t}}, \quad t \in \mathbb{R}^p.$$

(iii) Gaussian:  $w$  is a standard multivariate Gaussian density with chf

$$\tilde{w}(t) = e^{-\frac{1}{2} t^T t}, \quad t \in \mathbb{R}^p. \quad (5.3)$$

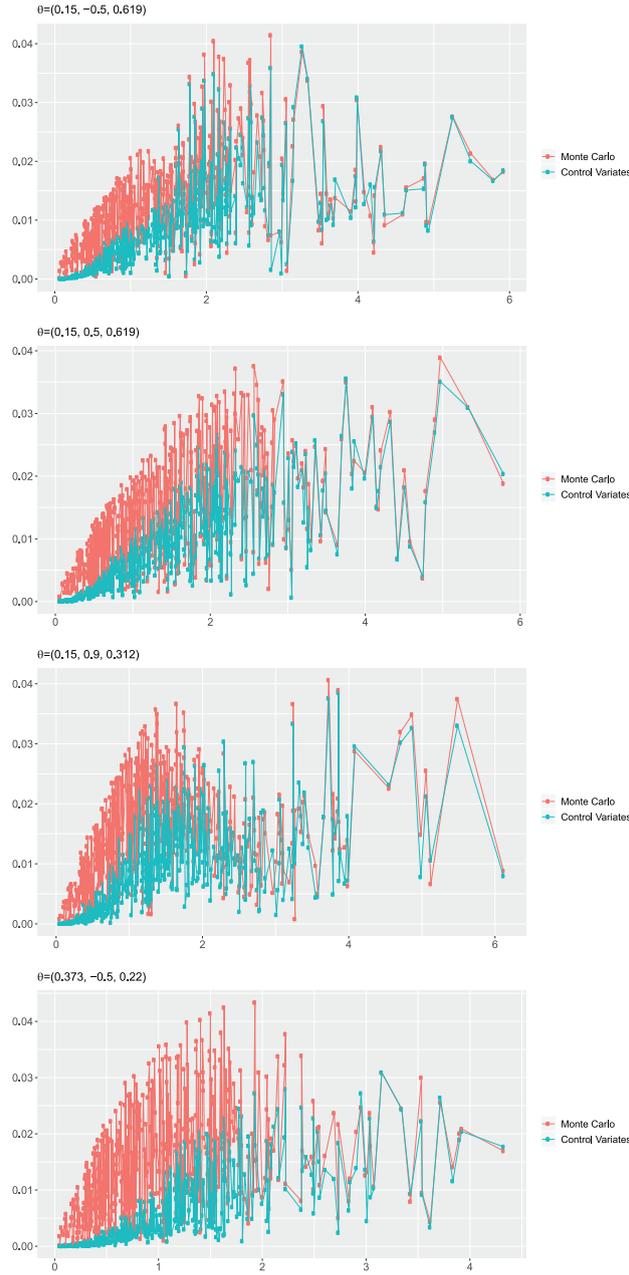


FIGURE 2. Poisson-AR model: Absolute errors  $\xi_H(t, \theta)$  (red) and  $\xi_H^{(cv)}(t, \theta)$  (green) for  $p = 3$  and  $H = 3000$  as in (4.4). We use 500 randomly generated values of  $t \in \mathbb{R}^3$  from the Laplace distribution (with chf as in (5.2) below), which are plotted against  $\sqrt{\text{Var}[\langle t, X_1(\theta) \rangle]}$

**Lemma 5.2.** Let  $Q_{n,H}(\theta)$  be as in (2.9) and  $w$  a weight function with Fourier transform  $\tilde{w}$ . Then

$$\begin{aligned}
 Q_{n,H}(\theta) &= \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n \tilde{w}(X_j - X_k) + \frac{1}{H^2} \sum_{j=1}^H \sum_{k=1}^H \tilde{w}(\tilde{X}_j(\theta) - \tilde{X}_k(\theta)) \\
 &\quad - \frac{1}{Hn} \sum_{k=1}^H \sum_{j=1}^n (\tilde{w}(X_j - \tilde{X}_k(\theta)) + \tilde{w}(\tilde{X}_k(\theta) - X_j)). \tag{5.4}
 \end{aligned}$$

Formula (5.4) is very useful, since it avoids the computation of a  $p$ -dimensional integral. Additionally, since the first double sum on the right-hand side of (5.4) does not depend on the argument  $\theta$ , for the optimization it can be ignored.

**Remark 5.3.** When evaluating the integrated weighted mean squared errors (2.9), (2.20), or (5.4) in practice, they need to be deterministic functions of  $\theta$ . This is enforced by taking a fixed seed for every  $j = 1, \dots, H$ , when simulating  $\tilde{X}_j(\theta)$  for different values of  $\theta \in \Theta$ .

In the following two examples we study the finite sample behavior of the estimators  $\hat{\theta}_{n,H}$  and  $\hat{\theta}_{n,H,k}^{(cv)}$ . We begin with a stationary Gaussian ARFIMA model, whose chf is explicitly known so that we can use the oracle estimator from Section 2.1. Afterwards we come back to the Poisson-AR process. We choose  $p = 3$ , since the three-dimensional chf contains sufficient information to identify the parameter of interest. We also choose  $H = 3000$ .

### 5.1. The ARFIMA Model

Let  $(X_j(\theta))_{j \in \mathbb{Z}}$  be the stationary Gaussian ARFIMA(0,  $d$ , 0) model

$$(1 - B)^d X_j(\theta) = Z_j(\theta), \quad j \in \mathbb{Z}, \quad (Z_j(\theta))_{j \in \mathbb{Z}} \stackrel{\text{iid}}{\sim} N(0, \sigma^2),$$

where  $B$  is the backshift operator, with parameter space  $\Theta$  being a compact subset of  $\{\theta = (d, \sigma) : d \in (-0.5, 0.5), \sigma > 0\}$ . Then the true chf of  $X_1(\theta) = (X_1(\theta), X_2(\theta), X_3(\theta))$  is given by  $\varphi(t, \theta) = e^{-\frac{1}{2}t^T \Gamma_3(\theta)t}$  for  $t \in \mathbb{R}^3, \theta \in \Theta$ , where the covariance matrix  $\Gamma_3(\theta)$  is explicitly known and identifies the parameter  $\theta$  uniquely; see for example, Pipiras and Taquq (2017), Corollary 2.4.4.

For the long-memory case, for each value of  $d \in \{0.05, \dots, 0.45\}$  we compare the new estimators with the MLE method as implemented in the R package `arfima`. Thus, for many  $\theta \in \Theta$ , we generate i.i.d. Gaussian random vectors with mean zero and covariance  $\Gamma_3(\theta)$  and use them to construct the simulation based estimator  $\hat{\theta}_{n,H}$ .

Since the chf  $\varphi(\cdot, \theta)$  is known in closed form, we are able to compute the oracle estimator  $\hat{\theta}_n$  from (2.4). To compute the integral appearing in (2.4) in closed form, we choose the weight function  $w(t) = (2\pi)^{-3/2} e^{-\frac{1}{2}t^T t}, t \in \mathbb{R}^3$ .

Then the integral in (2.4), which needs to be minimized with respect to the parameter  $\theta$ , can be evaluated similarly as in (5.4), giving for the chf being known, that  $Q_n(\theta)$  can be written as

$$\begin{aligned}
 &\int_{\mathbb{R}^3} \left| \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle} - e^{-\frac{1}{2}t^T \Gamma_3(\theta)t} \right|^2 w(t) dt \\
 &= \left( \det((2\Gamma_3(\theta) + I)^{-1}) \right)^{\frac{1}{2}} + \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \exp \left\{ -\frac{1}{2} (X_j - X_k)^T (X_j - X_k) \right\} \\
 &\quad - 2 \left( \det((\Gamma_3(\theta) + I)^{-1}) \right)^{\frac{1}{2}} \frac{1}{n} \sum_{j=1}^n \exp \left\{ -\frac{1}{2} X_j^T (\Gamma_3(\theta) + I)^{-1} X_j \right\}. \tag{5.5}
 \end{aligned}$$

We compare in Table AI the performance of the simulation based estimator  $\hat{\theta}_{n,H}$ , the oracle estimator  $\hat{\theta}_n$  in (2.3) based on the minimization of (5.5), and the MLE. We fixed  $\sigma = 1$  for all simulated sample paths used in the simulation study. For both  $\hat{\theta}_n$  and  $\hat{\theta}_{n,H}$ , we also estimate  $\sigma$  but report only the performance for the estimator of  $d$  which is the key parameter of interest in long-range dependence models. We notice that  $\hat{\theta}_{n,H}$  is comparable to the oracle estimator, so in this model there is no need to use control variates. When comparing both simulation based estimators, the RMSEs are almost the same for all  $d \geq 0.20$ . The MLE has a smaller RMSE, but both  $\hat{\theta}_n$  and  $\hat{\theta}_{n,H}$  have a smaller bias than the MLE. In the simulations, the density plots for the estimates of  $d$  with  $d \in \{0.25, 0.3\}$  look reasonably normal. On the other hand, the estimates when  $d$  is closer to 0.5 are rather skewed, which is expected due to the constraint  $d < 0.5$ . In this case a larger sample is needed to obtain more normal looking densities.

**Remark 5.4.** We also investigate the feasibility of our new estimation procedures for misspecified models. We take a Gaussian ARFIMA as the true model, but for the data we modify the distribution of its innovations. Specifically, we consider the two cases of ARFIMA models driven by noise with a Laplace distribution and with a Student's  $t$  distribution with 6 degrees of freedom. The estimation results under the two misspecification scenarios are shown in Tables AII and AIII of the Appendix. The quasi-oracle estimator is based on the Gaussian chf, and the quasi-MLE (QMLE) is found by maximizing the Gaussian likelihood, even though the data are in fact non-Gaussian. For both noise distributions, we see very little difference in the performance of the three estimators (QMLE compared with MLE) from the Gaussian ARFIMA scenario in Table AI. In particular, our estimator continues to have small bias and RMSE that is comparable to the oracle estimator and only slightly larger than that of the QMLE. Of course, it is known that the QMLE estimators behave asymptotically the same as the MLE when the data is Gaussian.

## 5.2. The Poisson-AR Process

The Poisson-AR model has been defined in Section 4.2. We conduct a simulation experiment in the same setting as in Table V in Davis and Rodriguez-Yam (2005) and Table III in Davis and Yau (2011). The results are shown in Table AIV of the Appendix for  $n = 400$  and nine different parameter settings, where we also classify the models by the corresponding index of dispersion  $D$  of the random variable  $e^{\beta + \alpha_1}$ , which assumes values in  $\{0.1, 1, 10\}$  as shown in Davis and Rodriguez-Yam (2005).

We compare both the simulation based estimator  $\hat{\theta}_{n,H}$  and control variates based estimator  $\hat{\theta}_{n,H,k}^{(cv)}$ . We fix  $H = 3000$ ,  $p = 3$  and the three-dimensional Laplace density as in (5.2) for  $w$ . To simulate i.i.d. observations of  $(X_1(\theta), X_2(\theta), X_3(\theta))$  we proceed as explained in Section 4.2. The simulation based estimator  $\hat{\theta}_{n,H}$  in (2.8) is computed via (5.4). Unfortunately, such a formula cannot be obtained for the control variates based estimator  $\hat{\theta}_{n,H,k}^{(cv)}$ , since the introduction of the correction  $\kappa_H$  in (2.18) introduces additional polynomial terms into  $Q_{n,H,k}^{(cv)}$  in (2.20). Thus, we resort to numerical integration to evaluate  $\hat{\theta}_{n,H,k}^{(cv)}$ .

Our findings are as follows. For  $D \in \{1, 0.1\}$ , the control variates based estimator  $\hat{\theta}_{n,H,k}^{(cv)}$  for  $k = 1$  presents smaller bias and RMSE than the simulation based estimator  $\hat{\theta}_{n,H}$  in most cases, in all others it is comparable. The smallest RMSE values are shaded in Table IV. Additionally, a significant improvement in the bias for estimating  $\phi$  is noticeable for  $\theta = (0.373, 0.500, 0.220)$  and  $\theta = (0.373, 0.900, 0.111)$ . This example shows the advantage of using control variates to improve the estimation of the model parameters. This is not surprising in view of the improved performance of estimating the characteristic function as seen in all three panels of Figure 2.

We compare now the control variates based estimator  $\hat{\theta}_{n,H,k}^{(cv)}$  in Table II of the Appendix, with the results for the consecutive pairwise likelihood (CPL) from Table III in Davis and Yau (2011), which is referred to as CPL1 in that article. The bias of  $\hat{\theta}_{n,H,k}^{(cv)}$  is smaller than that of CPL1 for the estimated  $\beta$  and  $\sigma$  for almost all cases, in all others it is comparable. For  $\phi$  the bias of  $\hat{\theta}_{n,H,k}^{(cv)}$  and CPL1 are comparable, except that  $\hat{\theta}_{n,H,k}^{(cv)}$  shows poor performance for estimating  $\phi$  for the true parameter  $(\beta, \phi, \sigma) = (0.373, 0.9, 0.111)$ . This is due to the fact that the simulated sample

paths contain a large number of zeros, giving very little information for the parameter estimation. The estimated values for  $\beta$  look normal for all parameter choices. The sampling distributions of the other parameter estimates look close to normal, except in the boundary. In particular, the density for the estimates of  $\phi$  when  $\phi = 0.9$  or  $\sigma \in \{0.22, 0.111\}$  and estimates of  $\sigma$  when  $\sigma \in \{0.22, 0.111\}$  show some asymmetry, deviating from normality. This is not unexpected because they are close to the boundary.

#### ACKNOWLEDGEMENTS

Thiago do Rêgo Sousa gratefully acknowledges support from the National Council for Scientific and Technological Development (CNPq – Brazil) and the TUM Graduate School. He also thanks the Statistics Department at Columbia University for its hospitality during his visit and takes pleasure to thank Viet Son Pham and Thibaut Vatter for helpful discussions. Davis' research was partially supported by NSF grant DMS 2015379 to Columbia University. Open Access funding enabled and organized by ProjektDEAL.

#### DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were analyzed or used in this study

#### REFERENCES

- Andrews DW. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society* **59**(3):817–858.
- Belomestny D, Comte F, Genon-Catalot V, Masuda H, Reiß M. 2015. *Estimation for discretely observed Lévy processes*. In *Lévy Matters IV. LN in Mathematics*, Vol. 2128 Springer: Cham.
- Bernstein DS. 2009. *Matrix Mathematics: Theory, Facts, and Formulas*, Princeton University Press: Princeton.
- Bianchi C, Cleur EM. 1996. Indirect estimation of stochastic differential equation models: some computational experiments. *Computational Economics* **9**(3):257–274.
- Brockwell P, Chadraa E, Lindner A, et al. . 2006. Continuous-time GARCH processes. *The Annals of Applied Probability* **16**(2):790–826.
- Brockwell PJ. 2001. Lévy-driven CARMA processes. *Annals of the Institute of Statistical Mathematics* **53**(1):113–124.
- Brockwell PJ, Davis RA. 2013. *Time Series: Theory and Methods*, Springer: New York.
- Campbell M. 1994. Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature. *Journal of the Royal Statistical Society. Series A* **157**(2):191–208.
- Carrasco M, Florens J-P. 2014. On the asymptotic efficiency of GMM. *Econometric Theory* **30**(2):372–406.
- Carrasco M, Kotchoni R. 2017. Efficient estimation using the characteristic function. *Econometric Theory* **33**(2):479–526.
- Carrasco M, Chernov M, Florens J-P, Ghysels E. 2007. Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics* **140**(2):529–573.
- Chan K, Ledolter J. 1995. Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association* **90**:242–252.
- Davis R, Dunsmuir WT, Wang Y. 2000. On autocorrelation in a Poisson regression model. *Biometrika* **87**(3):491–505.
- Davis RA, Rodriguez-Yam G. 2005. Estimation for state-space models based on a likelihood approximation. *Statistica Sinica* **15**:381–406.
- Davis RA, Yau CY. 2011. Comments on pairwise likelihood in time series models. *Statistica Sinica* **21**:255–277.
- Davis RA, Dunsmuir WT, Wang Y. 1999. *Modeling time series of count data*. In *Asymptotics, Nonparametrics, and Time Series*, Ghosh S (ed.) Marcel Dekker: New York, pp. 63–114.
- Davis RA, Matsui M, Mikosch T, Wan P. 2018. Applications of distance correlation to time series. *Bernoulli* **24**(4A):3087–3116.
- de Luna X, Genton MG. 2001. Robust simulation-based estimation of ARMA models. *Journal of Computational and Graphical Statistics* **10**(2):370–387.
- Do Rêgo Sousa T, Haug S, Klüppelberg C. 2019. Indirect inference for Lévy-driven continuous-time GARCH models. *Scandinavian Journal of Statistics* **46**(3):765–801.
- Doukhan P. 1994. *Mixing: properties and examples*. In *Springer LN in Statistics*, Vol. 85 Springer: Heidelberg.

- Fasen-Hartmann V, Kimmig S. 2020. Robust estimation of continuous-time ARMA models via indirect inference. *Journal of Time Series Analysis* **41**(5):765–801.
- Ferguson T. 1996. *A Course in Large Sample Theory*, Springer: Dordrecht.
- Feuerverger A. 1990. An efficiency result for the empirical characteristic function in stationary time-series models. *The Canadian Journal of Statistics* **18**(2):155–161.
- Forneron J-J. 2018. *Essays on Simulation-based Estimation*. PhD thesis, Columbia University: New York.
- Franq C, Meintanis SG. 2016. Fourier-type estimation of the power GARCH model with stable-Paretian innovations. *Metrika* **79**(4):389–424.
- Franq C, Zakoïan J-M. 2011. *GARCH Models: Structure, Statistical Inference and Financial Applications*, Wiley: West Sussex.
- Glynn PW, Szechtman R. 2002. *Some new perspectives on the method of control variates*. In *Monte Carlo and Quasi-Monte Carlo Methods 2000*, Fang K, Hickernell F, Niederreiter H (eds.) Springer: Berlin, pp. 27–49.
- Gourieroux C, Monfort A, Renault E. 1993. Indirect inference. *Journal of Applied Econometrics* **8**:S85–S118.
- Gourieroux C, Renault E, Touzi N. 2000. *Calibration by simulation for small sample bias correction*. In *Simulation-based Inference in Econometrics: Methods and Applications*, Mariano R, Schuermann T, Weeks MJ (eds.) Cambridge University Press: Cambridge, pp. 328–358.
- Gourieroux C, Phillips PCB, Yu J. 2010. Indirect Inference for dynamic panel models. *Journal of Econometrics* **157**(1):68–77.
- Hansen L. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* **50**(4):1029–1054.
- Ibragimov I, Linnik Y. 1971. *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff: Groningen.
- Jiang GJ. 1998. Estimation of jump-diffusion processes based on Indirect Inference. *IFAC Proceedings Volumes* **31**(16):385–390.
- Klenke A. 2013. *Probability Theory: A Comprehensive Course*, Springer: Berlin.
- Klüppelberg C, Lindner A, Maller R. 2004. A continuous-time GARCH process driven by a Lévy process: stationarity and second-order behaviour. *Journal of Applied Probability* **41**(3):601–622.
- Knight J, Yu J. 2002. Empirical characteristic function in time series estimation. *Econometric Theory* **18**(3):691–721.
- Knight JL, Satchell SE, Yu J. 2002. Estimation of the stochastic volatility model by the empirical characteristic function method. *Australian & New Zealand Journal of Statistics* **44**(3):319–335.
- Kotchoni R. 2012. Applications of the characteristic function-based continuum GMM in finance. *Computational Statistics & Data Analysis* **56**(11):3599–3622.
- Laurini MP, Hotta LK. 2013. Indirect inference in fractional short-term interest rate diffusions. *Mathematics and Computers in Simulation* **94**:109–126.
- Lindsay BG. 1988. Composite likelihood methods. *Contemporary Mathematics* **80**(1):221–239.
- McFadden D. 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* **57**:995–1026.
- Meintanis SG, Taufer E. 2012. Inference procedures for stable-Paretian stochastic volatility models. *Mathematical and Computer Modelling* **55**(3-4):1199–1212.
- Milovanovic G, Popovic B, Stojanovic V. 2014. An application of the ECF method and numerical integration in estimation of the stochastic volatility models. *Facta Universitatis, Mathematics and Informatics* **29**(3):295–311.
- Ndongo M, Kâ Diongue A, Diop A, Dossou-Gbété S. 2016. Estimation for seasonal fractional ARIMA with stable innovations via the empirical characteristic function method. *Statistics* **50**(2):298–311.
- Newey WK, West KD. 1994. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies* **61**(4):631–653.
- Oates CJ, Girolami M, Chopin N. 2017. Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society: Series B* **79**(3):695–718.
- Parthasarathy K. 1967. *Probability Measures on Metric Spaces*, Academic Press: New York.
- Pipiras V, Taqqu MS. 2017. *Long-Range Dependence and Self-Similarity*, Cambridge University Press: Cambridge.
- Portier F, Segers J. 2019. Monte Carlo integration with a growing number of control variates. *Journal of Applied Probability* **56**(4):1168–1186.
- Raknerud A, Skare Ø. 2012. Indirect inference methods for stochastic volatility models based on non-Gaussian Ornstein-Uhlenbeck processes. *Computational Statistics & Data Analysis* **56**(11):3260–3275.
- Robert C, Casella G. 2004. *Monte Carlo Statistical Methods*, Springer: New York.
- Smith AA. 1993. Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics* **8**(S1):S63–S84.
- Stelzer R. 2010. Multivariate COGARCH(1,1) processes. *Bernoulli* **16**(1):80–115.
- Wahlberg B, Welsh J, Ljung L. 2015. Identification of stochastic Wiener systems using Indirect Inference. *IFAC-PapersOnLine* **48**(28):620–625.
- White H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* **50**:1–26.

J. Yu. *Empirical characteristic function in time series estimation and a test statistic in financial modelling*. PhD thesis, Faculty of Graduate Studies, University of Western Ontario: London, Ontario, 1998.

Yu J. 2004. Empirical characteristic function estimation and its applications. *Econometric Reviews* **23**(2):93–123.

Zeger SL. 1988. A regression model for time series of counts. *Biometrika* **75**(4):621–629.

### APPENDIX A

Here we present the proofs of the main Theorems, as well as tables of results on the simulation study. Then, in Section A.1 we provide the proofs of Theorems 3.1, 3.2, and 3.3. Finally, we present in Section A.2 the tables summarizing the finite sample behavior of the simulation based estimators for ARFIMA models driven by noise from Gaussian, Laplace, and Student’s  $t$  distributions, and the Poisson-AR(1) model discussed in Section 5.

#### A.1. Proofs of the main results

In the following we define  $H = H(n)$  and  $\bar{H} = \bar{H}(n) = H(n)/n$ , but omit the argument  $n$  for notational simplicity. Throughout the letter  $c$  stands for any positive constant independent of the respective argument. Its value may change from line to line, but is not of particular interest. For a matrix with only real eigenvalues  $\lambda_{\min}(\cdot)$  denotes the smallest eigenvalue.

We often use the uniform SLLN, which guarantees for a continuous stochastic process  $(Z(t))_{t \in \mathbb{R}^p}$  satisfying  $\mathbb{E} \sup_{t \in K} |Z(t)| < \infty$  that  $\sup_{t \in K} |Z(t) - \mathbb{E}Z(t)| \xrightarrow{\text{a.s.}} 0$  as  $n \rightarrow \infty$  for every compact set  $K \subset \mathbb{R}^p$ . More precisely, we use the SLLN on the separable Banach space  $C(K)$ , the space of continuous functions on the compact set  $K \subset \mathbb{R}^p$ , endowed with the sup norm (see e.g., Theorem 16(a) in Ferguson (1996) or Theorem 9.4 in Parthasarathy (1967)).

TABLE AI. Comparison of the simulation based estimator  $\hat{\theta}_{n,H}$  for  $H = 3000$ , the oracle estimator  $\hat{\theta}_n$ , and the MLE for sample size  $n = 400$ . For all estimators we have taken  $p = 3$  with  $w$  the Gaussian density as in (5.3). Reported results are based on 500 replications

ARFIMA model driven by standard Gaussian noise									
	$d = 0.05$			$d = 0.10$			$d = 0.15$		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
$\hat{\theta}_{n,H}$	0.000	0.056	0.056	0.002	0.054	0.054	0.004	0.049	0.049
$\hat{\theta}_n$	-0.005	0.050	0.050	-0.004	0.047	0.047	-0.004	0.044	0.045
MLE	-0.015	0.040	0.043	-0.015	0.040	0.043	-0.016	0.040	0.043
	$d = 0.20$			$d = 0.25$			$d = 0.30$		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
$\hat{\theta}_{n,H}$	0.003	0.047	0.047	0.000	0.046	0.046	-0.003	0.048	0.048
$\hat{\theta}_n$	-0.004	0.045	0.045	-0.006	0.044	0.044	-0.007	0.046	0.047
MLE	-0.016	0.040	0.043	-0.017	0.039	0.043	-0.017	0.039	0.043
	$d = 0.35$			$d = 0.40$			$d = 0.45$		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
$\hat{\theta}_{n,H}$	-0.006	0.050	0.051	-0.013	0.051	0.052	-0.022	0.047	0.052
$\hat{\theta}_n$	-0.009	0.049	0.050	-0.013	0.051	0.052	-0.021	0.048	0.052
MLE	-0.019	0.039	0.043	-0.021	0.037	0.043	-0.027	0.034	0.043

TABLE AII. Comparison of the simulation based estimator  $\hat{\theta}_{n,H}$  for  $H = 3000$ , the quasi-oracle estimator  $\hat{\theta}_n$ , and the QMLE for sample size  $n = 400$ . For all estimators we have taken  $p = 3$  with  $w$  the Gaussian density as in (5.3). Reported are results based on 500 replications

ARFIMA model driven by standard Laplace noise									
	$d = 0.05$			$d = 0.10$			$d = 0.15$		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
$\hat{\theta}_{n,H}$	-0.004	0.062	0.062	-0.003	0.060	0.060	0.004	0.054	0.054
$\hat{\theta}_n$	-0.005	0.051	0.051	-0.003	0.049	0.049	0.001	0.048	0.047
QMLE	-0.012	0.043	0.045	-0.013	0.043	0.045	-0.013	0.043	0.045
	$d = 0.20$			$d = 0.25$			$d = 0.30$		
$\hat{\theta}_{n,H}$	0.008	0.049	0.050	0.012	0.051	0.053	0.012	0.049	0.051
$\hat{\theta}_n$	0.005	0.047	0.047	0.009	0.046	0.047	0.010	0.046	0.047
QMLE	-0.014	0.042	0.044	-0.014	0.042	0.044	-0.015	0.042	0.044
	$d = 0.35$			$d = 0.40$			$d = 0.45$		
$\hat{\theta}_{n,H}$	0.009	0.045	0.046	-0.004	0.042	0.042	-0.022	0.037	0.043
$\hat{\theta}_n$	0.006	0.044	0.044	-0.004	0.040	0.040	-0.023	0.035	0.042
QMLE	-0.016	0.041	0.044	-0.019	0.039	0.044	-0.025	0.035	0.043

Proof of Theorem 3.1. Let

$$Q(\theta) = \int_{\mathbb{R}^p} |\varphi(t, \theta_0) - \varphi(t, \theta)|^2 w(t) dt$$

be the candidate limiting function of  $Q_{n,H}(\theta)$ . For  $\delta > 0$  define the set

$$K_\delta = \{t \in \mathbb{R}^p : |t| \leq \delta\}. \tag{A1}$$

Since  $|e^{i\langle t, \tilde{X}_1(\theta) \rangle}| = 1$  for all  $\theta$  and  $t$ , and the random elements  $(\tilde{X}_j(\theta), \theta \in \Theta)_{j=1}^\infty$  are i.i.d., the uniform SLLN holds giving

$$\sup_{(t, \theta) \in \Theta \times K_\delta} \left| \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \tilde{X}_j(\theta) \rangle} - \varphi(t, \theta) \right| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty. \tag{A2}$$

In particular, for  $\theta = \theta_0$  we also have

$$\sup_{t \in K_\delta} \left| \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle} - \varphi(t, \theta_0) \right| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty. \tag{A3}$$

Applying the inequality  $\|a\|^2 - \|b\|^2 \leq 2\|a - b\|$  for  $a, b \in \mathbb{C}$ ,  $|a|, |b| \leq 1$  gives

$$\begin{aligned} |Q_{n,H}(\theta) - Q(\theta)| &= \int_{\mathbb{R}^p} \left| \left| \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle} - \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \tilde{X}_j(\theta) \rangle} \right|^2 - |\varphi(t, \theta_0) - \varphi(t, \theta)|^2 \right| w(t) dt \\ &\leq 2 \int_{\mathbb{R}^p} \left| \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle} - \varphi(t, \theta_0) + \varphi(t, \theta) - \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \tilde{X}_j(\theta) \rangle} \right| w(t) dt \end{aligned}$$

$$\begin{aligned}
 &\leq 2 \int_{\mathbb{R}^p} \left\{ \left| \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle} - \varphi(t, \theta_0) \right| + \sup_{\theta \in \Theta} \left| \varphi(t, \theta) - \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \tilde{X}_j(\theta) \rangle} \right| \right\} w(t) dt \\
 &\leq 2 \sup_{(t, \theta) \in \Theta \times K_\delta} \left\{ \left| \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle} - \varphi(t, \theta_0) \right| + \left| \varphi(t, \theta) - \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \tilde{X}_j(\theta) \rangle} \right| \right\} \\
 &\quad \times \int_{K_\delta} w(t) dt + 8 \int_{K_\delta^c} w(t) dt. \tag{A4}
 \end{aligned}$$

Applying  $\sup_{\theta \in \Theta}$  on both sides of (A4), using (A2) combined with (d.1), and taking the limit for  $\delta \downarrow 0$  gives

$$\sup_{\theta \in \Theta} |Q_{n,H}(\theta) - Q(\theta)| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty. \tag{A5}$$

Now we prove that  $Q(\theta) = 0$  if and only if  $\theta = \theta_0$ . Obviously  $Q(\theta_0) = 0$ . If  $\theta \neq \theta_0$ , then the distributions of  $X_1$  and  $\tilde{X}_1(\theta)$  are different and thus also their characteristic functions are different. Since characteristic functions are continuous, it follows that they are different at least on an interval with positive Lebesgue measure; hence  $Q(\theta) > 0$ . Therefore,  $Q(\theta)$  is uniquely minimized at  $\theta_0$  and this fact together with (A5) gives strong consistency of  $\hat{\theta}_{n,H}$ . ■

*Proof of Theorem 3.2.* We have that  $\widehat{\text{Var}}(\langle t, X_1 \rangle) = t^T \hat{\Gamma}_p t$ , with  $\hat{\Gamma}_p$  being the  $p$ -dimensional empirical covariance matrix of the observed time series  $(X_1, \dots, X_T)$  as in (2.21). Let  $k > 0$  be fixed and

$$Q^{(\text{cv})}(\theta) = \int_{\mathbb{R}^p} \left| \varphi(t, \theta_0) - \varphi(t, \theta) \right|^2 \frac{w(t)}{t^T \Gamma_p t} dt$$

be the candidate limiting function of  $Q_{n,H,k}^{(\text{cv})}(\theta)$  in (2.20), where  $\Gamma_p$  is the theoretical  $p$ -dimensional covariance matrix of the time series process  $(X_j)_{j \in \mathbb{Z}}$ .

Based on the definition of  $Q_{n,H,k}^{(\text{cv})}(\theta)$  in (2.20), we divide the domain of integration in the integrated mean squared error  $|Q_{n,H,k}^{(\text{cv})}(\theta) - Q^{(\text{cv})}(\theta)|$  into  $\{\widehat{\text{Var}}(\langle t, X_1 \rangle) < k\}$  and  $\{\widehat{\text{Var}}(\langle t, X_1 \rangle) \geq k\}$ , equivalently into  $L_n = \{t \in \mathbb{R}^p : t^T \hat{\Gamma}_p t < k\}$  and its complement  $L_n^c$ .

Recall also (2.17) and (2.18). Using  $|e^{ix}|=1$  for all  $x \in \mathbb{R}$ , together with  $|ab - cd| \leq |b||a - c| + |c||b - d|$  for  $a, b, c, d \in \mathbb{C}$  gives for the integral on  $L_n^c$ :

$$\begin{aligned}
 |Q_{n,H,k}^{(\text{cv})}(\theta) - Q^{(\text{cv})}(\theta)|_{L_n^c} &:= \int_{L_n^c} \left| \left| \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle} - \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \tilde{X}_j(\theta) \rangle} \right|^2 \frac{1}{t^T \hat{\Gamma}_p t} - \left| \varphi(t, \theta_0) - \varphi(t, \theta) \right|^2 \frac{1}{t^T \Gamma_p t} \right| w(t) dt \\
 &\leq \int_{L_n^c} \left| \left| \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle} - \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \tilde{X}_j(\theta) \rangle} \right|^2 - \left| \varphi(t, \theta_0) - \varphi(t, \theta) \right|^2 \right| \frac{1}{t^T \hat{\Gamma}_p t} w(t) dt \\
 &\quad + 4 \int_{L_n^c} \left| \frac{1}{t^T \hat{\Gamma}_p t} - \frac{1}{t^T \Gamma_p t} \right| w(t) dt. \tag{A6}
 \end{aligned}$$

By (a.3) and (c.1) it follows from Theorem 3(a) in Section 1.2.2 of Doukhan (1994) that

$$\left| \text{Cov}(X_0, X_j) \right| \leq 8\alpha_j^{\frac{1}{r}} (\mathbb{E} |X_1|^u)^{\frac{2}{u}} \rightarrow 0, \quad j \rightarrow \infty. \tag{A7}$$

Since  $\text{Var}(X_1) > 0$ , it follows from (A7) combined with Proposition 5.1.1 in Brockwell and Davis (2013) that  $\det(\Gamma_p) > 0$ , and therefore, the minimum eigenvalue  $\lambda_{\min}(\Gamma_p)$  of  $\Gamma_p$  is positive. Thus, for all  $t \in \mathbb{R}^p$ ,

$$t^T \Gamma_p t \geq \lambda_{\min}(\Gamma_p) |t|^2 > 0. \tag{A8}$$

By (a.2) and the ergodic theorem  $\hat{\Gamma}_p \xrightarrow{\text{a.s.}} \Gamma_p$  and, since the eigenvalues of a matrix are continuous functions of its entries (cf. Bernstein (2009), Fact 10.11.2), also  $\lambda_{\min}(\hat{\Gamma}_p) \xrightarrow{\text{a.s.}} \lambda_{\min}(\Gamma_p) > 0$ . It follows from (A8) and from the a.s. convergence of the eigenvalues that there exists  $N > 0$  such that

$$t^T \hat{\Gamma}_p t \geq |t|^2 \lambda_{\min}(\hat{\Gamma}_p) \geq |t|^2 \frac{\lambda_{\min}(\Gamma_p)}{2} > 0, \quad n \geq N. \tag{A9}$$

Thus, for  $t \in L_n^c$  we obtain

$$\left| \frac{1}{t^T \hat{\Gamma}_p t} - \frac{1}{t^T \Gamma_p t} \right| \leq \frac{2}{k \lambda_{\min}(\Gamma_p) |t|^2} |t^T (\Gamma_p - \hat{\Gamma}_p) t| \leq \frac{2|\Gamma_p - \hat{\Gamma}_p|}{k \lambda_{\min}(\Gamma_p)}. \tag{A10}$$

This together with (A10) gives the following upper bound for the right-hand side of (A6):

$$\int_{\mathbb{R}^p} \left\| \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle} - \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \bar{X}_j(\theta) \rangle} \right\|^2 - \left| \varphi(t, \theta_0) - \varphi(t, \theta) \right|^2 \frac{w(t)}{k} dt + \frac{8|\Gamma_p - \hat{\Gamma}_p|}{k \lambda_{\min}(\Gamma_p)} \int_{\mathbb{R}^p} w(t) dt. \tag{A11}$$

The first integral can be estimated as  $|Q_{n,H}(\theta) - Q(\theta)|$  in (A4) which tends to 0 uniformly for  $\theta \in \Theta$  provided that (d.1) holds. Since  $\hat{\Gamma}_p \xrightarrow{\text{a.s.}} \Gamma_p$ , also the second integral in (A11) tends 0 a.s. as  $n \rightarrow \infty$ .

We turn to the integrated mean squared error  $|Q_{n,H,k}^{(cv)}(\theta) - Q^{(cv)}(\theta)|$  on  $L_n$ . Let  $L = \{t \in \mathbb{R}^p : |t| \leq \sqrt{\frac{2k}{\lambda_{\min}(\Gamma_p)}}\}$ . The control variates correction used in (2.20) can be regarded as a continuous function  $g : \mathbb{R}^9 \mapsto \mathbb{R}^2$  whose entries are the arithmetic means defined in (4.1)-(4.2). By (c.3) and the uniform SLLN, each of these arithmetic means converge a.s. uniformly on  $L \times \Theta$  as  $n \rightarrow \infty$  and  $H \rightarrow \infty$ . Thus, it follows from the continuity of  $g$  and the continuous mapping theorem that

$$\sup_{(t,\theta) \in L \times \Theta} |\kappa_H(t, \theta)|^2 \xrightarrow{\text{a.s.}} 0. \tag{A12}$$

For  $n \geq N$  it follows from (A9) that  $L_n \subseteq L$  and thus using the inequality

$$\begin{aligned} \left| |a + b|^2 c - |d|^2 e \right| &\leq \left| |a + b|^2 - |d|^2 \right| |c| + |d|^2 |c - e| \\ &\leq (|a - d| + |b|)(4 + |b|) |c| + 4|c - e|, \end{aligned}$$

valid for  $a, b, c, d, e \in \mathbb{C}$  with  $|d| \leq 2$  gives

$$\begin{aligned} &\int_{L_n} \left\| \left( \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle} - \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \bar{X}_j(\theta) \rangle} \right) + \kappa_H(t, \theta) \right\|^2 \frac{1}{t^T \hat{\Gamma}_p t} - \left| \varphi(t, \theta_0) - \varphi(t, \theta) \right|^2 \frac{1}{t^T \Gamma_p t} w(t) dt \\ &\leq \int_L \left( \left| \frac{1}{n} \sum_{j=1}^n e^{i\langle t, X_j \rangle} - \varphi(t, \theta_0) \right| + \left| \frac{1}{H} \sum_{j=1}^H e^{i\langle t, \bar{X}_j(\theta) \rangle} - \varphi(t, \theta) \right| + |\kappa_H(t, \theta)| \right)^2 \frac{1}{t^T \Gamma_p t} w(t) dt \end{aligned}$$

$$\begin{aligned} & \times (4 + |\kappa_H(t, \theta)|) \frac{w(t)}{t^T \Gamma_p t} dt + 4 \int_L \left| \frac{1}{t^T \hat{\Gamma}_p t} - \frac{1}{t^T \Gamma_p t} \right| w(t) dt \\ & =: I_{1,n}(\theta) + I_{2,n}(\theta). \end{aligned}$$

From (A8), (A12), (A2), and (A3) with  $K_\delta = L$  for  $\delta = \sqrt{2k/\lambda_{\min}(\Gamma_p)}$ , and (d.4) it follows that  $\sup_{\theta \in \Theta} I_{1,n}(\theta) \xrightarrow{\text{a.s.}} 0$  as  $n \rightarrow \infty$ . Finally,  $\sup_{\theta \in \Theta} I_{2,n}(\theta) \xrightarrow{\text{a.s.}} 0$  by similar arguments as used in (A10) and (A11), since for  $t \in L$ , also applying (d.4),

$$\left| \frac{1}{t^T \hat{\Gamma}_p t} - \frac{1}{t^T \Gamma_p t} \right| \leq \frac{2}{(\lambda_{\min}(\Gamma_p))^2 |t|^4} |t^T (\Gamma_p - \hat{\Gamma}_p) t| \leq \frac{2|\Gamma_p - \hat{\Gamma}_p|}{(\lambda_{\min}(\Gamma_p))^2 |t|^2}$$

and

$$\int_{\mathbb{R}^p} \frac{w(t)}{|t|^2} dt < \infty. \quad \blacksquare$$

*Proof of Theorem 3.3.* By the definition of  $\hat{\theta}_{n,H}$  in (2.8) and under assumptions (a.1) and (b.2) we have

$$\nabla_\theta \mathcal{Q}_{n,H}(\hat{\theta}_{n,H}) = 0.$$

A Taylor expansion of order 1 of  $\nabla_\theta \mathcal{Q}_{n,H}$  around  $\theta_0$  gives

$$0 = \nabla_\theta \mathcal{Q}_{n,H}(\theta_0) + \nabla_\theta^2 \mathcal{Q}_{n,H}(\theta_n)(\hat{\theta}_{n,H} - \theta_0)$$

where  $\theta_n \xrightarrow{\text{a.s.}} \theta_0$  as  $n \rightarrow \infty$ . Therefore, asymptotic normality of  $\sqrt{n}(\hat{\theta}_{n,H} - \theta_0)$  will follow by the delta method, if we prove that as  $n \rightarrow \infty$ :

- (1)  $\sqrt{n} \nabla_\theta \mathcal{Q}_{n,H}(\theta_0)$  converges weakly to a multivariate normal random variable, and
- (2)  $\nabla_\theta^2 \mathcal{Q}_{n,H}(\theta_n)$  converges in probability to a non-singular matrix.

We start with the first point and compute the partial derivatives of  $\mathcal{Q}_{n,H}$ :

$$\begin{aligned} \frac{\partial}{\partial \theta^{(i)}} \mathcal{Q}_{n,H}(\theta) &= \frac{\partial}{\partial \theta^{(i)}} \left( \int_{\mathbb{R}^p} |\varphi_n(t) - \varphi_H(t, \theta)|^2 w(t) dt \right) \\ &= \int_{\mathbb{R}^p} \frac{\partial}{\partial \theta^{(i)}} (\Re(\varphi_n(t) - \varphi_H(t, \theta))^2 + \Im(\varphi_n(t) - \varphi_H(t, \theta))^2) w(t) dt \\ &= -2 \int_{\mathbb{R}^p} (\Re(\varphi_n(t) - \varphi_H(t, \theta)) \frac{\partial}{\partial \theta^{(i)}} \Re(\varphi_H(t, \theta)) \\ &\quad + \Im(\varphi_n(t) - \varphi_H(t, \theta)) \frac{\partial}{\partial \theta^{(i)}} \Im(\varphi_H(t, \theta))) w(t) dt, \quad i \in 1, \dots, q. \end{aligned} \tag{A13}$$

Recall that  $\varphi_n(t)$  and  $\varphi_H(t, \theta)$  denote the empirical characteristic functions of the observed blocks  $(X_1, \dots, X_n)$  as in (2.2) and of its Monte Carlo approximation  $(\tilde{X}_1(\theta), \dots, \tilde{X}_H(\theta))$  as in (2.7), respectively. Define the partial derivatives of the real and imaginary part of  $\varphi_H(t, \theta)$ :

$$b_H^{(i)}(t, \theta) = \frac{1}{H} \sum_{j=1}^H \begin{pmatrix} -\sin(\langle t, \tilde{X}_j(\theta) \rangle) \\ \cos(\langle t, \tilde{X}_j(\theta) \rangle) \end{pmatrix} \langle t, \frac{\partial}{\partial \theta^{(i)}} \tilde{X}_j(\theta) \rangle, \quad i = 1, \dots, q, \tag{A14}$$

and summarize them into

$$b_H(t, \theta) = \begin{pmatrix} (b_H^{(1)}(t, \theta))^T \\ \vdots \\ (b_H^{(q)}(t, \theta))^T \end{pmatrix}. \tag{A15}$$

Then consider

$$\begin{pmatrix} \mathfrak{R}(\varphi_n(t) - \varphi(t, \theta_0)) \\ \mathfrak{I}(\varphi_n(t) - \varphi(t, \theta_0)) \end{pmatrix} - \begin{pmatrix} \mathfrak{R}(\varphi_H(t, \theta) - \varphi(t, \theta_0)) \\ \mathfrak{I}(\varphi_H(t, \theta) - \varphi(t, \theta_0)) \end{pmatrix} =: g_n(t) - \tilde{g}_H(t, \theta). \tag{A16}$$

Abbreviate  $b_H(t) := b_H(t, \theta_0)$  and  $\tilde{g}_H(t) := \tilde{g}_H(t, \theta_0)$ . Then it follows from (A13), (A15) and (A16) that

$$\nabla_{\theta} Q_{n,H}(\theta_0) = 2 \int_{\mathbb{R}^p} b_H(t) g_n(t) w(t) dt - 2 \int_{\mathbb{R}^p} b_H(t) \tilde{g}_H(t) w(t) dt. \tag{A17}$$

We analyze the asymptotic behavior of the first term in (A17) in Lemma A.3. More precisely, we show there that  $\int_{K_{\delta}} b_H(t) g_n(t) w(t) dt$  for  $K_{\delta}$  as in (A1) converge in distribution to a  $q$ -dimensional Gaussian vector. Afterwards, Lemmas A.4 and A.5 show that as  $\delta \rightarrow \infty$ , componentwise in  $\mathbb{R}^q$ ,

$$\limsup_{n \rightarrow \infty} \text{Var} \left( \int_{K_{\delta}^c} b_H(t) \sqrt{n} g_n(t) w(t) dt \right) \rightarrow 0,$$

and

$$\text{and } \int_{K_{\delta}^c} \mathbb{E}[b_1(t)] G(t) w(t) dt \rightarrow 0$$

where  $G$  is a zero mean  $\mathbb{R}^2$ -valued Gaussian field. The formula given in (A17) tells us that the term  $\mathbb{E}[b_1(t, \theta)]$  will appear in the asymptotic covariance formula of the limiting distribution of the estimator. Therefore it is worth writing it in terms of the chf (2.5). ■

**Remark A.1.** For each  $i \in \{1, \dots, q\}$  and  $\theta \in \Theta$ , it follows from (A14) that

$$\begin{aligned} \mathbb{E}[b_1^{(i)}(t, \theta)] &= \mathbb{E} \left( \begin{pmatrix} -\sin(\langle t, \tilde{X}_j(\theta) \rangle) \\ \cos(\langle t, \tilde{X}_j(\theta) \rangle) \end{pmatrix} \langle t, \frac{\partial}{\partial \theta^{(i)}} \tilde{X}_j(\theta) \rangle \right) \\ &= \mathbb{E} \left( \frac{\partial}{\partial \theta^{(i)}} \cos(\langle t, \tilde{X}_j(\theta) \rangle), \frac{\partial}{\partial \theta^{(i)}} \sin(\langle t, \tilde{X}_j(\theta) \rangle) \right) \end{aligned} \tag{A18}$$

Since both sin and cos are bounded by 1 we can use (c.5) to interchange expectation and differentiation in (A18). This combined with (A15) gives

$$\mathbb{E}[b_1(t, \theta)] = \left( \frac{\partial}{\partial \theta} \mathbb{E} \cos(\langle t, \tilde{X}_j(\theta) \rangle), \frac{\partial}{\partial \theta} \mathbb{E} \sin(\langle t, \tilde{X}_j(\theta) \rangle) \right) = \left( \frac{\partial}{\partial \theta} \mathfrak{R}(\varphi(t, \theta)), \frac{\partial}{\partial \theta} \mathfrak{I}(\varphi(t, \theta)) \right) \tag{A19}$$

This remark will be used later in the proof of Theorem 3.3.

We show by a standard Chebyshev argument that the second term in (A17) converges in probability componentwise to 0 in (A48). The convergence of the second derivatives  $\nabla_{\theta}^2 Q_n(\theta_n)$  will be the topic of Lemma A.6. For the scalar products above we use the following bounds several times below.

**Lemma A.2.** Let  $\nu \geq 1$ ,  $t \in \mathbb{R}^p$ ,  $k, i \in \{1, \dots, q\}$  and  $j \in \mathbb{Z}$  be fixed and assume that (b.2) holds. Then the following bounds hold true.

(a) If  $\mathbb{E}|\nabla_\theta X_1(\theta)|^\nu < \infty$  for  $\theta \in \Theta$ , then there exists a constant  $c > 0$  such that

$$\mathbb{E}\left|\left\langle t, \frac{\partial}{\partial \theta^{(k)}} \tilde{X}_j(\theta) \right\rangle\right|^\nu \leq c|t|^\nu \mathbb{E}|\nabla_\theta X_1(\theta)|^\nu, \quad t \in \mathbb{R}^p. \tag{A20}$$

(b) If  $\mathbb{E}|\nabla_\theta^2 X_1(\theta)|^\nu < \infty$  for  $\theta \in \Theta$ , then there exists a constant  $c > 0$  such that

$$\mathbb{E}\left|\left\langle t, \frac{\partial}{\partial \theta^{(k)} \partial \theta^{(i)}} \tilde{X}_j(\theta) \right\rangle\right|^\nu \leq c|t|^\nu \mathbb{E}|\nabla_\theta^2 X_1(\theta)|^\nu, \quad t \in \mathbb{R}^p. \tag{A21}$$

The same bounds hold uniformly, taking expectations over  $\sup_{\theta \in \Theta}$  or over  $\sup_{t \in K}$  for some compact  $K \subset \mathbb{R}^p$  at both sides of (A20) and (A21), provided the corresponding expectations exist.

*Proof.* (a) Applying the Cauchy–Schwarz inequality for the inner product, the fact that  $(\tilde{X}_j(\theta), \theta \in \Theta) \stackrel{d}{=} (\tilde{X}_1(\theta), \theta \in \Theta) \stackrel{d}{=} (X_1(\theta), \theta \in \Theta)$ , bounding the  $L^2$ -norm by the  $L^1$ -norm, employing the inequality  $|\sum_{j=1}^p \beta_j|^\nu \leq p^{\nu-1} \sum_{j=1}^p |\beta_j|^\nu$  valid for  $\beta_1, \dots, \beta_p \in \mathbb{R}$  and  $\nu \geq 1$  gives

$$\begin{aligned} \mathbb{E}\left|\left\langle t, \frac{\partial}{\partial \theta^{(k)}} \tilde{X}_j(\theta) \right\rangle\right|^\nu &\leq |t|^\nu \mathbb{E}\left|\frac{\partial}{\partial \theta^{(k)}} \tilde{X}_j(\theta)\right|^\nu = |t|^\nu \mathbb{E}\left|\frac{\partial}{\partial \theta^{(k)}} X_1(\theta)\right|^\nu \\ &\leq |t|^\nu \mathbb{E}\left(\sum_{r=1}^p \left|\frac{\partial}{\partial \theta^{(k)}} X_r(\theta)\right|\right)^\nu \leq p^{\nu-1} |t|^\nu \sum_{r=1}^p \mathbb{E}\left|\frac{\partial}{\partial \theta^{(k)}} X_r(\theta)\right|^\nu \\ &\leq p^{\nu-1} |t|^\nu \sum_{r=1}^p \mathbb{E}|\nabla_\theta X_r(\theta)|^\nu = p^\nu |t|^\nu \mathbb{E}|\nabla_\theta X_1(\theta)|^\nu =: c|t|^\nu \mathbb{E}|\nabla_\theta X_1(\theta)|^\nu. \end{aligned} \tag{A22}$$

Part (b) follows by analogous calculations. ■

**Lemma A.3.** Under assumptions (a.2), (b.2), (a.3), (c.2) and (c.4) we have on the Borel sets of  $\mathbb{R}^q$ ,

$$\int_{K_\delta} b_H(t) \sqrt{n} g_n(t) w(t) dt \xrightarrow{d} \int_{K_\delta} \mathbb{E}[b_1(t)] G(t) w(t) dt, \quad n \rightarrow \infty, \tag{A23}$$

where  $G$  is an  $\mathbb{R}^2$ -valued Gaussian field.

*Proof.* Under assumptions (a.3) and (c.2), it follows from Lemma 4.1(2) in Davis *et al.* (2018) that  $\sqrt{n}(\varphi_n(\cdot) - \varphi(\cdot, \theta_0))$  converges in distribution on compact subsets of  $\mathbb{R}^p$  to a complex-valued Gaussian field  $\tilde{G}$ , equivalently the vector of real and imaginary part converge to a bivariate Gaussian field  $G$ . Since the random elements  $(\tilde{X}_j(\theta), \theta \in \Theta)_{j \in \mathbb{N}}$  are i.i.d. and the partial derivatives exist by (b.2), also  $(\tilde{X}_j(\theta_0), \nabla_\theta \tilde{X}_j(\theta_0))_{j \in \mathbb{N}}$  are i.i.d. Then it follows from the definitions (A14), (A15), and Lemma A.2 with  $K = K_\delta$  in combination with (c.4) that

$$\mathbb{E} \sup_{t \in K_\delta} |b_1(t)| \leq c \sup_{t \in K_\delta} |t| \mathbb{E}|\nabla_\theta X_1(\theta_0)| \leq c|\delta| \mathbb{E}|\nabla_\theta X_1(\theta_0)| < \infty. \tag{A24}$$

Hence, the uniform SLLN guarantees that

$$\sup_{t \in K_\delta} |b_H(t) - \mathbb{E}b_1(t)| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty.$$

Slutsky's theorem gives then  $b_H(\cdot)\sqrt{ng_n(\cdot, \theta_0)}$  convergences in distribution on compact subsets of  $\mathbb{R}^p$  to  $\mathbb{E}[b_1(\cdot)]G(\cdot)$  as  $n \rightarrow \infty$ . The result in (A23) follows from the continuity of the integral by another application of the continuous mapping theorem on  $C(K_\delta)$ . ■

**Lemma A.4.** Under assumptions (b.2), (c.4) and (d.2) we have componentwise in  $\mathbb{R}^q$ ,

$$\limsup_{n \rightarrow \infty} \text{Var} \left( \int_{K_\delta^c} b_H(t) \sqrt{n} g_n(t) w(t) dt \right) \rightarrow 0, \quad \delta \rightarrow \infty. \tag{A25}$$

*Proof.* Since  $b_H(\cdot)$  and  $g_n(\cdot)$  are independent and  $\mathbb{E}g_n(t) = 0$ , we have  $\mathbb{E}[b_H(t)g_n(t)] = 0$  for all  $t \in \mathbb{R}^p$ . An application of the Cauchy–Schwartz inequality for integrals gives

$$\begin{aligned} \text{Var} \left( \int_{K_\delta^c} b_H(t) \sqrt{n} g_n(t) w(t) dt \right) &= \mathbb{E} \left( \int_{K_\delta^c} b_H(t) \sqrt{n} g_n(t) w(t) dt \right)^2 \\ &\leq \left( \mathbb{E} \int_{K_\delta^c} |b_H(t)|^2 n |g_n(t)|^2 w(t) dt \right) \left( \int_{K_\delta^c} w(t) dt \right). \end{aligned} \tag{A26}$$

We first obtain a bound for the product between the first component  $g_{n,1}(\cdot)$  of  $g_n(\cdot)$  and the first component  $b_{H,1}^{(i)}(\cdot)$  of  $b_H^{(i)}(\cdot)$ . Define for  $t \in \mathbb{R}^p$

$$\begin{aligned} U_j(t) &= \cos(\langle t, X_j \rangle) - \mathfrak{R}(\varphi(t, \theta_0)) \\ V_j(t) &= -\sin(\langle t, \tilde{X}_j(\theta_0) \rangle) \langle t, \frac{\partial}{\partial \theta^{(i)}} \tilde{X}_j(\theta_0) \rangle, \quad j \in \mathbb{Z}. \end{aligned} \tag{A27}$$

Then,

$$g_{n,1}(t) = \frac{1}{n} \sum_{j=1}^n U_j(t) \quad \text{and} \quad b_{H,1}^{(i)}(t) = \frac{1}{H} \sum_{j=1}^H V_j(t), \quad t \in \mathbb{R}^p.$$

Under (a.3) it follows from Theorem 3(a) in Section 1.2.2 of Doukhan (1994) that for fixed  $t$ ,

$$|\text{Cov}(U_0(t), U_j(t))| \leq 8\alpha_j^{\frac{1}{r}} (\mathbb{E}|U_0(t)|^u)^{\frac{2}{u}}, \quad j \in \mathbb{N}, \tag{A28}$$

where  $u = \frac{2r}{(r-1)}$  and, thus, it follows from the stationarity of  $(U_j(t))_{j \in \mathbb{N}}$  combined with (A28) and the fact that  $|U_0(t)| \leq 2$  that

$$\begin{aligned} n \mathbb{E} \left| \frac{1}{n} \sum_{j=1}^n U_j(t) \right|^2 &= \frac{1}{n} \sum_{j=1}^n \mathbb{E} U_j^2(t) + \frac{2}{n} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \mathbb{E} |U_0(t) U_j(t)| \\ &\leq \mathbb{E} U_0^2(t) + 16 (\mathbb{E}|U_0(t)|^u)^{\frac{2}{u}} \sum_{j=1}^{\infty} \alpha_j^{1/r} \\ &\leq 4 + 64 \sum_{j=1}^{\infty} \alpha_j^{1/r} < \infty, \end{aligned} \tag{A29}$$

where the bound is independent of  $t$ . Recall that  $H = H(n) = \bar{H}(n)n$ . Under (c.4), it follows from the i.i.d. property of  $(V_j(t))_{j \in \mathbb{N}}$

$$n\mathbb{E}\left|\frac{1}{H}\sum_{j=1}^H V_j(t) - \mathbb{E}V_0(t)\right|^2 = n\mathbb{V}\text{ar}\left(\frac{1}{\bar{H}n}\sum_{j=1}^{\bar{H}n} V_j(t)\right) = \frac{\mathbb{E}V_1^2(t)}{\bar{H}(n)} \leq \frac{c|t|^2\mathbb{E}|\nabla_\theta X_1(\theta_0)|^2}{\bar{H}(n)} \leq \frac{c|t|^2}{\bar{H}(n)}. \tag{A30}$$

Using the fact that  $\left|\frac{1}{n}\sum_{j=1}^n U_j(t)\right| \leq 2$ , adding and subtracting  $\mathbb{E}V_0(t)$  with the inequality  $|a + b|^2 \leq 2(|a|^2 + |b|^2)$ , and (A30) gives

$$\begin{aligned} n\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n U_j(t)\right|^2 \left|\frac{1}{H}\sum_{j=1}^H V_j(t)\right|^2 &\leq 2n\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n U_j(t)\right|^2 (\mathbb{E}V_0(t))^2 + 8n\mathbb{E}\left|\frac{1}{H}\sum_{j=1}^H V_j(t) - \mathbb{E}V_0(t)\right|^2 \\ &\leq c\left(1 + \frac{|t|^2}{\bar{H}(n)}\right). \end{aligned} \tag{A31}$$

The calculations in (A29), (A30), and (A31) can now be applied to show that for all  $n \in \mathbb{N}$ ,

$$n\mathbb{E}|g_n(t)|^2 |b_H(t)|^2 \leq c\left(1 + \frac{|t|^2}{\bar{H}(n)}\right)$$

and, thus, it follows from (A26) together with (d.1) and (d.3) that

$$\limsup_{n \rightarrow \infty} \mathbb{V}\text{ar}\left(\int_{K_\delta^c} b_H(t)\sqrt{n}g_n(t)w(t)dt\right) \leq \limsup_{n \rightarrow \infty} \frac{c}{\bar{H}(n)} \int_{K_\delta^c} (1 + |t|^2)w(t)dt \int_{K_\delta^c} w(t)dt \rightarrow 0, \quad \delta \rightarrow \infty. \tag{A32}$$

**Lemma A.5.** Under assumptions (b.2), (d.2) and (c.4)

$$\int_{K_\delta^c} \mathbb{E}[b_1(t)]G(t)w(t)dt \xrightarrow{P} 0, \quad \delta \rightarrow \infty.$$

*Proof.* It follows from (A14), (A15), (c.4), and (A24)  $\mathbb{E}|b_1(t)| \leq c|t|\mathbb{E}|\nabla_\theta X_1(\theta_0)| < \infty$ . Now we find an upper bound for the variance of each component of  $G(t)$  for a fixed  $t$ . Let  $U_j(t)$  be as defined at the left-hand side of (A27) and notice that the first component of  $G(t)$  is the distributional limit of  $\frac{1}{\sqrt{n}}\sum_{j=1}^n U_j(t)$ . Since  $(U_j(t))_{j \in \mathbb{N}}$  is  $\alpha$ -mixing by (a.3), we can apply the CLT in Ibragimov and Linnik (1971) (Theorem 18.5.3 with  $\delta = 2/(r - 1)$ ) and find that the variance of the first component of  $G(t)$  is given by

$$\sigma_U^2 = \mathbb{E}[U_0^2(t)] + 2\sum_{j=1}^{\infty} \mathbb{E}[U_0(t)U_j(t)].$$

This combined with Theorem 3(a) in Section 1.2.2 of Doukhan (1994) and the fact that  $\mathbb{E}U_j(t) = 0$  and  $|U_j(t)| \leq 2$  for all  $j \in \mathbb{N}$  gives by (a.3) and (A28)

$$|\sigma_U^2| \leq 4 + \sum_{j=1}^{\infty} |\text{Cov}(U_0(t), U_j(t))| \leq 4 + 8\sum_{j=1}^{\infty} (2\alpha_j)^{1/r} (\mathbb{E}|U_0(t)|^u)^{\frac{2}{u}} \leq 4 + 64\sum_{j=1}^{\infty} (2\alpha_j)^{1/r}.$$

A similar calculation shows that the variance of the second component of  $G(t)$  is also bounded by a finite constant, which does not depend on  $t$ . Therefore,  $\mathbb{E}|G(t)| \leq c$ . This combined with (A24) and assumption (d.2) gives

$$\mathbb{E} \left| \int_{K_\delta^c} \mathbb{E}[b_1(t)]G(t)w(t)dt \right| \leq c\mathbb{E}|\nabla_\theta X_1(\theta_0)| \int_{K_\delta^c} |t|w(t)dt \rightarrow 0, \quad \delta \rightarrow \infty.$$

Since  $L^1$ -convergence implies convergence in probability the result follows. ■

This proves part (1) of the delta method. We now turn to part (2). To calculate the second derivatives of  $Q_{n,H}(\theta)$ , which exist by (b.2), we rewrite (A13) as

$$\begin{aligned} \frac{\partial}{\partial \theta^{(i)}} Q_{n,H}(\theta) &= -2 \int_{\mathbb{R}^d} \left\{ \left( \frac{1}{n} \sum_{j=1}^n \cos(\langle t, X_j(\theta) \rangle) - \frac{1}{H} \sum_{j=1}^H \cos(\langle t, \tilde{X}_j(\theta) \rangle) \right) \frac{\partial}{\partial \theta^{(i)}} \Re(\varphi_H(t, \theta)) \right. \\ &\quad \left. + \left( \frac{1}{n} \sum_{j=1}^n \sin(\langle t, X_j(\theta) \rangle) - \frac{1}{H} \sum_{j=1}^H \sin(\langle t, \tilde{X}_j(\theta) \rangle) \right) \frac{\partial}{\partial \theta^{(i)}} \Im(\varphi_H(t, \theta)) \right\} w(t)dt \\ &=: 2 \int_{\mathbb{R}^d} \{ i_{n,H}(t, \theta) j_{H,i}(t, \theta) - k_{n,H}(t, \theta) l_{H,i}(t, \theta) \} w(t)dt. \end{aligned}$$

For the second derivatives we calculate for every  $i, k \in \{1, \dots, q\}$ ,

$$\begin{aligned} \frac{\partial}{\partial \theta^{(k)} \partial \theta^{(i)}} Q_{n,H}(\theta) &= 2 \int_{\mathbb{R}^d} \left\{ j_{H,k}(t, \theta) j_{H,i}(t, \theta) + i_{n,H}(t, \theta) g_{H,k,i}(t, \theta) \right. \\ &\quad \left. + l_{H,k}(t, \theta) l_{H,i}(t, \theta) - k_{n,H}(t, \theta) h_{H,k,i}(t, \theta) \right\} w(t)dt, \end{aligned} \tag{A33}$$

where we summarize all quantities used in the following list:

$$\begin{aligned} i_{n,H}(t, \theta) &= \frac{1}{n} \sum_{j=1}^n \cos(\langle t, X_j \rangle) - \frac{1}{H} \sum_{j=1}^H \cos(\langle t, \tilde{X}_j(\theta) \rangle) \\ j_{H,i}(t, \theta) &= \frac{\partial}{\partial \theta^{(i)}} i_{n,H}(t, \theta) = \frac{1}{H} \sum_{j=1}^H \sin(\langle t, \tilde{X}_j(\theta) \rangle) \langle t, \frac{\partial}{\partial \theta^{(i)}} \tilde{X}_j(\theta) \rangle \\ g_{H,k,i}(t, \theta) &= \frac{\partial}{\partial \theta^{(k)}} j_{H,i}(t, \theta) \\ &= \frac{1}{H} \sum_{j=1}^H \cos(\langle t, \tilde{X}_j(\theta) \rangle) \langle t, \frac{\partial}{\partial \theta^{(k)}} \tilde{X}_j(\theta) \rangle \langle t, \frac{\partial}{\partial \theta^{(i)}} \tilde{X}_j(\theta) \rangle \\ &\quad + \sin(\langle t, \tilde{X}_j(\theta) \rangle) \langle t, \frac{\partial}{\partial \theta^{(k)} \partial \theta^{(i)}} \tilde{X}_j(\theta) \rangle \\ k_{n,H}(t, \theta) &= \frac{1}{n} \sum_{j=1}^n \sin(\langle t, X_j \rangle) - \frac{1}{H} \sum_{j=1}^H \sin(\langle t, \tilde{X}_j(\theta) \rangle) \\ l_{H,i}(t, \theta) &= -\frac{\partial}{\partial \theta^{(i)}} k_{n,H}(t, \theta) = \frac{1}{H} \sum_{j=1}^H \cos(\langle t, \tilde{X}_j(\theta) \rangle) \langle t, \frac{\partial}{\partial \theta^{(i)}} \tilde{X}_j(\theta) \rangle \\ h_{H,k,i}(t, \theta) &= \frac{\partial}{\partial \theta^{(k)}} l_{H,i}(t, \theta) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{H} \sum_{j=1}^H -\sin(\langle t, \tilde{X}_j(\theta) \rangle) \langle t, \frac{\partial}{\partial \theta^{(k)}} \tilde{X}_j(\theta) \rangle \langle t, \frac{\partial}{\partial \theta^{(i)}} \tilde{X}_j(\theta) \rangle \\
 &\quad + \cos(\langle t, \tilde{X}_j(\theta) \rangle) \langle t, \frac{\partial}{\partial \theta^{(k)} \partial \theta^{(i)}} \tilde{X}_j(\theta) \rangle.
 \end{aligned}$$

**Lemma A.6.** If the assumptions (a.2), (b.1), (b.2), (c.5), (d.3) hold and  $(\theta_n)_{n \in \mathbb{N}} \subset \Theta$  satisfying  $\theta_n \xrightarrow{\text{a.s.}} \theta_0$ , then for every  $k, i \in \{1, \dots, q\}$ , as  $n \rightarrow \infty$

$$\frac{\partial}{\partial \theta^{(k)} \partial \theta^{(i)}} Q_{n,H}(\theta_n) \xrightarrow{P} \int_{\mathbb{R}^p} (\mathbb{E}j_{1,k}(t, \theta_0) \mathbb{E}j_{1,i}(t, \theta_0) + \mathbb{E}l_{1,k}(t, \theta_0) \mathbb{E}l_{1,i}(t, \theta_0)) w(t) dt. \tag{A34}$$

*Proof.* We first prove that as  $n \rightarrow \infty$

$$\int_{\mathbb{R}^p} i_{n,H}(t, \theta_n) g_{H,k,i}(t, \theta_n) w(t) dt \xrightarrow{P} \int_{\mathbb{R}^p} \mathbb{E}i_{1,1}(\theta_0, t) \mathbb{E}g_{1,k,i}(\theta_0, t) w(t) dt. \tag{A35}$$

**Step 1: Uniform convergence on  $\Theta$ :** It follows from the i.i.d. property of the random elements  $(\tilde{X}_j(\theta), \theta \in \Theta)_{j \in \mathbb{N}}$  that the sequence  $(\tilde{X}_j(\theta), \nabla_{\theta} \tilde{X}_j(\theta), \nabla_{\theta}^2 \tilde{X}_j(\theta), \theta \in \Theta)_{j \in \mathbb{N}}$  is i.i.d. Lemma A.2 together with (c.5) gives the uniform bound

$$\mathbb{E} \sup_{\theta \in \Theta} |g_{1,k,i}(t, \theta)| \leq c(|t|^2 \mathbb{E} \sup_{\theta \in \Theta} |\nabla_{\theta} X_1(\theta)|^2 + |t| \mathbb{E} \sup_{\theta \in \Theta} |\nabla_{\theta}^2 X_1(\theta)|) < \infty,$$

and it follows from the uniform SLLN that for every fixed  $t \in \mathbb{R}^p$

$$\sup_{\theta \in \Theta} |g_{H,k,i}(t, \theta) - \mathbb{E}g_{1,k,i}(t, \theta)| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty. \tag{A36}$$

Similarly,

$$\sup_{\theta \in \Theta} \left| \frac{1}{H} \sum_{j=1}^H \cos(\langle t, \tilde{X}_j(\theta) \rangle) - \mathfrak{R}(\varphi(t, \theta)) \right| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty. \tag{A37}$$

Because of (a.2) the ergodic theorem gives

$$\frac{1}{n} \sum_{j=1}^n \cos(\langle t, X_j \rangle) \xrightarrow{\text{a.s.}} \mathfrak{R}(\varphi(t, \theta_0)), \quad n \rightarrow \infty. \tag{A38}$$

Therefore, (A37) combined with (A38) and the triangle inequality imply

$$\sup_{\theta \in \Theta} |i_{n,H}(t, \theta) - \mathbb{E}i_{1,1}(t, \theta)| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty. \tag{A39}$$

**Step 2: Pointwise convergence of  $i_{n,H}(t, \theta_n) g_{H,k,i}(t, \theta_n)$ :** The triangle inequality implies

$$\begin{aligned}
 &|i_{n,H}(t, \theta_n) g_{H,k,i}(t, \theta_n) - \mathbb{E}i_{1,1}(\theta_0, t) \mathbb{E}g_{1,k,i}(\theta_0, t)| \\
 &\leq |i_{n,H}(t, \theta_n) g_{H,k,i}(t, \theta_n) - \mathbb{E}i_{1,1}(t, \theta_n) \mathbb{E}g_{1,k,i}(t, \theta_n)| \\
 &\quad + |\mathbb{E}i_{1,1}(t, \theta_n) \mathbb{E}g_{1,k,i}(t, \theta_n) - \mathbb{E}i_{1,1}(\theta_0, t) \mathbb{E}g_{1,k,i}(\theta_0, t)| \\
 &\leq \sup_{\theta \in \Theta} \left\{ |i_{n,H}(t, \theta) g_{H,k,i}(t, \theta) - \mathbb{E}i_{1,1}(t, \theta) \mathbb{E}g_{1,k,i}(t, \theta)| \right\}
 \end{aligned}$$

$$+ |\mathbb{E}i_{1,1}(t, \theta_n)\mathbb{E}g_{1,k,i}(t, \theta_n) - \mathbb{E}i_{1,1}(\theta_0, t)\mathbb{E}g_{1,k,i}(\theta_0, t)|. \tag{A40}$$

Since  $\theta_n \xrightarrow{\text{a.s.}} \theta_0$  and the map  $\theta \mapsto \mathbb{E}i_{1,1}(t, \theta)\mathbb{E}g_{1,k,i}(t, \theta)$  is continuous in  $\Theta$ , (by (b.2) and (c.5)) it follows that the second term on the right-hand side of (A40) converges a.s. to zero. Additionally, since the uniform convergences on (A36) and (A39) imply the uniform convergence of the product  $i_{n,H}(t, \theta)g_{H,k,i}(t, \theta)$  on  $\Theta$  it follows that the first term on the right-hand side of (A40) also converges a.s. to zero.

**Step 3:  $L^1$ -convergence:** Since we have already shown a.s. convergence, it follows from Theorems 6.25(iii) and 6.19 in Klenke (2013) (with  $H(x) = |x|^{1+\epsilon}$ ) that  $L^1$ -convergence follows provided that

$$\sup_{n \in \mathbb{N}} \mathbb{E}|i_{n,H}(t, \theta_n)g_{H,k,i}(t, \theta_n)|^{1+\epsilon} < \infty$$

for some  $\epsilon > 0$ . Using the fact that  $|i_{n,H}(t, \theta_n)| \leq 2$  and the inequality  $|\frac{1}{n} \sum_{j=1}^n \beta_j|^{1+\epsilon} \leq \frac{1}{n} \sum_{j=1}^n |\beta_j|^{1+\epsilon}$ ,  $\beta_1, \dots, \beta_n \in \mathbb{R}$ , we obtain

$$\begin{aligned} \mathbb{E}|i_{n,H}(t, \theta_n)g_{H,k,i}(t, \theta_n)|^{1+\epsilon} &\leq 2^{1+\epsilon} \mathbb{E}|g_{H,k,i}(t, \theta_n)|^{1+\epsilon} \\ &\leq \frac{2^{1+\epsilon}}{H} \sum_{j=1}^H \mathbb{E} \left| \cos(\langle t, \tilde{X}_j(\theta_n) \rangle) \langle t, \frac{\partial}{\partial \theta^{(k)}} \tilde{X}_j(\theta_n) \rangle \langle t, \frac{\partial}{\partial \theta^{(i)}} \tilde{X}_j(\theta_n) \rangle \right. \\ &\quad \left. + \sin(\langle t, \tilde{X}_j(\theta_n) \rangle) \langle t, \frac{\partial}{\partial \theta^{(k)} \partial \theta^{(i)}} \tilde{X}_j(\theta_n) \rangle \right|^{1+\epsilon} \\ &\leq \frac{2^{1+\epsilon}}{H} \sum_{j=1}^H \mathbb{E} \left| \langle t, \frac{\partial}{\partial \theta^{(k)}} \tilde{X}_j(\theta_n) \rangle \langle t, \frac{\partial}{\partial \theta^{(i)}} \tilde{X}_j(\theta_n) \rangle + \langle t, \frac{\partial}{\partial \theta^{(k)} \partial \theta^{(i)}} \tilde{X}_j(\theta_n) \rangle \right|^{1+\epsilon}, \end{aligned} \tag{A41}$$

since  $|\cos(\cdot)|, |\sin(\cdot)| \leq 1$ . Now we use the inequality  $|a + b|^{1+\epsilon} \leq 2^\epsilon (|a|^{1+\epsilon} + |b|^{1+\epsilon})$  for  $a, b \in \mathbb{R}$ , assumption (c.5) for the uniform bound in Lemma A.2 and the fact that the sequence  $(\tilde{X}_j(\theta), \nabla_\theta \tilde{X}_j(\theta), \nabla_\theta^2 \tilde{X}_j(\theta), \theta \in \Theta)_{j \in \mathbb{N}}$  is i.i.d. to continue

$$\begin{aligned} &\leq 2^{1+2\epsilon} \frac{1}{H} \sum_{j=1}^H (\mathbb{E} \left| \langle t, \frac{\partial}{\partial \theta^{(k)}} \tilde{X}_j(\theta_n) \rangle \langle t, \frac{\partial}{\partial \theta^{(i)}} \tilde{X}_j(\theta_n) \rangle \right|^{1+\epsilon} + \mathbb{E} \left| \langle t, \frac{\partial}{\partial \theta^{(k)} \partial \theta^{(i)}} \tilde{X}_j(\theta_n) \rangle \right|^{1+\epsilon}) \\ &\leq c \frac{1}{H} \sum_{j=1}^H (|t|^{2(1+\epsilon)} \mathbb{E} |\nabla_\theta X_1(\theta_n)|^{2(1+\epsilon)} + |t|^{1+\epsilon} \mathbb{E} |\nabla_\theta^2 X_1(\theta_n)|^{1+\epsilon}) \\ &\leq c (|t|^{2(1+\epsilon)} \mathbb{E} \sup_{\theta \in \Theta} |\nabla_\theta X_1(\theta)|^{2(1+\epsilon)} + |t|^{1+\epsilon} \mathbb{E} \sup_{\theta \in \Theta} |\nabla_\theta^2 X_1(\theta)|^{1+\epsilon}) := v(t) < \infty. \end{aligned} \tag{A42}$$

**Step 4: Convergence of the random integrals:** Define the sequence of functions

$$v_n(t) = \mathbb{E}|i_{n,H}(t, \theta_n)g_{H,k,i}(t, \theta_n) - \mathbb{E}i_{1,1}(\theta_0, t)\mathbb{E}g_{1,k,i}(\theta_0, t)|, \quad t \in \mathbb{R}^p,$$

and recall that from the  $L^1$ -convergence showed in Step 3, for every  $t \in \mathbb{R}^p$  we have  $v_n(t) \rightarrow 0$  as  $n \rightarrow \infty$ . From the definition of the function  $v$  in the last line of (A42) it follows that  $\sup_{n \in \mathbb{N}} v_n(t) \leq 2v(t)$ . Additionally, assumption (d.3) implies that

$$\int_{\mathbb{R}^p} v(t)w(t)dt < \infty.$$

Therefore, it follows from Fubini’s Theorem and dominated convergence that

$$\begin{aligned} & \mathbb{E} \left| \int_{\mathbb{R}^p} (i_{n,H}(t, \theta_n)g_{H,k,i}(t, \theta_n) - \mathbb{E}i_{1,1}(\theta_0, t)\mathbb{E}g_{1,k,i}(\theta_0, t)) w(t)dt \right| \\ & \leq \mathbb{E} \int_{\mathbb{R}^p} |i_{n,H}(t, \theta_n)g_{H,k,i}(t, \theta_n) - \mathbb{E}i_{1,1}(\theta_0, t)\mathbb{E}g_{1,k,i}(\theta_0, t)| w(t)dt \\ & = \int_{\mathbb{R}^p} v_n(t)w(t)dt \rightarrow 0, \quad n \rightarrow \infty, \end{aligned} \tag{A43}$$

and therefore the convergence in probability of (A35) follows from the  $L^1$ -convergence in (A43).

The proofs for the other three remaining integrals on the right-hand side of (A33) follow along the same lines. The result in (A34) is then a consequence of the fact that for all  $t \in \mathbb{R}^p$ ,  $\mathbb{E}i_{1,1}(t, \theta_0) = \mathbb{E}k_{1,1}(t, \theta_0) = 0$ . ■

*Proof of Theorem 3.3.* We handle each term in (A17) separately. As a direct consequence of Theorem 3.1 and Lemmas A.3–A.6,

$$-2(\nabla_{\theta}^2 Q_{n,H}(\theta_n))^{-1} \int_{\mathbb{R}^p} b_H(t)\sqrt{n}g_n(t)w(t)dt \xrightarrow{d} N(0, Q^{-1}WQ^{-1}), \quad n \rightarrow \infty,$$

where  $Q = (Q_{k,i})_{k,i=1}^q$  with

$$Q_{k,i} = \int_{\mathbb{R}^p} (\mathbb{E}j_{1,k}(t, \theta_0)\mathbb{E}j_{1,i}(t, \theta_0) + \mathbb{E}l_{1,k}(t, \theta_0)\mathbb{E}l_{1,i}(t, \theta_0)) w(t)dt, \tag{A44}$$

$$W = \text{Var} \left( \int_{\mathbb{R}^p} \mathbb{E}[b_1(t)]G(t)w(t)dt \right),$$

and  $G$  being the  $\mathbb{R}^2$ -valued Gaussian field from Lemma A.3. For arbitrary  $k, r \in \{1, \dots, q\}$  we have

$$\begin{aligned} W_{k,r} &= \text{Cov} \left( \int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(t)]^T G(t)w(t)dt, \int_{\mathbb{R}^p} \mathbb{E}[b_1^{(r)}(t)]^T G(t)w(t)dt \right) \\ &= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(t)]^T \mathbb{E}[G(t)G(s)^T] \mathbb{E}[b_1^{(r)}(s)] w(t)w(s) dt ds. \end{aligned} \tag{A45}$$

Since  $(X_j)_{j \in \mathbb{N}}$  is  $\alpha$ -mixing by (a.3), we can apply the CLT in Ibragimov and Linnik (1971) (Theorem 18.5.3 with  $\delta = 2/(r - 1)$ ) and find that

$$\mathbb{E}[G(t)G(s)^T] = \mathbb{E}[F_1(t)F_1(s)^T] + 2 \sum_{j=2}^{\infty} \mathbb{E}[F_1(t)F_j(s)^T], \tag{A46}$$

where

$$F_j(t) = \begin{pmatrix} \cos(\langle t, X_j \rangle) - \Re(\varphi(t, \theta_0)) \\ \sin(\langle t, X_j \rangle) - \Im(\varphi(t, \theta_0)) \end{pmatrix}. \tag{A47}$$

Substituting (A46) and (A47) into (A45) gives with Fubini’s Theorem

$$W_{k,r} = \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \left\{ \mathbb{E}[b_1^{(k)}(t)]^T (\mathbb{E}[F_1(t)F_1(s)^T] + 2 \sum_{j=2}^{\infty} \mathbb{E}[F_1(t)F_j(s)^T]) \mathbb{E}[b_1^{(r)}(s)] w(t)w(s) \right\} dt ds$$

$$\begin{aligned}
 &= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(t)]^T \mathbb{E}[F_1(t)F_1(s)^T] \mathbb{E}[b_1^{(k)}(s)] w(t)w(s) dt ds \\
 &\quad + 2 \sum_{j=2}^{\infty} \int_{\mathbb{R}^{p^2}} \mathbb{E}[b_1^{(k)}(t)]^T \mathbb{E}[F_1(t)F_j(s)^T] \mathbb{E}[b_1^{(k)}(s)] w(t)w(s) dt ds \\
 &= \mathbb{E} \left( \int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(t)]^T F_1(t) w(t) dt \right)^2 \\
 &\quad + 2 \sum_{j=2}^{\infty} \mathbb{E} \left[ \left( \int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(t)]^T F_1(t) w(t) dt \right) \left( \int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(s)]^T F_j(s) w(s) ds \right) \right],
 \end{aligned}$$

which combined with Remark A.1 gives (3.4). By the same arguments of interchanging expectation and differentiation from Remark A.1 we obtain

$$\mathbb{E}j_{1,i}(t, \theta) = -\frac{\partial}{\partial \theta^{(i)}} \mathfrak{R}(\varphi(t, \theta)) \quad \text{and} \quad \mathbb{E}l_{1,i}(t, \theta) = \frac{\partial}{\partial \theta^{(i)}} \mathfrak{I}(\varphi(t, \theta)).$$

This together with (A44) gives

$$Q_{k,i} = \int_{\mathbb{R}^p} \left( \frac{\partial}{\partial \theta^{(k)}} \mathfrak{R}(\varphi(t, \theta_0)), \frac{\partial}{\partial \theta^{(k)}} \mathfrak{I}(\varphi(t, \theta_0)) \right) \left( \frac{\partial}{\partial \theta^{(i)}} \mathfrak{R}(\varphi(t, \theta_0)), \frac{\partial}{\partial \theta^{(i)}} \mathfrak{I}(\varphi(t, \theta_0)) \right)^T w(t) dt,$$

leading to (3.2).

The second term in (A17) is, up to a constant,

$$\int_{\mathbb{R}^p} b_H(t) \tilde{g}_H(t) w(t) dt.$$

It follows from the fact that  $(\tilde{X}_j(\theta_0))_{j \in \mathbb{N}} \stackrel{d}{=} (X_j)_{j \in \mathbb{N}}$  combined with (A32) that

$$\text{Var} \left( \int_{\mathbb{R}^p} b_H(t) \sqrt{n} \tilde{g}_H(t) w(t) dt \right) \leq \frac{c}{\bar{H}(n)} \left( \int_{\mathbb{R}^p} (1 + |t|^2) w(t) dt \right) \left( \int_{\mathbb{R}^p} w(t) dt \right) =: \frac{c}{\bar{H}(n)} \rightarrow 0, \tag{A48}$$

as  $n \rightarrow \infty$ . Thus (3.3) follows from Chebyshev’s inequality. ■

## A.2. Finite sample behavior of the estimators

A.2.1. ARFIMA models driven by noise from Gaussian, Laplace, and Student’s *t* distributions

A.2.2. Poisson-AR model

**Data sharing:** Data sharing is not applicable to this article as no datasets were analyzed or used in this study.

TABLE AIII. Comparison of the simulation based estimator  $\hat{\theta}_{n,H}$  for  $H = 3000$ , the quasi-oracle estimator  $\hat{\theta}_n$ , and the QMLE for sample size  $n = 400$ . For all estimators we have taken  $p = 3$  with  $w$  the Gaussian density as in (5.3). Reported results are based on 500 replications

ARFIMA model driven by standard Student's  $t$  noise with 6 degrees of freedom

	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
	$d = 0.05$			$d = 0.10$			$d = 0.15$		
$\hat{\theta}_{n,H}$	-0.002	0.063	0.063	0.005	0.059	0.060	0.008	0.053	0.054
$\hat{\theta}_n$	-0.002	0.052	0.052	-0.001	0.050	0.050	0.000	0.048	0.048
QMLE	-0.012	0.039	0.041	-0.012	0.039	0.041	-0.013	0.039	0.041
	$d = 0.20$			$d = 0.25$			$d = 0.30$		
$\hat{\theta}_{n,H}$	0.008	0.049	0.050	0.004	0.050	0.050	0.008	0.050	0.050
$\hat{\theta}_n$	0.001	0.047	0.047	0.002	0.047	0.047	0.001	0.047	0.047
QMLE	-0.013	0.039	0.041	-0.014	0.039	0.041	-0.014	0.039	0.041
	$d = 0.35$			$d = 0.40$			$d = 0.45$		
$\hat{\theta}_{n,H}$	0.002	0.049	0.049	-0.009	0.043	0.044	-0.029	0.038	0.048
$\hat{\theta}_n$	-0.004	0.046	0.046	-0.014	0.043	0.045	-0.031	0.038	0.049
QMLE	-0.016	0.038	0.041	-0.018	0.037	0.041	-0.024	0.033	0.041

TABLE AIV. Comparison of the simulation based estimator  $\hat{\theta}_{n,H}$  of (2.8) and the control variates based estimator  $\hat{\theta}_{n,H,k}^{(cv)}$  of (2.19) with  $k = 1$  for sample size  $n = 400$ . For all estimators we have taken  $H = 3000$ ,  $p = 3$  with  $w$  the Laplace density as in (5.2). Reported results are based on 500 replications. The models are classified by the index of dispersion  $D = e^{\beta + \alpha_1}$ . For each setting, the smallest RMSEs are shaded

Poisson-AR(1) model

	$\beta$	$\phi$	$\sigma$	$\beta$	$\phi$	$\sigma$	$\beta$	$\phi$	$\sigma$
$D = 10$									
TRUE	-0.613	-0.500	1.236	-0.613	0.500	1.236	-0.613	0.900	0.622
Bias( $\hat{\theta}_{n,H}$ )	-0.015	0.025	0.002	-0.012	0.014	-0.032	-0.016	-0.010	0.002
RMSE( $\hat{\theta}_{n,H}$ )	0.096	0.101	0.119	0.148	0.107	0.120	0.298	0.054	0.128
Bias( $\hat{\theta}_{n,H,k}^{(cv)}$ )	0.023	0.031	-0.007	0.006	0.002	-0.018	0.061	-0.007	-0.036
RMSE( $\hat{\theta}_{n,H,k}^{(cv)}$ )	0.102	0.129	0.122	0.138	0.098	0.098	0.285	0.049	0.132
$D = 1$									
TRUE	0.150	-0.500	0.619	0.150	0.500	0.619	0.150	0.900	0.312
Bias( $\hat{\theta}_{n,H}$ )	-0.004	0.024	-0.016	-0.006	0.005	-0.023	-0.016	-0.033	0.028
RMSE( $\hat{\theta}_{n,H}$ )	0.057	0.144	0.088	0.074	0.141	0.081	0.147	0.084	0.095
Bias( $\hat{\theta}_{n,H,k}^{(cv)}$ )	0.003	-0.011	-0.017	0.001	0.023	-0.019	0.003	-0.009	-0.012
RMSE( $\hat{\theta}_{n,H,k}^{(cv)}$ )	0.055	0.124	0.085	0.071	0.102	0.069	0.145	0.062	0.087
$D = 0.1$									
TRUE	0.373	-0.500	0.220	0.373	0.500	0.220	0.373	0.900	0.111
Bias( $\hat{\theta}_{n,H}$ )	-0.011	0.032	-0.045	-0.015	-0.322	-0.036	-0.019	-0.517	0.044
RMSE( $\hat{\theta}_{n,H}$ )	0.043	0.408	0.098	0.047	0.657	0.102	0.066	0.801	0.099
Bias( $\hat{\theta}_{n,H,k}^{(cv)}$ )	-0.002	0.056	-0.044	-0.003	-0.120	-0.038	-0.004	-0.310	0.031
RMSE( $\hat{\theta}_{n,H,k}^{(cv)}$ )	0.042	0.482	0.112	0.045	0.504	0.108	0.062	0.555	0.090