

# A Higher-Order Correct Fast Moving-Average Bootstrap for Dependent Data

Davide La Vecchia<sup>a</sup>, Alban Moor<sup>a</sup>, Olivier Scaillet<sup>b,\*</sup>

<sup>a</sup>Research Center for Statistics, Geneva School of Economics and Management, University of Geneva,  
Bd Pont-d'Arve 40, CH-1211 Geneva 4, Switzerland.

<sup>b</sup>Geneva Finance Research Institute, University of Geneva, and Swiss Finance Institute,  
Bd Pont-d'Arve 40, CH-1211 Geneva 4, Switzerland.

---

## Abstract

We develop theory of a novel fast bootstrap for dependent data. Our scheme deploys i.i.d. resampling of smoothed moment indicators. We characterize the class of parametric and semiparametric estimation problems for which the method is valid. We show the asymptotic refinements of the new procedure, proving that it is higher-order correct under mild assumptions on the time series, the estimating functions, and the smoothing kernel. We illustrate the applicability and the advantages of our procedure for M-estimation, generalized method of moments, and generalized empirical likelihood estimation. In a Monte Carlo study, we consider an autoregressive conditional duration model and we compare our method with other extant, routinely-applied first- and higher-order correct methods. The results provide numerical evidence that the novel bootstrap yields higher-order accurate confidence intervals, while remaining computationally lighter than its higher-order correct competitors. A real-data example on dynamics of trading volume of US stocks illustrates the empirical relevance of our method.

*JEL classification:* C12, C15, C22, C52, C58, G12.

*Keywords:* Fast bootstrap methods, Higher-order refinements, Generalized Empirical Likelihood, Confidence distributions, Mixing processes.

---

---

\*Corresponding author

*Email address:* `olivier.scaillet@unige.ch` (Olivier Scaillet)

## 1. Introduction

Inference based on first-order correct asymptotics can be misleading with confidence intervals having erratic probability coverage. It is especially true in the presence of serial dependence where first-order asymptotics often requires larger sample sizes than for i.i.d. data to apply. Resampling methods for time series help to obtain confidence intervals with better finite sample properties. Bootstrap methods for moment condition models have been extensively discussed under various dependence structures by, for example, [Hall and Horowitz \(1996\)](#), [Brown and Newey \(2002\)](#), [Inoue and Shintani \(2006\)](#), and [Davidson and MacKinnon \(2006\)](#). If bootstrap methods for  $m$ -dependent and strongly mixing data can achieve higher-order correctness ([Hall and Horowitz \(1996\)](#), [Inoue and Shintani \(2006\)](#)), they are computationally too intensive, once applied to heavy numerical estimation procedures. For a book-length review, see e.g. [Lahiri \(2010\)](#).

In this paper, we propose a novel fast bootstrap scheme, that we call the Fast Moving-average Bootstrap (FMB). The resampling method is computationally attractive while maintaining higher-order correctness of the inferential procedure for strongly mixing data. Our idea for building confidence regions for the parameter of interest is to realize that smoothing the moment indicators as in the Generalized Empirical Likelihood (GEL) literature permits to bootstrap them as if they were i.i.d. [Parente and Smith \(2018a\)](#) study the first-order validity of GEL test statistics based on a similar bootstrapping scheme, the Kernel Block Bootstrap (henceforth KBB); see [Parente and Smith \(2018b\)](#) and [Parente and Smith \(2021\)](#). Our approach differs from KBB in two significant aspects. First, our methodology does not require to solve the estimation problem at each bootstrap sample, lessening drastically the computational burden. Indeed, FMB is at least (except for a simple low-dimensional linear model) one thousand times faster, according to standard rules on bootstrap simulation errors ([Efron \(1987\)](#) Section 9, [Davison and Hinkley \(1997\)](#) Section 2.5.2). Second, we exploit an inversion technique to benefit from the kernel smoothing used in the studentization of our test statistic. The inversion is related to the standard percentile- $t$  bootstrap approach in the linear univariate case (see Example 1 below). The studentization relies on a simple sample variance of the smoothed moment indicators, which turns out to be asymptotically equivalent to a HAC estimator for the original moment indicators, as shown by [Smith \(2005\)](#). Together these inversion and studentization make our FMB inference amenable to

be shown higher-order correct. Our proof strategy is not directly applicable to KBB; its higher-order correctness remains a conjecture.

The already existing fast resampling methods usually hinge on the first-order von Mises expansion of the estimating function (Shao and Tu (1995), Davidson and MacKinnon (1999), Andrews (2002), Salibian-Barrera and Zamar (2002), Gonçalves and White (2004), Hong and Scaillet (2006), Salibian-Barrera et al. (2006), Salibian-Barrera et al. (2008), Camponovo et al. (2012), Camponovo et al. (2013), Armstrong et al. (2014), and Gonçalves et al. (2019)). It yields a fast approximation, but its inherent construction does not ensure higher-order correctness. Instead, our fast method relies on inversion, namely we identify the level sets of test statistics under the null hypothesis to obtain confidence regions for the parameter of interest (see Parzen et al. (1994) and Hu and Kalbfleisch (2000) for i.i.d. data). Furthermore, the FMB confidence regions are invariant to monotonic reparameterization, due to studentization of the moment indicators. It ensures stability of our method across varying parameter scales (DiCiccio and Efron (1996)).

We design FMB for GEL estimator to exploit its intrinsic smoothing, and as it provides a considerably wide theoretical framework on semiparametric estimation (Smith (2011)). As a consequence, the higher-order refinements achieved by our method ensue for the Empirical Likelihood (see Qin and Lawless (1994), Imbens (1996), Kitamura (1997)), the Exponential Tilting (Kitamura and Stutzer (1997), Imbens et al. (1998)), and the Continuously Updating Estimator (Hansen et al. (1996)). In addition to the KBB, other bootstrap methods already exist in the GEL literature. For instance, Bravo (2004) shows the higher-order correctness of the bootstrap for inference based on empirical likelihood with i.i.d. data, while Bravo (2005) shows consistency of the block bootstrap for empirical entropy tests in times series regressions with strongly mixing data. However, to our knowledge, there is no proof of higher-order correctness of the bootstrap for GEL in the literature yet.

Clearly, we can also apply FMB in the setting of M-estimation (Huber (1964)) and Generalized Method of Moment (Hansen (1982)), obtaining a fast version of the bootstrap methods derived in Hall and Horowitz (1996) for  $m$ -dependent data and Inoue and Shintani (2006) for strongly mixing data.

The structure of the paper is as follows. Section 2.1 is a simple introduction to the FMB algorithm in the univariate case. There, we also discuss connections between FMB and already existing resampling

schemes. In Section 2.2, we briefly present the GMM and GEL estimators for strongly mixing time series, using these frameworks as a tool to extend FMB to the multivariate setting. We itemize our assumptions and present the main theoretical results in Section 3. In Section 4, we give details on the implementation aspects of FMB, emphasizing the relation between the choice of the kernel and the properties of the long-run variance estimator. We also discuss connections with the recent literature on confidence distributions, that we use in our empirical application. We present our Monte Carlo experiments in Section 5, and a real data example in Section 6. We give the proofs in the Supplementary Material (available online).

## 2. FMB methodology

### 2.1. An introduction in the univariate case

Let  $\{X_t\}_{t \in \mathbb{Z}}$  be a stationary strongly mixing process in  $\mathbb{R}^d$ , observed at  $t = 1, \dots, T$ . We assume that the time series of interest satisfies Assumptions 1–7 in Section 3, which are standard in the bootstrap literature. Let  $\mathcal{B} \subset \mathbb{R}$  be the compact space of the parameter  $\beta$  and  $\mathbb{X}_t := \{X_{t_1}, \dots, X_{t_v}\}$  be a collection of vectors from the process  $\{X_t\}_{t \in \mathbb{Z}}$ . Consider the function  $g : \mathbb{R}^{dv} \times \mathcal{B} \rightarrow \mathbb{R}$  such that:

$$\mathbb{E}[g(\mathbb{X}_t, \beta_0)] = 0, \tag{1}$$

where the expectation  $\mathbb{E}$  is taken w.r.t. the true underlying distribution, unknown and depending on  $\beta_0$ . In the following, we use the shorthand notation  $g_t(\beta) := g(\mathbb{X}_t, \beta)$ .

The function  $g$  in (1) can be the (conditional) likelihood in full parametric models, or it can be obtained using the (conditional) moments and/or may depend on instrumental variables in semiparametric models. The collection of vectors  $\mathbb{X}_t$  typically contains information on the relation between the observations and the parameter characterizing the  $q$ -dimensional stationary distribution of a time series. More generally, we can exploit the knowledge in closed-form of the (conditional) moments to obtain (martingale) estimating functions for non-linear conditional autoregressive and heteroscedastic models or discretely observed diffusions. We refer to Godambe and Heyde (1987), Taniguchi and Kakizawa (2000), and Kessler et al. (2012) for book-length presentations.

Each function of the sequence  $\{g_t(\beta)\}_{t=1}^T$  is often defined using the innovations, which can be i.i.d. random variables or more generally martingale differences. Thus,  $\{g_t(\beta)\}_{t=1}^T$  exhibits less dependence than the original process  $\{X_t\}_{t=1}^T$ . Nevertheless, neglecting this temporal dependence has a serious impact on the performance of several inferential procedures, in particular it can affect the consistency of the bootstrap variance estimator and the higher-order accuracy of bootstrap confidence intervals.

To take automatically this aspect into account, we follow [Kitamura and Stutzer \(1997\)](#), [Otsu \(2006\)](#), [Guggenberger and Smith \(2008a\)](#), and [Smith \(2011\)](#), and we perform a convolution of the moment indicator  $g$  with the kernel  $k : \mathbb{R} \rightarrow \mathbb{R}$ , obtaining:

$$g_{T,t}(\beta) := B_T^{-1/2} \sum_{s=t-T}^{t-1} k\left(\frac{s}{B_T}\right) g_{t-s}(\beta), \quad (2)$$

where  $B_T$  is a bandwidth parameter, increasing in  $T$  and such that  $B_T/T \rightarrow 0$ . The convolution in (2) induces a HAC-type modification, ensuring consistency of the long-run variance estimation of the mean over time  $\bar{g}_T(\beta) := T^{-1} \sum_{t=1}^T g_{T,t}(\beta)$ ; see [Newey and West \(1987\)](#), [Andrews \(1991\)](#), and [Smith \(2005\)](#). Solving  $\bar{g}_T(\beta) = 0$  gives the just-identified univariate estimator  $\hat{\beta}$ . Hence, the estimator  $\hat{\beta}$  relies on a smoothed moment condition. Below, we explain how we can further exploit the convolution in (2) to derive our bootstrap.

Let us first give the intuition of our methodology for the construction of confidence interval (CI) for  $\beta_0$ ; more technical aspects are available in Sections 2.2 and 4.1. For ease of notation, we drop the subscript  $T$  from any estimator, whenever its dependence on the sample size is clear from the context.

To keep the exposition as simple as possible, we assume temporarily a one-to-one relationship between the parameter and the estimating function  $\bar{g}_T(\beta)$ , in an suitable subset of  $\mathcal{B}$ . Even though the probabilistic validity of the FMB CI does not depend on this condition (see e.g. [Lehmann \(1959\)](#), [Shao \(1999\)](#), [Hansen et al. \(1996\)](#) and [Guggenberger and Smith \(2008b\)](#)), this assumption allows us to explain easily why our resampling scheme does not need to solve the estimating equation for each bootstrap sample.

Intuitively, the construction of the FMB CI goes as follows. First, our bootstrap scheme provides a higher-order correct approximation of the distribution of a statistic  $\hat{S}(\beta_0)$ . This statistic is an asymptotically pivotal version of the estimating function  $T^{1/2}\bar{g}_T(\beta)$ , evaluated at the true parameter  $\beta_0$ . Second,

the one-to-one relationship allows us to map the quantile estimates of  $\hat{S}$  to CI limits in  $\mathcal{B}$ . This mapping is crucial to gain computational efficiency. Indeed, we use the computationally intensive part of the FMB algorithm to compute the distribution of simple mean-type statistic  $\hat{S}(\beta_0)$  which is much faster to compute than roots of  $T^{1/2}\bar{g}_T(\beta)$ , or numerical solutions to the estimating optimization problem (see Section 2.2). From an hypothesis testing point of view, FMB yields an approximation to the distribution of  $\hat{S}(\beta)$  under  $H_0 : \beta = \beta_0$ . Then, each  $\beta$  in the CI is in the non-rejection region of  $H_0$ .

The next example is a widely-applied model where the one-to-one condition is satisfied, since the considered function is monotonic. Below, we explain in Remark 2 how to adapt FMB to deal with general estimating functions, which do not necessarily satisfy the monotonicity condition.

**Example 1.** We consider an AR(1) process  $\{Y_t\}_{t=1}^T$ ,  $Y_t = \beta Y_{t-1} + \varepsilon_t$ , where  $|\beta| < 1$ , and  $\{\varepsilon_t\}_{t=1}^T$  is a white noise. The orthogonality of the innovations yields the moment indicators  $g_t(\beta) = (Y_t - \beta Y_{t-1})Y_{t-1}$ . For a given kernel  $k$ , smoothing these moment indicators leads to

$$g_{T,t}(\beta) = B_T^{-1/2} \sum_{s=t-T}^{t-1} k\left(\frac{s}{B_T}\right) Y_{t-s} Y_{t-s-1} - \beta B_T^{-1/2} \sum_{s=t-T}^{t-1} k\left(\frac{s}{B_T}\right) Y_{t-s-1}^2. \quad (3)$$

Equation (3) is linear in the parameter of interest  $\beta$ . Thus, neither taking the mean  $T^{1/2}\bar{g}_T(\beta)$  nor rescaling  $T^{1/2}\bar{g}_T(\beta)$  by a constant affect this linearity. As we build our statistic of interest by rescaling  $T^{1/2}\bar{g}_T(\beta)$ , the one-to-one condition is verified.

Now that the main principles of FMB are settled, we present the detailed algorithm underlying its numerical implementation. The statistic serving as basis for inference is the asymptotically pivotal quantity:

$$\hat{S} := T^{1/2}\bar{g}_T(\beta_0) / \hat{\sigma}, \quad (4)$$

where, for instance,  $\hat{\sigma} := \kappa_1^2(T\kappa_2)^{-1} \sum_{t=1}^T g_{T,t}^2(\hat{\beta})$  and  $\kappa_j := \int k(u)^j du$  for  $j = 1, 2$ . This statistic is a particular value of the function  $\hat{S} : \mathbb{R}^{d_v} \times \mathcal{B} \rightarrow \mathbb{R}$ ,  $\hat{S}(\beta) := T^{1/2}\bar{g}_T(\beta) / \hat{\sigma}$ , that we suppose strictly increasing on  $\mathcal{B}$ . The studentization in  $\hat{S}$  is crucial for FMB to be higher-order correct. In principle, we can apply other estimators of the long-run variance and we flag that each estimator  $\hat{\sigma}$  has its own bias, which is going to affect the properties (e.g. the accuracy) of FMB. We refer to Section 4.1 for further discussion.

Considering an i.i.d. bootstrap sample drawn from  $\{g_{T,t}(\hat{\beta})\}_{t=1}^T$ , say  $\{g_{T,t}^*\}_{t=1}^T$ , the bootstrap version of  $\hat{S}$  in (4) is:

$$S^* := T^{1/2} \bar{g}_T^* / \hat{\sigma}^*, \quad (5)$$

with  $\bar{g}_T^* := T^{-1} \sum_{t=1}^T g_{T,t}^*$  and  $\hat{\sigma}^{*2} := T^{-1} \sum_{t=1}^T g_{T,t}^{*2}$ . In (5), both the computed numerator and denominator rely on  $g_{T,t}^*$ , and thus avoid re-estimating the parameter on each bootstrap sample in order to make it fast.

Then, the algorithm of our FMB is made of five steps (lines 2-4, 5, 6-13, 14-15, 16-17).

**Algorithm 1.**

```

1: procedure FMB( $X_1, \dots, X_T, g, \alpha$ )      ▷ The FMB  $(1 - \alpha)$ -CI for  $\beta_0 \in \mathcal{B}$  s.t.  $\mathbb{E}[g(\mathbb{X}_t, \beta_0)] = 0$ .
2:   for  $t = 1, \dots, T$  do
3:      $g_{T,t}(\beta) \leftarrow B_T^{-1/2} \sum_{s=t-T}^{t-1} k(s/B_T) g_{t-s}(\beta)$ 
4:   end for
5:   Solve( $\sum_{t=1}^T g_{T,t}(\beta) = 0$ )  $\rightarrow \hat{\beta}$ 
6:   for  $r = 1, \dots, R$  do
7:     for  $t = 1, \dots, T$  do
8:        $g_{T,t}^* \leftarrow \mathbf{Draw}(g_{T,1}(\hat{\beta}), \dots, g_{T,T}(\hat{\beta}))$ 
9:     end for
10:     $\bar{g}_{T,r}^* \leftarrow T^{-1} \sum_{t=1}^T g_{T,t}^*$ 
11:     $\hat{\sigma}_r^* \leftarrow (T^{-1} \sum_{t=1}^T g_{T,t}^{*2})^{1/2}$ 
12:     $S_r^* \leftarrow T^{1/2} \bar{g}_{T,r}^* \hat{\sigma}_r^{*-1}$ 
13:  end for
14:   $\mathit{ConfidenceLevel} \leftarrow 1 - \alpha$ 
15:   $q^* \leftarrow \mathbf{Quantile}(S_1^*, \dots, S_R^*, \mathit{ConfidenceLevel})$ 
16:  Solve( $\hat{S}(\beta) = q^*$ )  $\rightarrow \mathit{UpperLimit}$ 
17:  return  $\mathit{UpperLimit}$       ▷ The upper limit of the one-sided FMB CI
18: end procedure

```

In Algorithm 1, we exploit the monotonicity assumption only in Step 5, where we invert the stu-

dentized estimating function  $\hat{S}(\beta)$ ; see [Hu and Kalbfleisch \(2000\)](#) for the use of a similar device.

Indeed, to derive the CI, the bootstrap procedure first provides  $(1 - \alpha)$ -quantile estimates of  $\hat{S}(\beta_0)$ , say  $q_{1-\alpha}^*$ . Then, a numerical method (e.g. Newton-Raphson or secant methods) defines a one-sided  $(1 - \alpha)$ -CI for  $\beta_0$  as  $[\beta_{\min}, \hat{q}_{1-\alpha}]$ , where  $\beta_{\min} := \min \mathcal{B}$  and the upper limit  $\hat{q}_{1-\alpha}$  solves  $\hat{S}(\hat{q}_{1-\alpha}) = q_{1-\alpha}^*$  in  $q_{1-\alpha}$ .

For a two-sided equal-tailed CI, we follow the same principles. We consider two real numbers  $s_1$  and  $s_2$  such that  $\mathbb{P}[\hat{S}(\beta_0) \leq s_1] = \alpha/2$  and  $\mathbb{P}[\hat{S}(\beta_0) > s_2] = \alpha/2$ . From FMB, we obtain the approximation<sup>1</sup>  $\mathbb{P}^*[s_1 < S^* \leq s_2] = \mathbb{P}[s_1 < \hat{S}(\beta_0) \leq s_2] + R_T$ , where  $R_T$  is an asymptotically negligible remainder term. Hence, we can compute  $s_1$  and  $s_2$  such that  $\mathbb{P}^*[s_1 < S^* \leq s_2] = 1 - \alpha$ . Then, the CI for  $\beta_0$  is  $\mathcal{C}_{1-\alpha} := (c_1, c_2]$ , with  $c_1 := \hat{S}^{-1}(s_1)$  and  $c_2 := \hat{S}^{-1}(s_2)$ , ensuring that  $\mathbb{P}[c_1 < \beta_0 \leq c_2] = 1 - \alpha + R_T$ . Under Assumptions 1–7 (Section 3), we can get  $R_T = o_p(T^{-1/2})$ , given a suitable choice of kernel  $k$  and bandwidth  $B_T$  (see Theorem 4 and discussion in Section 3). It implies that  $\mathcal{C}_{1-\alpha}$  is correct up to a higher order.

From the studentization in  $\hat{S}(\beta)$ , the CI limits  $c_1$  and  $c_2$  derived in Step 5 remain invariant to monotonic transformation of the parameter. This property is crucial for the bootstrap CI ([DiCiccio and Efron \(1996\)](#)), ensuring stability of FMB across varying parameter scale.

**Example 1 [cont'd].** *Let us see how the steps in Algorithm 1 specialize for the AR(1). In Step 1, we have  $g_{T,t}$  as in (3) and  $\bar{g}_T(\beta) = T^{-1} \sum_{t=1}^T B_T^{-1/2} \sum_{s=t-T}^{t-1} k(s/B_T)(Y_{t-s} - \beta Y_{t-s-1})Y_{t-s-1}$ . For Step 2, the estimator is available in closed form:*

$$\hat{\beta} = \left( \sum_{t=1}^T \sum_{s=t-T}^{t-1} k\left(\frac{s}{B_T}\right) Y_{t-s} Y_{t-s-1} \right) \left( \sum_{t=1}^T \sum_{s=t-T}^{t-1} k\left(\frac{s}{B_T}\right) Y_{t-s-1}^2 \right)^{-1}.$$

*For Step 3 and Step 4, we define  $S^*$  using (5), and we use i.i.d. resampling of  $\{g_{T,t}(\hat{\beta})\}_{t=1}^T$ . Since  $\hat{S}(\beta)$  is strictly decreasing, we switch the sign of  $S^*$  and  $\hat{S}(\beta)$  and proceed as in Step 5 of Algorithm 1 to build a one-sided CI for  $\beta_0$ . In this particular case of linear models, we can rewrite  $\hat{S}(\beta_0)$  as  $T^{1/2}(\hat{\beta} - \beta_0)/\hat{\varsigma}$ ,*

---

<sup>1</sup>As customary in the bootstrap literature,  $\mathbb{P}^*[X^* \leq x]$  denotes the empirical c.d.f. of any variable  $X^*$  generated by the bootstrap scheme (here  $S^*$ ). We give the general definition of the bootstrap probability measure  $\mathbb{P}^*$  in (14), Section 3.



where  $\hat{\varsigma} = \hat{W}^{-1}\hat{\sigma}$  and  $\hat{W} := T^{-1} \sum_{t=1}^T \partial g_{T,t}(\hat{\beta})/\partial \beta$ . Similarly, we can rewrite the bootstrap counterpart  $S^*$  as  $T^{1/2}(\beta^* - \hat{\beta})/\varsigma^*$ , where  $\beta^*$  is a bootstrap estimate,  $\varsigma^* = W^{*-1}\sigma^*$  and  $W^* := T^{-1} \sum_{t=1}^T \partial g_{T,t}^*(\hat{\beta})/\partial \beta$ . Then, the FMB CI is equivalent to  $[\hat{\beta} - q_{1-\alpha/2}^* \hat{\sigma}, \hat{\beta} - q_{\alpha/2}^* \hat{\sigma}]$ , where  $q^*$  are quantiles of  $S^*$ . In this representation, the FMB CI is similar to a percentile- $t$  bootstrap CI, up to our use of  $\hat{\beta}$  instead of  $\beta^*$  in the definition of  $\varsigma^*$ , avoiding to compute  $\beta^*$ , and in the multivariate case, to invert a matrix, for each bootstrap sample. This modification does not impact higher-order correctness, as shown in Section 3.

**Remark 2** (Step 5, Lines 16-17). When the function  $\hat{S}(\beta)$  is not monotonic, we can slightly modify the procedure if we want to obtain a simply connected CI, as opposed to a union of intervals. To this end, let us define  $\hat{Q}(\beta) := \hat{S}(\beta)^2$ . As an alternative statistic, we take the third-order Taylor expansion of  $\hat{Q}(\beta)$  around the root- $T$  consistent estimator  $\hat{\beta}$ . Namely, we define  $\tilde{Q}(\beta) := (\partial^2 \hat{Q}(\hat{\beta})/\partial \beta^2)(\beta - \hat{\beta})^2/2 + (\partial^3 \hat{Q}(\hat{\beta})/\partial \beta^3)(\beta - \hat{\beta})^3/6$ , the first two terms being zero. If we make use of  $\tilde{Q}(\beta)$  in a neighborhood  $\{\tilde{\beta} \in \mathcal{B} : \tilde{\beta} = \beta_0 + \bar{\delta}T^{-1/2}\}$ , for  $\bar{\delta} \in \mathbb{R}$  (Newey and McFadden (1994)), we can show that  $\tilde{Q}(\beta) = \hat{Q}(\beta) + O_p(T^{-1})$  in that neighborhood. Hence, FMB allows us to approximate  $\mathbb{P}[\tilde{Q}(\beta_0) \leq q_{1-\alpha}^*]$  by  $\mathbb{P}^*[S^{*2} \leq q_{1-\alpha}^*]$  with higher-order accuracy, as shown in Corollary 5 (Section 3). Then, we compute  $\mathcal{C}_{1-\alpha} := \{\beta \in \mathcal{B} : \tilde{Q}(\beta) \leq q_{1-\alpha}^*\}$  to get the desired higher-order correct  $(1-\alpha)$ -CI. From Section 9.1 in Newey and McFadden (1994), it should be clear that  $\bar{\delta}$  is not a tuning parameter to be chosen to apply FMB.

This modified FMB CI is simply connected with high probability when  $T$  is large enough. Indeed, we can show that  $\tilde{q} = \frac{4T}{27\sigma^2} \left[ \frac{\partial \bar{g}_T(\hat{\beta})^2}{\partial \beta} / \frac{\partial^2 \bar{g}_T(\hat{\beta})}{\partial \beta^2} \right]^2$  is the highest value such that the sublevel set  $\{\beta \in \mathcal{B} : \tilde{Q}(\beta) \leq \tilde{q}\}$  is still simply connected. It corresponds to the local maximum of a cubic polynomial (we provide a graphical illustration in Figure 2, Supplementary Material SM.9). Thus, the range of confidence level from which we can draw a simply connected set increases proportionally to the sample size  $T$ . In practice, we recommend to try using the  $\tilde{Q}$  approximation to guarantee the second-order correctness of the CI and connected CI with high probability. If this approach yields a disconnected CI because of a too small sample size  $T$ , the user can truncate the Taylor approximation at the quadratic term, which gives a simply connected CI w.p.1. This quadratic approximation does not guarantee higher-order correctness, but is prone to work better than the Gaussian approximation in practice. To summarise,

we face three possibilities. If we use  $\hat{Q}$ , we always get higher-order correctness, but not necessarily connected CI when monotonicity is not satisfied. If we use the cubic approximation  $\tilde{Q}$ , we get higher-order correctness and connected CI with high probability. If we use a quadratic truncation, while being asymptotically correct, higher-order correctness might be lost, but we ensure connected CI.

Therefore, we conclude that, even if the monotonicity condition is violated (or is not easy to check), we can choose a convenient statistic based on a cubic approximation and preserving the asymptotic properties of FMB. We point out that the proposed derivation of the CI only involves the (potentially numerical) computation of  $\hat{Q}'$ 's derivatives, whose evaluation is required only at the single point  $\hat{\beta}$ . As the shape of the CI is fully determined by these derivatives, it simplifies and speeds up the implementation of Step 5.

Some further remarks on the other steps Algorithm 1 are in order. First, Step 1 — Step 3 (Lines 2-13) hinge on bootstrapping the moment indicator evaluated at  $\hat{\beta}$ . It justifies the adjective “fast” in the name of our resampling scheme, and bears some similarities to the already existing fast bootstrap (henceforth FB) methods; see [Shao and Tu \(1995\)](#), [Davidson and MacKinnon \(1999\)](#), [Andrews \(2002\)](#), [Salibian-Barrera and Zamar \(2002\)](#), [Gonçalves and White \(2004\)](#), [Salibian-Barrera et al. \(2006\)](#), [Salibian-Barrera et al. \(2008\)](#), [Camponovo et al. \(2013\)](#), [Armstrong et al. \(2014\)](#), [Gonçalves et al. \(2019\)](#)), and to the estimating function bootstrap ([Parzen et al. \(1994\)](#), [Hu and Kalbfleisch \(2000\)](#)). However, FB methods typically rely on a first order von Mises expansion, which approximation error prevents the FB to be higher-order accurate.

Second, Step 3 (Line 12) computes the bootstrap statistic  $S^*$ , where the kernel  $k$  creates a block of moment indicators evaluated at  $\hat{\beta}$ . The block of  $g_t$  induced by the kernel is similar to a moving-average, as we emphasize in the name of our resampling scheme. The Moving Block Bootstrap (henceforth MBB) is the state-of-the-art higher-order correct alternative to FMB ([Götze and Künsch \(1996\)](#), [Lahiri \(1996\)](#)). For MBB, the blocks are defined at the level of the observations, whereas in our case the convolution is applied to the moment indicators.

In the same family of groupwise resampling schemes, FMB is even more reminiscent of the Tapered Block Bootstrap of [Paparoditis and Politis \(2001\)](#) (hereafter TBB), in the sense that we can view their tapered block as our moving-average kernel. The main difference is that our kernel has unbounded

support, in contradistinction with their block tapering window. It gives FMB an advantage in the studentization: it allows us to use the Quadratic Spectral (QS) kernel, which is optimal in terms of asymptotic mean squared error according to [Andrews \(1991\)](#). [Parente and Smith \(2018b\)](#) have already pointed out such an advantage for a KBB variance estimator. Yet, the TBB and the KBB approach of [Parente and Smith \(2018a\)](#) both require  $R$  bootstrap estimations. Thus, neither the TBB nor the KBB is fast and there is no result on their potential higher-order correctness. The higher-order correctness of the FMB approximation to the distribution of  $\hat{S}$  comes from jointly considering two ingredients: the smoothing of moment indicators and the studentization.<sup>2</sup> Taken in isolation, each ingredient does not allow to show higher-order correctness of the FMB CI.

To summarize, we itemize in [Table 1](#) the main features of the discussed bootstrap schemes. We only list methodologies that are specifically designed for dependent data.

Table 1: Properties of related bootstrap schemes for dependent data.

		HOC	
		Yes	No
Fast	Yes	FMB	FB
	No	MBB	TBB, KBB

We distinguish the Fast Moving-average Bootstrap (FMB), the Fast Bootstrap (FB), the Moving Block Bootstrap (MBB), the Tapered Block Bootstrap (TBB), and the Kernel Block Bootstrap (KBB) with respect to two features, namely computational speed (Fast) and higher-order correctness (HOC).

## 2.2. Over-identified case with multivariate parameter

In this section, we explain how FMB can yield higher-order correct inference on a multivariate parameter. In [Subsection 2.2.1](#), we consider the Generalized Method of Moments. Then, we extend the setting to Generalized Empirical Likelihood Estimation in [Subsection 2.2.2](#). The asymptotic refinements ([Section 3](#)) also hold for the standard GMM case and are not tied to the use of GEL.

---

<sup>2</sup>The higher-order correctness of the FMB CI comes from the higher-order correctness of the latter FMB distribution and from the inversion step (Step 5 with the monotonicity condition or the cubic approximation of [Remark 2](#)).

### 2.2.1. Generalized Method of Moments

Assume we have to conduct inference on the multivariate parameter  $\beta \in \mathcal{B} \subset \mathbb{R}^p$ , where  $\mathcal{B}$  is compact. We are given a random sample of  $\mathbb{X}_t = \{X_{t_1}, \dots, X_{t_v}\}$  observed at  $t = 1, \dots, T$ , and we define a set of moment conditions  $g : \mathbb{R}^{dv} \times \mathcal{B} \rightarrow \mathbb{R}^r$ , with  $r \geq p$ , such that  $\mathbb{E}[g(\mathbb{X}_t, \beta_0)] = 0$ . To handle the serial dependence, we define the smoothed moment indicator  $g_{T,t}(\beta)$  as in the univariate case

$$g_{T,t}(\beta) := B_T^{-1/2} \sum_{s=t-T}^{t-1} k\left(\frac{s}{B_T}\right) g_{t-s}(\beta),$$

and  $\bar{g}_T(\beta) := T^{-1} \sum_{t=1}^T g_{T,t}(\beta)$ . Estimating  $\beta_0$  via Generalized Method of Moments ([Hansen \(1982\)](#)) is the most popular approach in econometrics. In the next subsection, we discuss alternative estimators.

Step 1 — Step 3 of the FMB Algorithm 1 remain conceptually unchanged. As far as the bootstrap statistic is concerned, the principles of Step 4 and Step 5 stay the same as in Algorithm 1, the main change being that the asymptotically pivotal statistic becomes:

$$\hat{Q} := T \bar{g}_T(\beta_0)^\top \hat{\Omega}^{-1} \bar{g}_T(\beta_0), \quad (6)$$

where  $\hat{\Omega} = \kappa_1^2 (\kappa_2 T)^{-1} \sum_{t=1}^T \{g_{T,t}(\hat{\beta}) - \bar{g}_T(\hat{\beta})\} \{g_{T,t}(\hat{\beta}) - \bar{g}_T(\hat{\beta})\}^\top$  is a consistent estimator of the long-run covariance matrix of  $T^{1/2} \bar{g}_T(\beta_0)$ , of rank  $\nu = r$ ;<sup>3</sup> see Section 4.1. Standard results guarantee that  $\hat{Q}$  is asymptotically  $\mathcal{X}_\nu^2$ . As in the univariate case, the statistic of interest is a particular value of a function, here  $\hat{Q} : \mathbb{R}^{dv} \times \mathcal{B} \rightarrow \mathbb{R}$ ,

$$\hat{Q}(\beta) := T \bar{g}_T(\beta)^\top \hat{\Omega}^{-1} \bar{g}_T(\beta). \quad (7)$$

We define the GMM estimator as  $\hat{\beta} = \operatorname{argmin}_{\beta \in \mathcal{B}} \hat{Q}(\beta)$ . Similarly to the argument of Remark 2, we also define the cubic approximation centered on the root- $T$  consistent estimator  $\hat{\beta}$ :

$$\tilde{Q}(\beta) := \hat{Q}(\hat{\beta}) + (\beta - \hat{\beta})^\top \hat{H}(\beta - \hat{\beta})/2 + \left( (\beta - \hat{\beta}) \otimes (\beta - \hat{\beta}) \right)^\top \frac{\partial \operatorname{vec}(\hat{H})}{\partial \beta^\top} (\beta - \hat{\beta})/6, \quad (8)$$

where the matrix  $\hat{H} := \partial^2 \hat{Q}(\hat{\beta}) / \partial \beta^\top \partial \beta$ . It allows us to build simply connected level sets, yielding

---

<sup>3</sup>If the rank is lower than  $r$ , the covariance matrix is not invertible anymore and we resort to the generalized inverse, adapting the degrees of freedom of the  $\mathcal{X}^2$  distribution accordingly ([Moore \(1977\)](#)).

higher-order correct Confidence Region (henceforth CR), as shown in Corollary 5 (Section 3).

The bootstrap version of  $\hat{Q}$  is

$$Q^* := T^{-1} \sum_{t=1}^T \left\{ g_{T,t}^* - \bar{g}_T(\hat{\beta}) \right\}^\top \hat{\Omega}^{*-1} \sum_{t=1}^T \left\{ g_{T,t}^* - \bar{g}_T(\hat{\beta}) \right\}, \quad (9)$$

where  $\hat{\Omega}^* := T^{-1} \sum_{t=1}^T \left\{ g_{T,t}^* - \bar{g}_T(\hat{\beta}) \right\} \left\{ g_{T,t}^* - \bar{g}_T(\hat{\beta}) \right\}^\top$ , with the asterisk denoting the same i.i.d. resampling scheme as in Algorithm 1. Now we are ready to state the algorithm of our FMB in the over-identified case, made of five steps (lines 2-4, 5, 6-13, 14-15, 16-17).

**Algorithm 2.**

- 1: **procedure** FMB( $X_1, \dots, X_T, g, \alpha$ ) ▷ The FMB  $(1 - \alpha)$ -CR for  $\beta_0 \in \mathcal{B}$  s.t.  $\mathbb{E}[g(\mathbb{X}_t, \beta_0)] = 0$ .
- 2:   **for**  $t = 1, \dots, T$  **do**
- 3:      $g_{T,t}(\beta) \leftarrow B_T^{-1/2} \sum_{s=t-T}^{t-1} k(s/B_T) g_{t-s}(\beta)$
- 4:   **end for**
- 5:   **Argmin** $_{\beta \in \mathcal{B}} \hat{Q}(\beta) \rightarrow \hat{\beta}$  ▷ Using the function  $\hat{Q}(\beta)$  as in (7).
- 6:   **for**  $r = 1, \dots, R$  **do**
- 7:     **for**  $t = 1, \dots, T$  **do**
- 8:       $g_{T,t}^* \leftarrow \text{Draw}(g_{T,1}(\hat{\beta}), \dots, g_{T,T}(\hat{\beta}))$
- 9:     **end for**
- 10:      $\bar{g}_{T,r}^* \leftarrow T^{-1} \sum_{t=1}^T \left\{ g_{T,t}^* - \bar{g}_T(\hat{\beta}) \right\}$
- 11:      $\hat{\Omega}_r^* \leftarrow T^{-1} \sum_{t=1}^T \left\{ g_{T,t}^* - \bar{g}_T(\hat{\beta}) \right\} \left\{ g_{T,t}^* - \bar{g}_T(\hat{\beta}) \right\}^\top$
- 12:      $Q_r^* := T \bar{g}_{T,r}^{*\top} \hat{\Omega}_r^{*-1} \bar{g}_{T,r}^*$
- 13:   **end for**
- 14:    $\text{ConfidenceLevel} \leftarrow 1 - \alpha$
- 15:    $q^* \leftarrow \text{Quantile}(Q_1^*, \dots, Q_R^*, \text{ConfidenceLevel})$
- 16:    $\mathcal{C} \leftarrow \left\{ \beta \in \mathcal{B} : \hat{Q}(\beta) \leq q^* \right\}$
- 17:   **return**  $\mathcal{C}$
- 18: **end procedure**

A few remarks are in order. Step 3 uses (9), where we recenter the bootstrap statistic. Indeed,

the bootstrap expectation  $\mathbb{E}^* [g_{T,t}^*] = \bar{g}_T(\hat{\beta}) \neq 0$  in the over-identified case. Thus, we subtract its expectation from  $g_{T,t}^*$  to recenter the bootstrap variable. This operation is crucial to achieve higher-order accurate CR in Step 5 of Algorithm 2 (see e.g. Hall and Horowitz (1996)).

Moreover, in contradistinction with the already existing FB methods, Step 3—4 mimic the variability of the covariance estimator  $\hat{\Omega}$  (in (6)) to achieve higher-order refinements. To this end, we use  $\hat{\Omega}^*$  instead of  $\hat{\Omega}$  in the definition of  $Q^*$  (in (9)), such that we randomize the bootstrap covariance estimator across the different bootstrap samples, and do not keep it fixed at  $\hat{\Omega}$ . Similar comment applies to Algorithm 1.

Finally, to define the CR for  $\beta_0$ , we proceed similarly to Step 5 of Algorithm 1. We set  $q_{1-\alpha}^*$  such that  $\mathbb{P}^*[Q^* \leq q_{1-\alpha}^*] = 1 - \alpha$  and compute the CR as the subset  $\mathcal{C}_{1-\alpha} := \{\beta \in \mathcal{B} : \hat{Q}(\beta) \leq q_{1-\alpha}^*\}$ . Thus, we get  $\mathbb{P}[\beta_0 \in \mathcal{C}_{1-\alpha}] = \mathbb{P}[\hat{Q}(\beta_0) \leq q_{1-\alpha}^*] = \mathbb{P}^*[Q^* \leq q_{1-\alpha}^*] + R_T = 1 - \alpha + R_T$ . Under Assumptions 1—7 (Section 3), we show in Theorem 4 that the remainder  $R_T$  can be at most of order  $o_p(T^{-1/2})$ , given a suitable choice of kernel  $k$  and bandwidth  $B_T$ , which implies that  $\mathcal{C}_{1-\alpha}$  is correct up to a higher order.

**Remark 3** (Step 5, Lines 16-17). *If the higher-order correctness of FMB is guaranteed independently of the CR shape, they are generally not elliptical and they might fail to be simply connected (they can come in several pieces or contain holes). Although this irregularity is customary for small to moderate sample sizes, it can still make the results difficult to interpret. The monotonicity condition aforementioned is one of the possible ways out. As a more general solution, we propose to use the cubic approximation  $\tilde{Q}$  (as in (8)). Similarly to Remark 2, we have  $\tilde{Q}(\beta) = \hat{Q}(\beta) + O_p(T^{-1})$  in a neighborhood  $\{\check{\beta} \in \mathcal{B} : \check{\beta} = \beta_0 + \bar{\delta}T^{-1/2}\}$ , for  $\bar{\delta} \in \mathbb{R}^p$ . Thus, defining the modified FMB CR as  $\mathcal{C}_{1-\alpha} := \{\beta \in \mathcal{B} : \tilde{Q}(\beta) \leq q_{1-\alpha}^*\}$  preserves higher-order correctness, as shown in Corollary 5. The resulting CR is not simply connected for all sample sizes, but we can show that the range of confidence level  $(1-\alpha)$  leading to a simply connected CR is proportional to the sample size  $T$ . It is not an asymptotic property and the desired CR can already be simply connected for a small sample size, depending on the second derivative of the moment indicator  $g$ . If the sample size is too small for this modified CR to be simply connected and the user thoroughly needs this property, we recommend to truncate the  $\tilde{Q}$  approximation at the second term. The resulting CR is always simply connected and elliptical, but, while being asymptotically correct, we cannot guarantee its higher-order properties.*

### 2.2.2. Generalized Empirical Likelihood Estimation

FMB is not tied to a particular estimation method. In order to show its wide applicability, we consider a general estimation method, which includes (among others) a version of GMM. The Continuously Updating Estimator (CUE) (Hansen et al. (1996)), Empirical Likelihood (EL) (Qin and Lawless (1994), Imbens (1996), Kitamura (1997)), and the Exponential Tilting (ET) (Kitamura and Stutzer (1997), Imbens et al. (1998)) are all asymptotically equivalent to the (2S)GMM, but they tend to be less biased for small to moderate sample sizes (see Altonji and Segal (1996) for a Monte Carlo exploration, and Newey and Smith (2004), Anatolyev (2005) for theoretical insights). Putting EL, ET, and CUE under the same umbrella, Smith (2011) introduces the Generalized Empirical Likelihood (GEL) criterion for time series data. We briefly describe it before extending our FMB to this general setting.

Let  $\rho(\nu)$  be a concave function on an open interval  $\mathcal{V} \in \mathbb{R}$  containing 0. Writing  $\rho_\nu(\nu) := \partial^\nu \rho(\nu) / \partial \nu^\nu$  and  $\rho_\nu = \rho_\nu(0)$  for  $\nu = 0, 1$ , the function  $\rho(\nu)$  is standardized such that  $\rho_1 = -1$ . Defining a vector of auxiliary parameters  $\lambda \in \Lambda_T(\beta)$  with  $\Lambda_T(\beta) := \left\{ \lambda \in \mathbb{R}^r : \kappa B_T^{-1/2} \lambda^\top g_{T,t}(\beta) \in \mathcal{V} \right\}$ , and  $\kappa := \kappa_1 / \kappa_2$ , Smith (2011) defines the GEL criterion as:

$$\hat{P}(\beta, \lambda) = T^{-1} \sum_{t=1}^T \left[ \rho \left( \kappa B_T^{-1/2} \lambda^\top g_{T,t}(\beta) \right) - \rho_0 \right]. \quad (10)$$

To derive an estimator of  $\beta$ , we first optimize criterion (10) w.r.t.  $\lambda$  for a given  $\beta$ , so that  $\lambda(\beta) = \operatorname{argsup}_{\lambda \in \Lambda_T(\beta)} \hat{P}(\beta, \lambda)$ . Then, we define  $\hat{\beta}$  as the solution to  $\operatorname{argmin}_{\beta \in \mathcal{B}} \hat{P}(\beta, \lambda(\beta))$ .

Similarly to Khundi and Rilstone (2012) and Lee (2016), we are going to use the first order condition of the GEL criterion as a just-identified representation of the estimation problem to explain how to build the FMB statistic. Differentiating (10) w.r.t. to  $\lambda$  and  $\beta$ , we obtain:

$$T^{-1} \sum_{t=1}^T \rho_1 \left( \kappa B_T^{-1/2} \lambda(\beta)^\top g_{T,t}(\beta) \right) B_T^{-1/2} g_{T,t}(\beta) = 0, \quad (11)$$

$$T^{-1} \sum_{t=1}^T \rho_1 \left( \kappa B_T^{-1/2} \lambda(\beta)^\top g_{T,t}(\beta) \right) B_T^{-1/2} \frac{\partial g_{T,t}(\beta)^\top}{\partial \beta} \lambda(\beta) = 0. \quad (12)$$

We can see from (11) that  $\rho_1 \left( \kappa B_T^{-1/2} \lambda(\beta)^\top g_{T,t}(\beta) \right)$  gives weights to the observations such that the moment conditions in  $g$  are always enforced in a given sample. GEL estimators are equivalent to some

minimum discrepancy estimators based on the power-divergence family (see [Cressie and Read \(1984\)](#)), where the auxiliary vector parameter  $\lambda$  corresponds to the Lagrange multiplier enforcing this empirical moment condition. Thus, if the original moment conditions in  $g$  are correctly specified, the true (long-run) Lagrange multiplier is zero ( $\lambda_0 = 0$ ). Applying our FMB to this setting only requires to define a quadratic statistic from the GEL first order conditions ((11) and (12)) evaluated at the true value of the parameters. Yet, replacing  $\lambda = \lambda_0 = 0$  and  $\beta = \beta_0$  in (11) and (12) boils down to the original  $\bar{g}_T(\beta_0)$ . Thus, the natural extension of FMB for GEL estimators requires to take the asymptotically pivotal statistic  $\hat{Q}(\beta_0)$  as in the GMM case (6). As a consequence, FMB in the GEL setting is exactly the same as in the GMM setting (see Algorithm 2), up to the initial estimator  $\hat{\beta}$ . It is quite intuitive since, in absence of misspecification, the first order conditions of the GEL criterion convey the same information on  $\beta_0$  as the moment condition  $g$ .

### 3. Theory

In the next theorems, we state that FMB CI and CR are higher-order correct. By construction, the higher-order correctness of the FMB CI and CR entirely hinges on our bootstrap approximation of the distribution for the test statistics  $\hat{S}(\beta_0)$  (as in (4)) and  $\hat{Q}(\beta_0)$  (as in (6)).

We start by itemizing the assumptions and regularity conditions. In the following, we keep using the shorthand notations  $g_t = g_t(\beta_0)$  and  $g_{T,t} = g_{T,t}(\beta_0)$ . For any vector  $V \in \mathbb{R}^n$ , we write  $\|V\| = (v_1^2 + \dots + v_n^2)^{1/2}$ , where  $v_j$  is the  $j$ -th element of  $V$ . We make use of generic constants  $C$ ,  $\delta$  and  $\epsilon$ , whose value can differ from an expression to another. We define  $\{g_t\}_{t \in \mathbb{Z}}$  on the probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Let  $\{\mathcal{D}_t\}_{t \in \mathbb{Z}}$  be a given sequence of sub-sigma-fields of  $\mathcal{A}$ , and  $\mathcal{D}_a^b = \sigma\{\{\mathcal{D}_j : a \leq j \leq b\}\}$ . A straightforward example is to take  $\mathcal{D}_t := \sigma\{g_t\}$ , but it is not always the most efficient choice to check the assumptions below (see [Götze and Hipp \(1983\)](#) and [Götze and Hipp \(1994\)](#) for practical examples). The higher-order correctness of FMB is subject to the following conditions ([Götze and Künsch \(1996\)](#), [Lahiri \(2010\)](#)), which we assume to hold for  $\{g_t\}_{t \in \mathbb{Z}}$ :

**Assumption 1.**  $\mathbb{E}[g_t(\beta)] = 0, t = 1, 2, \dots$  only for  $\beta = \beta_0$  on the compact parameter space  $\mathcal{B}$ . Moreover,  $\hat{\beta} \xrightarrow{a.s.} \beta_0$  and  $\|\hat{\beta} - \beta_0\| = O_p(T^{-1/2})$  as  $T \rightarrow \infty$ .

**Assumption 2.**  $\mathbb{E}\left[\|g_t\|^{s+\delta}\right] < \infty$  for a positive  $s \geq 8$ ,  $t = 1, 2, \dots$ , and  $\delta > 0$ .



**Assumption 3.** *There exists a constant  $\delta > 0$  such that for  $t, m = 1, 2, \dots$  and  $m > \delta^{-1}$ , we can approximate  $g_t$  by a  $\mathcal{D}_{t-m}^{t+m}$ -measurable random vector  $g_{t,m}^\dagger$ , such that  $\mathbb{E} \left\| g_t - g_{t,m}^\dagger \right\| \leq \delta^{-1} \exp(-\delta m)$ .*

**Assumption 4.** *There exists a constant  $\delta > 0$  such that  $\alpha(m) = \sup |\mathbb{P}[A \cap B] - \mathbb{P}[A] \mathbb{P}[B]| \leq \delta^{-1} \exp(-\delta m)$ , for all  $t, m = 1, 2, \dots$  and  $A \in \mathcal{D}_{-\infty}^t, B \in \mathcal{D}_{t+m}^\infty$ .*

**Assumption 5.** *There exists a constant  $\delta > 0$  such that for all  $t, m = 1, 2, \dots$ ,  $\delta^{-1} < m < t$ , and all  $\tau \in \mathbb{R}^r$  with  $\|\tau\| \geq \delta$ ,  $\mathbb{E} [|\mathbb{E} [\exp(i\tau^\top (g_{t-m} + \dots + g_{t+m})) | \mathcal{D}_k : k \neq t]|] \leq \exp(-\delta)$ , and*

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \text{Var} \left[ \sum_{t=1}^T g_t \right] > 0. \quad (13)$$

**Assumption 6.** *There exists a constant  $\delta > 0$  such that for all  $t, m, p = 1, 2, \dots$  and  $A \in \mathcal{D}_{t-p}^{t+p}$ ,  $\mathbb{E} [|\mathbb{P}[A | \mathcal{D}_k : k \neq t] - \mathbb{P}[A | \mathcal{D}_k : 0 < |t-k| \leq m+p]|] \leq \delta^{-1} \exp(-\delta m)$ .*

**Assumption 7.** *Defining  $f(\beta) := \limsup_{T \rightarrow \infty} \sup_{b < \|\tau\| < e^{\delta T}} |T^{-1} \sum_{t=1}^T \exp(i\tau^\top g_{T,t}(\beta))|$ , there exist constants  $b$  and  $\delta > 0$  such that  $f(\beta_0) < 1$  and is continuous at  $\beta_0$  a.s.*

Assumption 1 is an identification condition. It is necessary also because we evaluate the moment conditions at  $\hat{\beta}$  in the bootstrap samples. Since we are going to prove the validity and higher-order correctness of FMB using the  $(s-2)$ -th order Edgeworth expansion for the mean of [Götze and Hipp \(1994\)](#), we require the moments in Assumption 2 to be defined. Assumption 3 ensures that the process  $\{g_t\}$  is close enough to another process  $\{g_{t,m}^\dagger\}$ , measurable w.r.t. sub-sigma-fields belonging to the sequence  $\{\mathcal{D}_t\}$ , whose dependence structure is controlled by the mixing condition in Assumption 4. Assumption 5 is the conditional Cramér condition of [Götze and Hipp \(1994\)](#) for weakly dependent process. Assumption 6 ensures that we can approximate the probability of  $A \in \mathcal{D}_{t-p}^{t+p}$  given  $\{\mathcal{D}_k : k \neq t\}$  with an exponentially increasing accuracy, as the information in  $\{\mathcal{D}_k : 0 < |t-k| \leq m+p\}$  increases with  $m$ . We need Assumption 7 on the regularity of the bootstrap characteristic function in order for the appropriate Cramér condition to hold for  $S^*$  and  $Q^*$ . We refer to [Götze and Hipp \(1983\)](#) for a general overview of the processes in agreement with our assumptions. As an example, the OLS moment indicators for the autoregressive parameters of a stationary AR( $p$ ) process  $Y_t = \sum_{i=1}^p \theta_i Y_{t-i} + e_t = \sum_{j=0}^\infty w_j e_{t-j}$  satisfy Assumptions 1–7, when  $e_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$  and  $|w_j| \leq \delta^{-1} \exp(-\delta j) \forall j \in \mathbb{N}$ ,  $\delta > 0$ .

We verify the assumptions for this example in the Supplementary Material (SM.10). We can check along the same lines that an ACD( $v,0$ ) model, and in particular the ACD(1,0) used in our Monte Carlo simulations, satisfies Assumptions 1–7.

Under the stated assumptions, we prove (see Supplementary Material available online) the following two theorems, in which we denote by  $\mathbb{P}^*$  the bootstrap probability measure given the data:

$$\mathbb{P}^*(A) := T^{-1} \sum_{t=1}^T \mathbb{I}_A\{g_{T,t}(\hat{\beta})\}, \quad (14)$$

where  $A$  is a set and  $\mathbb{I}_A\{V\} = 1$  if  $V \in A$  and 0 otherwise.

**Theorem 4.** *Under Assumptions 1–7,  $\mathbb{E} \left[ \|g_t\|^{\bar{q}s+\delta} \right] < \infty$ , for  $\delta > 0$ ,  $s \geq 8$ , and  $\bar{q} \geq 3$ ,  $\log T = o(B_T)$ , we have (i) for  $S^*$  as in (5) and  $\hat{S}$  as in (4):*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}^*[S^* \leq x] - \mathbb{P}[\hat{S} \leq x] \right| = o_p(T^{-1/2}) + O_p(B_T^{-q}) + O_p(B_T/T),$$

and (ii) for  $Q^*$  as in (9) and  $\hat{Q}$  as in (6):

$$\sup_{x \in \mathbb{R}^+} \left| \mathbb{P}^*[Q^* \leq x] - \mathbb{P}[\hat{Q} \leq x] \right| = o_p(T^{-1/2}) + O_p(B_T^{-q}) + O_p(B_T/T),$$

where  $q$  is the Parzen exponent of  $k^*$ , where  $k^*(a) := \kappa_2^{-1} \int k(b-a)k(b)db$ .

It is now apparent that the kernel  $k$  impacts the bootstrap accuracy through the variance estimators in  $\hat{S}$  and  $\hat{Q}$ . As discussed by Parzen (1957) and Andrews (1991), the bias of this estimator depends on the smoothness of the kernel  $k^*$  at zero. The kernel  $k^*(a) := \kappa_2^{-1} \int k(b-a)k(b)db$  is induced by the self-convolution of the smoothing kernel  $k$ . This bias is of order  $O(B_T^{-q})$ , where  $q$  is the Parzen exponent of  $k^*$ , namely the maximal natural number such that  $\lim_{a \rightarrow 0} (1 - k^*(a))/|a|^q$  is finite. The bias is minimal for the rectangular kernel. However, as discussed in Section 4.1, the resulting estimate  $\hat{\Omega}$  is not necessarily positive semi-definite. In contrast, the QS kernel has an optimal Parzen exponent  $q = 2$  over all the kernels giving positive semi-definite estimators. Thus, the covariance matrix estimator with QS kernel has asymptotic MSE of order  $O(B_T/T) + O(B_T^{-2})$ . Consequently,  $B_T$  must grow faster than  $T^{1/4}$  and slower than  $T^{1/2}$ , for the bootstrap error to be  $o_p(T^{-1/2})$ . In Section 2, we define the FMB CR by

$\mathcal{C}_{1-\alpha} = \left\{ \beta \in \mathcal{B} : \hat{Q}(\beta) \leq q_{1-\alpha}^* \right\}$  (or alternatively using  $\tilde{Q}(\beta)$  as in (8)), where  $q_{1-\alpha}^*$  is the approximation of  $\hat{Q}(\beta_0)$  quantiles by FMB. Thus,  $\mathbb{P}[\beta_0 \in \mathcal{C}_{1-\alpha}] = \mathbb{P}[\hat{Q}(\beta_0) \leq q_{1-\alpha}^*] = \mathbb{P}^*[Q^* \leq q_{1-\alpha}^*] + R_T = 1 - \alpha + R_T$ . Theorem 4 gives the order  $R_T = o_p(T^{-1/2}) + O_p(B_T^{-q}) + O_p(B_T/T)$ . If we select a kernel  $k$  and a bandwidth  $B_T$  such that  $O_p(B_T^{-q}) + O_p(B_T/T) = o_p(T^{-1/2})$ , it implies that  $\mathcal{C}_{1-\alpha}$  is higher-order correct by construction, as the (first-order) Gaussian approximation is at best of order  $O(T^{-1/2})$ . If  $B_T = CT^\gamma$ ,  $C > 0$ , we get the conditions  $q > 1$  and  $(2q)^{-1} < \gamma < 1/2$ . Similar arguments apply for Algorithm 1. Part (ii) extends the higher-order correctness of the i.i.d. bootstrap of Hu and Kalbfleisch (2000) in the just-identified multivariate parameter case (see their Remark 10) to the over-identified case with time-dependent data.

When the monotonicity condition discussed in Section 2 (Example 1) is violated (or is uneasy to check), we may use the alternative FMB CI or CR defined as  $\mathcal{C}_{1-\alpha} := \left\{ \beta \in \mathcal{B} : \tilde{Q}(\beta) \leq q_{1-\alpha}^* \right\}$  (see Remark 2 and (8)), which are simply connected when  $T$  is large enough and centered at  $\hat{\beta}$ . The following corollary (see the Supplementary Material for its proof) states the higher-order correctness of this alternative version of FMB CI and CR, based on the  $\tilde{Q}$  statistic.

**Corollary 5.** *Under Assumptions 1–7,  $\mathbb{E} \left[ \|g_t\|^{\bar{q}s+\delta} \right] < \infty$ , for  $\delta > 0$ ,  $s \geq 8$ , and  $\bar{q} \geq 3$ ,  $\log T = o(B_T)$ , we have for  $\tilde{Q}$  as in (8) and  $Q^*$  as in (9):*

$$\sup_{x \in \mathbb{R}^+} \left| \mathbb{P}^*[Q^* \leq x] - \mathbb{P}[\tilde{Q} \leq x] \right| = o_p(T^{-1/2}) + O_p(B_T^{-q}) + O_p(B_T/T),$$

where  $q$  is the Parzen exponent of  $k^*$ .

As a consequence, we have  $\mathbb{P}[\beta_0 \in \mathcal{C}_{1-\alpha}] = \mathbb{P}[\tilde{Q}(\beta_0) \leq q_{1-\alpha}^*] = \mathbb{P}^*[Q^* \leq q_{1-\alpha}^*] + R_T = 1 - \alpha + R_T$ , with  $R_T = o_p(T^{-1/2}) + O_p(B_T^{-q}) + O_p(B_T/T)$ . Thus, similarly to the CI and CR based on  $\hat{S}$  and  $\hat{Q}$ , the CI and CR based on  $\tilde{Q}$  are higher-order correct if we select a kernel  $k$  such that  $O_p(B_T^{-q}) + O_p(B_T/T) = o_p(T^{-1/2})$ .

#### 4. Implementation aspects

#### 4.1. Consistent covariance matrix estimation

In this section, we give the necessary details on the appropriate way to estimate the long-run variance in the statistics  $\hat{S}$  (as in (4)) and  $\hat{Q}$  (as in (6)). We treat the general case of  $\hat{\Omega}$  (as in (6)), from which the univariate case can be easily deduced. Following Andrews (1991) and Smith (2011), we can obtain several estimators  $\hat{\Omega}$  using different types of kernels. Let us give some examples of kernels useful in the implementation of FMB and their respective properties.

**Example 6** (Truncated kernel). *The so-called truncated kernel is  $k(x) = 1$ , if  $|x| \leq 1$ , and 0, if  $|x| > 1$ . Considering  $m_T$  such that  $B_T = (2m_T + 1)/2$ , we write  $g_{T,t}(\beta) = 2(2m_T + 1)^{-1} \sum_{s=\max[t-T, -m_T]}^{\min[t-1, m_T]} g_{t-s}(\beta)$ . The corresponding long-run variance estimator follows directly from the definition of  $\hat{\Omega}$  and has minimal asymptotic mean squared error (see Andrews (1991)), but does not guarantee the resulting variance estimator to be positive semi-definite. The spectral window generator of the truncated kernel is its Fourier transform  $K(\lambda) = \pi^{-1}[\sin(\lambda)/\lambda]$ . The corresponding induced kernel is the Bartlett kernel  $k^*(x) = 1 - |x/2|$  for  $|x| \leq 2$ , 0 for  $|x| > 2$ , as its spectral window generator is  $K^*(\lambda) = \pi^{-1}[\sin(\lambda)/\lambda]^2$ . According to Andrews (1991), the corresponding optimal bandwidth parameter is  $m_T^{\text{opt}} = O(T^{1/3})$ , and the standardizing constants are  $\kappa_1 = 2$  and  $\kappa_2 = 2$ .*

**Example 7** (Quadratic Spectral kernel). *Among the available kernels ensuring the long-run variance estimator to be positive semi-definite, Andrews (1991) points out the optimal QS kernel  $k_{QS}^*$ , as well as the respective optimal bandwidth  $B_T^{\text{opt}} = O(T^{1/5})$ . From the relationship  $K^*(\lambda) = (2\pi/\kappa_2) |K(\lambda)|^2$  and the inverse Fourier transform, Smith (2011) identifies the kernel*

$$k_J(x) := \begin{cases} \frac{1}{x} J_1\left(\frac{6\pi x}{5}\right) \left(\frac{5\pi}{8}\right)^{1/2} & \text{if } x \neq 0, \\ \frac{3\pi}{5} \left(\frac{5\pi}{8}\right)^{1/2} & \text{if } x = 0, \end{cases}$$

$$J_\nu(z) := \frac{z^\nu}{2^\nu} \sum_{j=0}^{\infty} (-1)^j \frac{z^{2j}}{2^{2j} j! \Gamma(\nu + j + 1)},$$

inducing the QS kernel by self-convolution:  $k_{QS}^*(a) = (1/\kappa_2) \int k_J(b-a)k_J(b)db$ . The standardizing constants are  $\kappa_1 = (5\pi/2)^{1/2}$  and  $\kappa_2 = 2\pi$ .

Thereby, we preferably use  $k_J(x)$  of Example 7 in (2) to get the estimator  $\hat{\Omega}$ . Indeed, both from theory and simulations, the QS kernel is the optimal induced kernel in terms of asymptotic mean squared

error ([Andrews \(1991\)](#)), among all kernels giving positive semi-definite long-run variance estimators. The optimal bandwidth  $B_T = O(T^{1/5})$  minimising the mean-squared error of the variance estimator with  $k_J(x)$  (see Example 5) does not satisfy the condition  $(1 - \epsilon)/q < \gamma < \epsilon$  for any  $\epsilon \in (0, 1/2]$  (see Section 3) since  $q = 2$  for the QS kernel. Then, we can take  $\gamma = 1/3$ , so that we match the condition for  $q = 2$ .

[Wilhelm \(2015\)](#) has already exhibited a discrepancy between the optimal choice for the HAC variance estimator and the optimal choice for a GMM point estimator. In his case, the optimal bandwidth for the point estimate is of the same order as the one minimizing the mean-squared error of the nonparametric plugin estimate, while the constants of proportionality are significantly different. In our case, the order is even different if we want to achieve higher-order correctness for FMB. It makes the data-driven bandwidth selection a difficult, yet important, problem. In practice, it is sound to apply the guidelines of the HAC estimators or to adapt the ones of MBB (see Chapter 7 in [Lahiri \(2010\)](#) for a comprehensive discussion).

Alternatively, we may use the flat-top kernel version of the QS kernel, see [Politis \(2011\)](#), to get an even faster rate of convergence for the estimated long-run variance. Unfortunately, self-convolution of a kernel  $k$  cannot induce a flat-top kernel  $k^*$ . Indeed, we know that it cannot be the case that  $U = X + Y$ , where the random variable  $U$  is uniformly distributed on  $[0, 1]$  (the flat-top part) and the random variables  $X$  and  $Y$  are independent and identically distributed (see Exercise 4.14.20 and its proof by contradiction in [Grimmett and Stirzaker \(2001\)](#)). As a consequence, if we want to benefit from the smaller bias of the flat-top kernel, we should use a different kernel for the original statistic and the bootstrap one. A potential modification of FMB is to decouple the kernel  $k^*$  used in the variance estimator of the original statistic, say a flat-top kernel, and the kernel  $k$  used for the smoothed moment indicators. This version of FMB also achieves higher-order correctness since we maintain the asymptotic pivotal nature of the test statistics.

#### 4.2. From confidence regions to confidence intervals

FMB user can manipulate the higher-order correct CR  $\mathcal{C}_{1-\alpha}$  to obtain CR for a subset of parameters, or CI for a single parameter. In this section, we will consider separately the (possibly multivariate) parameter of interest  $\beta_0^{(1)}$ , and the (possibly multivariate) nuisance parameter  $\beta_0^{(2)}$ . Without loss of

generality, we write the partition in the order  $\beta = (\beta^{(1)\top}, \beta^{(2)\top})^\top$ .

We define the CR for  $\beta_0^{(1)}$  as  $\mathcal{C}_{1-\alpha}^{(1)} := \{\beta^{(1)} : \hat{Q}(\beta^{(1)}, \hat{\beta}^{(2)}) \leq q_{1-\alpha}^*\}$ . This manipulation preserves higher-order correctness when  $\hat{Q}(\beta_0^{(1)}, \hat{\beta}^{(2)}) \leq \hat{Q}(\beta_0)$  *a.s.*, in the sense that it ensures  $\mathbb{P}[\beta_0^{(1)} \in \mathcal{C}_{1-\alpha}^{(1)}] = \mathbb{P}[\hat{Q}(\beta_0^{(1)}, \hat{\beta}^{(2)}) \leq q_{1-\alpha}^*] \geq \mathbb{P}[\hat{Q}(\beta_0) \leq q_{1-\alpha}^*] = \mathbb{P}^*[Q^* \leq q_{1-\alpha}^*] + R_T = 1 - \alpha + R_T$ . This condition is satisfied, for instance, when  $\hat{Q}$  is a monotonic function of the norm  $\|\beta - \hat{\beta}\|$ , since  $\|(\beta_0^{(1)\top}, \hat{\beta}^{(2)\top})^\top - \hat{\beta}\| \leq \|\beta_0 - \hat{\beta}\|$ . It is also satisfied when the two sets of parameters  $\beta^{(1)}$  and  $\beta^{(2)}$  can be estimated independently from each other. For instance, in a two-step OLS estimation, it is the case in our ACD(1,0) example of Section 5, by the Frisch-Waugh-Lovell Theorem, since we can rewrite the ACD(1,0) as an AR(1) process with orthogonal regressors.

If the inequality  $\hat{Q}(\beta_0^{(1)}, \hat{\beta}^{(2)}) \leq \hat{Q}(\beta_0)$  fails to be true almost surely, we cannot guarantee the higher-order correctness of  $\mathcal{C}_{1-\alpha}^{(1)}$ . Nevertheless, the inequality  $\mathbb{P}[\hat{Q}(\beta_0^{(1)}, \hat{\beta}^{(2)}) \leq q_{1-\alpha}^*] \geq \mathbb{P}[\hat{Q}(\beta_0) \leq q_{1-\alpha}^*]$  remains true for large enough  $T$ , as long as  $\hat{Q}(\beta_0^{(1)}, \hat{\beta}^{(2)}) \xrightarrow{\mathcal{D}} \mathcal{X}_{\dim(\beta_0^{(1)})}$  and  $\hat{Q}(\beta_0) \xrightarrow{\mathcal{D}} \mathcal{X}_p$ . This general result implies that  $\mathcal{C}_{1-\alpha}^{(1)}$  contains  $\beta_0^{(1)}$  at least with probability  $1 - \alpha$  asymptotically, but not always with higher-order refinements.

The drawback of no guarantee of higher-order refinements is inherent to the existing information on a multivariate parameter. The difficult task of reducing a CR for  $\beta$  to a CR for  $\beta^{(1)}$  is not directly entangled with the FMB, but more with the nature of dependence between  $\beta^{(1)}$  and  $\beta^{(2)}$  induced by the moment condition themselves. However, there exists a general way to build CR for  $\beta_0^{(1)}$  while preserving higher-order refinements. Indeed, defining the profile statistic  $\bar{Q}(\beta^{(1)}) := \inf_{\beta^{(2)}} \hat{Q}(\beta^{(1)}, \beta^{(2)})$ , we get  $\bar{Q}(\beta_0^{(1)}) \leq \hat{Q}(\beta_0)$  *a.s.* by construction. Thus, by the same argument, the alternative CR  $\bar{\mathcal{C}}_{1-\alpha}^{(1)} := \{\beta^{(1)} : \bar{Q}(\beta^{(1)}) \leq q_{1-\alpha}^*\}$  preserves the higher-order refinements. This property comes with a cost, as  $\bar{\mathcal{C}}_{1-\alpha}^{(1)}$  is generally more conservative than  $\mathcal{C}_{1-\alpha}^{(1)}$  and it might be heavy to compute in high dimension.

#### 4.3. From confidence intervals to a confidence curve

Let us now make connections to the concept of Confidence Distributions (CD) that we use in our empirical application. It aims at answering the following question: can we also use a distribution function, or a “distribution estimator”, to estimate or test for a parameter of interest in frequentist inference in the style of a Bayesian posterior? (see the review paper by [Xie and Singh \(2013\)](#)). Natural point estimators include the median, the mean, and the maximum of the CD density ([Singh et al.](#)

(2005)). That “distribution estimator” is named CD in agreement with the terminology coined by Efron (1998), and traces back to the fiducial distribution of Fisher (1930), albeit being a purely frequentist concept. It was introduced by Schweder and Hjort (2002) and its asymptotic extension by Singh et al. (2005) (see also Xie et al. (2011), Veronese and Melilli (2015), and the book-length presentation of Schweder and Hjort (2016)). Example 2.4 of Singh et al. (2005) discusses how a bootstrap distribution can yield a valid asymptotic CD, and Section 2.3.3 of Xie and Singh (2013) how studentization can transmit higher-order accuracy in the i.i.d. case. Paralleling these recent developments in fiducial inference theory (see also the review paper of Hannig et al. (2016)), we can exploit our FMB to produce a fast methodology to build an asymptotically higher-order correct CD as a by-product. Let us define the functions  $H_S(\beta) := \mathbb{P}[\hat{S} \leq \hat{S}(\beta)]$  and its FMB counterpart  $H_S^*(\beta) := \mathbb{P}^*[S^* \leq \hat{S}(\beta)]$ , for  $\beta \in \mathcal{B}$ .

**Corollary 8.** *Under Assumptions 1–7,  $\mathbb{E}[\|g_t\|^{\bar{q}s+\delta}] < \infty$ , for  $\delta > 0$ ,  $s \geq 8$ , and  $\bar{q} \geq 3$ ,  $\log T = o(B_T)$ , we have the uniform error bound:  $\sup_{\beta \in \mathcal{B}} |H_S^*(\beta) - H_S(\beta)| = o_p(T^{-1/2}) + O_p(B_T^{-q}) + O_p(B_T/T)$ , and  $H_S^*(\beta)$  is an asymptotic CD.*

We omit the proof since the uniform error bound follows immediately from the proof of Theorem 4. The second statement comes from the two conditions of Definition 1.1. of Singh et al. (2005) being met, namely  $H_S^*(\beta)$  is a cdf, and  $H_S^*(\beta_0)$  is uniformly distributed on the unit interval when  $T$  goes to infinity. Here, as clarified by Pitman (1957), we follow indeed the frequentist view. In  $H_S(\beta)$  and  $H_S^*(\beta)$ , randomness is not coming from the (non-random) parameter  $\beta$ , but from  $\hat{S}$  and  $S^*$ .

As described in Singh et al. (2005) (see also Fraser (1961), Xie and Singh (2013)), we can also use CD to get  $p$ -values. For example, the classical bootstrap  $p$ -value of  $H_0 : \beta_0 \leq \beta$  versus  $H_1 : \beta_0 > \beta$  corresponds to  $H_S^*(\beta)$ , and the classical equal-tail bootstrap  $p$ -value of  $H_0 : \beta_0 = \beta$  versus  $H_1 : \beta_0 \neq \beta$  corresponds to  $2 \min\{H_S^*(\beta), 1 - H_S^*(\beta)\}$ .

These  $p$ -values also benefit from higher-order correctness. Collecting them for different values of  $\beta$  yields the so-called confidence curve  $CV^*(\beta) := 2 \min\{H_S^*(\beta), 1 - H_S^*(\beta)\}$ , introduced by Birnbaum (1961) (see Xie and Singh (2013) and Hannig et al. (2016) for illustrations). We can view this graphical tool as a piled-up form of two-sided CI of equal tails, at all confidence levels. We provide an example of such a plot in Figure 1 for our empirical application in Section 6, where we compare CI given by our FMB and first-order Gaussian asymptotics. Finally, Coudin and Dufour (2020) show how we can design

a Hodges-Lehmann-type point estimator (Hodges and Lehmann (1963)) when a CD is constructed from a hypothesis test.

There exist analogue multivariate CD, for instance Definitions 5.1 and 5.2 in Singh et al. (2007). Similarly to the univariate case, we can apply FMB to achieve higher-order accuracy, as long as these multivariate confidence distributions are based on the test statistic  $\hat{Q}(\beta_0)$ .

## 5. Monte Carlo experiments

To illustrate the applicability of FMB, we consider a simulation exercise on constructing CR for the parameters of an Autoregressive Conditional Duration (ACD) model (Engle and Russel (1998)). It is a model typically applied for the analysis of high-frequency data in finance, and more generally to model positive variables (e.g. volatility or volume) via a multiplicative error model; see e.g. Hautsch (2012) for a recent book-length presentation.

The duration is defined as the time lag between two consecutive events occurrence, namely  $x_\ell := t_\ell - t_{\ell-1}$ . Clearly,  $x_\ell > 0$ , for any  $\ell \in \mathbb{T}$ . We model  $\mathbb{E}(x_\ell | x_{\ell-1}, \dots, x_1) = m_\ell(x_{\ell-1}, \dots, x_1; \beta) := m_\ell$ , assuming the model  $x_\ell = m_\ell \varepsilon_\ell$ , with  $\varepsilon_\ell \stackrel{i.i.d.}{\sim} \mathcal{E}(1)$  for any  $\ell$ , with  $\mathcal{E}(1)$  being an exponential random variable with mean one. Specifically, for the ACD(1, 1) specification, we have

$$x_\ell = \varepsilon_\ell m_\ell, \quad \text{with} \quad m_\ell = \omega + \beta_1 x_{\ell-1} + \beta_2 m_{\ell-1}, \quad \ell \in \mathbb{Z}, \quad (15)$$

for  $\omega > 0$ ,  $\beta_1, \beta_2 \in \mathbb{R}^+$  and  $\beta_1 + \beta_2 < 1$ . When we take  $\beta_2 = 0$ , the ACD(1,0) model is in agreement with Assumptions 1–7 (in Section 3). Thus, we start our numerical experiment with this specification, conducting inference on  $\beta := (\omega, \beta_1)^\top$ . We apply the optimal estimating functions of Li and Turtle (2000) and a moment condition which does not assume any specific functional form for the underlying innovation density, but relies on the unconditional expectation of  $x_\ell$ . Therefore, given a random sample of durations  $(x_1, \dots, x_\ell, \dots, x_T)$ , the vector of moment conditions for the  $\ell$ -th observation is  $g_\ell(\beta) := (g_{1,\ell}(\beta), g_{2,\ell}(\beta), g_{3,\ell}(\beta))^\top$ , with  $g_{1,\ell}(\beta) := ((x_\ell - m_\ell) / m_\ell^2)(\partial m_\ell / \partial \omega)$ ,  $g_{2,\ell}(\beta) := ((x_\ell - m_\ell) / m_\ell^2)(\partial m_\ell / \partial \beta_1)$ , and  $g_{3,\ell}(\beta) := x_\ell - (1 - \beta_1)^{-1}$  (Li and Turtle (2000)). Thus, we are in the over-identified case with  $r = 3$  for  $p = 2$ .

In a second step, to get numerical insights on the applicability of our FMB, we extend our Monte



Carlo experiment to a general ACD(1,1) model. The latter is non-markovian in the observations and does not fit our setting *stricto sensu*, since we assume the vector of observations in (1) to be finite to ease notations and proofs. However, taking a large number  $v$  of lagged durations in an ACD( $v,0$ ) model is close to an ACD(1,1) model, and it meets our current theoretical framework.

We conduct inference on  $\beta := (\omega, \beta_1, \beta_2)^\top$  and our moment conditions for the  $\ell$ -th observation is  $g_\ell(\beta) := (g_{1,\ell}(\beta), g_{2,\ell}(\beta), g_{3,\ell}(\beta), g_{4,\ell}(\beta))^\top$ , with  $g_{1,\ell}(\beta) := ((x_\ell - m_\ell)/m_\ell^2)(\partial m_\ell/\partial\omega)$ ,  $g_{2,\ell}(\beta) := ((x_\ell - m_\ell)/m_\ell^2)(\partial m_\ell/\partial\beta_1)$ ,  $g_{3,\ell}(\beta) := ((x_\ell - m_\ell)/m_\ell^2)(\partial m_\ell/\partial\beta_2)$ , and  $g_{4,\ell}(\beta) := x_\ell - (1 - \beta_1 - \beta_2)^{-1}$ . We stay in the over-identified case with  $r = 4$  for  $p = 3$ .

In the following, we compute by Monte Carlo simulations the coverage of FMB CR. We label the results  $\hat{Q}_{FMB}$  for the FMB using  $\hat{Q}$ ,  $\tilde{Q}_{FMB}$  for the FMB using the cubic approximation of Remark 3 and  $\tilde{Q}_{FMB,2}$  for the FMB using the quadratic approximation of Remark 3. To validate numerically our theoretical results, we compare the coverage of these FMB CR to some standard first-order correct alternatives. The first one, labeled as  $\hat{Q}_{\mathcal{X}_r^2}$ , defines CR as contours of the same statistic than FMB, but making use of the (first-order correct)  $\mathcal{X}_r^2$  asymptotic distribution to compute the rejection probabilities. The second one is the standard elliptical contour of an asymptotically  $\mathcal{X}_p^2$  distributed Wald statistic (labeled as  $W_{\mathcal{X}_p^2}$ ), whose covariance matrix is a HAC estimator with bandwidth  $B_T$ .

To compare with a state-of-the-art competitor of FMB in terms of higher-order correctness, we also show the coverage of CR yielded by MBB. We adapt the MBB of Inoue and Shintani (2006) to a Wald statistic, which is, in turn, an adaptation of Götze and Künsch (1996). As the statistic defining the CR has to be asymptotically pivotal to obtain higher-order correctness, we choose to apply MBB to the Wald statistic, which we label as  $W_{MBB}$ . We show a detailed CPU time comparison between FMB and MBB in the Supplementary Material (Table 6). In general, FMB appears to be at least  $10^3$  times faster than MBB as expected.

In Table 2, we display the empirical coverages for the ACD(1,0) model, for  $B_T = 3$  and  $B_T = 5$ . In Table 3, we show the same outputs for the ACD(1,1) specification.

Table 2: Empirical coverage of CR, a comparison between first and higher-order correct methods.

		$B_T = 3$			$B_T = 5$		
Coverages:		0.90	0.95	0.99	0.90	0.95	0.99
$T = 250$	$\hat{Q}_{FMB}$	0.92	0.94	0.98	0.89	0.93	0.96
	$\tilde{Q}_{FMB}$	0.92	0.95	0.98	0.90	0.94	0.97
	$\tilde{Q}_{FMB,2}$	0.92	0.95	0.98	0.90	0.93	0.97
	$\hat{Q}_{\chi_r^2}$	0.91	0.93	0.97	0.88	0.90	0.96
	$W_{MBB}$	0.93	0.97	1.00	0.93	0.97	1.00
	$W_{\chi_p^2}$	0.84	0.90	0.96	0.80	0.86	0.92
$T = 500$	$\hat{Q}_{FMB}$	0.93	0.96	0.98	0.92	0.95	0.98
	$\tilde{Q}_{FMB}$	0.93	0.95	0.98	0.92	0.95	0.98
	$\tilde{Q}_{FMB,2}$	0.92	0.96	0.99	0.91	0.95	0.99
	$\hat{Q}_{\chi_r^2}$	0.92	0.95	0.98	0.91	0.94	0.97
	$W_{MBB}$	0.95	0.98	1.00	0.94	0.98	1.00
	$W_{\chi_p^2}$	0.86	0.91	0.97	0.84	0.89	0.96

The true values of the unknown parameters of the ACD(1,0) are  $\omega = 1.5$  and  $\beta_1 = 0.25$ .

Table 3: Empirical coverage of CR, a comparison between first and higher-order correct methods.

		$B_T = 3$			$B_T = 5$		
Coverages:		0.90	0.95	0.99	0.90	0.95	0.99
$T = 250$	$\hat{Q}_{FMB}$	0.88	0.93	0.98	0.87	0.90	0.96
	$\tilde{Q}_{FMB}$	0.89	0.92	0.95	0.87	0.90	0.94
	$\tilde{Q}_{FMB,2}$	0.82	0.87	0.94	0.80	0.85	0.93
	$\hat{Q}_{\chi_r^2}$	0.86	0.91	0.96	0.85	0.89	0.95
	$W_{MBB}$	0.97	0.99	1.00	0.97	0.99	1.00
	$W_{\chi_p^2}$	0.73	0.79	0.88	0.71	0.77	0.86
$T = 500$	$\hat{Q}_{FMB}$	0.91	0.95	0.99	0.90	0.94	0.98
	$\tilde{Q}_{FMB}$	0.91	0.94	0.97	0.91	0.94	0.97
	$\tilde{Q}_{FMB,2}$	0.85	0.90	0.96	0.84	0.89	0.95
	$\hat{Q}_{\chi_r^2}$	0.90	0.93	0.97	0.89	0.93	0.97
	$W_{MBB}$	0.98	0.99	1.00	0.97	0.99	1.00
	$W_{\chi_p^2}$	0.78	0.83	0.92	0.77	0.83	0.92

The true values of the unknown parameters of the ACD(1,1) are  $\omega = 1.5$ ,  $\beta_1 = 0.25$  and  $\beta_2 = 0.25$ .

In line with our theoretical results, we observe that FMB performs well compared to its first-order correct competitors: the coverages are typically closer to their nominal level. It generally remains true for the  $\tilde{Q}_{FMB}$  version of FMB CR, whose coverage is often very close to the one of the original FMB CR based on  $\hat{Q}_{FMB}$ . The coverage of the quadratic  $\tilde{Q}_{FMB,2}$  version of FMB CR is generally further away from the original FMB.

The Wald statistic seems to yield very erratic CR (which is additionally confirmed by unreported plots). The MBB version of the Wald statistic improves slightly on the asymptotic  $\chi_p^2$  distribution, without being convincing though.

Finally, the FMB presented here does not take advantage of all the potential fine-tuning, and this should leave room for practical improvement. First, we might improve FMB if the long-run variance  $\Omega(\beta_0)$  is estimated with a less biased version of variance estimator, for instance carrying out a prewhitening step ([Andrews and Monahan \(1992\)](#)), or using a flat-top kernel ([Politis \(2011\)](#)) as discussed in [Section 4.1](#). It should yield a smaller bootstrap error, as shown in [Theorem 4](#). Second, we stress that the moment indicators do not have necessarily the same dependence structure. Thus, smoothing the multivariate moment indicators with different bandwidths might further improve the coverage of FMB.

## 6. Real data application

In this section, we illustrate how FMB performs on real data. We look at daily volumes of stock transaction (in millions), modeled with the same exponential ACD as in [Section 5](#) (see [\(15\)](#)). We focus on data available online (*Yahoo! Finance*), for five stocks in three different sectors, namely bank, technology, and food. We compute the CR (for parameters  $\omega$ ,  $\beta_1$  and  $\beta_2$ ), before the subprime crisis (2005), during the crisis (2008), and the current period (2018). The sample size of each period corresponds to the number of trading days, namely  $T = 252$  up to negligible variations from year to year. Before diving into a deeper analysis, we briefly describe the data at hand in [Table 4](#) below.

[Table 4](#) illustrates the larger variability of the volumes of transaction during 2008, as measured by the standard deviation (SD) and the interquartile range (IQR). The high skewness (SKN) and excess of kurtosis (KURT) typically indicate that a higher-order correct inferential procedure might be required

Table 4: Summary statistics of volumes of transaction.

<i>Year 2005</i>	Min	Max	Med	Mean	IQR	SD	SKN	KURT
BA	4.52	42.05	10.67	11.24	4.98	4.47	2.36	11.18
JPM	3.77	25.47	9.93	10.6	4.11	3.25	0.94	1.31
MSF	27.21	187.38	63.53	66.61	20.94	20.23	1.78	6.40
KO	3.96	39.75	11.01	11.86	3.83	4.15	2.64	12.16
UL	0.16	3.36	0.49	0.59	0.31	0.39	2.91	12.78
<i>Year 2008</i>	Min	Max	Med	Mean	IQR	SD	SKN	KURT
BA	22.59	322.73	63.16	78.63	59.52	49.01	1.66	3.28
JPM	12.34	194.07	41.96	48.99	29.64	25.90	1.93	5.72
MSF	16.88	291.14	78.50	84.17	38.81	35.51	1.55	4.73
KO	5.32	79.21	23.35	25.26	12.77	10.63	1.62	4.20
UL	0.26	5.19	0.79	1.08	0.73	0.83	2.09	4.85
<i>Year 2018</i>	Min	Max	Med	Mean	IQR	SD	SKN	KURT
BA	22.97	165.88	62.21	67.91	29.16	24.15	1.29	2.02
JPM	6.49	41.31	13.90	15.17	6.43	5.53	1.60	3.63
MSF	13.66	111.24	27.61	31.59	14.23	13.40	1.74	4.82
KO	4.79	32.48	11.91	12.52	4.32	4.13	1.32	2.67
UL	0.33	4.88	0.90	1.05	0.54	0.64	2.96	11.54

We consider the companies Bank of America (BA), JP Morgan (JPM), Microsoft (MSF), Coca-Cola (KO) and Unilever (UL). In the summary, Med stands for the median, IQR for the inter-quantile range, SD for the standard deviation, SKN for skewness, and KURT for excess of kurtosis.

in finite samples.

To investigate further the impact that asymmetry and fat tails may have on the conducted inference, we compute the FMB and the asymptotic normal (Asy) CR of nominal coverage  $(1 - \alpha) = 95\%$ , for the ACD(1,1) parameters  $\omega$ ,  $\beta_1$  and  $\beta_2$  at each period. As FMB yields higher-order accurate CR by inverting probabilities of the test statistic  $\hat{Q}(\beta_0) = \hat{Q}(\omega_0, \beta_{1,0}, \beta_{2,0})$ , we represent these trivariate CR by slicing them at the estimates  $\hat{\omega}$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$  (Table 5). Namely, we cut the CR by fixing the parameters that are not of interest to their estimated values. To keep Table 5 concise, we do not report the estimate  $\hat{\omega}$  and the intervals for the parameter  $\omega$ . We can deduce the former from Table 4.<sup>4</sup>

Table 5: Analysis of volumes of transaction.

Asset		2005		2008		2018	
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
JPM	Est	0.402	0.132	0.691	0.122	0.459	0.345
	Asy	[0.384,0.420]	[0.114,0.150]	[0.673,0.709]	[0.104,0.140]	[0.448,0.469]	[0.335,0.355]
	FMB	[0.357,0.442]	[0.085,0.178]	[0.621,0.726]	[0.060,0.164]	[0.429,0.483]	[0.318,0.372]
BA	Est	0.366	0.534	0.624	0.323	0.538	0.141
	Asy	[0.361,0.371]	[0.529,0.540]	[0.618,0.629]	[0.318,0.328]	[0.521,0.554]	[0.125,0.157]
	FMB	[0.345,0.381]	[0.512,0.549]	[0.594,0.641]	[0.295,0.344]	[0.491,0.584]	[0.102,0.192]
MSF	Est	0.271	0.340	0.595	0.272	0.570	0.268
	Asy	[0.261,0.281]	[0.330,0.350]	[0.586,0.604]	[0.262,0.281]	[0.559,0.582]	[0.257,0.280]
	FMB	[0.242,0.297]	[0.307,0.373]	[0.562,0.623]	[0.243,0.301]	[0.536,0.596]	[0.238,0.294]
KO	Est	0.202	0.377	0.488	0.371	0.392	0.382
	Asy	[0.190,0.213]	[0.365,0.389]	[0.479,0.496]	[0.363,0.380]	[0.384,0.399]	[0.375,0.390]
	FMB	[0.166,0.233]	[0.344,0.426]	[0.458,0.513]	[0.343,0.400]	[0.363,0.420]	[0.358,0.413]
UL	Est	0.331	0.529	0.571	0.290	0.435	0.483
	Asy	[0.320,0.343]	[0.518,0.541]	[0.555,0.586]	[0.275,0.305]	[0.428,0.441]	[0.477,0.490]
	FMB	[0.297,0.365]	[0.509,0.571]	[0.513,0.599]	[0.236,0.321]	[0.413,0.456]	[0.464,0.503]

We consider the companies Bank of America (BA), JP Morgan (JPM), Microsoft (MSF), Coca-Cola (KO) and Unilever (UL). We use the Exponential Tilting estimator (Est), a particular case of GEL, and the benchmark intervals are the first-order asymptotic Gaussian (Asy). All the intervals are equal-tailed and have conditional nominal coverage of 95% when we fix the other parameters at their estimated values.

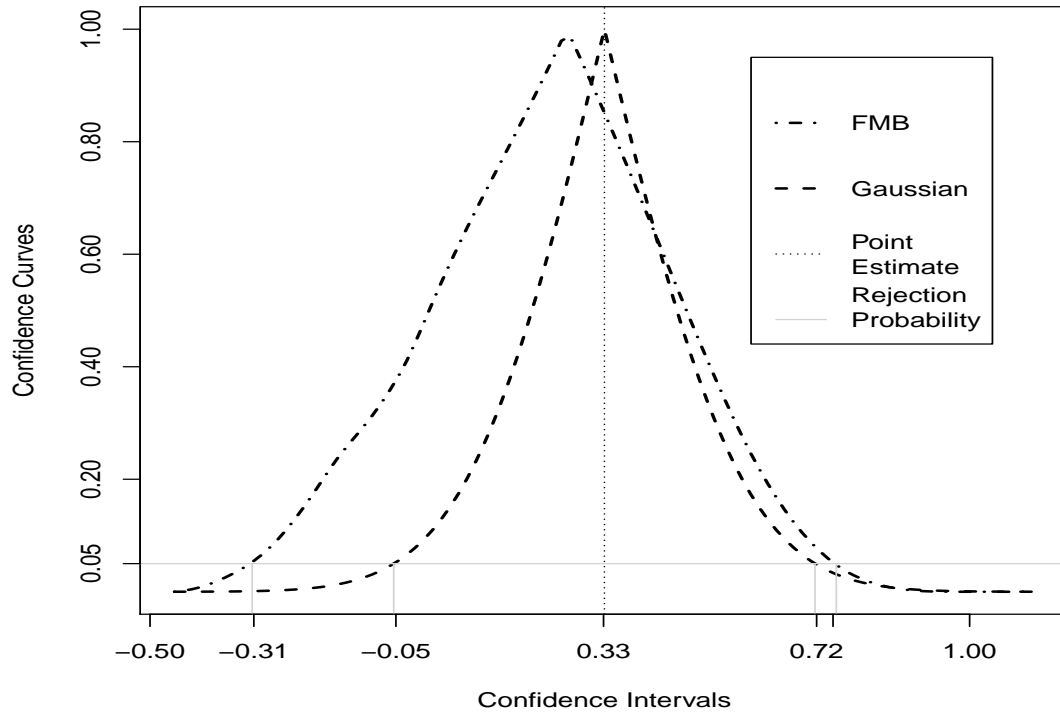
A few comments are in order. First of all, the different sectors exhibit very diverse reactions to the events happening in 2008. For instance, the food sector seems to be the most stable, while financial sector undergoes a huge variability, as we could expect. We can observe this either comparing non-

<sup>4</sup>Using Section 5 and volumes of transaction  $\{x_t\}_{t=1}^T$  instead of durations, we have  $\hat{\omega} \approx (1 - \hat{\beta}_1 - \hat{\beta}_2)T^{-1} \sum_{\ell=1}^T x_\ell$ , from the moment condition based on  $g_{3,t}$ . We report the sample mean  $T^{-1} \sum_{\ell=1}^T x_\ell$  in Table 4.

critical periods to the crisis, or comparing the estimates and their CI before and after the crisis. For instance, the estimates for Unilever are almost the same before and after the crisis, as if the company has recovered the same volume behaviour. Coca-Cola looks equally stable with respect to the parameter  $\beta_2$ , which is almost unchanged after the crisis. Second, the estimate  $\hat{\beta}_1$ , respectively  $\hat{\beta}_2$ , seems to be larger, respectively smaller, during the crisis period. It is expected since  $\beta_1$  reflects the sudden trading reactions due to changes in the expectations by the market participants during the crisis period. Thus, this feature of 2008 corresponds to an increase of the impact of news (shocks) on the volumes of transaction (via the parameter  $\beta_1$ ), relative to persistence (via the parameter  $\beta_2$ ). Finally, we see that the FMB CI are longer than the first-order correct Gaussian CI. It is in line with our Monte Carlo experiments, as available in Section 5. Indeed, as the CR are defined by level sets of  $\hat{Q}$ , a longer CI corresponds to an adaptation of FMB to a skewed or fat-tailed distribution. Since the CI obtained by Gaussian approximation are typically shorter, we conclude that the routinely applied first-order asymptotic theory tends to underestimate the rejection probability, whereas FMB stays conservative. Our experience underpinned by several Monte Carlo simulations makes us expect that the distribution of  $\hat{Q}(\beta_0)$  is more skewed or fat-tailed than the chi-squared; see the comparison between  $\hat{Q}_{\chi^2}$  and FMB in Table 3.

Following our discussion on CD (Subsection 4.3), we illustrate here the link between our FMB CR and our previous definition of asymptotic confidence distribution  $H_S^*(\beta)$ , via the confidence curve  $CV^*(\beta)$ . Among the alternative ways to represent the former CR, marginalization allows us to build unconditional CI. Stacking the CR at different coverages  $1 - \alpha$  leads to a center-outward confidence curve for the multidimensional parameter  $CV(\omega, \beta_1, \beta_2)$ . For each CI, we integrate out the two parameters that are not of interest in  $CV(\omega, \beta_1, \beta_2)$ . This yields a different confidence curve  $CV^*(\beta)$  for each  $\beta \in \{\omega, \beta_1, \beta_2\}$ , whose level sets give the equal-tailed CI. As an illustration of graphical use of these confidence curves (defined in Section 3), Figure 1 reports a comparison between the FMB and Gaussian CI based on the FMB and Gaussian confidence curves. Again we observe that FMB is much more conservative.

Figure 1: Confidence Curves of the FMB and Gaussian approximation.



Confidence curve for the parameter  $\beta_1$  of Unilever in 2005, with point estimate  $\hat{\beta}_1 = 0.33$ . The sample size is  $T = 252$ , and the nominal coverage of the confidence intervals is  $(1 - \alpha) = 95\%$ . The flat solid line is the rejection probability level 5%, and its crossing with the confidence curves gives the FMB 95% equal-tailed CI  $[-0.31, 0.75]$  and Gaussian 95% equal-tailed CI  $[-0.05, 0.72]$ .



## Acknowledgement

We would like to thank the editor, the co-editor, and the referee for constructive criticism and numerous suggestions which have led to substantial improvements over the previous versions. We thank A.-P. Fortin, E. Paparoditis, D. Politis, and R. J. Smith for helpful comments and discussion, as well as participants in the North American Summer Meeting of the Econometric Society (Seattle, 2019), Geneva Finance Research Institute seminar, Workshop on Statistical Learning (Geneva, 2019-20), online RCEA Time Series Workshop 2021, and online IAAE Conference 2021.

## References

- Altonji JG, Segal LM. Small-sample bias in GMM estimation of covariance structures. *Journal of Business & Economic Statistics* 1996;14(3):353–66.
- Anatolyev S. GMM, GEL, serial correlation, and asymptotic bias. *Econometrica* 2005;73:983–1002.
- Andrews DW. Heteroscedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 1991;59(3):817–58.
- Andrews DW. Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica* 2002;69:119–62.
- Andrews DW, Monahan JC. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 1992;60(4):953–66.
- Armstrong TB, Bertanha M, Hong H. A fast resample method for parametric and semiparametric models. *Journal of Econometrics* 2014;179:128–33.
- Birnbaum A. Confidence curves: an omnibus technique for estimation and testing statistical hypotheses. *Journal of the American Statistical Association* 1961;56:246–9.
- Bravo F. Empirical likelihood based inference with applications to some econometrics models. *Econometric Theory* 2004;20(2):231–64.

- Bravo F. Blockwise empirical entropy tests for time series regressions. *Journal of Time Series Analysis* 2005;26(2):185–210.
- Brown BW, Newey WK. Generalized Method of Moments, efficient bootstrapping, and improved inference. *Journal of Business & Economic Statistics* 2002;20(4):507–17.
- Camponovo L, Scaillet O, Trojani F. Robust subsampling. *Journal of Econometrics* 2012;167(1):197–210.
- Camponovo L, Scaillet O, Trojani F. Predictability hidden by anomalous observations. Working paper; 2013.
- Chandra TK, Ghosh JK. Valid asymptotic expansion for the likelihood ratio statistic and other perturbed chi-square variables. *Sankhya, the Indian Journal of Statistics A* 1979;41:22–47.
- Chibisov DM. An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. *Theory of Probability and its Applications* 1972;17:620–30.
- Coudin E, Dufour JM. Finite-sample generalized confidence distributions and sign-based robust estimators in median regressions with heterogeneous dependent errors. *Econometric Reviews* 2020;39(8):763–91.
- Cressie N, Read T. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B (Methodological)* 1984;46:440–64.
- Davidson R, MacKinnon JG. Bootstrap testing in nonlinear models. *International Economic Review* 1999;40:487–508.
- Davidson R, MacKinnon JG. Bootstrap methods in econometrics. In: *Palgrave Handbook of Econometrics: Vol.1 Econometric Theory*. Palgrave-Macmillan; 2006. p. 812–38.
- Davison A, Hinkley D. *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics, 1997.
- DiCiccio T, Efron B. Bootstrap confidence intervals. *Statistical Science* 1996;11(3):189–212.

- Efron B. Better bootstrap confidence intervals. *Journal of the American Statistical Association* 1987;397(82):171–85.
- Efron B. R. A. Fisher in the 21st century. *Statistical Science* 1998;13(2):95–114.
- Engle RF, Russel JR. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 1998;66(5):1127–62.
- Fisher RA. Inverse probability. *Mathematical Proceedings of the Cambridge Philosophical Society* 1930;26(4):528–35.
- Fraser DAS. On fiducial inference. *The Annals of Mathematical Statistics* 1961;32(3):661–76.
- Godambe VP, Heyde CC. Quasi-likelihood and optimal estimation. *International Statistical Review* 1987;55(3):231–44.
- Gonçalves S, Hounyo U, Patton A, Shephard K. Bootstrapping two-stage quasi-maximum likelihood estimators of time series models. Working paper, McGill University; 2019.
- Gonçalves S, White H. Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics* 2004;119(1):199–219.
- Götze F, Hipp C. Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 1983;64(2):211–39.
- Götze F, Hipp C. Asymptotic distribution of statistics in time series. *The Annals of Statistics* 1994;22(4):2062–88.
- Götze F, Künsch H. Second-order correctness of the blockwise bootstrap for stationary observations. *The Annals of Statistics* 1996;24(5):1914–33.
- Grimmett G, Stirzaker D. *One Thousand Exercises in Probability*. Oxford University Press, 2001.
- Guggenberger P, Smith RJ. Generalized empirical likelihood test in time series models with potential identification failure. *Journal of Econometrics* 2008a;142(1):134–61.

- Guggenberger P, Smith RJ. Generalized empirical likelihood test in time series models with potential identification failure. *Journal of Econometrics* 2008b;142(1):134–61.
- Hall P, Horowitz J. Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* 1996;64(4):891–916.
- Hannig J, Iyer H, Lai RCS, Lee TCM. Generalized fiducial inference: a review and new results. *Journal of the American Statistical Association* 2016;515(111):1346–61.
- Hansen LP. Large sample properties of generalized method of moments estimators. *Econometrica* 1982;50(22):1029–54.
- Hansen LP, Heaton J, Yaron A. Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics* 1996;14(3):262–80.
- Hautsch N. *Econometrics of Financial High-frequency Data*. Springer, 2012.
- Hodges JL, Lehmann EL. Estimates of location based on rank tests. *The Annals of Mathematical Statistics* 1963;34(2):598–611.
- Hong H, Scaillet O. A fast subsampling method for nonlinear dynamic models. *Journal of Econometrics* 2006;133:557–78.
- Hu F, Kalbfleisch JD. The estimating function bootstrap. *The Canadian Journal of Statistics* 2000;28(3):449–81.
- Huber PJ. Robust estimation of location parameter. *The Annals of Mathematical Statistics* 1964;35(1):73–101.
- Imbens G. One-step estimators for over-identified generalized method of moments models. *The Review of Economic Studies* 1996;64:359–83.
- Imbens G, Spady R, Johnson P. Information theoretic approaches to inference in moment condition models. *Econometrica* 1998;66:333–57.

- Inoue A, Shintani M. Bootstrapping GMM estimators for time series. *Journal of Econometrics* 2006;133:531–55.
- Kessler M, Lindner A, Sorensen M. *Statistical Methods for Stochastic Differential Equations*. Chapman and Hall/CRC, 2012.
- Khundi G, Rilstone P. Edgeworth expansions for GEL estimators. *Journal of Multivariate Analysis* 2012;106:118–46.
- Kitamura Y. Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics* 1997;25(5):2084–102.
- Kitamura Y, Stutzer M. An information-theoretic alternative to generalized method of moments estimation. *Econometrica* 1997;65(4):861–74.
- Lahiri SN. On Edgeworth expansion and Moving Block Bootstrap for studentized M-estimators in multiple linear regression models. *Journal of Multivariate Analysis* 1996;56(1):42–59.
- Lahiri SN. *Resampling Methods for Dependent Data*. Springer Series in Statistics, 2010.
- Lee S. Asymptotic refinements of a misspecification-robust bootstrap for GEL estimators. *Journal of Econometrics* 2016;192:86–104.
- Lehmann EL. *Testing Statistical Hypotheses*. New-York: John Wiley and Sons, 1959.
- Li DX, Turtle HJ. Semiparametric ARCH models: an estimating function approach. *Journal of Business & Economic Statistics* 2000;18(2):174–86.
- Moore DS. Generalized inverses, Wald’s method, and the construction of chi-squared tests of fit. *Journal of the American Statistical Association* 1977;357(72):131–7.
- Newey WK, McFadden D. Chapter 36, large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*; 1994. p. 2111 –245.
- Newey WK, Smith RJ. Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 2004;72(1):219–55.

- Newey WK, West KD. A simple, positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 1987;55:703–8.
- Otsu T. Generalized empirical likelihood inference for nonlinear and time series models under weak identification. *Econometric Theory* 2006;22:513–27.
- Paparoditis E, Politis DN. Tapered block bootstrap. *Biometrika* 2001;88(4):1105–19.
- Parente P, Smith R. Generalised empirical likelihood kernel block bootstrapping. REM working paper 055-2018; 2018a.
- Parente P, Smith R. Kernel block bootstrap. CEMMAP working paper CWP48/18; 2018b.
- Parente PMDC, Smith RJ. Quasi-maximum likelihood and the kernel block bootstrap for nonlinear dynamic models. *Journal of Time Series Analysis* 2021;42(4):377–405.
- Parzen E. On consistent estimates of the spectrum of a stationary time series. *The Annals of Mathematical Statistics* 1957;28(2):329–48.
- Parzen MI, Wei LJ, Ying Z. A resampling method based on pivotal estimating functions. *Biometrika* 1994;81(2):341–50.
- Pitman EJG. Statistics and science. *Journal of the American Statistical Association* 1957;279(52):322–30.
- Politis DN. Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory* 2011;27(4):703–44.
- Qin J, Lawless J. Empirical likelihood and general estimating equations. *The Annals of Statistics* 1994;22(1):300–25.
- Rio E. About the Lindeberg method for strongly mixing sequences. *ESAIM: Probability and Statistics* 1997;1:35–61.
- Salibian-Barrera M, Van Aelst S, Willems G. Principal components analysis based on multivariate MM-estimators with fast and robust bootstrap. *Journal of the American Statistical Association* 2006;475(101):1198–211.

- Salibian-Barrera M, Van Aelst S, Willems G. Fast and robust bootstrap. *Statistical Methods and Applications* 2008;17(1):41–71.
- Salibian-Barrera M, Zamar R. Bootstrapping robust estimates of regression. *The Annals of Statistics* 2002;30(2):556–82.
- Schweder T, Hjort NL. Confidence and likelihood. *Scandinavian Journal of Statistics* 2002;29:309–32.
- Schweder T, Hjort NL. *Confidence, Likelihood, Probability: Statistical Inference with Confidence Distributions*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016.
- Shao J. *Mathematical Statistics*. New-York: Springer, 1999.
- Shao J, Tu D. *The Jackknife and Bootstrap*. Springer Series in Statistics, Vol.85, 1995.
- Singh K, Xie M, Strawderman WE. Combining information from independent sources through confidence distributions. *The Annals of Statistics* 2005;33(1):159–83.
- Singh K, Xie M, Strawderman WE. Confidence Distribution (CD): Distribution estimator of a parameter. Institute of Mathematical Statistics, Lecture Notes - Monograph Series 2007;54:132–50.
- Smith RJ. Automatic positive semidefinite HAC covariance matrix and GMM estimation. *Econometric Theory* 2005;21(1):158–70.
- Smith RJ. GEL criteria for moment condition models. *Econometric Theory* 2011;27:1192–235.
- Taniguchi M, Kakizawa Y. *Asymptotic Theory of Statistical Inference for Time Series*. Springer New York, 2000.
- Veronese P, Melilli E. Fiducial and confidence distributions for real exponential families. *Scandinavian Journal of Statistics* 2015;42(2):471–84.
- Wilhelm D. Optimal bandwidth selection for robust generalized method of moments estimation. *Econometric Theory* 2015;31:1054–77.

Xie M, Singh K. Confidence distribution, the frequentist distribution estimator of a parameter: a review. *International Statistical Review* 2013;81(1):3–39.

Xie M, Singh K, Strawderman W. Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association* 2011;106:320–33.



# A Higher-Order Correct Fast Moving-Average Bootstrap for Dependent Data

## Online Supplementary Material

Davide La Vecchia, Alban Moor and Olivier Scaillet

---

This Supplementary Material contains three sections. In Section 1, we give an appendix containing the proof of Theorem 4. Section 2 contains the proofs of Lemmas, the key points of the derivation of the Edgeworth expansion for the quadratic statistics  $\hat{Q}$  and  $Q^*$ , as well as the proof of Corollary 5. In Section 3, we provide the detailed comparison of CPU time between FMB and MBB (Table 6), a complement of Tables 2 and 3 with a smaller sample size  $T = 150$  in Table 7 and a graphical illustration of the use of  $\tilde{Q}$  as in Remark 2.

---

### 1. Appendix : Proof of Theorem 4

By construction, the higher-order correctness of FMB CI and CR (for  $\beta_0$ ) entirely hinges on FMB applied to the test statistics  $\hat{S}(\beta_0)$  (see (4)) and  $\hat{Q}(\beta_0)$  (see (6)). Therefore, it follows from Theorem 4 as in Section 3.

The outline of the proof goes as follows. In Appendix A.1 below, we derive an Edgeworth expansion for  $\hat{S}(\beta_0)$  and  $\hat{Q}(\beta_0)$ . In Appendix A.2, we derive a similar Edgeworth expansion for the bootstrap counterparts  $S^*$  and  $Q^*$ . In Appendix A.3, we show that the difference between the two Edgeworth expansions is of order  $o_p(T^{-1/2}) + O_p(B_T^{-q}) + O_p(B_T/T)$ . The first term in this difference is smaller than the usual  $O_p(T^{-1/2})$  of the standard first-order asymptotics (central limit theorem). The improvement is essentially due to FMB being able to approximate accurately the third moment of the statistics, whilst CLT approximates only the first two moments. It yields higher-order correctness.

The second and third term of the difference has the same order than the bias and variance of the variance estimator, which scales  $\hat{S}(\beta_0)$  and  $\hat{Q}(\beta_0)$ . We get the order  $O_p(B_T^{-q}) + O_p(B_T/T)$ , where  $q$  is the Parzen exponent of the induced kernel  $k^*$  (see Section 4.1). With  $B_T = o(T^{1/2})$ ,  $q$  has to be larger than one for the overall error of FMB to be  $o_p(T^{-1/2})$ . We have to consider this aspect in the choice of the kernel  $k$  and of the bandwidth  $B_T$  entering the construction of the estimator  $\hat{\Omega}$  (see Section 3 for further details).

### A.1. Edgeworth expansion of the original sample statistics

In this section, we derive the Edgeworth expansions of  $\hat{S}(\beta_0)$  and  $\hat{Q}(\beta_0)$ . We state the following:

**Theorem 9.** *Under Assumptions 1–6, with  $s \geq 8$  and  $\log T = o(B_T)$ , we get the Edgeworth expansions:*

$$\Upsilon_{S,T}^\dagger(x) = \Phi(x) + T^{-1/2}p_1(x, \mathcal{K}_S^1)\phi(x) + (B_T/T)p_2(x, \mathcal{K}_S^2)\phi(x), \quad (16)$$

$$\Upsilon_{Q,T}^\dagger(x) = F_{\mathcal{X}_r^2}(x) + (B_T/T)p_Q(x, \mathcal{K}_Q)f_{\mathcal{X}_r^2}(x). \quad (17)$$

with uniform error bound:

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left[ \hat{S} \leq x \right] - \Upsilon_{S,T}^\dagger(x) \right| = o(B_T/T) + O(B_T^{-q}) + O(B_T/T), \quad (18)$$

$$\sup_{x \in \mathbb{R}_+} \left| \mathbb{P} \left[ \hat{Q} \leq x \right] - \Upsilon_{Q,T}^\dagger(x) \right| = o(B_T/T) + O(B_T^{-q}) + O(B_T/T). \quad (19)$$

In the expansions,  $p_1$  is an even polynomial in  $x$ , and  $p_2$  and  $p_Q$  are odd polynomials in  $x$ . These polynomials depend on vectors  $\mathcal{K}_S^1$ ,  $\mathcal{K}_S^2$ , and  $\mathcal{K}_Q$ , respectively containing the first three cumulants of  $\hat{S}$  and  $\hat{Q}$ . The uppercase  $\Phi$  and  $F_{\mathcal{X}_r^2}$  are respectively the c.d.f. of a  $\mathcal{N}(0, 1)$  and a  $\mathcal{X}_r^2$  random variable, the lowercase  $\phi$  and  $f_{\mathcal{X}_r^2}$  stand for their densities.

To prove the statements, we remark that  $\hat{S}(\beta_0)$  bears some similarity to the studentized smooth function of means of Götze and Künsch (1996). Therefore, our proof mainly relies on the same strategy as in Götze and Künsch paper. However, we have to discuss two important distinctions.

First, Götze and Künsch (1996) derive an Edgeworth expansion considering a class of studentizing factor of the form  $\hat{\zeta}^2 = \kappa_1^2 \sum_{s=1}^{T-1} k^*(s/B_T) \hat{\Gamma}_s(\hat{\beta})$ , with  $\hat{\Gamma}_s(\hat{\beta}) := T^{-1} \sum_{t=1}^{T-s} g_t(\hat{\beta}) g_{t+s}^\top(\hat{\beta})$ , as introduced by Andrews (1991). Our definition of the studentizing factor is different, but asymptotically

equivalent, since

$$\hat{\sigma}^2 = \kappa_1^2 \sum_{s=1-T}^{T-1} k_T^*(s/B_T) \hat{\Gamma}_s(\hat{\beta}),$$

where  $k_T^*(s/B_T) := (\kappa_2 B_T)^{-1} \sum_{t=\max[1-T, 1-T+s]}^{\min[T-1, T-1+s]} k(t-s/B_T) k(t/B_T)$  is a consistent approximation of  $k^*(s/B_T)$  by Riemann sum (Smith (2005)). In particular, the bias and variance of both studentizing factor have the same order, respectively  $O(B_T^{-q})$  and  $O(B_T/T)$ , where  $q$  is the Parzen exponent (see Section 3). As a consequence, both studentizing factors act equivalently on the error bound of the Edgeworth expansion for  $\hat{S}(\beta_0)$  (see (25)).

Second, from the convolution step of Equation (2), we are interested in  $T^{1/2}\bar{g}_T = T^{-1/2} \sum_{t=1}^T g_{T,t}$  instead of  $T^{1/2}\bar{g} = T^{-1/2} \sum_{t=1}^T g_t$ , as in Götze and Künsch (1996). Thus, we have to derive a valid Edgeworth expansion under this modification. To this end, we rewrite

$$T^{1/2}\bar{g}_T = T^{-1/2} \sum_{t=1}^T B_T^{-1/2} \sum_{s=t-T}^{t-1} k(s/B_T) g_{t-s} = T^{-1/2} \sum_{t=1}^T g_t B_T^{-1/2} \sum_{s=1-t}^{T-t} k(s/B_T).$$

In this representation, the kernel smoothing induces a tapering window  $w(t) := B_T^{-1/2} \sum_{s=1-t}^{T-t} k(s/B_T)$  on the summand time series in such a way that  $T^{1/2}\bar{g}_T = T^{-1/2} \sum_{t=1}^T w(t)g_t$ . Hence, we need to check that the regularity conditions given by Götze and Hipp (1994) hold for the tapered moment indicators, when they are assumed to be true for the original ones (Assumptions 1–6).

### Proof of Theorem 9.

To check Assumptions 1–6 for  $\{w(t)g_t\}$ , we label as “RC  $i$ ” the regularity conditions defined in Assumptions  $i$ ; for instance “RC 1” represents the regularity condition of Assumption 1.

When we assume that RC 1 and RC 2 hold for the process  $\{g_t\}$ , we can easily verify that they are true also for  $\{w(t)g_t\}$ , given that  $w(t) < \infty, \forall t$ . To check RC 3, consider that

$$\mathbb{E} \left\| w(t)g_t - w(t)g_{t,m}^\dagger \right\| \leq \left| \max_{t=1, \dots, T} w(t) \right| \mathbb{E} \left\| g_t - g_{t,m}^\dagger \right\| \leq \left| \max_{t=1, \dots, T} w(t) \right| \delta^{-1} \exp(-\delta m)$$

by Assumption 3 on  $\{g_t\}$ . Thus, the exponential rate of decay of the approximation error is not affected by tapering, and we can always take the process  $\{w(t)g_{t,m}^\dagger\}$  to approximate  $\{w(t)g_t\}$ . Without loss of generality, let us take  $\mathcal{D}_t := \sigma\{g_{t,0}^\dagger\}$ , and note that both  $w(t)g_{t,0}^\dagger$  and  $g_{t,0}^\dagger$  are  $\mathcal{D}_t$ -measurable. Then, RC

4 and RC 6 follow immediately, by means of Assumptions 4 and 6 on the same sigma-fields  $\mathcal{D}_t$ . The verification of RC 5 can be found in Section 2 of the Supplementary Material. Here, we summarize the result in Lemma 10:

**Lemma 10.** *If RC 5 holds for  $\{g_t\}_{t=1}^T$ , then it holds for  $\{w(t)g_t\}_{t=1}^T$ , with  $w(t) = B_T^{-1/2} \sum_{s=1-t}^{T-t} k(s/B_T)$ .*

Therefore, the validity of Assumptions 1–6 and Lemma 10 imply that we can suitably approximate the probability distribution of  $T^{1/2}\bar{g}_T$  by an Edgeworth expansion of the same kind as the one for  $T^{1/2}\bar{g}$ , when the latter exists. To derive the Edgeworth expansion for  $\hat{S}$ , we have to adapt the expansion for  $T^{1/2}\bar{g}_T$ , as defined in Götze and Hipp (1994), to accommodate our studentization as in (4). To this end, we modify the proof of Götze and Künsch (1996) to consider the tapering related to  $\{w(t)\}_{t=1}^T$ . To define our Edgeworth expansion, we need the following quantities:

$$\pi_T^\dagger := T^{-1} \sum_{t=1}^T w(t) \sum_{j=1}^{T-B_T} \sum_{s=-B_T}^{B_T} k_T^*(s/B_T) \mathbb{E}[g_t g_j g_{j+s}], \quad (20)$$

$$V_T := \sum_{s=-B_T}^{B_T} k_T^*(s/B_T) T^{-1} \sum_{j=1}^{T-B_T} \left( g_j(\hat{\beta}) g_{j+s}(\hat{\beta}) - \mathbb{E}[g_0 g_s] \right), \quad (21)$$

$$\sigma_T^{\dagger 2} := \sum_{s=-T}^T \left( 1 - \frac{|s|}{T} \right) w(s) \mathbb{E}[g_0 g_s], \quad (22)$$

$$\tau_{1T}^2 := \sum_{s=-T}^T k_T^*(s/B_T) \mathbb{E}[g_0 g_s], \quad (23)$$

$$\mu_{3,T}^\dagger := T^2 \mathbb{E}[\bar{g}_T^3]. \quad (24)$$

We take the expansion  $\hat{\sigma}^2 = \sigma_T^{\dagger 2} + V_T + (\tau_{1T}^2 - \sigma_T^{\dagger 2}) + O_p(B_T/T)$ , where  $(\tau_{1T}^2 - \sigma_T^{\dagger 2})$  reflects the bias of the estimator  $\hat{\sigma}$ . Then, we approximate  $\hat{\sigma}^{-1}$  by the linear part of the Taylor series  $\hat{\sigma}^{-1} = \sigma_T^{\dagger -1} + \sum_{j=1}^{\infty} h^{(j)}(\sigma_T^{\dagger 2})(\hat{\sigma}^2 - \sigma_T^{\dagger 2})^j/j!$ , where  $h(x) = x^{-1/2}$  and  $h^{(j)}$  is its  $j$ -th derivative, getting  $\hat{\sigma}^{-1} = \sigma_T^{\dagger -1} - (1/2)(V_T + (\tau_{1T}^2 - \sigma_T^{\dagger 2}) + O_p(B_T/T))\sigma_T^{\dagger -3}$ . Taking the product with  $T^{1/2}\bar{g}_T$ , we obtain:

$$\hat{S} = T^{1/2}\bar{g}_T\sigma_T^{\dagger -1} - (1/2)T^{1/2}\bar{g}_TV_T\sigma_T^{\dagger -3} - (1/2)T^{1/2}\bar{g}_T(\tau_{1T}^2 - \sigma_T^{\dagger 2})\sigma_T^{\dagger -3} + O_p(B_T/T), \quad (25)$$

$$= E_T - (1/2)T^{1/2}\bar{g}_T(\tau_{1T}^2 - \sigma_T^{\dagger 2})\sigma_T^{\dagger -3} + O_p(B_T/T), \quad (26)$$

where  $E_T := T^{1/2}\bar{g}_T\sigma_T^{\dagger-1} - (1/2)T^{1/2}\bar{g}_TV_T\sigma_T^{\dagger-3}$ . Following the standard argument of [Chibisov \(1972\)](#), the Edgeworth expansion of  $\hat{S}$  coincides with the one of  $E_T$  up to the order  $O(T^{1/2}\bar{g}_T(\tau_{1T}^2 - \sigma_T^{\dagger 2})\sigma_T^{\dagger-3} + B_T/T)$ . As  $T^{1/2}\bar{g}_T$  is  $O_p(1)$ , the term  $T^{1/2}\bar{g}_T(\tau_{1T}^2 - \sigma_T^{\dagger 2})\sigma_T^{\dagger-3}$  is of order  $O_p(|\tau_{1T}^2 - \sigma_T^{\dagger 2}|) = O_p(B_T^{-q})$ , and corresponds to the bias of the variance estimator.

To assess the order of the FMB approximation error, it is more convenient to work directly with the Fourier transform of the Edgeworth expansion; see [Götze and Künsch \(1996\)](#), p. 1919. Therefore, we define the Edgeworth expansion of  $\hat{S}$ , say  $\Upsilon_{S,T}^\dagger$ , in terms of its Fourier transform. We do so by collecting the expansion of  $E_T$  and the error terms discussed above:

$$\begin{aligned} \mathcal{E}_T^\dagger(\tau) &:= \int \exp(i\tau^\top x) d\Upsilon_{S,T}^\dagger(x) \\ &= \left\{ 1 + \frac{1}{\sqrt{T}} \frac{1}{\sigma_T^{\dagger 3}} \left[ \left( \frac{\mu_{3,T}^\dagger}{6} - \frac{\pi_T^\dagger}{2} \right) (i\tau)^3 - \frac{(i\tau) \pi_T^\dagger}{2} \right] + O(B_T^{-q}) + O(B_T/T) \right\} \exp\left(-\frac{\tau^2}{2}\right). \end{aligned} \quad (27)$$

Invoking Esseen Lemma, we get the approximation error of  $\Upsilon_{S,T}^\dagger$  as in [Theorem 9](#), Equation (16).

We provide also a sketch for the derivation of Edgeworth expansion for  $\hat{Q}$ . The detailed calculation is available in Section 2 of the Supplementary Material. To get an Edgeworth expansion for  $\hat{Q}$ , namely (17) in [Theorem 9](#), we make use of the univariate Edgeworth expansion  $\Upsilon_{S,T}^\dagger$ , explicitly defined in (16).

First, when  $T^{1/2}\bar{g}_T$  is multivariate of dimension  $r$ , expansions of the same form as in [Theorem 9](#) Equation (16) hold (up to constants independent of  $T$ ) for any linear combination  $T^{1/2}v^\top\bar{g}_T/(v^\top\hat{\Omega}v)^{1/2}$  ([Götze and Künsch \(1996\)](#)).

Second, as  $\hat{\Omega}$  is symmetric positive semi-definite by construction, we have a unique symmetric positive semi-definite square root  $\hat{\Omega}^{1/2}$ , which admits an inverse. Thus, we have a vector  $\hat{Q}^{1/2} := \hat{\Omega}^{-1/2}T^{1/2}\bar{g}_T$ , such that  $\hat{Q}^{1/2}\hat{Q}^{1/2} = \hat{Q}$ . Projecting the vector  $T^{1/2}\bar{g}_T$  onto the orthonormal eigenvectors of  $\hat{\Omega}$ , we get  $\hat{Q}^{1/2} = \Lambda^{-1/2}P^\top T^{1/2}\bar{g}_T = (T^{1/2}\bar{g}_T^\top v_1/\lambda_1^{1/2}, \dots, T^{1/2}\bar{g}_T^\top v_r/\lambda_r^{1/2})^\top$ , where  $\{\lambda_1, \dots, \lambda_r\}$  are the eigenvalues of  $\hat{\Omega}$  corresponding to its normalized eigenvectors  $\{v_1, \dots, v_r\}$ ,  $\Lambda := \text{diag}(\lambda_1, \dots, \lambda_r)$ , and  $P := (v_1, \dots, v_r)$ . As  $T^{1/2}\bar{g}_T^\top v_j/\lambda_j^{1/2} = T^{1/2}v^\top\bar{g}_T/(v^\top\hat{\Omega}v)$  when we choose  $v = v_j$  for each  $j = 1, \dots, r$ , we directly see that there exist expansions of the same form as in [Theorem 9](#) Equation (16) for each element of  $\hat{Q}^{1/2}$ . Furthermore, taking any vector  $c$  such that  $\|c\| = 1$ , there exists a univariate expansion

of the same form as in Theorem 9 Equation (16) for  $c^\top \hat{Q}^{1/2} / (c^\top I_r c)^{1/2} = c^\top \hat{Q}^{1/2}$ , as the variance estimator of  $\hat{Q}^{1/2}$  is the  $r$ -dimensional identity matrix  $I_r$  by definition. By the Cramér-Wold device, the characteristic function of  $\hat{Q}^{1/2}$  is  $\mathcal{E}_r(\tau c) := \mathbb{E}[\exp(i\tau c^\top \hat{Q}^{1/2})]$ , where  $\tau$  is a scalar. As a consequence, there exists an expansion of the same form as in (27) for  $\mathcal{E}_r(\tau c)$ . Taking the inverse Fourier transform of this approximation, we approximate the probability distribution of  $\hat{Q}^{1/2}$  by a multivariate Edgeworth expansion  $\Upsilon_{r,T}(z) := \Phi_r(z) + T^{-1/2} p_1(z) \phi_r(z) + (B_T/T) p_2(z) \phi_r(z)$ , where  $\phi_r(z)$  is the normal density on  $\mathbb{R}^r$  with mean zero and variance  $I_r$  and  $\Phi_r(z)$  is the corresponding cdf. In this expansion,  $p_1$  is an even polynomial in  $z$  and  $p_2$  is an odd polynomial in  $z$ . The error of this approximation is by construction the same as the univariate one, as it is built with the same type of variance estimator.

Finally, we consider the statistic of interest  $\hat{Q}$  and we work on a variable transform. Indeed,

$$\sup_{x \in \mathbb{R}^r} \left| \int_{t \leq x} d\Upsilon_{r,T} - \int_{t \leq y} dF_{\hat{Q}^{1/2}} \right| = o(B_T/T) + O(B_T^{-q}) + O(B_T/T)$$

implies  $\sup_{y \in \mathbb{R}_+} \left| \int_{\{t: t^\top t \leq y\}} d\Upsilon_{r,T} - \int_{t \leq y} dF_{\hat{Q}} \right| = o(B_T/T) + O(B_T^{-q}) + O(B_T/T)$ . Along the same lines of Chandra and Ghosh (1979) (see Section 2 of the Supplementary Material), we identify an expansion  $\Upsilon_{Q,T}^\dagger$  such that  $\sup_{y \in \mathbb{R}_+} \left| \int_{t \leq y} d\Upsilon_{Q,T}^\dagger - \int_{t \leq y} dF_{\hat{Q}} \right| = o(B_T/T) + O(B_T^{-q}) + O(B_T/T)$ . In this expansion, the elementary probability measure is  $\mathcal{X}_r^2$  instead of the Gaussian  $\Phi_r$ , and the term of order  $T^{-1/2}$  disappears because  $p_1$  is even in  $z$ .

In principle, we can find explicitly  $p_1$  by inverting the Fourier transform in (27). However, we do not need  $p_1$  in this proof since we work directly with the Fourier transform. Likewise, we only give the polynomials  $p_2$  and  $p_Q$  formally (implicitly), as we do not need their explicit form in the sequel of the proof. What matters is the order of error to which they correspond, namely  $O(B_T/T)$ .

## A.2. Edgeworth expansion of the bootstrap sample statistics

In the sequel,  $\mathbb{E}^*$  denotes the expectation under the bootstrap probability measure as in (14). We state the following:

**Theorem 11.** Under Assumptions 1, 2, 7 and if  $\mathbb{E} \left[ \|g_t\|^{\bar{q}s+\delta} \right] < \infty$ , for  $\delta > 0$ ,  $s \geq 8$ , and  $\bar{q} \geq 3$ :

$$\Upsilon_{S,T}^*(x) = \Phi(x) + T^{-1/2}p_1(x, \mathcal{K}_S^{1*})\phi(x) + T^{-1}p_2(x, \mathcal{K}_S^{2*})\phi(x), \quad (28)$$

$$\Upsilon_{Q,T}^*(x) = F_{\mathcal{X}_r^2}(x) + T^{-1}p_Q(x, \mathcal{K}_Q^*)f_{\mathcal{X}_r^2}(x), \quad (29)$$

with uniform error bound:

$$\sup_{x \in \mathbb{R}} |\mathbb{P}^* [S^* \leq x] - \Upsilon_{S,T}^*(x)| = o_p(T^{-1}), \quad (30)$$

$$\sup_{x \in \mathbb{R}_+} |\mathbb{P}^* [Q^* \leq x] - \Upsilon_{Q,T}^*(x)| = o_p(T^{-1}), \quad (31)$$

where  $\mathcal{K}_S^{1*}$ ,  $\mathcal{K}_S^{2*}$  and  $\mathcal{K}_Q^*$  are the bootstrap counterparts of the vectors  $\mathcal{K}_S^1$ ,  $\mathcal{K}_S^2$  and  $\mathcal{K}_Q$ .

**Proof of Theorem 11.** To prove the theorem, we need to define the Edgeworth expansion for the bootstrap statistic. To this end, we need the same arguments as for the original statistic. Namely, we need to check RC 1–6 for the bootstrap sample  $\{g_{T,t}^*\}_{t=1}^T$ , conditionally on  $\{g_t(\beta_0)\}_{t=1}^T$ , uniformly on a set whose probability tends to one. When  $p = r$  (exactly-identified moment conditions), RC 1 holds by the nature of the estimating equations, as  $\mathbb{E}^* \left[ g_{T,t}^* \right] = \bar{g}_T(\hat{\beta}) = 0$ . In the case of over-identification, we have to recenter the bootstrap statistic so that RC 1 holds. For RC 2, we have  $\mathbb{E}^* \left[ \left\| g_{T,t}^* \right\|^s \right] = T^{-1} \sum_{t=1}^T \|g_{T,t}\|^s$ . For the unconditional moments, we get  $\mathbb{E} \left[ \mathbb{E}^* \left[ \left\| g_{T,t}^* \right\|^s \right] \right] = T^{-1} \sum_{t=1}^T \mathbb{E} [\|g_{T,t}\|^s] = \mathbb{E} [\|g_{T,t}\|^s] < \infty$ , from Assumption 2 on  $g_t$  and by Minkowski inequality. Then, assuming  $\mathbb{E} [\|g_t\|^{\bar{q}s}] < \infty$ , for  $\bar{q} \geq 3$ , we obtain  $\mathbb{E}^* \left[ \left\| g_{T,t}^* \right\|^s \right] - \mathbb{E} \left[ \mathbb{E}^* \left[ \left\| g_{T,t}^* \right\|^s \right] \right] = O_p(T^{-1/2})$ . As a consequence,  $\mathbb{E}^* \left[ \left\| g_{T,t}^* \right\|^s \right]$  is bounded with probability tending to one.

RC 3, 4, and 6, trivially hold by independence of the  $g_{T,t}^*$  under the bootstrap probability measure as in (14), when  $\mathcal{D}_t := \sigma\langle g_{T,t}^* \rangle$  for  $t = 1, \dots, T$ . Finally, as the bootstrap random variables are i.i.d, we can show that the standard Cramér condition holds asymptotically, instead of checking RC 5. Thus, the Cramér condition in Assumption 7 is sufficient to verify RC 5 for the bootstrap process, with the same sub-sigma-fields  $\mathcal{D}_t$ . Indeed, from Assumption 7, we have :

$$f(\beta_0) = \limsup_{T \rightarrow \infty} \sup_{b < |\tau| < e^{\delta T}} |T^{-1} \sum_{t=1}^T \exp(i\tau g_{T,t}(\beta_0))| < 1, \text{ a.s.}$$

We have to make sure that the same limit holds when  $g_{T,t}(\beta)$  is evaluated at  $\hat{\beta}$  instead of  $\beta_0$ , since we resample the  $\{g_{T,t}(\hat{\beta})\}_{t=1}^T$  in FMB. From Assumption 7, there exists a constant  $c > 0$  such that  $\sup_{\|\beta-\beta_0\|\leq c} f(\beta) < 1$ . As  $\lim_{T\rightarrow\infty} \|\hat{\beta} - \beta_0\| \leq c$ , we have a.s.  $f(\hat{\beta}) \leq \sup_{\|\beta-\beta_0\|\leq c} f(\beta) < 1$ , which verifies the necessary Cramér condition.

We construct the Edgeworth expansion  $\Upsilon_{S,T}^*$  of the bootstrap statistic  $S^*$  in the same way as the one for the original sample statistic  $\hat{S}$  in (27) above. Thus, we need to define the bootstrap counterpart of the quantities (20)—(24):

$$\pi_T^* := T^{-1} \sum_{t=1}^T \sum_{j=1}^T \sum_{s=1}^T \mathbb{E}^* [g_{T,t}^* g_{T,j}^* g_{T,s}^*], \quad (32)$$

$$\sigma_T^{*2} := \mathbb{E}^* \left[ (T^{1/2} \tilde{g}_T^*)^2 \right], \quad (33)$$

$$\tau_{1T}^{*2} := \sum_{s=-T}^T \mathbb{E}^* [g_0^* g_s^*], \quad (34)$$

$$\mu_{3,T}^* := T^{1/2} \mathbb{E}^* \left[ (T^{1/2} \tilde{g}_T^*)^3 \right]. \quad (35)$$

Then, we can define the bootstrap Edgeworth expansion via its Fourier transform:

$$\begin{aligned} \mathcal{E}_T^*(\tau) &:= \int \exp(i\tau^\top x) d\Upsilon_{S,T}^*(x) \\ &= \left\{ 1 + \frac{1}{\sqrt{T}} \frac{1}{\sigma_T^{*3}} \left[ \left( \frac{\mu_{3,T}^*}{6} - \frac{\pi_T^*}{2} \right) (i\tau)^3 - \frac{(i\tau) \pi_T^*}{2} \right] + O_p(|\sigma_T^{*2} - \tau_{1T}^{*2}|) + O_p(1/T) \right\} \exp\left(-\frac{\tau^2}{2}\right). \end{aligned}$$

We have the same error bound as in Theorem 9 up to four aspects. First,  $\Upsilon_{S,T}^*$  is a random measure because it is conditional to the original sample. Second, by independence of the bootstrap variables, we have  $\sigma_T^{*2} = \tau_{1T}^{*2}$ , so the contribution of the  $|\sigma_T^{*2} - \tau_{1T}^{*2}|$  term to the error disappears. Third, in contrary to (27), we have the order  $O_p(1/T)$  instead of  $O_p(B_T/T)$ , as it is an Edgeworth expansion for the studentized mean of i.i.d. random variables. Finally, by independence of the  $\{g_{T,t}^*\}$ , we have



$\pi_T^* = T^{-1} \sum_{t=1}^T \sum_{j=1}^T \sum_{s=1}^T \mathbb{E}^* [g_{T,t}^* g_{T,j}^* g_{T,s}^*] = T^2 \mathbb{E}^* [\bar{g}_T^{*3}] = \mu_{3,T}^*$ . Altogether, it leads to:

$$\begin{aligned} \mathcal{E}_T^*(\tau) &= \left\{ 1 + \frac{1}{\sqrt{T}} \frac{1}{\sigma_T^{*3}} \left[ \left( \frac{\mu_{3,T}^*}{6} - \frac{\mu_{3,T}^*}{2} \right) (i\tau)^3 - \frac{(i\tau) \mu_{3,T}^*}{2} \right] + O_p(1/T) \right\} \exp\left(-\frac{\tau^2}{2}\right) \\ &= \left\{ 1 + \frac{\mu_{3,T}^*}{\sqrt{T} \sigma_T^{*3}} \left[ -\frac{(i\tau)^3}{3} - \frac{(i\tau)}{2} \right] + O_p(1/T) \right\} \exp\left(-\frac{\tau^2}{2}\right). \end{aligned} \quad (36)$$

To derive an Edgeworth expansion for  $Q^*$ , we proceed along the same lines as for the original sample statistic  $\hat{Q}$ , working with univariate Edgeworth expansions for studentized linear combinations of the form  $T^{1/2} v^\top \bar{g}_T^*$ .

### A.3. Higher-order correctness of FMB

To prove part (i) of Theorem 4, we essentially need the convergence of the terms of order  $T^{-1/2}$  in  $\Upsilon_{S,T}^\dagger$  (as in (16)) and  $\Upsilon_{S,T}^*$  (as in (28)) to the same quantities. Indeed, in view of Theorems 9 and 11, the condition  $\text{plim}_{T \rightarrow \infty} \sup_{x \in \mathbb{R}} |p_1(x, \mathcal{K}_S^{1*}) - p_1(x, \mathcal{K}_S^1)| = 0$ , or equivalently  $\mathcal{K}_S^{1*} = \mathcal{K}_S^1 + o_p(1)$ , is sufficient to show higher-order refinements of FMB. We give the details of this convergence in the proof of Lemma 12 in Section 2 of the Supplementary Material.

**Lemma 12.** *Under Assumptions 1–7 and if  $\mathbb{E} \left[ \|g_t\|^{\bar{q}^{s+\delta}} \right] < \infty$ , for  $\delta > 0$ ,  $s \geq 8$ , and  $\bar{q} \geq 3$ :*

$$\sup_{x \in \mathbb{R}} |\Upsilon_{S,T}^\dagger(x) - \Upsilon_{S,T}^*(x)| = o_p(T^{-1/2}) + O_p(B_T^{-q}) + O_p(B_T/T).$$

Collecting the error bounds of Theorem 9, Theorem 11, and Lemma 12, we use the triangular inequality to get:

$$\begin{aligned} \sup_{x \in \mathbb{R}} \left| \mathbb{P}^* [S^* \leq x] - \mathbb{P} [\hat{S} \leq x] \right| &\leq \sup_{x \in \mathbb{R}} \left| \mathbb{P}^* [S^* \leq x] - \Upsilon_{S,T}^*(x) \right| + \sup_{x \in \mathbb{R}} \left| \Upsilon_{S,T}^*(x) - \Upsilon_{S,T}^\dagger(x) \right| \\ &\quad + \sup_{x \in \mathbb{R}} \left| \Upsilon_{S,T}^\dagger(x) - \mathbb{P} [\hat{S} \leq x] \right|, \\ &= o_p(T^{-1/2}) + O_p(B_T^{-q}) + O_p(B_T/T). \end{aligned}$$

The proof of part (ii) of Theorem 4 is more direct, as the even polynomial  $p_1$  in Theorems 9 and 11 deletes the  $T^{-1/2}$  term from the Edgeworth expansion of  $\hat{Q}$ . Hence, only the higher-order

terms remain. As a consequence, we immediately verify that  $\sup_{x \in \mathbb{R}^+} |\Upsilon_{Q,T}^*(x) - \Upsilon_{Q,T}^\dagger(x)|$  is of order  $O_p(B_T^{-q}) + O_p(B_T/T)$ . Then, we prove part (ii) making use of Theorems 9, 11, and the triangular inequality:

$$\begin{aligned} \sup_{x \in \mathbb{R}^+} \left| \mathbb{P}^*[Q^* \leq x] - \mathbb{P}[\hat{Q} \leq x] \right| &\leq \sup_{x \in \mathbb{R}^+} |\mathbb{P}^*[Q^* \leq x] - \Upsilon_{Q,T}^*(x)| + \sup_{x \in \mathbb{R}^+} |\Upsilon_{Q,T}^*(x) - \Upsilon_{Q,T}^\dagger(x)| \\ &\quad + \sup_{x \in \mathbb{R}^+} \left| \Upsilon_{Q,T}^\dagger(x) - \mathbb{P}[\hat{Q} \leq x] \right|, \\ &= O_p(B_T^{-q}) + O_p(B_T/T). \end{aligned}$$

## 2. Miscellaneous proofs

### SM.1. Proof of Lemma 10

Without loss of generality, take  $D_t := \sigma\langle g_{t,0}^\dagger \rangle$ . Then, let us define  $\mathcal{F}_m$  as the set of index  $k \in \{-m, \dots, m\}$  such that  $g_{t+k}$  is  $\{\mathcal{D}_j : j \neq t\}$ -measurable. We can now rewrite RC 5 as follows:

$$\begin{aligned} \exp(-\delta) &\geq \mathbb{E}[|\mathbb{E}[\exp(i\tau^\top(g_{t-m} + \dots + g_{t+m})) | \mathcal{D}_j : j \neq t]|], \\ &= \mathbb{E}[|\mathbb{E}[\exp(i\tau^\top \sum_{k \in \mathcal{F}_m} g_{t+k}) \exp(i\tau^\top \sum_{\ell \in \mathcal{F}_m^c} g_{t+\ell}) | \mathcal{D}_j : j \neq t]|], \\ &= \mathbb{E}[|\exp(i\tau^\top \sum_{k \in \mathcal{F}_m} g_{t+k}) \mathbb{E}[\exp(i\tau^\top \sum_{\ell \in \mathcal{F}_m^c} g_{t+\ell}) | \mathcal{D}_j : j \neq t]|], \\ &= |\mathbb{E}[\exp(i\tau^\top \sum_{\ell \in \mathcal{F}_m^c} g_{t+\ell}) | \mathcal{D}_j : j \neq t]| \mathbb{E}[|\exp(i\tau^\top \sum_{k \in \mathcal{F}_m} g_{t+k})|], \\ &= |\mathbb{E}[\exp(i\tau^\top \sum_{\ell \in \mathcal{F}_m^c} g_{t+\ell}) | \mathcal{D}_j : j \neq t]|. \end{aligned}$$

Therefore, we have to show that  $|\mathbb{E}[\exp(i\tau^\top \sum_{\ell \in \mathcal{F}_m^c} g_{t+\ell}) | \mathcal{D}_j : j \neq t]| \leq \exp(-\delta)$  implies  $|\mathbb{E}[\exp(i\tau^\top \sum_{\ell \in \mathcal{F}_m^c} w(t+\ell)g_{t+\ell}) | \mathcal{D}_j : j \neq t]| \leq \exp(-\delta)$ . First, the expectation is taken with respect to the same conditional probability measure, say  $F(g_{t+\ell} : \ell \in \mathcal{F}_m^c)$  for convenience of notation. Then,  $|\mathbb{E}[\exp(i\tau^\top \sum_{\ell \in \mathcal{F}_m^c} w(t+\ell)g_{t+\ell}) | \mathcal{D}_j : j \neq t]| = |\int \exp(i\tau^\top \sum_{\ell \in \mathcal{F}_m^c} w(t+\ell)g_{t+\ell}) dF(g_{t+\ell} : \ell \in \mathcal{F}_m^c)|$ . It is now clear that the weights  $w(t+\ell)$  are rescaling horizontally the same characteristic function. As  $w(t) \neq 0, \forall t$ , the complex modulus of the characteristic function never reaches unity (corresponding to the infimum  $\delta = 0$ ). Thus, the range of  $\mathbb{E}[\exp(i\tau^\top \sum_{\ell \in \mathcal{F}_m^c} g_{t+\ell}) | \mathcal{D}_j : j \neq t]$  is the same than the range of

$\mathbb{E}[\exp(i\tau^\top \sum_{\ell \in \mathcal{F}_m^c} w(t+\ell)g_{t+\ell}) | \mathcal{D}_j : j \neq t]$ . As a consequence, the upper bound of their complex modulus is the same,  $\forall \|\tau\| > 0$ .

For the second part of the condition, it is immediate that  $\liminf_{T \rightarrow \infty} T^{-1} \text{Var} \left[ \sum_{t=1}^T w(t)g_t \right] > 0$ , if the same condition holds for  $\{g_t\}_{t=1}^T$ , as  $w(t) \neq 0, \forall t$ .

### SM.2. Edgeworth expansion for the quadratic statistics $\hat{Q}$ and $Q^*$

For the completeness of the presentation, we supplement here the arguments of the paragraph before Theorem 9, following closely the lines of Chandra and Ghosh (1979) with their notation. The changes concern the order of approximations in the Edgeworth expansion and the specification of the statistic of interest.

First, consider the Edgeworth expansion  $\Upsilon_{r,T}(z) := \Phi_r(z) + T^{-1/2}p_1(z)\phi_r(z) + (B_T/T)p_2(z)\phi_r(z)$ , which approximates the distribution of the  $r$ -dimensional statistic  $\hat{Q}^{1/2} = \Lambda^{-1/2}P^\top T^{1/2}\bar{g}_T = (T^{1/2}\bar{g}_T^\top v_1/\lambda_1^{1/2}, \dots, T^{1/2}\bar{g}_T^\top v_r/\lambda_r^{1/2})^\top$ , with an error of order  $o(B_T/T) + O(|\tau_{1T}^2 - \sigma_T^{\dagger 2}|)$ . For ease of notation, let us take  $Z := \hat{Q}^{1/2}$ . In the paragraph before Theorem 9, we obtain the bound  $\sup_{y \in \mathbb{R}_+} |\int_{\{t: t \leq y\}} d\Upsilon_{r,T} - \int_{t \leq y} dF_{\hat{Q}}| = o(B_T/T) + O(|\tau_{1T}^2 - \sigma_T^{\dagger 2}|)$ . Thus, we need to identify an expansion of the form  $\Upsilon_{Q,T}^\dagger(x) = F_{\mathcal{X}_r^2}(x) + (B_T/T)p_Q(x, \mathcal{K}_Q)f_{\mathcal{X}_r^2}(x)$  such that  $\sup_{y \in \mathbb{R}_+} |\int_{t \leq y} d\Upsilon_{Q,T}^\dagger - \int_{t \leq y} dF_{\hat{Q}}| = o(B_T/T) + O(|\tau_{1T}^2 - \sigma_T^{\dagger 2}|)$ .

For  $s = 5$ , our  $(s - 3)$ -order Edgeworth expansion  $\Upsilon_{r,T}(z)$  has a density of the form  $\tilde{v}_{r,T}(z) := [1 + T^{-1/2}\tilde{p}_1(z) + (B_T/T)\tilde{p}_2(z)]\phi_r(z)$ , where the degree of each term in the polynomials  $\tilde{p}_1(z)$  and  $\tilde{p}_2(z)$  is respectively odd and even.

Now, we explain how to obtain the expansion  $\Upsilon_{Q,T}^\dagger(v)$ , with density of the form  $\tilde{v}_{Q,T}^\dagger(v) := [1 + (B_T/T)\tilde{p}_Q(v)]f_{\mathcal{X}_r^2}(v)$ . This density is similar to  $\tilde{v}_{r,T}(z)$ , except for the  $\mathcal{X}_r^2$  measure replacing the Gaussian  $\Phi_r$  measure, and the cancellation of the term of order  $T^{-1/2}$ .

Consider the multivariate polar transformation  $T_1$ , which sends  $Z$  to  $(R, \beta)^\top := (R, \beta^{(1)}, \dots, \beta^{(r-1)})^\top$  via  $z_1 = R \prod_{i=1}^{r-1} \cos(\beta^{(i)})$  and  $z_j = R \sin(\beta^{(r-j+1)}) \prod_{i=1}^{r-j} \cos(\beta^{(i)})$ , where  $2 \leq j \leq r$ ,  $R \in \mathbb{R}_+^*$  is the radius, and  $-\pi/2 < \beta^{(i)} < \pi/2$  for  $i = 1, \dots, r-2$ ,  $0 \leq \beta^{(r-1)} \leq 2\pi$  are the angles. Then, for a vector of non-negative integers  $A := (a_1, \dots, a_r)$ , we write  $\mathcal{R}(A) = \mathcal{R}(R, \beta, Z) = R^{a_0} \prod_{i=1}^r (z_i/R)^{a_i} = \prod_{i=1}^r z_i^{a_i}$ , where  $a_0 = \sum_{i=1}^r a_i$ . We will use notation  $\mathcal{R}(A)$  even when  $a_0 \neq \sum_{i=1}^r a_i$ . We say that  $\mathcal{R}(A)$  is odd if at least one element of  $A$  is odd, and more generally, we say that the expression  $R^{a_0} \prod_{i=1}^{r-1} \cos(\beta^{(i)})^{a_i} \sin(\beta^{(i)})^{b_i}$  is odd if at least one of  $\{b_1, \dots, b_{r-1}, a_{r-1}\}$  is odd. The Jacobian of  $T_1$

is  $R^{r-1} \prod_{i=1}^{r-2} \cos(\beta^{(i)})^{r-i-1}$ , say  $R^{r-1}J(\beta)$  for ease of notation, and that an odd  $\mathcal{R}(A)$  implies an odd  $\mathcal{R}(A)J(\beta)$ . Finally, we write  $R_{i,j}(R, \beta, Z)$  a finite sum of constant multiples of terms of the form  $\mathcal{R}(A)$ , and say that  $R_{i,j}(R, \beta, Z)$  is odd if every such  $\mathcal{R}(A)$  are odd. In the following proof, various expressions of the form  $R_{i,j}(R, \beta, Z)$  occur with the two following properties. First,  $R_{i,j}(R, \beta, Z)$  happens to be odd when  $j$  is odd. Second, if  $j$  is even and  $R_{i,j}(R, \beta, Z)$  includes some (constant multiple) of  $\mathcal{R}(A)$  which fails to be odd, then the power  $a_0$  in the corresponding  $\mathcal{R}(A)$  will be even. Let us define the set  $M_T := \{Z : \|Z\|^2 < (s-1) \log T\}$ , and for any  $\bar{B} \subset \mathbb{R}$ , write  $\bar{B}_T := \{Z : Z^\top Z/2 \in \bar{B}\}$ . It is sufficient to exhibit an expansion  $\tilde{v}_{Q,T}^\dagger(v)$  such that  $\sup\{|\int_{\{\bar{B}_T \cap M_T\}} \tilde{v}_{r,T} - \int_{\{\bar{B}\}} \tilde{v}_{Q,T}^\dagger| : \bar{B} \in \mathcal{B}\} = o(B_T/T) + O(|\tau_{1T}^2 - \sigma_T^{\dagger 2}|)$ , where  $\mathcal{B}$  are the Borel sets on  $\mathbb{R}$ . Applying  $T_1$ , we get

$$\int_{\{\bar{B}_T \cap M_T\}} \tilde{v}_{r,T} = \int_{T_1(\bar{B}_T \cap M_T)} R^{r-1}J(\beta)[1 + T^{-1/2}R_{1,1}(R, \beta, Z) + (B_T/T)R_{1,2}(R, \beta, Z)] \exp(-R^2/2).$$

Then, let us apply the transformation  $T_2(R, \beta, Z) := (R', \beta, Z)$ , where  $R' := (T_1^{-1}(R, \beta, Z)^\top T_1^{-1}(R, \beta, Z)/2)^{1/2}$ . It can be shown that  $R = R'(1 + T^{-1/2}R_{3,1}(R, \beta, Z) + (B_T/T)R_{3,2}(R, \beta, Z) + o(B_T/T) + O(|\tau_{1T}^2 - \sigma_T^{\dagger 2}|))$  uniformly on  $T_2T_1(M_T)$ . The derivative of  $\hat{Q}$  with respect to the radius  $R$  is

$$\partial(T_1^{-1}(R, \beta, Z)^\top T_1^{-1}(R, \beta, Z)/2)/\partial R = 2R(1 + T^{-1/2}R_{4,1}(R, \beta, Z) + (B_T/T)R_{4,2}(R, \beta, Z)),$$

and the Jacobian of  $T_2$  is  $\partial R/\partial R' = 2R'(\partial(T_1^{-1}(R, \beta, Z)^\top T_1^{-1}(R, \beta, Z)/2)/\partial R)^{-1}$ . As a consequence, we obtain:

$$\begin{aligned} \int_{\{\bar{B}_T \cap M_T\}} \tilde{v}_{r,T} = & \int_{\{(R, \beta, Z) : (R')^2 \in \bar{B}\} \cap T_2T_1(M_T)} \left[ (R')^{r-1}J(\beta)[1 + T^{-1/2}R_{5,1}(R', \beta, Z) \right. \\ & \left. + (B_T/T)R_{5,2}(R', \beta, Z)] \exp(-(R')^2/2) \right] + o(B_T/T) + O(|\tau_{1T}^2 - \sigma_T^{\dagger 2}|), \end{aligned}$$

uniformly on the Borel subsets on  $\mathbb{R}$ . Finally, as  $T_2T_1(M_T) \subset \{(R', \beta, Z) : (R')^2 < ((s-3)/2) \log T\}$ ,

we have:

$$\int_{\{\bar{B}_T \cap M_T\}} \tilde{v}_{r,T} = \int_{\{(R', \beta, Z): (R')^2 \in \bar{B}\}} \left[ (R')^{r-1} J(\beta) [1 + T^{-1/2} R_{5,1}(R', \beta, Z) + (B_T/T) R_{5,2}(R', \beta, Z)] \exp(-(R')^2/2) \right] + o(B_T/T) + O(|\tau_{1T}^2 - \sigma_T^{\dagger 2}|),$$

uniformly on the Borel subsets on  $\mathbb{R}$ . As  $R_{5,1}(R', \beta, Z)$  is odd,  $R_{5,1}(R', \beta, Z) \exp(-(R')^2/2)$  integrates to zero uniformly on the Borel subsets on  $\mathbb{R}$ . Therefore, we get an expansion of the form  $\Upsilon_{Q,T}^\dagger(x) = F_{\chi_r^2}(x) + (B_T/T) p_Q(x, \mathcal{K}_Q) f_{\chi_r^2}(x)$  such that  $\sup_{y \in \mathbb{R}_+} |\int_{t \leq y} d\Upsilon_{Q,T}^\dagger - \int_{t \leq y} dF_{\hat{Q}}| = o(B_T/T) + O(|\tau_{1T}^2 - \sigma_T^{\dagger 2}|)$ . For an explicit method to determine the polynomials, see Remark 2.6 of [Chandra and Ghosh \(1979\)](#). The arguments are the same for the bootstrap quadratic statistic  $Q^*$ , except that we have to replace the orders  $o(B_T/T)$  by  $o(1/T)$ .

*Proof of Lemma 12*

Let us define:

$$\pi_T := T^{-1} \sum_{t=1}^T \sum_{j=1}^{T-B_T} \sum_{k=-B_T}^{B_T} k_{B_T}^*(k) \mathbb{E} [g_t g_j g_{j+k}], \quad (37)$$

$$\mu_{3,T} := T^2 \mathbb{E} [\bar{g}^3], \quad \mu_{3,\infty} := \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \mathbb{E} [g_0 g_i g_j]. \quad (38)$$

For computational convenience, let us consider the auxiliary signed measure  $\Upsilon_{S,\infty}$ , defined by its Fourier transform:

$$\begin{aligned} \mathcal{E}_\infty(\tau) &:= \int \exp(i\tau^\top x) d\Upsilon_{S,\infty}(x) \\ &= \left\{ 1 + \frac{\mu_{3,\infty}}{\sqrt{T}\sigma_\infty^3} \left[ -\frac{1}{3} (i\tau)^3 - \frac{1}{2} (i\tau) \right] \right\} \exp\left(-\frac{\tau^2}{2}\right). \end{aligned}$$

We can see that  $\mathcal{E}_\infty$  differs from  $\mathcal{E}_T^\dagger$  and  $\mathcal{E}_T^*$  by the variance and the third moment. Yet, we can control those differences by the following asymptotic results:

**Lemma 13.** *Under Assumptions 1 – 4, we have: (i)  $\mu_{3,T}^\dagger = \mu_{3,\infty} + O(T^{-1})$ , (ii)  $\pi_T^\dagger = \mu_{3,\infty} + O(T^{-1})$  and (iii)  $\sigma_T^{\dagger 2} = \sigma_\infty + o(1)$ .*

Thus, by Esseen Lemma:

$$\sup_x \left| \Upsilon_{S,\infty}(x) - \Upsilon_{S,T}^\dagger(x) \right| \leq C \int_{|\tau| \leq T^{1/2+\epsilon}} |\mathcal{E}_\infty(\tau) - \mathcal{E}_T(\tau)| |\tau|^{-1} d\tau + o\left(T^{-1/2}\right) + O(B_T^{-q}) + O(B_T/T). \quad (39)$$

Furthermore, for the bootstrap statistic, we have the following asymptotic convergences of the variance and the third moment:

**Lemma 14.** *Under Assumptions 1 – 4, we have: (i)  $\mu_{3,T}^* = \mu_{3,\infty} + O_p(T^{-1/2}) + O(B_T/T) + o(1)$  and (ii)  $\sigma_T^{*2} = \sigma_\infty^2 + O_p(B_T^{-q}) + O_p((B_T/T)^{1/2})$ .*

Therefore, we have by the same argument:

$$\sup_x \left| \Upsilon_{S,\infty}(x) - \Upsilon_{S,T}^*(x) \right| \leq C \int_{|\tau| \leq T^{1/2+\epsilon}} \left| \mathcal{E}_\infty(\tau) - \hat{\mathcal{E}}_T(\tau) \right| |\tau|^{-1} d\tau + o_p(T^{-1/2}). \quad (40)$$

*SM.4. Proof of Lemma 13*

(i) First,  $\pi_T = \mu_{3,T} + O(T^{-1})$  and  $\pi_T^\dagger = \mu_{3,T}^\dagger + O(T^{-1})$  (Götze and Künsch (1996)). Then,

$$\begin{aligned} |\pi_T - \pi_T^\dagger| &= \left| T^{-1} \sum_{t=1}^T \sum_{j=1}^{T-l} \sum_{k=1}^l k_{B_T}^*(k) \mathbb{E}[g_t(\beta_0) g_j(\beta_0) g_{j+k}(\beta_0)] \right. \\ &\quad \left. - T^{-1} \sum_{t=1}^T w(t) \sum_{j=1}^{T-l} \sum_{k=1}^l k_{B_T}^*(k) \mathbb{E}[g_t(\beta_0) g_j(\beta_0) g_{j+k}(\beta_0)] \right| \\ &= T^{-1} \left| \sum_{t=1}^T (1 - w(t)) \sum_{j=1}^{T-l} \sum_{k=1}^l k_{B_T}^*(k) \mathbb{E}[g_t(\beta_0) g_j(\beta_0) g_{j+k}(\beta_0)] \right| \\ &= O(B_T/T). \end{aligned}$$

Thus,  $|\mu_{3,T} - \mu_{3,T}^\dagger| \leq |\mu_{3,T} - \pi_T| + |\pi_T - \pi_T^\dagger| + |\pi_T^\dagger - \mu_{3,T}^\dagger| = O(T^{-1}) + O(B_T/T) + O(T^{-1})$ . As  $\mu_{3,T}$  tends to  $\mu_{3,\infty}$  by definition, the proof is concluded.

(ii) Recall that  $|\mu_{3,T} - \mu_{3,T}^\dagger| = O(B_T/T)$  and  $\pi_T^\dagger = \mu_{3,T}^\dagger + O(T^{-1})$ . Then, the lemma follows by triangular inequality, and as  $\mu_{3,T}$  tends to  $\mu_{3,\infty}$  by definition.

(iii)  $\sigma_T^{\dagger 2} = \text{Var}[T^{1/2} \bar{g}_T] = \sigma_\infty + o(1)$  by definition.

SM.5. Proof of Lemma 14

(i) We have:

$$\begin{aligned}\mu_{3,T}^* &= T^{1/2} \mathbb{E}^* \left[ T^{1/2} \bar{g}_T^{*3} \right] = T^{-1} \sum_{r=1}^T \sum_{s=1}^T \sum_{t=1}^T \mathbb{E}^* \left[ g_{T,r}^* g_{T,s}^* g_{T,t}^* \right] \\ &= T^{-1} \sum_{t=1}^T \mathbb{E}^* \left[ g_{T,t}^{*3} \right] = T^{-1} \sum_{t=1}^T g_{T,t}^3 = \mathbb{E} \left[ T^{-1} \sum_{t=1}^T g_{T,t}^3 \right] + O_p(T^{-1/2}).\end{aligned}$$

Now,  $\mathbb{E} \left[ T^{-1} \sum_{t=1}^T g_{T,t}^3 \right] = \mathbb{E} \left[ g_{T,s_1}^3 \right] + O(B_T/T)$ , where  $s_1 := \lfloor T/2 \rfloor$ . Thus,  $\mu_{3,T}^* = \mathbb{E} \left[ g_{T,s_1}^3 \right] + O_p(T^{-1/2}) + O(B_T/T)$ . Then, considering that  $g_{T,s_1}$  is a standardized weighted sum and from any CLT for strongly mixing process under Lindeberg conditions (see e.g. Rio (1997)),  $\mathbb{E} \left[ g_{T,s_1}^3 \right] = \mu_{3,\infty} + o(1)$ .

(ii) Recall that  $\sigma_T^{*2} = \mathbb{E}^* \left[ T \bar{g}_T^{*2} \right] = T^{-1} \sum_{t=1}^T g_{T,t}^2$ . Thus,  $\sigma_T^{*2}$  has the form of an automatically positive semi-definite HAC estimator (Smith (2005)). From the properties of HAC-type estimators (see Section 4.1), we have:  $\sigma_T^{*2} = \sigma_\infty^2 + O_p(B_T^{-q}) + O_p((B_T/T)^{1/2})$ , where  $q$  is the Parzen exponent. As a consequence, the bootstrap variance converges to the long-run one.

SM.6. Proof of Corollary 5

From the Taylor expansion defining  $\tilde{Q}$  in (8) we have  $\tilde{Q}(\beta_0) = \hat{Q}(\beta_0) + R_T$ , where  $\mathbb{P}[R_T > \delta_T] = \delta_T$  for some positive sequence  $\delta_T = o(T^{-1})$ . It follows that:

$$\begin{aligned}\mathbb{P}[\tilde{Q}(\beta_0) \leq x] &= \mathbb{P}[\hat{Q}(\beta_0) + R_T \leq x] \\ &= \mathbb{P}[\hat{Q}(\beta_0) + R_T \leq x \mid |R_T| \leq \delta_T] \mathbb{P}[|R_T| \leq \delta_T] \\ &\quad + \mathbb{P}[\hat{Q}(\beta_0) + R_T \leq x \mid |R_T| > \delta_T] \mathbb{P}[|R_T| > \delta_T] \\ &= \mathbb{P}[\hat{Q}(\beta_0) + R_T \leq x \mid |R_T| \leq \delta_T] (1 - \delta_T) \\ &\quad + \mathbb{P}[\hat{Q}(\beta_0) + R_T \leq x \mid |R_T| > \delta_T] \delta_T \\ &= \mathbb{P}[\hat{Q}(\beta_0) + R_T \leq x \mid |R_T| \leq \delta_T] + C\delta_T \\ &\geq \mathbb{P}[\hat{Q}(\beta_0) \leq x - \delta_T] + C\delta_T.\end{aligned}$$

Following the same argument, we have:

$$\begin{aligned}\mathbb{P}[\tilde{Q}(\beta_0) \leq x] &= \mathbb{P}[\hat{Q}(\beta_0) + R_T \leq x \mid |R_T| \leq \delta_T] + C\delta_T \\ &\leq \mathbb{P}[\hat{Q}(\beta_0) \leq x + \delta_T] + C\delta_T.\end{aligned}$$

As both  $\mathbb{P}[\hat{Q}(\beta_0) \leq x]$  and  $\mathbb{P}[\tilde{Q}(\beta_0) \leq x]$  are bounded above and below up to the  $C\delta_T = o(T^{-1})$  term, we get:

$$\sup_{x \in \mathbb{R}^+} |\mathbb{P}[\tilde{Q}(\beta_0) \leq x] - \mathbb{P}[\hat{Q}(\beta_0) \leq x]| \leq \sup_{x \in \mathbb{R}^+} |\mathbb{P}[\hat{Q}(\beta_0) \leq x - \delta_T] - \mathbb{P}[\hat{Q}(\beta_0) \leq x + \delta_T] + o(T^{-1})| \quad (41)$$

$$\leq \sup_{x \in \mathbb{R}^+} |\mathbb{P}[\hat{Q}(\beta_0) \leq x - \delta_T] - \mathbb{P}[\hat{Q}(\beta_0) \leq x + \delta_T]| + o(T^{-1}) \quad (42)$$

$$= o(T^{-1}), \quad (43)$$

by continuity of the probability distribution. Then, the result follows from an application of Theorem [4](#).

### 3. Monte Carlo

#### *SM.7. CPU time*



Table 6: Comparison of CPU time (in seconds) between FMB and MBB.

---



---

$ACD(1, 0)$			
		$B_T = 3$	$B_T = 5$
$T = 250$	$\hat{Q}_{FMB}$	0.72	0.65
	$W_{MBB}$	362.61	286.29
$T = 500$	$\hat{Q}_{FMB}$	0.75	0.68
	$W_{MBB}$	617.46	467.86
$ACD(1, 1)$			
		$B_T = 3$	$B_T = 5$
$T = 250$	$\hat{Q}_{FMB}$	0.74	0.74
	$W_{MBB}$	1518.59	1104.58
$T = 500$	$\hat{Q}_{FMB}$	0.78	1.21
	$W_{MBB}$	2963.06	1751.53

---



---

The number of bootstrap samples is  $R = 2500$  for both methods.

*SM.8. Coverages for CR of ACD(1,0) and ACD(1,1) with  $T = 150$ .*

The results we show in Table 7 must be interpreted with caution since more than 10% of the estimators did not converge for this sample size in the Monte Carlo simulations.

Table 7: Coverage of CR, a comparison between first and higher-order correct methods.

		ACD(1,0)					
		$B_T = 3$			$B_T = 5$		
Coverages:		0.90	0.95	0.99	0.90	0.95	0.99
$T = 150$	$\hat{Q}_{FMB}$	0.88	0.92	0.96	0.85	0.89	0.94
	$\tilde{Q}_{FMB}$	0.89	0.92	0.96	0.88	0.91	0.96
	$\tilde{Q}_{FMB,2}$	0.88	0.92	0.97	0.85	0.90	0.96
	$\hat{Q}_{\chi_r^2}$	0.85	0.9	0.94	0.84	0.87	0.93
	$W_{MBB}$	0.9	0.94	1.00	0.92	0.96	1.00
	$W_{\chi_p^2}$	0.78	0.85	0.92	0.74	0.80	0.88
			ACD(1,1)				
		$B_T = 3$			$B_T = 5$		
$T = 150$	$\hat{Q}_{FMB}$	0.87	0.92	0.96	0.83	0.88	0.93
	$\tilde{Q}_{FMB}$	0.87	0.9	0.94	0.84	0.88	0.91
	$\tilde{Q}_{FMB,2}$	0.78	0.84	0.91	0.72	0.79	0.88
	$\hat{Q}_{\chi_r^2}$	0.83	0.88	0.93	0.80	0.85	0.90
	$W_{MBB}$	0.95	0.98	1.00	0.97	0.99	1.00
	$W_{\chi_p^2}$	0.67	0.73	0.83	0.62	0.69	0.79

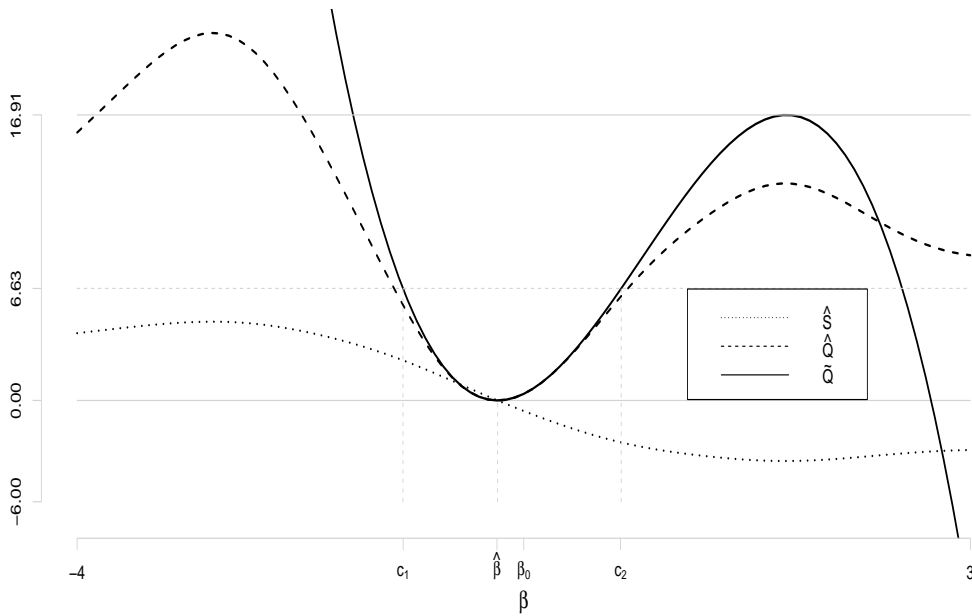
The true values of the unknown parameters of the ACD(1,0) and ACD(1,1) are  $\omega = 1.5$  and  $\beta_1 = 0.25$  and  $\beta_2 = 0.25$ .

*SM.9. Graphical illustration of the use of  $\tilde{Q}$  as in Remark 2.*

To illustrate the behavior of the approximation  $\tilde{Q}$ , we consider the problem of estimating the location parameter  $\beta_0$  of a Cauchy distribution, from an i.i.d. sample of size  $T = 30$ . This example does not correspond to our inferential setting of interest because of its lack of finite moments, but it has the advantage to provide a clear picture of the worst-case scenario, where the studentized score function  $\hat{S}(\beta_0)$  fails to be one-to-one in  $\beta_0$ , making the direct inversion technique impossible. Figure 2 shows the plot of  $\hat{S}$ ,  $\hat{Q}$ , and its approximation  $\tilde{Q}$ . The local maximum of  $\tilde{Q}$  (16.91 for this sample) is the maximum

level allowing to define a simply connected CI as level set. As discussed in Remark 2, this maximum increases proportionally to the sample size  $T$ , by definition of  $\tilde{Q}$  as a cubic polynomial. In this example, we draw a line at 6.63 — the 99% percentile of a chisquare distribution with one degree of freedom — to illustrate the applicability of  $\tilde{Q}$  already with  $T = 30$ .

Figure 2: Graphical illustration of Remark 2.



The studentized score  $\hat{S}(\beta_0)$  of the Cauchy location parameter  $\beta_0$ , its square  $\hat{Q}(\beta_0)$  and the approximation by  $\tilde{Q}(\beta_0)$ . In this example, the bounds of the CI are be  $c_1 = -0.81$  and  $c_2 = 0.65$ .

*SM.10. Verification of Assumptions 1–7 for the AR(p) example.*

As an example of process where Assumptions 1–7 are satisfied, we take the OLS moment indicators for the autoregressive parameters of an AR(p) process  $Y_t = \sum_{k=1}^p \theta_k Y_{t-k} + e_t = \sum_{j=0}^{\infty} w_j e_{t-j}$ , with  $e_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$  and  $|w_j| \leq \delta^{-1} \exp(-\delta j)$ . Namely,  $g_t = g(Y_t, \dots, Y_{t-p}; \theta)$  is the  $p$ -dimensional moment condition vector with entries  $g_i(Y_t, \dots, Y_{t-p}; \theta) = Y_{t-i}(Y_t - \sum_{k=1}^p \theta_k Y_{t-k})$ ,  $\forall i = 1, \dots, p$ . We define the sub-sigma-fields  $D_t := \sigma\langle e_t, \dots, e_{t-p} \rangle$ . The innovations  $e_t$  being i.i.d. Gaussian, the verification of Assumptions 1 and 2 follows from the orthogonality conditions of OLS and from the existence of all the moments.

We verify the exponential decay in Assumptions 4 and 6 as the strongly mixing coefficient of Assumption 4 and the Markov-type condition of Assumption 6 drop to zero for independent random variables. For Assumption 3, there exists a constant  $\delta > 0$  such that for  $t, m = 1, 2, \dots$  and  $m > \delta^{-1}$ , we can approximate  $Y_t$  by a  $\mathcal{D}_{t-m}^{t+m}$ -measurable random vector  $Y_{t,m}^\dagger$ , such that  $\mathbb{E} \left\| Y_t - Y_{t,m}^\dagger \right\| \leq \delta^{-1} \exp(-\delta m)$ , for instance by taking  $Y_{t,m}^\dagger = \sum_{j=0}^m w_j e_{t-j}$ . Then,  $\mathbb{E} \left\| g(Y_t, \dots, Y_{t-p}; \theta) - g(Y_{t,m}^\dagger, \dots, Y_{t-p,m}^\dagger; \theta) \right\| \leq \sum_{i=1}^p \mathbb{E} |g_i(Y_t, \dots, Y_{t-p}; \theta) - g_i(Y_{t,m}^\dagger, \dots, Y_{t-p,m}^\dagger; \theta)|$ . Therefore, it is sufficient to show that each element of the latter sum is decreasing exponentially with  $m$ , to get a process  $g_{t,m}^\dagger = g(Y_{t,m}^\dagger, \dots, Y_{t-p,m}^\dagger; \theta)$  approximating  $g_t$  with exponentially decaying error, as required by Assumption 3. Using a Taylor expansion, we can write  $g_i(Y_t, \dots, Y_{t-p}; \theta) - g_i(Y_{t,m}^\dagger, \dots, Y_{t-p,m}^\dagger; \theta) = Z_i^\top D_t^\dagger + D_t^{\dagger\top} A D_t^\dagger, \forall i = 1, \dots, p$ , where  $Z_i$  is a vector of Gaussian random variables,  $A$  is a matrix of deterministic constants and  $D_t^\dagger$  is a vector with entries  $Y_{t-i} - Y_{t-i,m}^\dagger, \forall i = 1, \dots, p$ . Then, using Cauchy-Schwarz inequality on  $Z_i^\top D_t^\dagger$  allows to verify directly Assumption 3 on  $g_t$  by making use of the exponential decay of each element in  $D_t^\dagger$ . We can verify Assumptions 5 and 7 by using the Riemann-Lebesgue lemma and the continuity of the function  $g, g_t$  having always a density.