Taylor & Francis
Taylor & Francis Group

Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Bootstrapping Two-Stage Quasi-Maximum Likelihood Estimators of Time Series Models

Sílvia Gonçalves, Ulrich Hounyo, Andrew J. Patton & Kevin Sheppard

To cite this article: Sílvia Gonçalves, Ulrich Hounyo, Andrew J. Patton & Kevin Sheppard (2023) Bootstrapping Two-Stage Quasi-Maximum Likelihood Estimators of Time Series Models, Journal of Business & Economic Statistics, 41:3, 683-694, DOI: 10.1080/07350015.2022.2058949

To link to this article: https://doi.org/10.1080/07350015.2022.2058949

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2023, VOL. 41, NO. 3, 683–694
https://doi.org/10.1080/07350015.2022.2058949

Taylor & Francis
Taylor & Francis Group

Check for updates

Bootstrapping Two-Stage Quasi-Maximum Likelihood Estimators of Time Series Models

Sílvia Gonçalves[a], Ulrich Hounyo[b], Andrew J. Patton[c], and Kevin Sheppard[d]

[a]Department of Economics, McGill University, Montreal, Quebec, Canada; [b]Department of Economics, University at Albany, SUNY, New York, NY; [c]Department of Economics, Duke University, Durham, NC; [d]Department of Economics, University of Oxford, Oxford, UK

ABSTRACT
This article provides results on the validity of bootstrap inference methods for two-stage quasi-maximum likelihood estimation involving time series data, such as those used for multivariate volatility models or copula-based models. Existing approaches require the researcher to compute and combine many first- and second-order derivatives, which can be difficult to do and is susceptible to error. Bootstrap methods are simpler to apply, allowing the substitution of capital (CPU cycles) for labor (keeping track of derivatives). We show the consistency of the bootstrap distribution and consistency of bootstrap variance estimators, thereby justifying the use of bootstrap percentile intervals and bootstrap standard errors.

ARTICLE HISTORY
Received December 2020
Accepted March 2022

KEYWORDS
Copulas; Inference; Multivariate GARCH

1. Introduction

Many econometric models are estimated in stages in order to make them more computationally tractable. For example, copula-based models often have the marginal distribution parameters estimated in the first N stages, and then the copula parameters estimated in a final stage (see, e.g., Joe 1995; Patton 2006). Similarly, the famous Dynamic Conditional Correlation (DCC) GARCH model of Engle (2002) first estimates univariate GARCH models for each variable, and then the correlation matrix component is estimated. In empirical finance, the ubiquitous Fama-MacBeth (1973) procedure involves estimating assets' exposures to risk factors in a first stage, before estimating risk premia associated with those factors in a second stage. In all cases, valid inference on parameters in the final stage must account for estimation error accumulated in earlier stages.

Standard methods to account for estimation error from earlier estimation stages requires computing many first- and second-order derivatives of the overall objective function.[1] In most cases, it is difficult or tedious to obtain these analytically, and many authors instead rely on numerical derivatives. Both of these approaches are susceptible to error: human, in the first case; numerical, in the second. Inference based on the bootstrap is an attractive alternative, allowing the substitution of capital (CPU cycles) for labor (keeping track of derivatives). See, for example, Cochrane (2001, chap. 15.2) for bootstrap inference in two-pass regressions and Patton (2012) for bootstrap inference in copula-based models.

Some related results are available in the literature. For example, Chen, Linton, and van Keilegom (2003) and Armstrong, Bertanha, and Hong (2014) consider the bootstrap for two-step nonlinear parametric and semiparametric models, and Cattaneo, Jansson, and Ma (2019) considers bootstrap inference for two-step GMM estimators, but all of these articles consider only iid data. Bootstrap methods for *one*-step quasi-maximum likelihood estimators (QMLE) have been considered (e.g., Gonçalves and White 2004) for both iid and time series data, but no general results are available for multistep QMLE. We address this gap and show bootstrap validity for such applications under very general time series dependence and heterogeneity.

We consider two approaches for bootstrap inference. The first jointly resamples the contributions to the quasi log-likelihood functions in the various stages of estimation. Since model misspecification at any stage can induce time series dependence in the scores of each model, we rely on the moving blocks bootstrap (MBB) of Künsch (1989) and Liu and Singh (1992). This approach involves re-estimating the model on each bootstrapped objective function, and this "fully optimized" approach is thus computationally demanding.

The second approach is a fast resampling method that avoids any optimization problem in the bootstrap world. In particular, our proposal is to resample the score function underlying the asymptotic linear representation of the final-step QMLE, evaluated at the parameters estimated in the real data. This is related to the resampling method proposed in Armstrong, Bertanha, and Hong (2014), however, in their approach the first-stage model is re-estimated on each bootstrap sample (which has advantages in the semiparametric settings considered in that

[1]For a textbook discussion of inference in multi-stage estimation see White (1994) and Newey and MacFadden (1994).

CONTACT Sílvia Gonçalves silvia.goncalves@mcgill.ca Department of Economics, McGill University, Leacock Building, Room 506, 855 Sherbrooke Street West, Montreal, Quebec, H3A 2T7 Canada.
A MATLAB toolbox for this article is available at https://sites.google.com/site/ulrichounyo.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

article) while our proposed method requires no estimation in the bootstrap world.[2]

We prove two sets of results for both bootstrap methods. For simplicity, we provide formal results only for the two-step QMLE.[3] First, we show the consistency of the bootstrap distribution of the final-stage parameter estimator for the usual limit distribution of that parameter, justifying the construction of bootstrap percentile intervals. We consider regularity conditions that allow for time series dependence and heterogeneity of unknown form, extending the work of Gonçalves and White (2004) who show the asymptotic validity of the MBB for inference on one-step QMLE. Second, we prove the consistency of bootstrap standard errors. Although these are very popular in applied work, their consistency does not automatically follow from the consistency of the bootstrap distribution. This was recently emphasized by Hahn and Liao (2020), who proved that bootstrap standard errors for smooth functions of iid data may lead to conservative inference. Here, we show that the bootstrap variance estimator of the two-step QMLE is consistent by verifying a certain uniform integrability condition. We follow the approach of Kato (2011) and Cheng (2015) and rely on empirical process theory to prove our results. The results of those two articles apply to one-step M-estimators with iid data and do not cover time series applications. They also do not cover two-step QMLE, even for the iid data. Similarly, although the results of Gonçalves and White (2005) allow for time series dependence, they are specific to the one-step least squares estimator. Thus, no results appear to be available regarding the consistency of bootstrap variance estimators of one or multi-stage QMLE with general time series dependence.

We present Monte Carlo simulations to illustrate the usefulness of our results. We consider the estimation of a bivariate copula model, where estimation is done in stages and the parameter of interest is the copula parameter, which is estimated in the final (third) stage. In addition to the standard confidence interval that relies on analytical standard errors, we consider two types of bootstrap-based intervals: intervals that rely on bootstrap standard errors, but use the normal critical value, and bootstrap percentile intervals. We find that all methods provide similarly good coverage probabilities, even for the smaller sample sizes. This is in agreement with the theory of the bootstrap since none of the methods promises asymptotic refinements. However, we find that the competing inference methods differ in their confidence interval lengths. In particular, the intervals based on the fully optimized bootstrap method tend to be narrower than the intervals based on either the fast resampling method or the usual asymptotic approach using analytical standard errors. We show that this is due to the fully optimized bootstrap standard error having a smaller mean squared error than the competing methods. Thus, although more computationally intensive than the fast resampling method, the fully optimized bootstrap intervals have better finite sample properties.

The rest of the article is organized as follows. In Section 2, we present the framework and provide an example of a two-step QMLE based on the bivariate copula model. In Section 3, we describe our two bootstrap methods and prove their consistency in Section 4. Section 5 contains the simulation results and Section 6 concludes. Assumptions are collected in Appendix A. The supplementary materials contains the proofs of the bootstrap distribution consistency and the proof of the bootstrap variance consistency results. This supplementary materials also contains two general bootstrap consistency theorems for two-step M and GMM bootstrap estimators, which could be of independent interest as their high level conditions can be verified for any bootstrap scheme. Our focus on the two-stage QMLE estimator based on the moving blocks bootstrap in the main text is justified by the fact that this covers the main applications in finance we have in mind.

2. Framework

2.1. Two-Stage Quasi Maximum Likelihood Estimation

Suppose $\{X_t : \Omega \to \mathbb{R}^l,\ t \in \mathbb{N}\}$ denotes a sequence of $\mathbb{R}^l$-valued random vectors defined on some probability space $(\Omega, \mathcal{F}, P)$. Let $\Theta = \mathcal{A} \times \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are compact subsets of finite dimensional Euclidean spaces. Given data $\{X_t : t = 1, \ldots, n\}$, our goal is to estimate a parameter vector $\beta_0 \in \mathcal{B} \subset \mathbb{R}^p$ by a two-stage quasi-maximum likelihood (2QMLE) estimator. We focus on two-stage QMLE for ease of presentation, but our results generalize directly to multi-stage QMLEs. (e.g., our simulation study considers a three-stage QMLE problem.)

In the first step, we estimate $\alpha_0 \in \mathcal{A} \subset \mathbb{R}^k$ with

$$\hat{\alpha}_n = \arg\max_{\alpha \in \mathcal{A}} Q_{1n}(\alpha),$$

where $Q_{1n}(\alpha) \equiv n^{-1}\sum_{t=1}^{n} \log f_{1t}\left(X^t, \alpha\right)$, with $X^t \equiv (X_1, \ldots, X_{t-1}, X_t)$, for some quasi-likelihood function $f_{1t}\left(X^t, \alpha\right) : \mathbb{R}^{lt} \times \mathcal{A} \to \mathbb{R}^+$. To simplify the notation, we sometimes write $f_{1t}(\alpha) \equiv f_{1t}\left(X^t, \alpha\right)$. A similar notation is used for any other function of X^t throughout.

In the second step, we estimate β with

$$\hat{\beta}_n = \arg\max_{\beta \in \mathcal{B}} Q_{2n}\left(\hat{\alpha}_n, \beta\right),$$

where $Q_{2n}(\alpha, \beta) \equiv n^{-1}\sum_{t=1}^{n} \log f_{2t}\left(X^t, \alpha, \beta\right)$, for a conditional quasi-likelihood function $f_{2t}(\alpha, \beta) \equiv f_{2t}\left(X^t, \alpha, \beta\right) : \mathbb{R}^{lt} \times \mathcal{A} \times \mathcal{B} \to \mathbb{R}^+$. We allow for time heterogeneity in $f_{1t}(\alpha)$ and $f_{2t}(\alpha, \beta)$ (i.e., the functional forms may depend on t) and we also allow for the possibility that these functions depend on the past information up to time t (i.e., X^t is a vector of possibly growing dimension).

2.2. An Example: Copula Models

An example of time series models that are often estimated in multiple stages are copula-based multivariate models. These models combine separately estimated marginal distributions via a copula function to form a joint distribution. When the

[2]Other artcles that have proposed fast resampling methods include Davidson and MacKinnon (1999), Andrews (2002), Gonçalves and White (2004), Hong and Scaillet (2004), and La Vecchia, Moor and Scaillet (2020), among others. However, all these articles consider one-step M or GMM estimators.

[3]These results can easily be generalized to multistage QMLE. In our Monte Carlo simulations, we estimate a copula model involving a three-stage QMLE.

parameters that characterize the marginal distributions are different from those that characterize the copula density function, estimation and inference can be done in stages. Our results can be useful in this context.

To illustrate, let $X_t \equiv (X_{1t}, X_{2t})'$ denote a random vector whose joint conditional density we would like to model. By the usual decomposition, we can write

$$\log g(X_1, \ldots, X_n, \theta) = \sum_{t=1}^{n} \log g_t\left(X_t|\mathcal{F}^{t-1}, \theta\right),$$

where $g_t\left(X_t|\mathcal{F}^{t-1}, \theta\right)$ is the conditional density function of X_t given $\mathcal{F}^{t-1}$. Suppose $X_{it}|\mathcal{F}^{t-1} \sim G_{it}(\alpha_i)$, some distribution function parameterized by a set of parameters α_i with density function $g_{it}(\alpha_i)$. The joint (conditional) pdf of X_t is then given by

$$g_t\left(X_t|\mathcal{F}^{t-1}, \theta\right) = g_{1t}(X_{1t}, \alpha_1)\, g_{2t}(X_{2t}, \alpha_2)\, c_t\Big(G_{1t}(X_{1t}, \alpha_1),\; G_{2t}(X_{2t}, \alpha_2), \beta\Big),$$

where $c_t(\cdot, \cdot, \beta)$ is a copula density function parameterized by β, and $\theta = (\alpha_1, \alpha_2, \beta)'$ denotes the full set of parameters. It follows that the joint log-likelihood function can be written as

$$\begin{aligned} &\log g(X_1, \ldots, X_n, \theta) \\ &= \sum_{t=1}^{n} \log g_{1t}\left(X_{1t}|\mathcal{F}^{t-1}, \alpha_1\right) + \sum_{t=1}^{n} \log g_{2t}\left(X_{2t}|\mathcal{F}^{t-1}, \alpha_2\right) \\ &\quad + \sum_{t=1}^{n} \log c_t\left(G_{1t}(X_{1t}, \alpha_1), G_{2t}(X_{2t}, \alpha_2)\,|\mathcal{F}^{t-1}, \beta\right). \end{aligned}$$

When the parameters characterizing the marginals and the copula function are separable (i.e., the parameters that enter one marginal do not enter another marginal nor the copula function and there are no cross equation restrictions), these parameters can be estimated in stages. In particular, first estimate α_i by QMLE:

$$\hat{\alpha}_{in} = \arg\max_{\alpha_i \in \mathcal{A}} \sum_{t=1}^{n} \log g_{it}\left(X_{it}|\mathcal{F}^{t-1}, \alpha_i\right), \text{ for } i = 1, 2,$$

and then estimate the copula parameters β in a second stage by

$$\hat{\beta}_n = \arg\max_{\beta \in \mathcal{B}} \sum_{t=1}^{n} \log c_t\big(G_{1t}\big(X_{1t}, \hat{\alpha}_{1n}\big), G_{2t}\big(X_{2t}, \hat{\alpha}_{2n}\big)\,|\mathcal{F}^{t-1}, \beta\big).$$

Thus, in our previous notation,

$$Q_{2n}(\hat{\alpha}_n, \beta) = \sum_{t=1}^{n} \log f_{2t}\left(X^t, \hat{\alpha}_n, \beta\right), \quad \text{where} \quad \hat{\alpha}_n = \left(\hat{\alpha}_{1n}, \hat{\alpha}_{2n}\right)',$$

and

$$f_{2t}\left(X^t, \hat{\alpha}_n, \beta\right) \equiv c_t\left(G_{1t}\left(X_{1t}, \hat{\alpha}_{1n}\right), G_{2t}\left(X_{2t}, \hat{\alpha}_{2n}\right)|\mathcal{F}^{t-1}, \beta\right).$$

The contributions to this quasi-log-likelihood function depend on the sample of $X_t = (X_{1t}, X_{2t})'$ up to time t through the integral probability transforms $G_{it}\left(X_{it}, \hat{\alpha}_{in}\right)$.

2.3. Asymptotic Properties of Two-Stage QMLE

We now briefly review the asymptotic properties of the two-stage QMLE to assist in later explaining the properties that the bootstrap needs to have in order to be asymptotically valid. These properties are well known in the literature, see, for example, White (1994), Newey and McFadden (1994), and Wooldridge (1994).

Let α_0 be the unique maximizer of $\bar{Q}_1(\alpha) \equiv \lim_{n\to\infty} E(Q_{1n}(\alpha))$ on $\mathcal{A}$ and let β_0 be the unique maximizer of $\bar{Q}_2(\alpha_0, \beta) \equiv \lim_{n\to\infty} E(Q_{2n}(\alpha_0, \beta))$ on $\mathcal{B}$.[4] Then, under Assumption A in Appendix A, we can show that $\hat{\beta}_n \xrightarrow{P} \beta_0$ and

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \xrightarrow{d} N(0, C_0),$$

where the asymptotic covariance matrix $C_0 \equiv H_0^{-1} J_0 H_0^{-1}$, with

$$H_0 \equiv \lim_{n\to\infty} E\left(n^{-1} \sum_{t=1}^{n} \frac{\partial}{\partial \beta'} s_{2t}(\alpha_0, \beta_0)\right), \quad \text{such that}$$

$$s_{2t}(\alpha_0, \beta_0) \equiv \frac{\partial}{\partial \beta} \log f_{2t}(\alpha_0, \beta_0),$$

and

$$J_0 \equiv \lim_{n\to\infty} \operatorname{var}\left(n^{-\frac{1}{2}} \sum_{t=1}^{n} \left(s_{2t}(\alpha_0, \beta_0) - F_0 A_0^{-1} s_{1t}(\alpha_0)\right)\right),$$

where

$$s_{1t}(\alpha_0) \equiv \frac{\partial}{\partial \alpha} \log f_{1t}(\alpha_0),$$

$$A_0 \equiv \lim_{n\to\infty} E\left(n^{-1} \sum_{t=1}^{n} \frac{\partial}{\partial \alpha'} s_{1t}(\alpha_0)\right),$$

and

$$F_0 \equiv \lim_{n\to\infty} E\left(n^{-1} \sum_{t=1}^{n} \frac{\partial}{\partial \alpha'} s_{2t}(\alpha_0, \beta_0)\right).$$

As this result shows, the impact of the first stage estimation of α_0 is not negligible asymptotically except when $F_0 = 0$. This implies that we need to adjust the standard errors of $\hat{\beta}_n$ for the added estimation uncertainty of $\hat{\alpha}_n$. Although a consistent estimator of J_0 can be obtained by applying a HAC (heteroscedasticity and autocorrelation covariance) estimator to $\left\{s_{2t}\left(\hat{\alpha}_n, \hat{\beta}_n\right) - \hat{F}_n \hat{A}_n^{-1} s_{1t}\left(\hat{\alpha}_n\right)\right\}$ (where $\hat{F}_n$ and $\hat{A}_n$ are consistent estimators of F_0 and A_0) in practice the bootstrap is often used. Our goal is to provide a set of conditions that justify this practice in time series applications.

[4] Under general heterogeneity and time series dependence, α_0 and β_0 could depend on n. We omit the index n to simplify the notation.

3. Bootstrap Methods

The asymptotic validity of the bootstrap depends on its ability to mimic the asymptotic variance-covariance matrix of $\hat{\beta}_n$. The form of J_0 suggests that the bootstrap should replicate the time series dependence and the heterogeneity properties of the score vector $\left\{s_t(\alpha_0, \beta_0) \equiv s_{2t}(\alpha_0, \beta_0) - F_0 A_0^{-1} s_{1t}(\alpha_0)\right\}$. Model misspecification at any stage can induce time series dependence and our approach is to use a block bootstrap. In particular, we rely on the moving blocks bootstrap (MBB) of Künsch (1989) and Liu and Singh (1992). See also Gonçalves and White (2002, 2004, 2005) for the validity of the MBB under general time series dependence and heterogeneity. Although we focus on the MBB, other bootstrap methods which are robust to serial dependence could be used. Examples include the tapered block bootstrap of Paparoditis and Politis (2002) and the kernel block bootstrap of Parente and Smith (2021). The validity of these alternative bootstrap schemes could be established by verifying the high level conditions given in Theorems S4.3 and S4.4 in the supplementary materials.

We consider two different methods. One is based on resampling the contributions to the likelihood functions $\{f_{1t}(\alpha)\}$ (which yields a bootstrap QMLE $\hat{\alpha}_n^*$) and $\{f_{2t}(\hat{\alpha}_n^*, \beta)\}$ (which is optimized over β to yield $\hat{\beta}_n^*$). The same bootstrap indices obtained with the MBB are used across the two stages, ensuring that this method mimics the time series dependence of the extended score. Because it requires two (or multiple) sets of maximization, this method may be computationally intensive. For this reason, we also propose another bootstrap method which resamples directly the estimated score $s_t\left(\hat{\alpha}_n, \hat{\beta}_n\right) \equiv s_{2t}\left(\hat{\alpha}_n, \hat{\beta}_n\right) - \hat{F}_n \hat{A}_n^{-1} s_{1t}(\hat{\alpha}_n)$. Our simulations show that this method yields wider confidence intervals than the fully optimized bootstrap method. In particular, the fast resampling standard errors have larger mean squared errors compared to the fully optimized standard errors, especially for the smaller sample sizes. This translates into wider confidence intervals for the parameter of interest.

Both methods involve resampling certain functions of the data using the MBB to obtain new indices, which can be described as follows. For a generic time series $\{Z_t : t = 1, \ldots, n\}$, let $\ell = \ell_n \in \mathbb{N}$ $(1 \le \ell < n)$ be a block length. Define $B_{t,\ell} = \{Z_t, Z_{t+1}, \ldots, Z_{t+\ell-1}\}$ as the block of ℓ consecutive observations starting at Z_t ($\ell = 1$ corresponds to the standard iid bootstrap). For simplicity take $n = k\ell$. The MBB draws $k = n/\ell$ blocks randomly with replacement from the set of overlapping blocks $\{B_{1,\ell}, \ldots, B_{n-\ell+1,\ell}\}$. Letting $I_1, \ldots, I_k$ be iid random variables distributed uniformly on $\{0, \ldots, n-\ell\}$, we have $\{Z_t^* = Z_{\tau_t}, t = 1, \ldots, n\}$, where τ_t is defined as $\{\tau_t\} \equiv \{I_1 + 1, \ldots, I_1 + \ell, \ldots, I_k + 1, \ldots, I_k + \ell\}$.

3.1. The Fully Optimized Bootstrap Method

The first method we consider requires resampling the contributions to the two (or more) likelihood functions f_{1t} and f_{2t} and then computing $\hat{\alpha}_n^*$ and $\hat{\beta}_n^*$ using these resampled likelihood functions. More specifically, the bootstrap analogue of $\hat{\alpha}_n$ is given by

$$\hat{\alpha}_n^* = \arg\max_{\alpha \in \mathcal{A}} Q_{1n}^*(\alpha),$$

where $Q_{1n}^*(\alpha) \equiv n^{-1} \sum_{t=1}^n \log f_{1t}^*(\alpha)$, and $f_{1t}^*(\alpha) = f_{1,\tau_t}(\alpha) \equiv f_{1,\tau_t}(X^{\tau_t}, \alpha)$ is a resampled version of $f_{1t}(\alpha) \equiv f_{1t}\left(X^t, \alpha\right)$, where the indices τ_t are chosen by the bootstrap. Thus, we resample the functions $f_{1t}(\alpha)$ rather than the data directly. However, when $f_{1t}(\alpha) = f_1(Z_t, \alpha)$ where the function f_{1t} does not depend on t and is a function of $Z_t \equiv \left(X_t, X_{t-1}, \ldots, X_{t-k}\right)'$ for some finite $k \ge 0$, resampling $f_{1t}(\alpha)$ is equivalent to resampling the vector Z_t, that is, $f_{1t}^*(\alpha) \equiv f_{1t}^*\left(X^{*t}, \alpha\right) = f_1\left(Z_t^*, \alpha\right) = f_1\left(Z_{\tau_t}, \alpha\right)$.

The second-step bootstrap estimator $\hat{\beta}_n^*$ is obtained as

$$\hat{\beta}_n^* = \arg\max_{\beta \in \mathcal{B}} Q_{2n}^*\left(\hat{\alpha}_n^*, \beta\right),$$

where $Q_{2n}^*\left(\hat{\alpha}_n^*, \beta\right) \equiv n^{-1} \sum_{t=1}^n \log f_{2t}^*\left(\hat{\alpha}_n^*, \beta\right)$, with $f_{2t}^*\left(\hat{\alpha}_n^*, \beta\right) = f_{2,\tau_t}\left(\hat{\alpha}_n^*, \beta\right) \equiv f_{2,\tau_t}\left(X^{\tau_t}, \hat{\alpha}_n^*, \beta\right)$. Thus, we resample the functions $f_{2t}(\alpha, \beta) \equiv f_{2t}\left(X^t, \alpha, \beta\right)$ evaluated at $\alpha = \hat{\alpha}_n^*$ using the *same* indices τ_t used in computing $\hat{\alpha}_n^*$. Resampling both functions f_{1t} and f_{2t} with the same set of indices is important because this will preserve the form of dependence between the two functions. In particular, this will guarantee that the bootstrap is able to mimic the dependence in the score vector $s_t\left(\hat{\alpha}_n, \hat{\beta}_n\right)$. If instead we used two different sets of indices, say τ_{1t} and τ_{2t}, generated independently of each other, this would induce an independence between f_{1t}^* and f_{2t}^* which would not necessarily exist for the original functions.

3.2. A Fast Resampling Method

Bootstrapping multi-stage extremum estimators can be computationally intensive as this may require solving multiple optimization problems for each resample. For this reason, we also consider a fast resampling method for bootstrapping two-step QMLE which has a closed form expression and avoids any numerical optimization. To describe this estimator, let

$$\hat{H}_n = n^{-1} \sum_{t=1}^n \frac{\partial}{\partial \beta'} s_{2t}\left(\hat{\alpha}_n, \hat{\beta}_n\right),$$

$$\hat{A}_n = n^{-1} \sum_{t=1}^n \frac{\partial}{\partial \alpha'} s_{1t}(\hat{\alpha}_n), \quad \text{and}$$

$$\hat{F}_n = n^{-1} \sum_{t=1}^n \frac{\partial}{\partial \alpha'} s_{2t}\left(\hat{\alpha}_n, \hat{\beta}_n\right).$$

The fast resample two-step QMLE is given by

$$\hat{\beta}_{1,n}^* = \hat{\beta}_n - \hat{H}_n^{-1} n^{-1} \sum_{t=1}^n s_t^*\left(\hat{\alpha}_n, \hat{\beta}_n\right),$$

where $s_t^*\left(\hat{\alpha}_n, \hat{\beta}_n\right)$ is a resampled version of

$$s_t\left(\hat{\alpha}_n, \hat{\beta}_n\right) \equiv s_{2t}\left(\hat{\alpha}_n, \hat{\beta}_n\right) - \hat{F}_n \hat{A}_n^{-1} s_{1t}(\hat{\alpha}_n),$$

that is, $s_t^*\left(\hat{\alpha}_n, \hat{\beta}_n\right) = s_{\tau_t}\left(\hat{\alpha}_n, \hat{\beta}_n\right) \equiv s_{2\tau_t}\left(\hat{\alpha}_n, \hat{\beta}_n\right) - \hat{F}_n \hat{A}_n^{-1} s_{1\tau_t}(\hat{\alpha}_n)$. Thus, $\hat{\beta}_{1,n}^*$ is a one-step bootstrap QMLE which updates $\hat{\beta}_n$ using the estimated Hessian $\hat{H}_n$ and the bootstrap scores $s_t^*\left(\hat{\alpha}_n, \hat{\beta}_n\right)$, evaluated at $\hat{\alpha}_n$ and $\hat{\beta}_n$.

A special case of $\hat{\beta}_{1,n}^*$ is a version of the one-step bootstrap QMLE considered in Gonçalves and White (2004) (henceforth GW(2004)) in the context of one-stage QMLE. In that article, $\hat{\beta}_n$ does not depend on a first stage estimator $\hat{\alpha}_n$, implying that $s_{2t}\left(\hat{\alpha}_n, \hat{\beta}_n\right) = s_{2t}\left(\hat{\beta}_n\right)$ and $s_t\left(\hat{\alpha}_n, \hat{\beta}_n\right) = s_{2t}\left(\hat{\beta}_n\right)$. The only difference with respect to GW(2004) in this case is that our proposal only resamples the estimated scores and does not involve resampling the contributions to the Hessian $\hat{H}_n$ (instead, their one-step bootstrap QMLE involves resampling both; see also Davidson and MacKinnon (1999) and Andrews (2002) who proposed k-step bootstrap methods that resample the contributions to the Hessian and the score vector at each iteration, starting from the original estimators).

$\hat{\beta}_{1,n}^*$ is also related to a fast resampling approach proposed by Armstrong, Bertanha, and Hong (2014) in the context of a two-step GMM estimator with iid data, where the first step is a potentially nonparametric estimator (see also Chen, Linton, and van Keilegom 2003; Chen and Liao 2015). In our context, it amounts to

$$\tilde{\beta}_{1,n}^* = \hat{\beta}_n - \hat{H}_n^{-1} n^{-1} \sum_{t=1}^{n} s_{2t}^*\left(\hat{\alpha}_n^*, \hat{\beta}_n\right).$$

There are two main differences between $\hat{\beta}_{1,n}^*$ and $\tilde{\beta}_{1,n}^*$. First, $\tilde{\beta}_{1,n}^*$ requires computing $\hat{\alpha}_n^*$ whereas $\hat{\beta}_{1,n}^*$ does not. Hence, $\tilde{\beta}_{1,n}^*$ only avoids the computational burden of the second step and not of the first step. Instead, our method avoids computing $\hat{\alpha}_n^*$ for each resample and therefore is computationally more attractive. Second, $\tilde{\beta}_{1,n}^*$ resamples the scores of the second-stage model (evaluated at $\left(\hat{\alpha}_n^*, \hat{\beta}_n\right)$), whereas our method involves resampling $s_t\left(\hat{\alpha}_n, \hat{\beta}_n\right) = s_{2t}\left(\hat{\alpha}_n, \hat{\beta}_n\right) - \hat{F}_n \hat{A}_n^{-1} s_{1t}\left(\hat{\alpha}_n\right)$. We can think of this vector as an "extended" version of the scores for the second-stage, extended by the term $-\hat{F}_n \hat{A}_n^{-1} s_{1t}\left(\hat{\alpha}_n\right)$. This term corrects for the added uncertainty due to the first step. We note that it would not be valid to resample $s_{2t}\left(\hat{\alpha}_n, \hat{\beta}_n\right)$ unless $\hat{F}_n = 0$.

4. Bootstrap Theory

We discuss two uses of the bootstrap for inference on β using $\hat{\beta}_n$. In Section 4.1 we consider using the bootstrap distribution of $\sqrt{n}\left(\hat{\beta}_n^* - \hat{\beta}_n\right)$ (or $\sqrt{n}\left(\hat{\beta}_{1,n}^* - \hat{\beta}_n\right)$) to approximate the quantiles of the distribution of $\sqrt{n}\left(\hat{\beta}_n - \beta_0\right)$. This approach underlies the construction of percentile bootstrap intervals for β. Even though it does not promise asymptotic refinements, it is empirically attractive as it does not require computing standard errors for $\hat{\beta}_n$. An alternative is to use the bootstrap to estimate standard errors, which we consider in Section 4.2.

4.1. Bootstrap Distribution Consistency for Nonstudentized Statistics

The first order asymptotic validity of the MBB based on the fully optimized bootstrap two-step QMLE $\hat{\beta}_n^*$ follows by showing that the bootstrap distribution of $\sqrt{n}\left(\hat{\beta}_n^* - \hat{\beta}_n\right)$ is consistent for the distribution of $\sqrt{n}\left(\hat{\beta}_n - \beta_0\right)$.This result requires we strengthen Assumption A as follows.

Assumption B. For some $r > 2$ and some $\delta > 0$,

B.1: i. $\{s_{1t}(\alpha_0)\}$ is $r+\delta$-dominated on $\mathcal{A}$ uniformly in t.
ii. $\{s_{2t}(\alpha_0, \beta_0)\}$ is $r+\delta$-dominated on $\mathcal{A} \times \mathcal{B}$ uniformly in t.
B.2: $\{V_t\}$ is an α-mixing sequence of size $-\frac{(2+\delta)(r+\delta)}{r-2}$.
B.3: i. The elements of $\{s_{1t}(\alpha)\}$ are $L_{2+\delta}$-NED on $\{V_t\}$ of size -1, uniformly on $(\mathcal{A}, \rho)$.
ii. The elements of $\{s_{2t}(\alpha, \beta)\}$ are $L_{2+\delta}$-NED on $\{V_t\}$ of size -1, uniformly on $(\mathcal{A} \times \mathcal{B}, \rho)$.
B.4: i. $n^{-1}\sum_{t=1}^{n} E\left(s_{1t}(\alpha_0)\right) E\left(s_{1t}(\alpha_0)\right)' = o\left(\ell_n^{-1}\right)$, where $\ell_n = o(n)$ and $\ell_n \to \infty$.
ii. $n^{-1}\sum_{t=1}^{n} E\left(s_{2t}(\alpha_0, \beta_0)\right) E\left(s_{2t}(\alpha_0, \beta_0)\right)' = o\left(\ell_n^{-1}\right)$, where $\ell_n = o(n)$ and $\ell_n \to \infty$.

These assumptions are weaker than those used by GW (2004) (see their Assumption 2.1) and are sufficient to show that a bootstrap CLT applies to $\left\{s_{2t}^*(\alpha_0, \beta_0) - F_0 A_0^{-1} s_{1t}^*(\alpha_0)\right\}$, as shown by Gonçalves and de Jong (2003). Assumption B.4 is a restatement of Assumption 2.2 of Gonçalves and White (2002) and is satisfied when the models are correctly specified or when the scores $\left\{s_{1t}\left(X^t, \alpha_0\right)\right\}$ and $\left\{s_{2t}\left(X^t, \alpha_0, \beta_0\right)\right\}$ are stationary (this follows if $\{X_t\}$ is a strictly stationary process, the likelihood functions $\left\{f_{1t}(\alpha)\right\}$ and $\left\{f_{2t}(\alpha, \beta)\right\}$ depend only on a finite number of lags of X_t and there is no time heterogeneity on $\left\{f_{1t}\right\}$ and $\left\{f_{2t}\right\}$). Under this assumption, the bootstrap covariance matrix of the scaled average of $\left\{s_{2t}^*(\alpha_0, \beta_0) - F_0 A_0^{-1} s_{1t}^*(\alpha_0)\right\}$ converges to J_0, the correct asymptotic covariance matrix of $\sqrt{n}\left(\hat{\beta}_n - \beta_0\right)$.

In the following theorem, and throughout, we let E^*, var* and P^* denote the bootstrap expectation, variance and probability measure induced by the resampling, conditional on the original sample.

Theorem 4.1. Let Assumption A as strengthened by Assumption B hold. If $\ell_n \to \infty$ and $\ell_n = o\left(n^{1/2}\right)$, then

$$\sup_{x \in \mathbb{R}^p} \left| P^*\left(\sqrt{n}\left(\hat{\beta}_n^* - \hat{\beta}_n\right) \le x\right) - P\left(\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \le x\right)\right| = o_P(1). \quad (1)$$

To prove Theorem 4.1, we verify the conditions of Theorem S2.4 in the supplementary materials. This result shows the consistency of the bootstrap distribution of a general two-step M-estimator $\hat{\beta}_n^*$ (based on an asymptotically linear first-step estimator $\hat{\alpha}_n^*$) under a set of bootstrap high level conditions (Assumption $\mathcal{B}^*$). We show that Assumption A strengthened by Assumption B verifies Assumption $\mathcal{B}^*$.

The first-order asymptotic validity of the fast resampling method is given in the next result. Its proof is a by-product of the proof of Theorem 4.1 and is omitted.

Theorem 4.2. Under the same assumptions as in Theorem 4.1,

$$\sup_{x \in \mathbb{R}^p} \left| P^*\left(\sqrt{n}\left(\hat{\beta}_{1,n}^* - \hat{\beta}_n\right) \le x\right) - P\left(\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \le x\right)\right| = o_P(1).$$

4.2. Bootstrap Distribution Consistency for Studentized Statistics

Here, we focus on testing hypotheses about β_0 based on Wald statistics.[5] Let $r : \mathcal{B} \rightarrow \mathbb{R}^q$, where $\mathcal{B} \subset \mathbb{R}^p$, $q \leq p$, be a continuously differentiable function on $\mathcal{B}$ such that $R_0 \equiv \frac{\partial}{\partial\beta'} r(\beta_0)$ has full row rank q. The Wald statistic for testing $\mathcal{H}_0 : r(\beta_0) = 0$ is

$$\mathcal{W}_n = n\hat{r}_n' \left(\hat{R}_n \hat{C}_n \hat{R}_n'\right)^{-1} \hat{r}_n,$$

where $\hat{r}_n = r\left(\hat{\beta}_n\right)$, $\hat{R}_n = \frac{\partial}{\partial\beta'} r\left(\hat{\beta}_n\right)$ and $\hat{C}_n = \hat{H}_n^{-1}\hat{J}_n\hat{H}_n^{-1}$ is consistent for C_0. In particular, $\hat{H}_n \equiv n^{-1}\sum_{t=1}^n \frac{\partial^2}{\partial\beta'\partial\beta} \log f_{2t}\left(\hat{\alpha}_n, \hat{\beta}_n\right)$ is a consistent estimator of H_0, and $\hat{J}_n$ is such that $\hat{J}_n - J_0 = o_P(1)$. We define the bootstrap Wald statistic as follows

$$\mathcal{W}_n^* = n\left(\hat{r}_n^* - \hat{r}_n\right)' \left(\hat{R}_n^* \hat{C}_n^* \hat{R}_n^{*\prime}\right)^{-1} \left(\hat{r}_n^* - \hat{r}_n\right).$$

For the fully optimized bootstrap method, we set $\hat{r}_n^* = r\left(\hat{\beta}_n^*\right)$, $\hat{R}_n^* = \frac{\partial}{\partial\beta'} r\left(\hat{\beta}_n^*\right)$ and

$$\hat{C}_n^* = \hat{H}_n^{*-1}\hat{J}_n^*\hat{H}_n^{*-1}, \text{ with} \tag{2}$$

$$\hat{H}_n^* = n^{-1}\sum_{t=1}^n \frac{\partial^2}{\partial\beta'\partial\beta} \log f_{2,\tau_t}\left(X^{\tau_t}, \hat{\alpha}_n^*, \hat{\beta}_n^*\right), \text{ and}$$

$$\hat{J}_n^* = k^{-1}\sum_{i=1}^k \left(\ell^{-1/2}\sum_{t=1}^\ell s_{I_{i+t}}\left(\hat{\alpha}_n^*, \hat{\beta}_n^*\right)\right)\left(\ell^{-1/2}\sum_{t=1}^\ell s_{I_{i+t}}\left(\hat{\alpha}_n^*, \hat{\beta}_n^*\right)\right)',$$

such that $s_{\tau_t}\left(\hat{\alpha}_n^*, \hat{\beta}_n^*\right) = s_{2\tau_t}\left(\hat{\alpha}_n^*, \hat{\beta}_n^*\right) - \hat{F}_n^*\hat{A}_n^{*-1}s_{1\tau_t}\left(\hat{\alpha}_n^*\right)$, where $\hat{A}_n^* = n^{-1}\sum_{t=1}^n \frac{\partial}{\partial\alpha'} s_{1\tau_t}\left(\hat{\alpha}_n^*\right)$ and $\hat{F}_n^* = n^{-1}\sum_{t=1}^n \frac{\partial}{\partial\alpha'} s_{2\tau_t}\left(\hat{\alpha}_n^*, \hat{\beta}_n^*\right)$.

For the fast resampling method, we set $\hat{r}_n^* = r\left(\hat{\beta}_{1,n}^*\right)$, $\hat{R}_n^* = \frac{\partial}{\partial\beta'} r\left(\hat{\beta}_{1,n}^*\right)$ and

$$\hat{C}_n^* = \hat{H}_n^{-1}\hat{J}_{1,n}^*\hat{H}_n^{-1}, \tag{3}$$

with $\hat{J}_{1,n}^* = k^{-1}\sum_{i=1}^k \left(\ell^{-1/2}\sum_{t=1}^\ell s_{I_{i+t}}\left(\hat{\alpha}_n, \hat{\beta}_n\right)\right)\left(\ell^{-1/2}\sum_{t=1}^\ell s_{I_{i+t}}\left(\hat{\alpha}_n, \hat{\beta}_n\right)\right)'$, such that $s_{\tau_t}\left(\hat{\alpha}_n, \hat{\beta}_n\right) = s_{2\tau_t}\left(\hat{\alpha}_n, \hat{\beta}_n\right) - \hat{F}_n\hat{A}_n^{-1}s_{1\tau_t}\left(\hat{\alpha}_n\right)$.

For the fast resampling approach proposed by Armstrong, Bertanha, and Hong (2014), we set $\hat{r}_n^* = r\left(\tilde{\beta}_{1,n}^*\right)$, $\hat{R}_n^* = \frac{\partial}{\partial\beta'} r\left(\tilde{\beta}_{1,n}^*\right)$ and

$$\hat{C}_n^* = \hat{H}_n^{-1}\widetilde{J}_{1,n}^*\hat{H}_n^{-1}, \tag{4}$$

where $\widetilde{J}_{1,n}^* = k^{-1}\sum_{i=1}^k \left(\ell^{-1/2}\sum_{t=1}^\ell s_{I_{i+t}}\left(\hat{\alpha}_n^*, \hat{\beta}_n\right)\right)\left(\ell^{-1/2}\sum_{t=1}^\ell s_{I_{i+t}}\left(\hat{\alpha}_n^*, \hat{\beta}_n\right)\right)'$, such that $s_{\tau_t}\left(\hat{\alpha}_n^*, \hat{\beta}_n\right) = s_{2\tau_t}\left(\hat{\alpha}_n^*, \hat{\beta}_n\right) - \hat{F}_n\hat{A}_n^{-1}s_{1\tau_t}\left(\hat{\alpha}_n^*\right)$.

Next we show that the sampling distribution of $\mathcal{W}_n$ is well approximated by the bootstrap distribution of $\mathcal{W}_n^*$. For this, we strengthen Assumption B4 as follows.

Assumption B.4′. $E\left(s_{1t}\left(\alpha_0\right)\right) = 0$ and $E\left(s_{2t}\left(\alpha_0, \beta_0\right)\right) = 0$ for all $t = 1, \ldots, n$.

This is a mild strengthening of Assumption B.4, which is satisfied whenever the score functions are not heterogeneous and/or the data $\{X_t\}$ are stationary.

Theorem 4.3. Let the assumptions of Theorem 4.1 as strengthened by Assumption B.4′ hold. Suppose further that $\mathcal{W}_n$ uses a consistent estimator of J_0. Then, under $\mathcal{H}_0$, if $\ell_n \rightarrow \infty$ and $\ell_n = o\left(n^{1/2}\right)$,

$$\sup_{x\in\mathbb{R}} \left|P^*\left(\mathcal{W}_n^* \leq x\right) - P\left(\mathcal{W}_n \leq x\right)\right| = o_P(1).$$

The proof of Theorem 4.3 is in the supplementary materials; it builds upon results in Gonçalves and White (2004, see, Theorem 3.1).

4.3. Bootstrap Variance Consistency

Bootstrap standard errors are often used in applied work as they are easy to compute, avoiding the need to look up complicated formulas. This is especially true in multistage estimation, where these formulas become quickly involved due to the need to keep track of the added uncertainty caused by each estimation stage. Instead, bootstrap standard errors are easily computed by Monte Carlo simulation. For instance, we can approximate the bootstrap variance estimator of the parameter $\hat{\beta}_{n,j}^*$, $\text{var}^*\left(\hat{\beta}_{n,j}^*\right)$, with the sample variance obtained across B replications of $\hat{\beta}_{n,j}^*$,

$$\frac{1}{B}\sum_{k=1}^B \left(\hat{\beta}_{n,j}^{*(k)} - \overline{\hat{\beta}_{n,j}^{*(k)}}\right)^2, \text{ where } \overline{\hat{\beta}_{n,j}^{*(k)}} = \frac{1}{B}\sum_{k=1}^B \hat{\beta}_{n,j}^{*(k)}.$$

The corresponding bootstrap standard error is the square root of this expression.

The previous results (Theorems 4.1 and 4.2) do not justify by themselves the consistency of bootstrap standard errors based on $\hat{\beta}_n^*$ or $\hat{\beta}_{1,n}^*$. The reason is that convergence in distribution of a random sequence does not imply convergence of moments. For instance, Ghosh et al. (1984) and Shao (1992) give examples of the inconsistency of bootstrap variance estimators for the sample median and smooth functions of sample means, respectively, in the iid context. Despite this, applied researchers routinely apply the bootstrap when computing standard errors. For a recent article emphasizing this point, see Hahn and Liao (2020), who show that bootstrap standard errors can lead to conservative inference.

The main goal of this section is to provide a theoretical justification for computing bootstrap standard errors in the context of two-step QMLE with time series data. The current

[5] The validity of the bootstrap for LM statistics follows from similar arguments. To conserve space and because Wald tests are more popular than LM tests, we only report results for Wald statistics.

bootstrap literature does not cover this case as it has either assumed iid data (as in Kato (2011) and Cheng (2015), who prove the consistency of bootstrap variance estimators for one-step M-estimators) or has considered time series least squares estimators, as in Gonçalves and White (2005). No results appear to be available for multistage QMLE, even for iid data.

Given our previous bootstrap distribution consistency results, a sufficient condition for showing the consistency of the corresponding bootstrap standard errors is to show that a uniform integrability condition holds. In particular, to show that $\text{var}^*\left(\sqrt{n}\hat{\beta}^*_n\right)$ is consistent, it suffices to show that $E^*\left|\sqrt{n}\left(\hat{\beta}^*_n - \hat{\beta}_n\right)\right|^{2+\delta} = O_P(1)$ for some small $\delta > 0$. Because $\hat{\beta}^*_{1,n}$ has a closed form expression, it is substantially easier to verify this condition for the fast resampling method than for the fully optimized bootstrap two-stage QMLE. For this reason, we focus on this estimator first.

We impose a smoothness condition on the vector of scores $\{s_{1t}\}$ and $\{s_{2t}\}$ which we did not need for bootstrap distribution consistency.

Assumption B.5.

i. $\{s_{1t}(\alpha)\}$ is Lipschitz continuous on $\mathcal{A}$, a.s.-P with Lipschitz functions $\{L_{1t}\}$ that satisfy the condition $n^{-1}\sum_{t=1}^{n} E(L_{1t})^{2+\delta} = O(1)$.
ii. $\{s_{2t}(\alpha,\beta)\}$ is Lipschitz continuous on $\mathcal{A}\times\mathcal{B}$, a.s.-$P$ with Lipschitz functions $\{L_{2t}\}$ satisfying the condition $n^{-1}\sum_{t=1}^{n} E(L_{2t})^{2+\delta} = O(1)$.

Theorem 4.4. Under Assumptions A and B strengthened by B.4′ and B.5, $\text{var}^*\left(\sqrt{n}\hat{\beta}^*_{1,n}\right) \xrightarrow{P} C_0$.

Next, we consider the fully optimized bootstrap estimator $\hat{\beta}^*_n$. Similarly to Theorem 4.4 , we prove the consistency of $\text{var}^*\left(\sqrt{n}\hat{\beta}^*_n\right)$ by relying on Theorem 4.1 and showing that $E^*\left|\sqrt{n}\left(\hat{\beta}^*_n - \hat{\beta}_n\right)\right|^{2+\delta} = O_P(1)$ for some small $\delta > 0$. Because $\hat{\beta}^*_n$ does not have a closed form expression, this condition is much harder to verify than for $\hat{\beta}^*_{1,n}$ and requires a different method of proof and a different set of assumptions.

Our proof and regularity conditions are inspired by 2011 and Cheng (2015). Kato (2011) shows the consistency of bootstrap moment estimators for M-estimators of parameteric models, whereas Cheng (2015) allows for semiparameteric models, where the parameter of interest is a finite dimensional parameter, but the model also contains a nuisance parameter that is potentially infinite dimensional. Both articles focus on one-step M-estimators and give sufficient conditions for bootstrap variance consistency that only cover iid data. Our contribution is to extend those results to two-stage M-estimation with time series data.

To present our regularity conditions, we need to introduce more notation. First, because our proof is based on showing that the unconditional moment of $\left|\sqrt{n}\left(\hat{\beta}^*_n - \hat{\beta}_n\right)\right|^{2+\delta}$ is finite, we need to introduce the joint probability measure $\mathbb{P} = P \times P^*$ that accounts for the two sources of randomness in $\hat{\beta}^*_n$: the randomness that comes from the original data (and which is described by P) and the randomness that comes from the resampling, conditional on the original sample (described by P^*). In the following, we write $\mathbb{E}$ to denote expected value with respect to $\mathbb{P}$. Second, to prove that $\mathbb{E}\left|\sqrt{n}\left(\hat{\beta}^*_n - \hat{\beta}_n\right)\right|^{2+\delta} < \infty$, we assume the uniform square integrability of the original two-step QMLE estimator (i.e., we assume that $E\left|\sqrt{n}\left(\hat{\beta}_n - \beta_0\right)\right|^{2+\delta} < \infty$) and provide regularity conditions that allows us to show that $\mathbb{E}\left|\sqrt{n}\left(\hat{\beta}^*_n - \beta_0\right)\right|^{2+\delta} < \infty$. We follow Kato (2011) and Cheng (2015) and use an argument that entails bounding the tail probability $\mathbb{P}\left(\left|\sqrt{n}\left(\hat{\beta}^*_n - \beta_0\right)\right| > u\right)$ for large u. This requires empirical process theory and maximal inequalities. In particular, we impose bounds on the L_p-moments (with $p > 2+\delta$) of the supremum of certain empirical processes which we describe next.

For any class of functions $\mathcal{F} = \{f_t\}$, define the empirical process $\mathbb{G}_n f = n^{-1/2}\sum_{t=1}^{n}\left(f_t - Ef_t\right)$ and let its norm be given by $\|\mathbb{G}_n\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}}\left|\mathbb{G}_n f\right|$.

Our assumptions are as follows.

Assumption B.6.

i. For any $(\alpha,\beta) \in \mathcal{A}\times\mathcal{B}$, the log-likelihood function $\log f_2(\cdot,\alpha,\beta)$ and its expectation $\bar{Q}_2(\alpha,\beta) \equiv E\left(\log f_2\left(X^t,\alpha,\beta\right)\right)$ are time invariant.
ii. There exists a positive constant K independent of β for which for all $\beta \in \mathcal{B}$, $\bar{Q}_2(\alpha_0,\beta) - \bar{Q}_2(\alpha_0,\beta_0) \le -K\|\beta-\beta_0\|^2$.
iii. Given $\eta > 0$, define the class of functions

$$\mathcal{N}_\eta = \{\log f_2(\alpha,\beta) - \log f_2(\alpha,\beta_0) : \|\beta-\beta_0\| \le \eta, (\alpha,\beta)\in\mathcal{A}\times\mathcal{B}\}.$$

Then, for some $p > 2+\delta$, and every $\eta > 0$, there exists a positive constant K such that

$$\left[E\left(\|\mathbb{G}_n\|^p_{\mathcal{N}_\eta}\right)\right]^{1/p} \le K\eta. \tag{5}$$

iv. The functions $\{\log f_2(\alpha,\beta)\}$, $\left\{\frac{\partial}{\partial\alpha'}\log f_2(\alpha,\beta)\right\}$ and $\left\{\frac{\partial}{\partial\alpha\partial\alpha'}\log f_2(\alpha,\beta)\right\}$ satisfy a Lipschitz continuity condition on $\mathcal{A}\times\mathcal{B}$, a.s.-$P$ with Lipschitz functions $\{L_t\}$, $\{L_{1t}\}$ and $\{L_{2t}\}$ such that $E|L_t|^p < \infty$, $E\left(|L_{1t}|^{\frac{\varepsilon}{\varepsilon-1}p}\right) < \infty$ and $E\left(|L_{2t}|^{\frac{\varepsilon}{\varepsilon-1}p}\right) < \infty$, respectively, for $p > 2+\delta$ as in (iii) and for some $\varepsilon > 1$.
v. The first-step estimator $\hat{\alpha}_n$ and its bootstrap analog $\hat{\alpha}^*_n$ are such that

$$E\left|\sqrt{n}\left(\hat{\alpha}_n - \alpha_0\right)\right|^{3\varepsilon p} = O(1) \quad \text{and}$$
$$\mathbb{E}\left|\sqrt{n}\left(\hat{\alpha}^*_n - \hat{\alpha}_n\right)\right|^{3\varepsilon p} = O(1), \tag{6}$$

where $\varepsilon > 1$ and $p > 2+\delta$ are as defined in (iv).
vi. $\sup_n E\left|\sqrt{n}\left(\hat{\beta}_n - \beta_0\right)\right|^{2+\delta} < \infty$.

Assumption B.6 (i) assumes that the log-likelihood functions f_{2t} and the population criterion function $\bar{Q}_2(\alpha,\beta) \equiv$

$E\left(\log f_2\left(X^t,\alpha,\beta\right)\right)$ are time invariant (the latter will follow from the first under stationarity of $\{X_t\}$).

To understand Assumptions B.6(ii)–(iv), suppose that β is a scalar and consider for example the generated regressor problem. In particular, suppose the following regression model $y_t = \beta q_t + e_t$, where the regressor $q_t = w_t\alpha$ is latent because we do not observe α . If $x_t = w_t\alpha + \eta_t$, where x_t and w_t are observed, we can obtain a consistent estimator of α by running an OLS regression of x_t on w_t. This yields a generated regressor $\hat{q}_t = w_t\hat{\alpha}$, which we can then use to obtain a consistent estimator of β. In terms of our notations, and letting $e_t|w_t, x_t \sim \text{iid}\, N\left(0,\sigma^2\right)$, the log-likelihood function underlying the second step is given by $\log f_{2t}(\alpha,\beta) = -\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left(y_t - w_t\alpha\beta\right)^2$ and $\bar{Q}_2(\alpha_0,\beta) = -\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}E\left(y_t - w_t\alpha_0\beta\right)^2$. This verifies condition B.6(i). To verify Condition B.6(ii), note that by a second-order Taylor expansion of $\bar{Q}_2(\alpha_0,\beta)$ around β_0, we get $\bar{Q}_2(\alpha_0,\beta) = \bar{Q}_2(\alpha_0,\beta_0) + \frac{\partial}{\partial\beta}\bar{Q}_2(\alpha_0,\beta_0)(\beta-\beta_0) + \frac{1}{2}\frac{\partial^2}{\partial\beta^2}\bar{Q}_2\left(\alpha_0,\ddot{\beta}\right)(\beta-\beta_0)^2$, where $\ddot{\beta}$ lies between β and β_0. Since (α_0,β_0) maximizes $\bar{Q}_2(\alpha,\beta)$, $\frac{\partial}{\partial\beta}\bar{Q}_2(\alpha_0,\beta_0) = 0$, implying that $\bar{Q}_2(\alpha_0,\beta) = \bar{Q}_2(\alpha_0,\beta_0) + \frac{1}{2}\frac{\partial^2}{\partial\beta^2}\bar{Q}_2\left(\alpha_0,\ddot{\beta}\right)(\beta-\beta_0)^2$. So, the condition will be satisfied if we can bound $\frac{\partial^2}{\partial\beta^2}\bar{Q}_2(\alpha_0,\beta)$ by a negative constant $-K$, for any value of β. This is true if $\bar{Q}_2(\alpha_0,\beta)$ is a quadratic function of β, as in the generated regressor problem. This is a strong condition since it imposes a global restriction on $\bar{Q}_2(\alpha_0,\beta)$, but it is crucial for controlling the tail probability $\mathbb{P}\left(\left|\sqrt{n}\left(\hat{\beta}_n^* - \beta_0\right)\right| > u\right)$, as Kato (2011) and Cheng (2015) note. A similar condition is also used by Nishiyama (2010) to prove the moment convergence of the original M-estimator.

Assumption B.6(iii) (see, Equation (5)) is a high level condition on the empirical process $\mathbb{G}_n$ Cheng (2015) relies on a similar assumption to show the consistency of bootstrap one-step moment estimators of any order $p \geq 1$ for iid data. This so-called L_p -maximal inequality condition can be verified under more primitive conditions involving the structure of the function class $\mathcal{N}_\eta$, for example, Cheng (2015) shows that (for one-step M-estimators with iid data) condition (5) is implied by a finite uniform entropy integral condition, which is verified when the functions in $\mathcal{N}_\eta$ are Lipschitz continuous. Our Assumption B.6(iii) adapts Cheng (2015)'s high level condition to the two-step QMLE context, where we use the function class $\mathcal{N}_\eta$.

To verify Condition B.6(iii) in the context of the generated regressor problem, note that $\log f_{2t}(\alpha,\beta) - \log f_{2t}(\alpha,\beta_0) = (w_t\alpha(\beta-\beta_0))\left(2y_t - w_t\alpha(\beta+\beta_0)\right)$. Given that β and $\beta_0 \in B$, a compact subset of $\mathbb{R}$, it follows that $\left|\log f_{2t}(\alpha,\beta) - \log f_{2t}(\alpha,\beta_0)\right| \leq C\left|\left|w_t\right|\left(\left|y_t\right| + \left|w_t\right|\right)\right|\left|\beta-\beta_0\right|$, for some sufficiently large constant C. Hence, the condition follows by Remark 2.3 of Kato (2011) (relying for instance on Lemma 2.14.1 of van der Vaart and Wellner (1996) provided that $E\left|\left|w_t\right|\left(\left|y_t\right| + \left|w_t\right|\right)\right|^p < \infty < \infty$, for some $p > 2+\delta$.

Assumptions B.6(iv) and (v) are new to the two-step estimators we treat here. Part (iv) imposes a Lipschitz continuity condition on the score and the Hessian of $\log f_{2t}(\alpha,\beta)$ with respect to α. Note that in the example of OLS with generated regressor, condition (iv) follows provided that $E\left(\left|w_t^2 + y_t^2\right|^{\frac{\varepsilon}{\varepsilon-1}p}\right) < \infty$, for $p > 2+\delta$ as in Assumption B.6(iii) and for some $\varepsilon > 1$.

Part (v) imposes uniform integrability conditions on the first-step estimator $\hat{\alpha}_n$ and its bootstrap analog $\hat{\alpha}_n^*$. Similarly, part (vi) assumes the uniform integrability condition on $\hat{\beta}_n$. These high level conditions could be derived from more primitive conditions such as the ones used by Cheng (2015) or Kato (2011), but we prefer to state them as high level conditions since our focus is on the second-step bootstrap estimator $\hat{\beta}_n^*$. It is nevertheless interesting to note that stronger than usual uniform square integrability conditions on the first-step estimators are imposed in order to verify the uniform square integrability condition on the second stage bootstrap estimator $\hat{\beta}_n^*$. In particular, we require the existence of a bit more than six moments for $\sqrt{n}\left(\hat{\alpha}_n - \alpha_0\right)$ and its bootstrap analogue. This is three times more than the number of moments for the second-step estimators $\hat{\beta}_n$ and $\hat{\beta}_n^*$. When the log-likelihood function $\log f_{2t}$ is quadratic in α and β, Assumption B.6 (v) can be weakened to $E\left|\sqrt{n}\left(\hat{\alpha}_n - \alpha_0\right)\right|^{2\varepsilon p} = O(1)$ and $\mathbb{E}\left|\sqrt{n}\left(\hat{\alpha}_n^* - \hat{\alpha}_n\right)\right|^{2\varepsilon p} = O(1)$. Under these assumptions, we can prove the following theorem.

Theorem 4.5. Suppose Assumptions A and B strengthened by Assumption B.6 holds. Then, for some $\delta > 0$, $\sup_n \mathbb{E}\left|\sqrt{n}\left(\hat{\beta}_n^* - \hat{\beta}_n\right)\right|^{2+\delta} < \infty$, implying that $\text{var}^*\left(\sqrt{n}\hat{\beta}_n^*\right) \xrightarrow{P} C_0$.

5. Monte Carlo Simulations

We here assess the properties of the bootstrap approximation proposed in Sections 3 and 4. We do so via detailed and realistic Monte Carlo simulations and we start by describing the design of the study. We consider a bivariate copula-based model. Each variable's marginal distribution is an AR(1)-GARCH(1,1) with standardized Student's t errors:

$$X_{it} = \phi_{0,i} + \phi_{1,i}X_{i,t-1} + \varepsilon_{it}, \quad \text{where } \varepsilon_{it} = \sigma_{it}\eta_{it},$$
$$\sigma_{it}^2 = \widetilde{\omega}_i + \widetilde{\alpha}_i\varepsilon_{i,t-1}^2 + \widetilde{\beta}_i\sigma_{i,t-1}^2 \text{ and } \eta_{it} \sim \text{iid}\, t(0,1,\nu_i).$$

We examine the case where the amount of dependence between the two variables X_1 and X_2 is related to the Clayton copula, with parameter $\beta = 1$, which roughly implies linear correlation of 0.5 (see Nelsen 1999) for more on this copula). We use parameters similar to those found in applied work (see, e.g., Oh and Patton 2013): for $i = 1,2$, $\left[\phi_{0,i},\phi_{1,i}\right] = [0,0.4]$, $\left[\widetilde{\omega}_i,\widetilde{\alpha}_i,\widetilde{\beta}_i\right] = [0.05, 0.05, 0.9]$, and $\nu_1 = \nu_2 = \nu$, such that $\nu \in \{6,10\}$.

Thus, we have two DGPs, which differ only in the value of the Student's t parameter ν, which control the thickness of the tail of the distributions. Note that when $\nu \to \infty$, this implies that $\eta \sim N(0,1)$. We generate repeated trials of length $n \in \{200, 500, 2500\}$ from these processes and conduct inference based on a misspecified model that assumes $\phi_{1,i} = 0$ (i.e., we estimate a constant mean-GARCH(1,1)-Student's t-Clayton copula model for each trial).

It is easy to see that our bivariate density models constructed using copulas can be partitioned into elements relating only to a marginal distribution and elements that relate only to the copula. As pointed out by Joe (1995) and Patton (2006), when such a partition is not possible, the familiar one-stage maximum likelihood estimator is the natural estimator to employ. However, when this partitioning is possible as in our simulation setting, great computational savings may be achieved by employing a multi-stage estimator. Therefore, in the following we consider the multi-stage maximum likelihood estimator (MSMLE).

Our estimation steps are:

1. Estimate the conditional mean of X_i, $i = 1, 2$, using the sample mean $n^{-1}\sum_{t=1}^{n} X_{it}$. Obtain the estimated residuals $\hat{\varepsilon}_{it} = X_{it} - n^{-1}\sum_{t=1}^{n} X_{it}$.
2. Estimate the conditional variance parameter using QML (with normal log-likelihood) and the residuals from step 1, conditioning on the realizations for $t = 1$. Obtain the estimated standardized residuals $\hat{\eta}_{it}$.
3. Estimate ν using ML and the standardized residuals from step 2. Obtain the estimated probability integral transforms (PITs) $\hat{G}_{it}$.
4. Estimate β using ML and the estimated PITs from step 3.

More generally in a multivariate d-dimensional application (with $d \geq 2$) there are a total of $3d + 1$ estimation steps: three steps for each marginal distribution, and 1 step for the copula. With misspecification as in our context, the scores are generally correlated, hence, justifying the use of blockwise bootstrap methods. Furthermore, for misspecified models, the QMLE is generally inconsistent for the copula true parameter β, instead confidence intervals for the copula pseudo-true parameter β_0 are obtained. We evaluate by simulation, the pseudo-true parameter using 1 million observations.

To generate the bootstrap data, we use the MBB. The number of Monte Carlo trials is 1000 with $B = 999$ bootstrap replications each. We implement three resampling methods: the fully optimized bootstrap procedure B1, the fast resampling approach B2 and the fast resampling approach proposed by Armstrong, Bertanha, and Hong (2014) B3. To select the block size, we rely on the asymptotic equivalence between the MBB and the Bartlett kernel variance estimators, and choose ℓ equal to the bandwidth chosen by Andrews's automatic procedure for the Bartlett kernel.

We consider three types of confidence intervals for the copula pseudo-true parameter β_0: asymptotic normal theory-based confidence intervals, computed by using the quantile of the standard normal distribution, bootstrap percentile confidence intervals, and bootstrap percentile-t confidence intervals, which use the bootstrap methods (B1, B2, and B3) to compute critical values for the nonstudentized and studentized statistics based on $\hat{\beta}_n$, respectively. The asymptotic normal theory-based confidence interval for β_0 is given by $\hat{\beta}_n \pm 1.96 \cdot \widehat{\text{SE}}\left(\hat{\beta}_n\right)$, where $\widehat{\text{SE}}\left(\hat{\beta}_n\right)$ is a consistent estimator of $\text{SE}\left(\hat{\beta}_n\right) = \sqrt{\text{var}\left(\hat{\beta}_n\right)}$. Four choices are used to compute $\widehat{\text{SE}}\left(\hat{\beta}_n\right)$. For our first choice of $\widehat{\text{SE}}\left(\hat{\beta}_n\right)$, we use the multi-stage maximum likelihood (MSML) standard errors estimator as described in detail in Section 3.1.1 in Patton (2012) (see Equation (41)). Then, we construct a MSMLE variance, asymptotic normal theory-based confidence interval for β_0 by using $\widehat{\text{SE}}\left(\hat{\beta}_n\right) = \widehat{\text{SE}}^{\text{MSML}}\left(\hat{\beta}_n\right)$, where $\widehat{\text{SE}}^{\text{MSML}}\left(\hat{\beta}_n\right)$ is the estimated MSML standard error of $\hat{\beta}_n$ (which has a sandwich form).

In our second, third and fourth choices of $\widehat{\text{SE}}\left(\hat{\beta}_n\right)$, we use the bootstrap approaches B1, B2, and B3. In particular, a fully optimized bootstrap procedure variance, asymptotic normal theory-based confidence interval for β_0 can be obtained with $\widehat{\text{SE}}\left(\hat{\beta}_n\right) = \widehat{\text{SE}}^{\text{B1}}\left(\hat{\beta}_n\right)$, where $\widehat{\text{SE}}^{\text{B1}}\left(\hat{\beta}_n\right)$ is the estimated standard error of $\hat{\beta}_n$ based on B1. Similarly, fast resampling procedure variance, asymptotic normal theory-based confidence intervals for β are obtained by using $\widehat{\text{SE}}\left(\hat{\beta}_n\right) = \widehat{\text{SE}}^{\text{B2}}\left(\hat{\beta}_n\right)$ and $\widehat{\text{SE}}\left(\hat{\beta}_n\right) = \widehat{\text{SE}}^{\text{B3}}\left(\hat{\beta}_n\right)$, where $\widehat{\text{SE}}^{\text{B2}}\left(\hat{\beta}_n\right)$ and $\widehat{\text{SE}}^{\text{B3}}\left(\hat{\beta}_n\right)$ are the estimated standard error of $\hat{\beta}_n$ based on B2 and B3, respectively. Note that, the standard errors $\widehat{\text{SE}}^{\text{B1}}\left(\hat{\beta}_n\right)$, $\widehat{\text{SE}}^{\text{B2}}\left(\hat{\beta}_n\right)$ and $\widehat{\text{SE}}^{\text{B3}}\left(\hat{\beta}_n\right)$ are obtained by computing the statistics $\sqrt{\frac{1}{B}\sum_{i=1}^{B}\left(\hat{\beta}_n^{*(i)} - \overline{\hat{\beta}_n^{*}}\right)^2}$ with $\overline{\hat{\beta}_n^{*}} = \frac{1}{B}\sum_{j=1}^{B}\hat{\beta}_n^{*(j)}$, where B is the number of bootstrap replications.

The second set of intervals we consider are bootstrap percentile confidence intervals, which are very simple to compute since they avoid the need to explicitly compute standard errors. For each resampling approach (B1, B2, and B3), a symmetric bootstrap percentile confidence intervals for β_0 is given by $\hat{\beta}_n \pm p_{95}^{*}$, where p_{95}^{*} is the 95% quantile of the bootstrap distribution of $|\hat{\beta}_n^{*} - \hat{\beta}_n|$.

Finally, we also consider bootstrap percentile-t confidence intervals. For each resampling approach, a symmetric bootstrap percentile-t confidence intervals for β_0 is given by $\hat{\beta}_n \pm q_{95}^{*}\widehat{\text{SE}}^{\text{MSML}}\left(\hat{\beta}_n\right)$, where q_{95}^{*} is the 95% quantile of the bootstrap distribution of $\sqrt{n}|\hat{\beta}_n^{*} - \hat{\beta}_n|/\sqrt{\hat{C}_n^{*}}$, with $\hat{C}_n^{*}$ as in (2), (3), and (4), for B1, B2, and B3, respectively.

Table 1 reports the coverage rates and lengths of 95% confidence intervals of the copula pseudo-true parameter β_0 for the two DGPs, respectively. Results in Table 1 are not too sensitive to the value of the Student's t parameter ν. For a given sample size, all intervals have approximately the same coverage rate, with only small differences among them. Both fast resampling procedures B2 and B3, and the MSML approach perform well even for the small sample size $n = 200$. Indeed, the evidence of presence of serial correlation in the scores is confirmed by the average value of the block sizes chosen by Andrews (1991) method, which is equal to 3.90 in our simulations. However, as Table 1 suggests, there are notable differences among the different methods when considering their confidence interval lengths. This table clearly shows that for all DGP's, the intervals based on B1 (either using the CLT-based or the bootstrap percentile and/or percentile-t approach) tend to display shorter intervals for the smaller sample sizes compared to CLT-MSML, B2 and the B3 intervals.

Note that all three asymptotic normal theory-based confidence intervals differ only by the way that the estimated standard errors of $\hat{\beta}_n$ have been computed. In order to gain

Table 1. Coverage rates and length confidence intervals of nominal 95% intervals for β_0.

	CLT				Percentile			Percentile-t		
	CLT-MSML	CLT-B1	CLT-B2	CLT-B3	B1	B2	B3	B1	B2	B3
Coverage rates										
					$\nu = 6$					
$n = 200$	90.70	91.40	90.60	90.70	90.80	90.00	90.30	91.70	91.80	91.90
$n = 500$	92.50	92.70	92.40	92.90	92.50	92.80	92.80	93.60	93.00	93.40
$n = 2500$	93.90	94.20	94.00	94.00	94.40	94.20	94.10	94.90	94.40	94.80
					$\nu = 10$					
$n = 200$	89.50	89.30	90.10	90.20	89.60	89.50	89.40	90.20	90.30	90.40
$n = 500$	92.20	92.60	92.50	92.70	92.30	92.70	92.50	93.00	93.00	93.10
$n = 2500$	94.20	94.90	94.50	94.20	94.60	94.70	94.80	95.50	94.70	95.60
Length confidence intervals										
					$\nu = 6$					
$n = 200$	0.66	0.61	0.65	0.63	0.61	0.68	0.64	0.60	0.68	0.64
$n = 500$	0.42	0.42	0.41	0.41	0.42	0.43	0.42	0.41	0.44	0.49
$n = 2500$	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
$\nu = 10$										
$n = 200$	0.70	0.61	0.70	0.66	0.62	0.73	0.63	0.61	0.73	0.68
$n = 500$	0.43	0.41	0.43	0.43	0.42	0.45	0.42	0.42	0.45	0.44
$n = 2500$	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.20

NOTE: CLT-MSML, CLT-B1, CLT-B2, and CLT-B3 -intervals based on the normal using estimated standard error based on the MSML, the B1, the B2, and the B3, respectively; B1 bootstrap intervals based on the fully optimized procedure, B2 bootstrap intervals based on the fast resampling procedure, B3 bootstrap intervals based on the fast resampling approach proposed by Armstrong, Bertanha, and Hong (2014). 1000 Monte Carlo trials with 999 bootstrap replications each. Pseudo-true parameters were calculated using one million observations. For $\nu = 6$, $\beta_0 = 0.99$, and $\beta_0 = 1.07$ when $\nu = 10$.

Table 2. Comparison of standard errors estimation of $\hat{\beta}_n$.

	$n = 200$				$n = 500$				$n = 2500$			
	MSML	B1	B2	B3	MSML	B1	B2	B3	MSML	B1	B2	B3
$\nu = 6$												
(MSE of estimated SE)$\cdot 10^3$	1.31	0.82	1.26	0.93	0.28	0.16	0.28	0.15	0.05	0.01	0.05	0.02
Ratio of SE over the true value	1.12	1.01	1.14	0.99	1.00	1.00	0.99	0.99	1.00	0.99	1.00	1.01
$\nu = 10$												
(MSE of estimated SE)$\cdot 10^3$	7.03	0.96	6.31	3.28	0.42	0.16	0.31	0.19	0.17	0.02	0.15	0.03
Ratio of SE over the true value	1.11	1.01	1.12	1.02	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99

NOTE: This table provides the MSE of estimated standard errors and the ratio of estimated standard errors over the true value (simulation based) of standard errors. We compute the ratio of standard error of an estimator $\hat{\theta}$ over the true value as: $S^{-1}\sum_{i=1}^{S} \frac{\hat{\mathrm{SE}}_i^j\left(\hat{\theta}\right)}{\mathrm{SE}\left(\hat{\theta}\right)}$, where S is the number of Monte Carlo replications, $i = 1, \ldots, S$, $j =$ MSML, B1, B2, B3 thus, $\hat{\mathrm{SE}}_i^j\left(\hat{\theta}\right)$ is the estimated value of the standard error of an estimator $\hat{\theta}$ on the ith Monte Carlo replication obtained by using the method j. $\mathrm{SE}\left(\hat{\theta}\right)$ is defined as $\mathrm{SE}\left(\hat{\theta}\right) = \sqrt{S^{-1}\sum_{i=1}^{S}\left(\hat{\theta}_i - S^{-1}\sum_{s=1}^{S}\hat{\theta}_s\right)^2}$, with $\hat{\theta}_i$ the estimated value of the parameter θ on the ith Monte Carlo replication. Similarly, we compute the MSE as $S^{-1}\sum_{i=1}^{S}\left(\hat{\mathrm{SE}}_i^j\left(\hat{\theta}\right) - \mathrm{SE}\left(\hat{\theta}\right)\right)^2$. Simulations where done with 1000 Monte Carlo trials with 999 bootstrap replications each.

further insight into the "relatively" good performance of these asymptotic normal theory-based confidence intervals in finite samples, we compute the ratio of the estimated standard error over the true value and the mean-square error (MSE) of the estimated standard errors. The results are presented in Table 2. For small sample sizes, on average MSML and B2 overestimate the standard errors, with the ratio of estimated standard error over the true value above 1. For instance, when $n = 200$ and $\nu = 6$, the ratio of estimated standard error over the true value based on MSML and B2 are 1.12 and 1.14 for MSML and B2, respectively, whereas this ratio is 1.01 for B1 and 0.99 for B3. Consequently, the length of confidence intervals based on estimated standard error from MSML and B2 are larger than the one based on B1 and/or B3.

The gains associated with the bootstrap methods can be quite substantial when the main goal of the researcher and/or practitioner is to estimate the standard errors. The results in Table 2 are in favor of the bootstrap particularly for small sample sizes. More specifically, the full resampling method B1 is better than using MSML and/or B2 standard errors. For small samples, the bootstrap method B1 estimates the standard error of the copulas parameter estimator $\hat{\beta}_n$ more precisely than the MSML, B2 and B3 approaches. For large sample sizes, we have approximately the same performance for all three methods. For instance, when $n = 200$ and $\nu = 10$, the MSE of the estimated standard errors of $\hat{\beta}_n$ based on MSML, B1, B2, and B3 are $7.03 \cdot 10^{-3}$, $0.96 \cdot 10^{-3}$, $6.31 \cdot 10^{-3}$, and $3.28 \cdot 10^{-3}$, respectively. Whereas, for $n = 2,500$ and $\nu = 10$, the MSE become $0.17 \cdot 10^{-3}$, $0.02 \cdot 10^{-3}$, $0.15 \cdot 10^{-3}$, and $0.03 \cdot 10^{-3}$, respectively. Thus we see that although all four methods MSML, B1, B2, and B3 are asymptotically equivalent, and the full resampling method B1 may be computationally much more demanding, in small samples, the improved estimates of the standard errors based on B1 may outweigh its computational cost. Overall, the performance of B2 is comparable to that of MSML, B3 outperforms B2, whereas B1 outperforms B2, B3 and MSML and provides more accurate estimators of the standard errors, especially when the sample size is small.

6. Conclusion

This article proposes and theoretically justifies bootstrap methods for inference on nonlinear dynamic models that are estimated in two (or more) stages of quasi-maximum likelihood estimation (QMLE). In particular, we show the consistency of the bootstrap distribution of the two-step QML estimator using dependence and heterogeneity conditions similar to those used by Gonçalves and White (2004) for the one-step case. In addition, we also prove the consistency of bootstrap standard errors for two-step QMLE, a result that does not seem to be available even for iid data. This justifies the standard practice of computing bootstrap standard errors instead of computing analytical standard errors, which quickly becomes cumbersome in the multistage QMLE context. Our simulation results show that intervals based on bootstrap standard errors or bootstrap percentile intervals obtained with the fully optimized method that resamples the log-likelihood functions jointly are shorter on average than intervals based only on asymptotic theory or on the fast resampling method we propose. Thus, although more computationally demanding, the fully optimized bootstrap method has better finite sample properties than the other methods we consider.

Appendix A

In this appendix, we provide a set of primitive assumptions under which the asymptotic theory of the two-step QMLE (consistency and asymptotic distribution) follows. These assumptions extends the assumptions of GW (2004) to the two-step QMLE context and are used to prove our bootstrap results.

Assumption A.

A.1: Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. The observed data are a realization of a stochastic process $\left\{X_t : \Omega \to \mathbb{R}^l, t \in \mathbb{N}\right\}$, with

$$X_t(\omega) = W_t(\ldots, V_{t-1}(\omega), V_t(\omega), V_{t+1}(\omega), \ldots),$$

$V_t : \Omega \to \mathbb{R}^v$, and $W_t : \times_{\tau=-\infty}^{\infty} \mathbb{R}^v \to \mathbb{R}^l$ is such that X_t is measurable for t.

A.2: i. The functions $\left\{f_{1t}\left(X^t, \alpha\right)\right\}$ are such that $f_{1t}(\cdot, \alpha)$ is measurable for each $\alpha \in \mathcal{A}$, where $\mathcal{A}$ is a compact subset of $\mathbb{R}^k$, $f_{1t}\left(X^t, \cdot\right)$ is continuous on $\mathcal{A}$, a.s.-P, and $f_{1t}\left(X^t, \cdot\right)$ is twice continuously differentiable on *int* $(\mathcal{A})$, a.s-P.

ii. The functions $\left\{f_{2t}\left(X^t, \alpha, \beta\right)\right\}$ are such that $f_{2t}(\cdot, \alpha, \beta)$ is measurable for each $(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$, where $\mathcal{B}$ is a compact subset of $\mathbb{R}^p$, $f_{2t}\left(X^t, \cdot, \cdot\right)$ is continuous on $\Theta = \mathcal{A} \times \mathcal{B}$, a.s.-$P$, and $f_{2t}\left(X^t, \cdot, \cdot\right)$ is twice continuously differentiable on *int* (Θ), a.s-P.

A.3: i. α_0 is the unique maximizer of $\bar{Q}_1(\alpha) \equiv \lim_{n\to\infty} E\left(Q_{1n}(\alpha)\right)$ on $\mathcal{A}$.

ii. β_0 is the unique maximizer of $\bar{Q}_2(\alpha_0, \beta) \equiv \lim_{n\to\infty} E\left(Q_{2n}(\alpha_0, \beta)\right)$ on $\mathcal{B}$.

iii. $\theta_0 = (\alpha_0, \beta_0)$ is interior to $\Theta = \mathcal{A} \times \mathcal{B}$.

A.4: i. The functions $\left\{\log f_{1t}\left(X^t, \alpha\right)\right\}$ and $\left\{\frac{\partial}{\partial \alpha'} s_{1t}\left(X^t, \alpha\right)\right\}$ are Lipschitz continuous on $\mathcal{A}$, a.s.-P, where $s_{1t}\left(X^t, \alpha\right) \equiv \frac{\partial}{\partial \alpha} \log f_{1t}\left(X^t, \alpha\right)$.

ii. The functions $\left\{\log f_{2t}\left(X^t, \alpha, \beta\right)\right\}$, $\left\{\frac{\partial}{\partial \beta'} s_{2t}\left(X^t, \alpha, \beta\right)\right\}$ and $\left\{\frac{\partial}{\partial \alpha'} s_{2t}\left(X^t, \alpha, \beta\right)\right\}$ are Lipschitz continuous on $\mathcal{A} \times \mathcal{B}$, a.s.-$P$, where $s_{2t}\left(X^t, \alpha, \beta\right) \equiv \frac{\partial}{\partial \beta} \log f_{2t}\left(X^t, \alpha, \beta\right)$.

A.5: For some $r > 2$,

i. The functions $\left\{\log f_{1t}\left(X^t, \alpha\right)\right\}$, $\left\{s_{1t}\left(X^t, \alpha\right)\right\}$ and $\left\{\frac{\partial}{\partial \alpha'} s_{1t}\left(X^t, \alpha\right)\right\}$ are r-dominated on $\mathcal{A}$ uniformly in t.

ii. The functions $\left\{\log f_{2t}\left(X^t, \alpha, \beta\right)\right\}$, $\left\{s_{2t}\left(X^t, \alpha, \beta\right)\right\}$, $\left\{\frac{\partial}{\partial \beta'} s_{2t}\left(X^t, \alpha, \beta\right)\right\}$ and $\left\{\frac{\partial}{\partial \alpha'} s_{2t}\left(X^t, \alpha, \beta\right)\right\}$ are r -dominated on $\Theta = \mathcal{A} \times \mathcal{B}$ uniformly in t.

A.6: $\{V_t\}$ is an α-mixing sequence of size $-\frac{2r}{r-2}$, with $r > 2$.

A.7: The elements of

i. $\left\{\log f_{1t}\left(X^t, \alpha\right)\right\}$ and $\left\{\frac{\partial}{\partial \alpha'} s_{1t}\left(X^t, \alpha\right)\right\}$ are L_2-NED on $\{V_t\}$ of size $-\frac{1}{2}$, and those of $\left\{s_{1t}\left(X^t, \alpha\right)\right\}$ are L_2-NED on $\{V_t\}$ of size -1, uniformly on $(\mathcal{A}, \rho)$, where ρ is a metric on $\mathbb{R}^k$;

ii. $\left\{\log f_{2t}\left(X^t, \alpha, \beta\right)\right\}$, $\left\{\frac{\partial}{\partial \beta'} s_{2t}\left(X^t, \alpha, \beta\right)\right\}$ and $\left\{\frac{\partial}{\partial \alpha'} s_{2t}\left(X^t, \alpha, \beta\right)\right\}$ are L_2 -NED on $\{V_t\}$ of size $-\frac{1}{2}$, and those of $\left\{s_{2t}\left(X^t, \alpha, \beta\right)\right\}$ are L_2-NED on $\{V_t\}$ of size -1, uniformly on $(\mathcal{A} \times \mathcal{B}, \rho)$, where ρ is a metric on $\mathbb{R}^k \times \mathbb{R}^p$.

A.8: i. $A_0 \equiv \lim_{n\to\infty} E\left(n^{-1} \sum_{t=1}^n \frac{\partial}{\partial \alpha'} s_{1t}\left(X^t, \alpha_0\right)\right)$ is nonsingular and $B_0 \equiv \lim_{n\to\infty} \operatorname{var}\left(n^{-\frac{1}{2}} \sum_{t=1}^n s_{1t}\left(X^t, \alpha_0\right)\right)$ is positive definite.

ii. $H_0 \equiv \lim_{n\to\infty} E\left(n^{-1} \sum_{t=1}^n \frac{\partial}{\partial \beta'} s_{2t}\left(X^t, \alpha_0, \beta_0\right)\right)$ is nonsingular,

$$J_0 \equiv \lim_{n\to\infty} \operatorname{var}\left(n^{-\frac{1}{2}} \sum_{t=1}^{n} \left(s_{2t}\left(X^t, \alpha_0, \beta_0\right) - F_0 A_0^{-1} s_{1t}\left(X^t, \alpha_0\right)\right)\right)$$

is positive definite, and $F_0 \equiv \lim_{n\to\infty} E\left(n^{-1} \sum_{t=1}^n \frac{\partial}{\partial \alpha'} s_{2t}\left(X^t, \alpha_0, \beta_0\right)\right) < \infty$.

Supplementary Materials

The supplemental appendix contains all proofs.

References

Andrews D. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858. [691]

——— (2002), "Higher-Order Improvements of a Computationally Attractive k-step Bootstrap for Extremum Estimators," *Econometrica* 70, 119–162. [684,687]

Armstrong, T. B., Bertanha, M., and Hong, H. (2014), "A Fast Resample Method for Parametric and Semiparametric Models," *Journal of Econometrics*, 179, 128–133. [683,687,688,691,692]

Cattaneo, M. D., Jansson, M., and Ma, X. (2019), "Two-Step Estimation and Inference with Possibly Many Included Covariates," *Review of Economic Studies*, 86, 1095–1122. [683]

Cochrane, J. H. (2001), *Asset Pricing*, Princeton, NJ: Princeton University Press. [683]

Chen, X., and Liao, Z. (2015), "Sieve Semiparametric Two-step GMM with Weakly Dependent Data," *Journal of Econometrics*, 189, 163–186. [687]

Chen, X., Linton, O., and van Keilegom, I. (2003), "Estimation of Semiparametric Models When the Criterion Function is not Smooth," *Econometrica*, 71, 1591–1608. [683,687]

Cheng, G. (2015), "Moment Consistency of the Exchangeably Weighted Bootstrap for Semiparametric M estimation," *Scandinavian Journal of Statistics*, 42, 665–684. [684,689,690]

Davidson, R., and MacKinnon, J. (1999), "Bootstrap Testing in Nonlinear Models," *International Economic Review*, 40, 487–508. [684,687]

Engle, R. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business & Economic Statistics*, 20, 339–350. [683]

Fama, E.F. and MacBeth, J.D. (1973), "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636. [683]

Ghosh, M., Parr, W .C., Singh, K., and Babu, G. J. (1984), "A Note on Bootstrapping the Sample Median," *Annals of Statistics*, 12, 1130–1135. [688]

Gonçalves, S., and de Jong, R. (2003), "Consistency of the Stationary Bootstrap Under Weak Moment Conditions," *Economics Letters*, 81, 273–278. [687]

Gonçalves, S., and White, H. (2002), "The Bootstrap of the Mean for Dependent Heterogeneous Arrays," *Econometric Theory*, 18, 1367–1384. [686,687]

——— (2004), "Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models," *Journal of Econometrics*, 119, 199–219. [683,684,686,687,688,693]

——— (2005), "Bootstrap Standard Error Estimates for Linear Regressions," *Journal of the American Statistical Association*, 100, 970–979. [684,686,689]

Hahn, J., and Liao, Z. (2020), "Bootstrap Standard Error Estimates and Inference," *Econometrica*, 89, 1963–1977. [684,688]

Hong, H., and Scaillet, O. (2004), "A Fast Subsampling Method for Nonlinear Dynamic Models," *Journal of Econometrics*, 133, 557–578. [684]

Kato, K. (2011), "A Note on Moment Convergence of Bootstrap M Estimators," *Statistics & Decisions* 28, 51–61. [684,689,690]

Künsch, H. R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241. [683,686]

Joe, H (1995), "Asymptotic Efficiency of the Two-Stage Estimation Method for Copula-Based Models," *Journal of Multivariate Analysis*, 94, 401–419. [683,691]

La Vecchia, D., Moor, A., and Scaillet, O. (2020), "A Higher-Order Correct Fast Moving-Average Bootstrap for Dependent Data," University of Geneva, mimeo. [684]

Liu, R. Y., and Singh, K. (1992), "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of the Bootstrap*, eds. R. LePage, and L. Billiard, pp. 224–248, New York: Wiley. [683,686]

Nishiyama, Y. (2010), "Moment Convergence of *M* Estimators," *Statistica Neerlandica* 64, 505–507. [690]

Nelsen, R. B. (1999), *An Introduction to Copulas*, Lecture notes in statistics, New York: Springer. [690]

Newey, W. K., and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics* (Vol. IV), eds. R. Engle and D. McFadden, pp. 2111–2245, Amsterdam: North Holland. [683,685]

Oh, D. H., and Patton, A. J. (2013), "Simulated Method of Moments Estimation for Copula-based Multivariate Models," *Journal of the American Statistical Association*, 108, 689–700. [690]

Parente, P., and Smith, R. J. (2021), "Quasi-Maximum Likelihood and the Kernel Block Bootstrap for Nonlinear Dynamic Models," *Journal of Time Series Analysis*, 42, 377–405. [686]

Patton, A. J., (2006), "Estimation of Multivariate Models for Time Series of Possibly Different Lengths," *Journal of Applied Econometrics*, 21, 14–173. [683,691]

——— (2012), "Copula Methods for Forecasting Multivariate Time Series," in *Handbook of Economic Forecasting* (Vol. 2), eds. G. Elliott and A. Timmermann, pp. 899–960, Oxford: Elsevier. [683,691]

Paparoditis, E., and Politis, D. (2002), "The Tapered Block Bootstrap for General Statistics from Stationary Sequences," *Econometrics Journal*, 5, 131–148. [686]

Shao, J. (1992), "Bootstrap Variance Estimators with Truncation," *Statistics and Probability Letters*, 15, 95–101. [688]

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer-Verlag. [8] [690]

White, H. (1994), *Estimation, Inference and Specification Analysis*. Econometric Society Monographs 22, Cambridge: Cambridge University Press. [683,685]

Wooldridge, J. (1994), "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics* (vol. IV), eds. R. Engle and D. McFadden, pp. 2639–2738, Amsterdam: North Holland. [685]