Title

Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models

Authors
Goncalves, Silvia
White, Halbert

Publication Date
2000-12-01

2000-32

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

MAXIMUM LIKELIHOOD AND THE BOOTSTRAP FOR NONLINEAR DYNAMIC MODELS

BY

SÍLVIA GONÇALVES

AND

HALBERT WHITE

DISCUSSION PAPER 2000-32
DECEMBER 2000

Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models

Sílvia Gonçalves
Université de Montréal
C.P.6128, succ. Centre-Ville
Montréal, QC H3C 3J7
Canada
Tel: (514) 343 6556
Email: silvia.goncalves@umontreal.ca

Halbert White
University of California, San Diego
9500 Gilman Drive, La Jolla
California, 92093-0508
USA
Tel: (858) 534-3502
Email: hwhite@weber.ucsd.edu

Abstract

The bootstrap is an increasingly popular method for performing statistical inference. This paper provides the theoretical foundation for using the bootstrap as a valid tool of inference for quasi-maximum likelihood estimators (QMLE). We provide a unified framework for analyzing bootstrapped extremum estimators of nonlinear dynamic models for heterogeneous dependent stochastic processes. We apply our results to two block bootstrap methods, the moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992) and the stationary bootstrap of Politis and Romano (1994), and prove the first order asymptotic validity of the bootstrap approximation to the true distribution of QML estimators. Further, these block bootstrap methods are shown to provide heteroskedastic and autocorrelation consistent standard errors for the QMLE, thus extending the already large literature on robust inference and covariance matrix estimation. We also consider bootstrap testing. In particular, we prove the first order asymptotic validity of the bootstrap distribution of a suitable bootstrap analog of a Wald test statistic for testing hypotheses.

Keywords: block bootstrap, quasi maximum likelihood estimator, nonlinear dynamic model, near epoch dependence, Wald test.

1. Introduction

The bootstrap is a powerful and increasingly utilized method for obtaining confidence intervals and performing statistical inference. Nevertheless, despite its power and increasing popularity, results establishing the general applicability of the bootstrap to the quasi-maximum likelihood estimator (QMLE) or generalized method of moments (GMM) estimator commonly used in econometrics are currently available only under restrictive assumptions, such as stationarity. A main purpose of this paper is therefore to provide results that establish the first order asymptotic validity of the bootstrap for the data generating processes, models and estimators analyzed by Gallant and White (1988) and Pötscher and Prucha (1991): extremum estimators for nonlinear dynamic models of stochastic processes near epoch dependent (NED) on an underlying mixing process. We discuss primarily QML estimators both for concreteness and because there are fewer results in this area. We apply our results to two common block bootstrap procedures, the moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992) and the stationary bootstrap of Politis and Romano (1994a).

In pursuing this goal, we make a number of distinct but related contributions, both theoretical and practical: (1) we prove the consistency of the moving blocks and stationary bootstrap distributions of the bootstrapped QML estimators; (2) we provide new general purpose tools useful for analyzing the application of any bootstrap method to extremum estimators, such as the QMLE or GMM estimator, analogous to those of Gallant and White (1988); (3) we show that the moving blocks and stationary bootstrap provide new HAC covariance matrix estimators for the QMLE, and (4) we prove the first order asymptotic validity of the bootstrap distribution of a new bootstrap analog of a Wald test statistic for testing general nonlinear hypotheses on the pseudo-true parameters of a nonlinear dynamic model. Similar results hold for our new bootstrap Lagrange Multiplier statistic.

The plan of the paper is as follows. In section 2, we provide background and further motivation for the focus of this paper. Section 3 provides results applicable to extremum estimators generally and applies these to the QMLE under the nonparametric block bootstrap methods. Section 4 considers bootstrap testing. Section 5 concludes. Detailed assumptions and mathematical proofs of results in Section 3 appear in Appendix A; proofs for Section 4 are contained in Appendix B.

2. Background and Motivation

The bootstrap has gained increased popularity as an inference tool in econometrics since the seminal paper of Efron in 1979. Although the early work in the bootstrap literature applied only to statistics obtained from a sequence of independent identically distributed (i.i.d.) random variables (e.g., Bickel and Freedman (1981), Singh (1981)), the failure of Efron's bootstrap in non-i.i.d. settings was soon recognized (cf. Singh, 1981) and several bootstrap methods were suggested to overcome the problem. One approach was to reduce the problem to an approximate i.i.d. setting by bootstrapping the residuals of some parametric model. See for example Freedman (1981) and Liu (1988) for linear regression and Freedman (1984) and Efron and Tibshirani (1986) for ARMA models. More recently, several block bootstrap methods that do not require fitting a parametric model first were developed as a way to capture the dependence structure of the observed data in the "resampled" data. Examples are the moving blocks bootstrap (MBB) of Künsch (1989) and Liu and Singh (1992) and the stationary bootstrap (SB) of Politis and Romano (1994a). These methods amount to resampling blocks of observations instead of individual observations and were first proposed for stationary strong mixing observations.

The goal of this paper is to extend the estimation context in which bootstrap methods are validly available. In particular, Politis and Romano (1994b) demonstrate the validity of the stationary bootstrap for minimum distance estimation with stationary mixing processes. Fitzenberger (1997) establishes the validity of the moving blocks bootstrap for instrumental variables estimation of linear models of heterogeneous mixing processes. Although these results represent important advances, they do not apply generally to a considerable portion of the estimation procedures encountered in econometrics, which are applications of quasi-maximum likelihood methods or the generalized method of moments. Establishing the validity of the bootstrap for these estimation procedures is thus of great interest.

For these methods, existing results are available only for an array of special cases. For example, Hahn (1996) establishes the first order asymptotic validity of Efron's bootstrap for the GMM in the i.i.d. case. Asymptotic refinements for bootstrapped GMM estimators are studied by Hall and Horowitz (1996) in a stationary ergodic context. Recently, Andrews (1999) has extended Hall and Horowitz's results by establishing higher-order improvements of the bootstrap for nonlinear extremum estimators, thus including GMM and ML estimators. As in Hall and Horowitz (1996), Andrews (1999) assumes a stationary ergodic data generating process for which the true moment conditions for the extremum estimator are uncorrelated after finitely many lags.

In this paper, we provide conditions ensuring the first-order asymptotic validity of the moving blocks bootstrap and the stationary bootstrap when applied to extremum estimators (e.g., QML or GMM) of nonlinear dynamic models for processes NED on an underlying mixing process. To accomplish this goal we rely on the results of the bootstrap of the sample mean for NED processes established in Gonçalves and White (2000a).

We focus on the nonparametric block bootstrap (MBB and SB) to define the bootstrapped QML estimator, although we provide results that cover bootstrap methods generally. The nonparametric block bootstrap does not depend on a particular parametric model to generate the resamples and it is therefore robust to model misspecification. In the context of nonlinear dynamic models, it amounts to resampling observations on a vector valued array containing all the dependent and explanatory variables that enter the log-likelihood for a given observation, where lagged dependent variables may be included as explanatory variables. In order to capture the serial dependence in the data, blocks of observations on the vector valued data are resampled instead of individual observations.

By extending the applicability of bootstrap methods to a context comparable to the general setting of Gallant and White (1988) or Pötscher and Prucha (1991), we show that the bootstrap applies not only to a useful range of data generating processes, but also to a useful variety of modeling and estimation procedures relevant to economics and finance.

We do not attempt here to obtain asymptotic refinements. Our goal is rather to prove the consistency of the block bootstrap estimators of the QMLE sampling distribution for a broad class of models and data generating processes. In particular, we avoid the stationarity assumption and restrictive memory conditions used by Hall and Horowitz (1996) and by Andrews (1999) in studying higher order properties of the bootstrap. Among other things, we prove that the block bootstrap distribution of the QMLE converges weakly to the distribution of the QMLE in probability. As a consequence, bootstrap confidence intervals have asymptotic coverage probability equal to the nominal coverage probability. Further, the MBB and SB methods are shown to provide heteroskedastic and autocorrelation consistent standard errors for the QML estimator, thus extending the already large literature on robust inference and covariance matrix estimation.

An important application of the bootstrap is in hypothesis testing. The consistency of the bootstrap distribution of the QMLE is a first step to proving the validity of the block bootstrap methods for studentized statistics and hence for hypothesis testing. In particular, we propose and show the first order asymptotic validity of a new bootstrap Wald test for testing restrictions on the pseudo-true parameters of a nonlinear dynamic model. A variance estimator is typically required to studentize the statistic. For block bootstrap methods with dependent data, the choice of the variance estimators used to studentize the bootstrap and the original statistic is crucial, if second order improvements are to be expected (see Davison and Hall (1993), Götze and Künsch (1996) and Lahiri (1996, 1999b)). In particular, for smooth functions of means of stationary mixing data, to studentize the bootstrap statistic Götze and Künsch (1996) suggest a variance estimator that exploits the independence of the bootstrap blocks and that can be interpreted as the sample variance of the bootstrap block means. (Lahiri (1996) and Lahiri (1999b) also use this variance estimator in the context of the MBB for M-estimators and the SB for the smooth function model, respectively) To studentize the original statistic, Götze and Künsch (1996) use a kernel variance estimator with rectangular weights and warn that triangular weights will destroy second-order properties of the block bootstrap. Here, to studentize the bootstrap Wald test statistic we use the multivariate analog of the Götze and Künsch (1996) variance estimator, adapted to the

QMLE context. We prove that our new bootstrap Wald statistic is asymptotically distributed as a chi-squared distribution in probability, which implies that to first order we can use the block bootstrap to estimate the critical values of the Wald test statistic. Similar results hold for an analogous bootstrap Lagrange Multiplier statistic.

Bootstrapping extremum estimators for nonlinear models often requires solving a large number of nonlinear optimization problems, one for each resample, which may be computationally very demanding. Davidson and MacKinnon (1999) have recently proposed a class of k-step bootstrap estimators which only requires a small number of Newton or quasi-Newton steps starting from the estimate based on the original sample. The higher order properties of these attractive bootstrap estimators have been subsequently established by Andrews (1999) for the stationary ergodic context. Davidson and MacKinnon's (1999) one-step bootstrap estimator is easily obtained by a closed-form expression that only requires resampling the scores and the Hessian matrix of the model evaluated at the QMLE based on the original sample. A closely related computationally even simpler bootstrap estimator was suggested by Shao and Tu (1995, Section 8.1.2) which only requires resampling the scores at the QMLE for the original sample. We show that the one-step bootstrapped QML estimators of Davidson and MacKinnon (1999) and of Shao and Tu (1995) are first-order asymptotically equivalent to the standard (i.e. fully-optimized) block bootstrap QML estimator when the data is NED on a mixing process.

3. The Bootstrap for Extremum Estimators and Application to the QMLE

In this section we provide conditions ensuring the first order asymptotic validity of the bootstrap for quasi-maximum likelihood estimation (QMLE) methods under a setup identical to that used by Gallant and White (1988). The goal is to conduct inference on a certain parameter of interest θ_n^o from a given observed sample $X_{n1}, \ldots, X_{nn}$ assumed to be near epoch dependent on an underlying mixing process. We let X_{nt} denote a vector that contains both the explanatory variables and the dependent variables that enter the likelihood of observation t and we define $\{X_{nt}\}$ to be NED on a mixing process $\{V_t\}$ provided $E\left(X_{nt}^2\right) < \infty$ and $v_k \equiv \sup_{n,t} \left\| X_{nt} - E_{t-k}^{t+k}\left(X_{nt}\right)\right\|_2$

tends to zero at an appropriate rate. Here and in what follows, $\|X_{nt}\|_p \equiv (E|X_{nt}|^p)^{1/p}$ denotes the L_p norm and $E_{t-k}^{t+k}(\cdot) \equiv E\left(\cdot|\mathcal{F}_{t-k}^{t+k}\right)$, where $\mathcal{F}_{t-k}^{t+k} \equiv \sigma(V_{t-k}, \ldots, V_{t+k})$ is the σ-field generated by $V_{t-k}, \ldots, V_{t+k}$. In particular, if $v_k = O\left(k^{-a-\delta}\right)$ for some $\delta > 0$ we say $\{X_{nt}\}$ is NED of size $-a$. The sequence $\{V_t\}$ is assumed to be strong mixing; analogous results could be derived under the assumption of uniform mixing. We define the strong or α-mixing coefficients as usual, i.e. $\alpha_k \equiv \sup_m \sup_{\{A\in\mathcal{F}_{-\infty}^m, B\in\mathcal{F}_{m+k}^{\infty}\}} |P(A\cap B) - P(A)P(B)|$, and we require $\alpha_k \to 0$ as $k \to \infty$ at an appropriate rate.

The statistical properties of the QMLE $\hat{\theta}_n$ (consistency and asymptotic normality) are established in Gallant and White (1988) under the possibility of model misspecification. For this context, we establish the usefulness of two commonly used block bootstrap procedures, the moving blocks bootstrap and the stationary bootstrap, to approximate the sampling distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta_n^o\right)$. Given the original sample $X_{n1}, \ldots, X_{nn}$, a bootstrap version of $\hat{\theta}_n$, say $\hat{\theta}_n^*$, is obtained by considering the QMLE for the bootstrap resamples $X_{n1}^*, \ldots, X_{nn}^*$. We show that the bootstrap approximation to the sampling distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta_n^o\right)$ given by the distribution of $\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$, conditional on $X_{n1}, \ldots, X_{nn}$, is asymptotically normal with the correct asymptotic covariance matrix, in probability. We also give an explicit formula for the bootstrap covariance matrix of the scaled mean of the resampled scores evaluated at $\hat{\theta}_n$ which is heteroskedasticity and autocorrelation consistent (HAC), in that it converges in probability to the covariance matrix of the scaled mean of the scores evaluated at θ_n^o.

We follow Lahiri (1999a) in describing the block bootstrap methods. Let $\ell = \ell_n \in \mathbb{N}$ $(1 \leq \ell < n)$ denote the (expected) length of the blocks and let $B_{t,\ell} = \{X_{nt}, X_{n,t+1}, \ldots, X_{n,t+\ell-1}\}$ be the block of ℓ consecutive observations starting at X_{nt}; $\ell = 1$ corresponds to the standard bootstrap. Assume for simplicity that $n = k\ell$. The MBB resamples $k = n/\ell$ blocks randomly with replacement from the set of $n - \ell + 1$ overlapping blocks $\{B_{1,\ell}, \ldots, B_{n-\ell+1,\ell}\}$. Thus, if we let $I_{n1}, \ldots, I_{nk}$ be i.i.d. random variables uniformly distributed on $\{0, \ldots, n-\ell\}$, the MBB pseudo-time series $\left\{X_{nt}^{*(1)}, t = 1, \ldots, n\right\}$ is the result of arranging the elements of the k resampled blocks $B_{I_{n1}}, \ldots, B_{I_{nk}}$ in a sequence: $X_{n1}^{*(1)} = X_{n,I_{n1}+1}, \ldots, X_{n\ell}^{*(1)} = X_{n,I_{n1}+\ell}$,

$X^{*(1)}_{n,\ell+1} = X_{n,I_{n2}+1}, \ldots, X^{*(1)}_{n,kl} = X_{n,I_{nk}+\ell}$. Here and throughout, we use the superscript (1) in $X^{*(1)}_{n,t}$ to denote the bootstrap samples obtained by the MBB. Similarly, we will use the superscript (2) to denote bootstrap samples obtained by the SB resampling scheme.

Unlike the MBB, the stationary bootstrap resamples blocks of random size. Let $p = \ell^{-1}$ be a given number in $(0,1)$; $p = 1$ corresponds to the standard bootstrap. Let $L_{n1}, L_{n2}, \ldots$ be conditionally i.i.d. random variables having the geometric distribution with parameter p so that the probability of the event $\{L_{n1} = k\}$ is $(1-p)^{k-1}p$ for $k = 1, 2, \ldots$. Independent of $\{X_{nt}\}$ and of $\{L_{ni}\}$, let $I_{n1}, I_{n2}, \ldots$ be i.i.d. random variables having the uniform distribution on $\{1, \ldots, n\}$. The SB pseudo-time series $\left\{X^{*(2)}_{nt}\right\}$ can be obtained by joining the resampled blocks $B_{I_{n1},L_{n1}}, B_{I_{n2},L_{n2}}, \ldots, B_{I_{nK},L_{nK}}$, where $K = \inf\{k \geq 1 : L_{n1} + \ldots + L_{nk} \geq n\}$. Thus, the stationary bootstrap amounts to resampling blocks of observations of random length, where each block size has a geometric distribution with parameter p and expected length equal to $\frac{1}{p} = \ell$. We shall require $\ell = \ell_n$ to tend to infinity at an appropriate rate, which is equivalent to letting $p = p_n$ tend to zero. Hence, on average the lengths of the SB blocks tend to infinity with n as also happens with the (fixed) MBB blocks lengths.

In contrast to the MBB resampling method, the stationary bootstrap resample is (conditionally) a strictly stationary process (Politis and Romano, 1994a), hence the name stationary bootstrap.

A convenient way to formalize any bootstrap procedure is as follows. Given the underlying probability space $(\Omega, \mathcal{F}, P)$, we observe a sample $X_{n1}(\omega), \ldots, X_{nn}(\omega)$ of size n from a given realization $\omega \in \Omega$, where $X_{nt}(\omega)$ assumes values in $\mathbb{R}^l$, $l \in \mathbb{N}$. We let $(\Lambda, \mathcal{G})$ be a measurable space and for each $\omega \in \Omega$ and $n \in \mathbb{N}$, we let $P^*_{n,\omega}$ denote the probability measure induced by the bootstrap on $(\Lambda, \mathcal{G})$. The bootstrap is a method to generate a new time series on $\left(\Lambda, \mathcal{G}, P^*_{n,\omega}\right)$ conditional on the data $X_{n1}(\omega), \ldots, X_{nn}(\omega)$ from which we can obtain as many (re-) samples as we want.

For each $n \in \mathbb{N}$, let $\tau_{nt} : \Lambda \to \{1, 2, \ldots, n\}$ denote a random index generated by the resampling scheme for each $t = 1, 2, \ldots$. For example, for the standard bootstrap, $\{\tau_{nt}\}$ is a sequence of i.i.d. random variables uniformly distributed on $\{1, 2, \ldots, n\}$, independent of $\{X_{nt}\}$. For the MBB described above, $\{\tau_{nt}\} \equiv \left\{\tau_{nt}^{(1)}\right\} = \{I_{n1}+1, \ldots, I_{n1}+\ell, I_{n2}+1, \ldots\}$, where $\{I_{ni}\}$ are conditionally i.i.d. uniform on $\{0, \ldots, n-\ell\}$ and ℓ is the fixed block length. For the SB, $\{\tau_{nt}\} \equiv \left\{\tau_{nt}^{(2)}\right\} = \{I_{n1}, I_{n1}+1, \ldots, I_{n1}+L_{n1}-1, I_{n2}, I_{n2}+1, \ldots\}$, where $\{I_{ni}\}$ are conditionally i.i.d. uniform on $\{1, 2, \ldots, n\}$ and $\{L_{ni}\}$ are i.i.d. random variables having the geometric distribution with parameter p, independent of $\{X_{nt}\}$ and $\{I_{ni}\}$.

We define the stochastic process induced by the bootstrap as the mapping $X_n^*(\cdot, \omega) \equiv \{X_{nt}^*(\cdot, \omega), t = 1, 2, \ldots\} : \Lambda \to \mathbb{R}^{l\infty}$, $\omega \in \Omega$, $n = 1, 2, \ldots$ and $l \in \mathbb{N}$, such that for $(\lambda, \omega) \in \Lambda \times \Omega$,

$$X_{nt}^*(\lambda, \omega) = X_{n, \tau_{nt}(\lambda)}(\omega), \qquad t = 1, 2, \ldots .$$

As Lemma A.1 in Appendix A establishes, $X_{nt}^*(\cdot, \omega)$ is $\mathcal{G}$-measurable and thus it indeed defines a stochastic process on $\left(\Lambda, \mathcal{G}, P_{n,\omega}^*\right)$. In this context, a resample obtained by the bootstrap is a sample of size m (say) corresponding to a given realization $X_n^*(\lambda, \omega)$ of the stochastic process $X_n^*(\cdot, \omega)$, conditional on the original sample $X_{n1}(\omega), \ldots, X_{nn}(\omega)$. We will let $m = n$ throughout.

We first establish the ability of the block bootstrap methods to provide HAC covariance estimators for the QMLE.

The QML estimator $\hat{\theta}_n$ solves the problem

$$\max_{\Theta} L_n(X_n^n, \theta) \equiv n^{-1} \sum_{t=1}^{n} \log f_{nt}\left(X_n^t, \theta\right), \quad n = 1, 2, \ldots,$$

where $X_n^t \equiv (X_{n1}', \ldots, X_{nt}')'$, $t = 1, 2, \ldots, n$, and θ is an element of Θ, a compact subset of $\mathbb{R}^p$, $p \in \mathbb{N}$. Thus, X_n^t is a vector valued array that contains all the relevant explanatory and dependent variables that enter f_{nt}, the "quasi-likelihood function" for observation t. The function L_n is the "quasi-log-likelihood function".

The asymptotic properties (consistency, asymptotic normality) of $\hat{\theta}_n$ have been established by Gallant and White (1988) under certain regularity assumptions, which we collect in Appendix A for convenience. Gallant and White's (1988) data generating assumption assumes a singly indexed stochastic process $\{X_t\}$. Nevertheless, as they remark in page 9 of their book, their results can easily be extended to doubly indexed arrays by relying on weak laws of large numbers for double arrays $\{X_{nt}\}$ (such as Andrews' (1988) weak law of large numbers for uniformly integrable L_1-mixingales) under even weaker conditions than those used in their work. Our assumptions A.1-A.10 are the doubly indexed versions of Gallant and White's (1988) regularity conditions. Thus, under assumptions A.1-A.10, Theorem 5.7 of Gallant and White (1988) in particular shows that[1] $B_n^{o-1/2} A_n^o \sqrt{n}\left(\hat{\theta}_n - \theta_n^o\right) \Rightarrow N\left(0, I_p\right),$ where $\Rightarrow$ denotes convergence in distribution. The asymptotic covariance matrix of $\hat{\theta}_n$ is thus $C_n^o \equiv A_n^{o-1} B_n^o A_n^{o-1}$, where $A_n^o \equiv E\left[\nabla^2 L_n\left(X_n^n, \theta_n^o\right)\right]$ and $B_n^o \equiv \text{var}\left(n^{-1/2} \sum_{t=1}^n \nabla \log f_{nt}\left(X_n^t, \theta_n^o\right)\right)$. Because A_n^o and B_n^o are generally unknown, consistent estimators are needed in order for the normal approximation to be useful in constructing asymptotic confidence intervals or testing hypotheses. A consistent estimator of A_n^o is $\hat{A}_n = \nabla^2 L_n\left(X_n^n, \hat{\theta}_n\right)$, as proven in Theorem 6.1 in Gallant and White (1988) under the same assumptions that deliver the asymptotic normality result. The consistent estimation of B_n^o is a more challenging task. Gallant and White (1988) establish the consistency of a particular estimator of B_n^o when the data $\{X_{nt}\}$ is NED on a mixing process and misspecification is allowed. Alternative consistent estimators of B_n^o are given by the block bootstrap variance estimators, as we now prove.

We first consider an infeasible estimator based on θ_n^o. We then build on this estimator to obtain a feasible estimator. Accordingly, let $s_{nt}^o \equiv \nabla \log f_{nt}\left(X_n^t, \theta_n^o\right)$ so that $B_n^o \equiv \text{var}\left(n^{-1/2} \sum_{t=1}^n s_{nt}^o\right)$. For $j = 1, 2$, let $B_{n,j}^o \equiv \text{var}^*\left(n^{-1/2} \sum_{t=1}^n s_{nt}^{o*(j)}\right)$ denote the block bootstrap covariance matrix estimators based on the resampled scores $\left\{s_{nt}^{o*(j)} \equiv \nabla \log f_{n,\tau_{nt}^{(j)}}\left(X_n^{\tau_{nt}^{(j)}}, \theta_n^o\right)\right\}$ evaluated at θ_n^o, where $\left\{\tau_{nt}^{(1)}\right\}$ and $\left\{\tau_{nt}^{(2)}\right\}$ denote the random indexes chosen according to the MBB and the SB, respectively. Here and throughout we let E^* (var*) denote expectation (variance) with respect to the bootstrap probability measure, conditional on $X_{n1}, \ldots, X_{nn}$.

[1] Note that θ_n^o here plays the same role as θ_n^* in Gallant and White (1988). The change in notation is due to the fact that we reserve the superscript star for the bootstrap, as is usual in the bootstrap literature.

The following formula is available to compute $B^o_{n,1}$ (Künsch, 1989, Theorems 3.1 and 3.4):

$$B^o_{n,1} = \sum_{t=1}^{n} \alpha_n(t) \left(s^o_{nt} - \overline{s}^o_{\alpha,n}\right) \left(s^o_{nt} - \overline{s}^o_{\alpha,n}\right)' \tag{3.1}$$

$$+ \sum_{\tau=1}^{\ell-1} \left(1 - \frac{\tau}{\ell}\right) \sum_{t=1}^{n-\tau} \beta_n(t,\tau) \left[\left(s^o_{nt} - \overline{s}^o_{\alpha,n}\right)\left(s^o_{n,t+\tau} - \overline{s}^o_{\alpha,n}\right)' + \left(s^o_{n,t+\tau} - \overline{s}^o_{\alpha,n}\right)\left(s^o_{nt} - \overline{s}^o_{\alpha,n}\right)'\right],$$

where $\overline{s}^o_{\alpha,n} \equiv \sum_{t=1}^{n} \alpha_n(t) s^o_{nt}$. The weights $\alpha_n(t)$ and $\beta_n(t,\tau)$ are given as follows (cf. Künsch, 1989, expressions (3.2) and (3.7)):

$$\alpha_n(t) = \begin{cases} \frac{t}{\ell(n-\ell+1)} & \text{if } t \in \{1, \ldots, \ell-1\} \\ \frac{1}{n-\ell+1} & \text{if } t \in \{\ell, \ldots, n-\ell+1\} \\ \frac{n-t+1}{\ell(n-\ell+1)} & \text{if } t \in \{n-\ell+2, \ldots, n\}, \end{cases} \tag{3.2}$$

$$\beta_n(t,\tau) = \begin{cases} \frac{t}{(\ell-|\tau|)(n-\ell+1)} & \text{if } t \in \{1, \ldots, \ell-|\tau|-1\} \\ \frac{1}{n-\ell+1} & \text{if } t \in \{\ell-|\tau|, \ldots, n-\ell+1\} \\ \frac{n-t-|\tau|+1}{(\ell-|\tau|)(n-\ell+1)} & \text{if } t \in \{n-\ell+2, \ldots, n-|\tau|\}, \end{cases} \tag{3.3}$$

where $\sum_{t=1}^{n} \alpha_n(t) = 1$ and $\sum_{t=1}^{n-|\tau|} \beta_n(t,\tau) = 1$.

Similarly, we can compute $B^o_{n,2}$ without resampling (Politis and Romano, 1994a, Lemma 1):

$$B^o_{n,2} = n^{-1} \sum_{t=1}^{n} \left(s^o_{nt} - \overline{s}^o_n\right)\left(s^o_{nt} - \overline{s}^o_n\right)' \tag{3.4}$$

$$+ \sum_{\tau=1}^{n-1} b_n(\tau)\, n^{-1} \sum_{t=1}^{n-\tau} \left[\left(s^o_{nt} - \overline{s}^o_n\right)\left(s^o_{n,t+\tau} - \overline{s}^o_n\right)' + \left(s^o_{n,t+\tau} - \overline{s}^o_n\right)\left(s^o_{nt} - \overline{s}^o_n\right)'\right],$$

where $\overline{s}^o_n \equiv n^{-1} \sum_{t=1}^{n} s^o_{nt}$, and

$$b_n(\tau) = \left(1 - \frac{\tau}{n}\right)(1-p)^{\tau} + \frac{\tau}{n}(1-p)^{n-\tau}, \tag{3.5}$$

with $p = p_n \equiv \ell_n^{-1}$.

Gonçalves and White (2000a) give sufficient conditions for the consistency of the MBB and SB covariance matrix estimators of the sample mean of a vector NED array on a mixing process. In particular, an application of their Theorem 3.1 shows that under the NED assumption, $B^o_{n,j}$ is consistent for $B^o_n + U^o_{n,j}$, where $U^o_{n,j} = \text{var}^*\left(n^{-1/2} \sum_{t=1}^{n} [E(s^o_{nt})]^{*(j)}\right)$ and $\left\{[E(s^o_{nt})]^{*(j)}\right\}$ is a resample of $\{E(s^o_{nt})\}$ obtained by the MBB if $j = 1$ and by the SB if $j = 2$. The bias terms $U^o_{n,j}$ can be written explicitly as a function of $\{E(s^o_{nt})\}$ (see Lemma 2.1 of Gonçalves and White, 2000a).

To obtain a feasible estimator, we consider instead the block bootstrap covariance matrix estimators $\hat{B}_{n,j} \equiv \mathrm{var}^*\left(n^{-1/2}\sum_{t=1}^{n}\hat{s}_{nt}^{*(j)}\right)$ that are based on the resampled estimated scores $\left\{\hat{s}_{nt}^{*(j)} \equiv \nabla\log f_{n,\tau_{nt}^{(j)}}\left(X_n^{\tau_{nt}^{(j)}},\hat{\theta}_n\right)\right\}$ obtained from $\left\{\hat{s}_{nt} = \nabla\log f_{nt}\left(X_n^t,\hat{\theta}_n\right)\right\}$ by the MBB and by the SB resampling schemes. Again, no resampling is necessary to compute $\hat{B}_{n,j}$:

$$\hat{B}_{n,1} = \sum_{t=1}^{n}\alpha_n(t)\left(\hat{s}_{nt}-\overline{\hat{s}}_{\alpha,n}\right)\left(\hat{s}_{nt}-\overline{\hat{s}}_{\alpha,n}\right)' \tag{3.6}$$
$$+\sum_{\tau=1}^{\ell-1}\left(1-\frac{\tau}{\ell}\right)\sum_{t=1}^{n-\tau}\beta_n(t,\tau)\left[\left(\hat{s}_{nt}-\overline{\hat{s}}_{\alpha,n}\right)\left(\hat{s}_{n,t+\tau}-\overline{\hat{s}}_{\alpha,n}\right)'+\left(\hat{s}_{n,t+\tau}-\overline{\hat{s}}_{\alpha,n}\right)\left(\hat{s}_{nt}-\overline{\hat{s}}_{\alpha,n}\right)'\right],$$

where $\alpha_n(t)$ and $\beta_n(t,\tau)$ are defined in (3.2) and (3.3), $\overline{\hat{s}}_{\alpha,n} \equiv n^{-1}\sum_{t=1}^{n}\alpha_n(t)\,\hat{s}_{nt}$, and

$$\hat{B}_{n,2} = n^{-1}\sum_{t=1}^{n}\hat{s}_{nt}\hat{s}_{nt}' + \sum_{\tau=1}^{n-1}b_n(\tau)\,n^{-1}\sum_{t=1}^{n-\tau}\left[\hat{s}_{nt}\hat{s}_{nt+\tau}'+\hat{s}_{nt+\tau}\hat{s}_{nt}'\right], \tag{3.7}$$

with $b_n(\tau)$ as defined in (3.5).

Careful argument establishes that the difference between $\hat{B}_{n,j}$ and $B_{n,j}^o$ converges to zero in probability, which implies the consistency of $\hat{B}_{n,j}$ for $B_n^o+U_{n,j}^o$. Moreover, $\hat{B}_{n,j}$ is positive semidefinite by construction. To obtain this result, we strengthen assumptions A.1-A.10 in Appendix A in the following way:

Assumption 3.1

3.1.a) $\left\{s_{nt}\left(X_n^t,\theta\right) \equiv \nabla\log f_{nt}\left(X_n^t,\theta\right)\right\}$ is $3r$-dominated on Θ uniformly in $n, t = 1, 2, \ldots, r > 2$.

3.1.b) The elements of $\left\{s_{nt}\left(X_n^t,\theta\right) \equiv \nabla\log f_{nt}\left(X_n^t,\theta\right)\right\}$ are NED on $\{V_t\}$ of size $-\frac{2(r-1)}{r-2}$ uniformly on (Θ,ρ).

Theorem 3.1. *Given Assumptions A.1-A.10 as strengthened by Assumption 3.1, if* $\ell_n \to \infty$ *and* $\ell_n = o\left(n^{1/2}\right)$, *then for* $j = 1, 2$,

$$\hat{B}_{n,j} - \left(B_n^o + U_{n,j}^o\right) \xrightarrow{P} 0,$$

where $U_{n,j}^o = \mathrm{var}^*\left(n^{-1/2}\sum_{t=1}^{n}\left[E\left(s_{nt}^o\right)\right]^{*(j)}\right)$.

Theorem 3.1 gives sufficient conditions for the consistency of the block bootstrap variance estimators $\hat{B}_{n,j}$ under the possibility of misspecification. By allowing for general heterogeneous dependent data generating processes and the possibility of misspecified models, it adds to the already extensive literature on heteroskedastic and autocorrelation consistent (HAC) covariance matrix estimation (see e.g. Eicker (1967), White (1980), Hansen (1992a), White (1984), White and Domowitz (1984), Newey and West (1987), Gallant and White (1988), Andrews (1991), Andrews and Monahan (1992), Den Haan and Levin (1997), and de Jong and Davidson (2000)).

Assumption 3.1 ensures that the elements of the double array $\{s_{nt}^o\}$ satisfy Assumption 2.1 of Gonçalves and White (2000a). Except for Assumption 3.1.a), we use the same assumptions as Gallant and White's consistency result (Theorem 6.8(b), 1988). However, their result only requires $\left\{s_{nt}\left(X_n^t, \theta\right)\right\}$ to be $2r$-dominated on Θ uniformly in $n, t = 1, 2, \ldots, r > 2,$ rather than $3r$-dominated as we assume here.

In the presence of heterogeneous observations and arbitrary misspecification, the stationary bootstrap is not consistent for B_n^o, but instead for $B_n^o + U_{n,j}^o$. A sufficient condition for the bias term $\left\{U_{n,j}^o\right\}$ to vanish is that $E\left(s_{nt}^o\right)$ be zero for $t = 1, 2, \ldots, n,$ $n = 1, 2, \ldots$. Most of the literature on robust covariance matrix estimation has adopted this assumption. One exception is Gallant and White (1988). As they point out (Gallant and White, 1988, p.102), this condition is true if for example the model is correctly specified or if $\{X_{nt}\}$ is stationary and the model embodies no regime changes (i.e. $f_{nt}\left(\cdot, \theta\right) = f\left(\cdot, \theta\right)$ for all n, t).

We next investigate the asymptotic properties (consistency and asymptotic normality) of the bootstrapped QMLE for generic bootstrap procedures.

We use the following notation (see Hahn, 1996, p. 190, for similar notation). For any bootstrap statistic $T_n^*\left(\cdot, \omega\right)$ we write $T_n^*\left(\cdot, \omega\right) \to 0 \; prob - P_{n,\omega}^*,\; a.s. - P$ if $T_n^*\left(\cdot, \omega\right)$ converges to zero in probability$-P_{n,\omega}^*$ for almost all ω, i.e. if for any $\varepsilon > 0$ there exists $F \in \mathcal{F}$ with $P\left(F\right) = 1$ such that for all ω in F, $\lim_{n\to\infty} P_{n,\omega}^*\left[\lambda : \left|T_n^*\left(\lambda, \omega\right)\right| > \varepsilon\right] = 0$. We write $T_n^*\left(\cdot, \omega\right) \to 0 \; prob - P_{n,\omega}^*$, $prob{-}P$ if for any $\varepsilon > 0$ and for any $\delta > 0$, $\lim_{n\to\infty} P\left[\omega : P_{n,\omega}^*\left[\lambda : \left|T_n^*\left(\lambda, \omega\right)\right| > \varepsilon\right] > \delta\right] = 0$. Using a subsequence argument (e.g. Billingsley, 1995, Theorem 20.5), $T_n^*\left(\cdot, \omega\right) \to 0 \; prob{-}P_{n,\omega}^*, prob{-}P$

is equivalent to having that for any subsequence $\{n'\}$ there exists a further subsequence $\{n''\}$ such that $T^*_{n''}(\cdot,\omega)$ $prob - P^*_{n'',\omega}$, $a.s. - P$. We write[2] $T^*_n(\cdot,\omega) \Rightarrow^{d_{P^*_{n,\omega}}} N(0,1)$ $prob - P$ when for every subsequence there exists a further subsequence for which weak convergence takes place almost surely. This subsequence argument will often be used in our proofs for the bootstrap. For example, to prove that $T^*_n(\cdot,\omega) \Rightarrow^{d_{P^*_{n,\omega}}} N(0,1)$ $prob - P$, we consider an arbitrary subsequence indexed by $\{n'\}$ and prove that there exists a further subsequence $\{n''\}$ for which $T^*_{n''}(\cdot,\omega) \Rightarrow^{d_{P^*_{n'',\omega}}} N(0,1)$ for fixed ω in a set with probability one. Because for such ω we have that $T^*_{n''}(\cdot,\omega)$ is a random variable in the usual sense, this can be accomplished by applying a standard central limit theorem to $T^*_{n''}(\cdot,\omega)$ under the bootstrap probability measure $P^*_{n'',\omega}$.

For a general bootstrap method, we define the bootstrapped QMLE, $\hat{\theta}^*_n(\cdot,\omega)$, $\omega \in \Omega$, as the solution to

$$\max_{\Theta} L^*_n(\cdot,\omega,\theta) \equiv n^{-1}\sum_{t=1}^{n} \log f^*_{nt}(\cdot,\omega,\theta), \quad n = 1,2,\ldots,$$

where, for each $n = 1,2,\ldots$ and for each $\theta \in \Theta$, $\{f^*_{nt}(\cdot,\omega,\theta), t = 1,\ldots,n\}$ is a bootstrap resample of size n from $\left\{f_{nt}\left(X^t_n(\omega),\theta\right), t = 1,\ldots,n\right\}$, i.e. $f^*_{nt}(\cdot,\omega,\theta) = f_{n,\tau_{nt}(\cdot)}\left(X^{\tau_{nt}(\cdot)}_n(\omega),\theta\right)$, and $\tau_{nt} : \Lambda \to \{1,2,\ldots,n\}$ is a random index chosen according to a given bootstrap resampling scheme, $t = 1,2,\ldots,n$, e.g. the MBB or the SB. As above, we let $\left\{\tau^{(j)}_{nt}\right\}$ denote the random index generated by the MBB if $j = 1$ and by the SB if $j = 2$. Consequently, $\hat{\theta}^{*(j)}_n(\cdot,\omega)$ will denote the MBB and the SB QMLE's, for $j = 1$ and $j = 2$, respectively.

Given assumptions A.1 and A.2, the existence of $\hat{\theta}^*_n(\cdot,\omega)$ as a measurable-$\mathcal{G}$ function for each n and almost all ω follows by Lemma 2 of Jennrich (1969). Because for almost all ω $\hat{\theta}^*_n(\cdot,\omega)$ is a random variable in $\left(\Lambda,\mathcal{G},P^*_{n,\omega}\right)$, we can study its stochastic properties.

The next result helps in establishing the consistency of a general bootstrapped QMLE. The same heuristics that deliver the consistency of $\hat{\theta}_n$ for θ^o_n apply in the bootstrap context. Specifically, if for all ω in a set with probability approaching one, $L^*_n(\cdot,\omega,\theta)$ tends (almost surely or in probability) to $L_n(\omega,\theta)$, then we should expect $\hat{\theta}^*_n(\cdot,\omega)$, which maximizes $L^*_n(\cdot,\omega)$, to tend

[2] We follow Giné and Zinn (1989, p. 688) to define weak convergence in probability; in particular, we use the distance induced by the sup norm of distribution functions to metrize weak convergence.

to $\hat{\theta}_n(\omega)$, which maximizes $L_n(\omega, \theta)$. As the next lemma shows, the consistency of $\hat{\theta}_n$ for θ_n^o, where θ_n^o is assumed to be identifiably unique[3], ensures that $\hat{\theta}_n(\omega)$ is identifiably unique for all n sufficiently large and ω in a set with probability one or approaching one.

Lemma 3.1 (Identifiable uniqueness of $\hat{\theta}_n$ for all n sufficiently large). *Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and let Θ be a compact subset of $\mathbb{R}^p$, $p \in \mathbb{N}$. Let $\left\{Q_n : \Omega \times \Theta \to \overline{\mathbb{R}}\right\}$ be a sequence of random functions continuous on Θ a.s. $-$ P, and let $\hat{\theta}_n = \arg\max_{\Theta} Q_n(\cdot, \theta)$ a.s. $-$ P. If $Q_n(\cdot, \theta) - \overline{Q}_n(\theta) \to 0$ a.s. $-$ P uniformly on Θ and if $\left\{\overline{Q}_n : \Theta \to \overline{\mathbb{R}}\right\}$ has identifiably unique maximizers $\{\theta_n^o\}$ on Θ, then the sequence $\left\{\hat{\theta}_n\right\}$ is identifiably unique on Θ with respect to $\{Q_n\}$ a.s. $-$ P, i.e. there exists $F \in \mathcal{F}$, $P(F) = 1$, such that given any $\varepsilon > 0$ and some $\delta(\varepsilon) > 0$, for each $\omega \in F$, there exists $N(\omega, \varepsilon) < \infty$ such that*

$$\sup_{n \geq N(\omega,\varepsilon)} \left[\max_{\eta^c\left(\hat{\theta}_n, \varepsilon\right)} Q_n(\omega, \theta) - Q_n\left(\omega, \hat{\theta}_n\right) \right] \leq -\delta(\varepsilon) < 0,$$

where $\eta_n^c\left(\hat{\theta}_n, \varepsilon\right)$ is the compact complement of $\eta\left(\hat{\theta}_n, \varepsilon\right) \equiv \left\{\theta \in \Theta : \left|\theta - \hat{\theta}_n\right| < \varepsilon\right\}$ for $n = 1, 2, \ldots$.

If instead $Q_n(\cdot, \theta) - \overline{Q}_n(\theta) \to 0$ prob $-$ P uniformly on Θ then for any subsequence of $\left\{\hat{\theta}_n\right\}$, say $\left\{\hat{\theta}_{n'}\right\}$, there exists a further subsequence $\left\{\hat{\theta}_{n''}\right\}$ such that $\left\{\hat{\theta}_{n''}\right\}$ is identifiably unique with respect to $\{Q_{n''}\}$ a.s. $-$ P.

By applying Lemma 3.1, we can obtain the following fundamental consistency result in the bootstrap context, which can be used to prove the consistency of bootstrapped estimators for the case of the QMLE as well as for other extremum estimators such as GMM.

Lemma 3.2. *Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and let Θ be a compact subset of $\mathbb{R}^p$, $p \in \mathbb{N}$. Let $\left\{Q_n : \Omega \times \Theta \to \overline{\mathbb{R}}\right\}$ be a sequence of random functions such that*

a1) *$Q_n(\cdot, \theta) : \Omega \to \overline{\mathbb{R}}$ is measurable-$\mathcal{F}$ for each $\theta \in \Theta$, $n = 1, 2, \ldots$;*

a2) *$Q_n(\omega, \cdot) : \Theta \to \overline{\mathbb{R}}$ is continuous on Θ a.s. $-$ P (i.e. for almost all ω), $n = 1, 2, \ldots$.*

[3] The definition of identifiable uniqueness was introduced by Domowitz and White (1982). For convenience, we restate it in the Appendix as Assumption A.3.

Let $\hat{\theta}_n = \arg\max_\Theta Q_n(\cdot, \theta)$ a.s. $- P$ be measurable and assume there exists $\{\overline{Q}_n : \Theta \to \overline{\mathbb{R}}\}$ with identifiably unique maximizers $\{\theta_n^o\}$ such that

a3) *$Q_n(\cdot, \theta) - \overline{Q}_n(\theta) \to 0$ as $n \to \infty$ prob $- P$ uniformly on Θ. Then,*

$$\text{(A)} \qquad \hat{\theta}_n - \theta_n^o \to 0 \ prob - P.$$

Let $(\Lambda, \mathcal{G})$ be a measurable space, and for each $\omega \in \Omega$ and $n \in \mathbb{N}$, let $\left(\Lambda, \mathcal{G}, P_{n,\omega}^*\right)$ be a complete probability space. Let $\left\{Q_n^* : \Lambda \times \Omega \times \Theta \to \overline{\mathbb{R}}\right\}$ be a sequence of functions such that

b1) *$Q_n^*(\cdot, \omega, \theta) : \Lambda \to \overline{\mathbb{R}}$ is measurable-$\mathcal{G}$ for each (ω, θ) in $\Omega \times \Theta$, $n = 1, 2, \ldots$;*

b2) *$Q_n^*(\lambda, \omega, \cdot) : \Theta \to \overline{\mathbb{R}}$ is continuous on Θ a.s.$-P$ (i.e. for all λ and almost all ω), $n = 1, 2, \ldots$.*

Let $\left\{\hat{\theta}_n^* : \Lambda \times \Omega \to \Theta,\ n = 1, 2, \ldots\right\}$ be a sequence of random functions such that for each $\omega \in \Omega$, $\hat{\theta}_n^*(\cdot, \omega) : \Lambda \to \Theta$ is measurable-$\mathcal{G}$ and $\hat{\theta}_n^*(\cdot, \omega) = \arg\max_\Theta Q_n^*(\cdot, \omega, \theta)$ a.s. $- P$. Assume

b3) *$Q_n^*(\cdot, \omega, \theta) - Q_n(\omega, \theta) \to 0$ as $n \to \infty$ prob $- P_{n,\omega}^*$, prob $- P$ uniformly on Θ. Then,*

$$\text{(B)} \qquad \hat{\theta}_n^*(\cdot, \omega) - \hat{\theta}_n(\omega) \to 0,\ prob - P_{n,\omega}^*,\ prob - P.$$

Theorem 3.2. *Let Assumptions A.1, A.2, A.3, A.5(i), A.6(i), A.7 and A.8(i) hold. Then, $\hat{\theta}_n - \theta_n^o \to 0$ prob $- P$. If we assume further that $\ell_n \to \infty$ and $\ell_n = o(n)$ then for $j = 1, 2$, $\hat{\theta}_n^{*(j)}(\cdot, \omega) - \hat{\theta}_n(\omega) \to 0$ prob $- P_{n,\omega}^*$, prob $- P$.*

The same set of assumptions that delivers the weak consistency of the QMLE $\hat{\theta}_n$ for θ_n^o also delivers the weak consistency of the block bootstrapped QMLE's $\hat{\theta}_n^{*(j)}(\cdot, \omega)$ for θ_n^o, for all ω in a set with probability approaching one, provided $\ell_n = o(n)$ and $\ell_n \to \infty$. (for the SB, this is equivalent to requiring that $np_n \to \infty$ and $p_n \to 0$). In particular, we don't need to impose the stronger condition $\ell_n = o\left(n^{1/2}\right)$ used in Theorem 3.1 to obtain the present result. To prove Theorem 3.2 we apply Lemma 3.2 with

$$\begin{aligned} Q_n(\cdot, \theta) &= L_n(X_n^n(\cdot), \theta) \equiv n^{-1} \sum_{t=1}^n \log f_{nt}\left(X_n^t(\cdot), \theta\right), \\ \bar{Q}_n(\theta) &= E\left[L_n(X_n^n, \theta)\right], \end{aligned}$$

and, for $j = 1, 2$,

$$Q_n^{*(j)}(\cdot, \omega, \theta) = L_n^{*(j)}(\cdot, \omega, \theta) \equiv n^{-1} \sum_{t=1}^{n} \log f_{n, \tau_{nt}^{(j)}(\cdot)} \left(X_n^{\tau_{nt}^{(j)}(\cdot)}(\omega), \theta \right).$$

To study higher-order properties of the bootstrap, Andrews (1999) suggests a recentering of the criterion function to define bootstrapped extremum estimators. In particular, his criterion function is

$$L_n^{*(j)}(\cdot, \omega, \theta) - n^{-1} \sum_{t=1}^{n} E^* \left(\hat{s}_{nt}^{*(j)} \right)' \theta.$$

The recentering term $E^* \left(\hat{s}_{nt}^{*(j)\prime} \theta \right)$ is intended to yield bootstrap population first-order conditions that are zero at $\hat{\theta}_n$. This is relevant for the MBB since $E^* \left[n^{-1} \sum_{t=1}^{n} \hat{s}_{nt}^{*(1)} \right] = \sum_{t=1}^{n} \alpha_n(t) \hat{s}_{nt}$, which in general is not zero. For the SB, $E^* \left[n^{-1} \sum_{t=1}^{n} \hat{s}_{nt}^{*(2)} \right] = n^{-1} \sum_{t=1}^{n} \hat{s}_{nt} = 0$, where the second equality holds by the first-order conditions for $\hat{\theta}_n$, and no recentering is needed. Nevertheless, here we do not need to recenter the moving blocks bootstrap log-likelihood function because it does not impact the first-order properties of the bootstrap QMLE $\hat{\theta}_n^{*(1)}$. It is easy to show that if $\ell_n = o(n)$ $E^* \left[L_n^{(1)}(\cdot, \omega, \theta) \right] = L_n(\omega, \theta) + O_P\left(\frac{\ell_n}{n} \right)$. By a law of large numbers, for ω in a set with probability tending to one, $L_n^{*(1)}(\cdot, \omega, \theta) - E^* \left[L_n^{(1)}(\cdot, \omega, \theta) \right] \stackrel{P_{n,\omega}^*}{\to} 0$. These two facts imply that the conditions of Lemma 3.2 (in particular, condition b3)) are satisfied with our choice of $Q_n^*(\cdot, \omega, \theta)$ for the moving blocks bootstrap provided $\ell_n = o(n)$.

We next investigate the first order asymptotic validity of the bootstrap to approximate the sampling distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta_n^o \right)$. The bootstrap approximation to the true sampling distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta_n^o \right)$ is given by the distribution of $\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n \right)$ conditional on the original sample $X_{n1}, \ldots, X_{nn}$. Typically, we can show that $B_n^{o-1/2} A_n^o \sqrt{n}\left(\hat{\theta}_n - \theta_n^o \right)$ has an asymptotic normal distribution. If we can show that the limiting distribution of $B_n^{o-1/2} A_n^o \sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n \right)$ is also the standard normal distribution, conditional on $X_{n1}, \ldots, X_{nn}$, then the bootstrap approximation is appropriately close to the true sampling distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta_n^o \right)$.

The next lemma is used in the proof of our approximation theorem (Theorem 3.3 below). Its usefulness extends to other applications as the assumptions can be verified for other bootstrap procedures and for other extremum estimators.

Lemma 3.3. *Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and let Θ be a compact subset of $\mathbb{R}^p$, $p \in \mathbb{N}$. Let $\left\{Q_n : \Omega \times \Theta \to \overline{\mathbb{R}}\right\}$ be a sequence of random functions such that*

a1) *$Q_n(\cdot, \theta) : \Omega \to \overline{\mathbb{R}}$ is measurable-$\mathcal{F}$ for each $\theta \in \Theta$, $n = 1, 2, \dots$;*

a2) *$Q_n(\omega, \cdot) : \Theta \to \overline{\mathbb{R}}$ is continuously differentiable of order 2 on Θ a.s. $-P$, $n = 1, 2, \dots$.*

Let $\hat{\theta}_n = \arg\max_\Theta Q_n(\cdot, \theta)$ a.s. $-P$ be measurable such that $\hat{\theta}_n - \theta_n^o \to 0$ as $n \to \infty$ prob $-P$, where $\{\theta_n^o\}$ is interior to Θ uniformly in n. Suppose there exists a nonstochastic sequence of $p \times p$ matrices $\{B_n^o\}$ that is $O(1)$ and uniformly positive definite such that

a3) *$B_n^{o-1/2} \sqrt{n} \nabla Q_n(\cdot, \theta_n^o) \Rightarrow N(0, I_p)$.*

Suppose further that there exists a sequence $\{A_n : \Theta \to \mathbb{R}^{p \times p}\}$ such that $\{A_n\}$ is continuous on Θ uniformly in n and

a4) *$\nabla^2 Q_n(\cdot, \theta) - A_n(\theta) \to 0$ as $n \to \infty$ prob $-P$ uniformly on Θ,*

where $\{A_n^o \equiv A_n(\theta_n^o)\}$ is $O(1)$ and uniformly nonsingular. Then,

$$B_n^{o-1/2} A_n^o \sqrt{n} \left(\hat{\theta}_n - \theta_n^o\right) \Rightarrow N(0, I_p). \tag{A}$$

Let $(\Lambda, \mathcal{G})$ be a measurable space, and for each $\omega \in \Omega$ and $n \in \mathbb{N}$, let $\left(\Lambda, \mathcal{G}, P_{n,\omega}^*\right)$ be a complete probability space. Let $\left\{Q_n^* : \Lambda \times \Omega \times \Theta \to \overline{\mathbb{R}}\right\}$ be a sequence of functions such that

b1) *$Q_n^*(\cdot, \omega, \theta) : \Lambda \to \overline{\mathbb{R}}$ is measurable-$\mathcal{G}$ for each (ω, θ) in $\Omega \times \Theta$, $n = 1, 2, \dots$;*

b2) *$Q_n^*(\lambda, \omega, \cdot) : \Theta \to \overline{\mathbb{R}}$ is continuously differentiable of order 2 on Θ a.s. $-P$, $n = 1, 2, \dots$.*

For each $n = 1, 2, \dots$, let $\hat{\theta}_n^*(\cdot, \omega) = \arg\max_\Theta Q_n^*(\cdot, \omega, \theta)$ a.s. $-P$ be measurable such that $\hat{\theta}_n^*(\cdot, \omega) - \hat{\theta}_n(\omega) \to 0$ as $n \to \infty$, prob $-P_{n,\omega}^*$, prob $-P$. Assume further that

b3) *$B_n^{o-1/2} \sqrt{n} \nabla Q_n^*\left(\cdot, \omega, \hat{\theta}_n(\omega)\right) \Rightarrow^{d_{P_{n,\omega}^*}} N(0, I_p)$ in prob $-P$;*

b4) $\nabla^2 Q_n^* (\cdot, \omega, \theta) - \nabla^2 Q_n (\omega, \theta) \to 0$ *as* $n \to \infty$ *prob* $- P_{n,\omega}^*$, *prob* $- P$ *uniformly on* Θ. *Then,*

$$\text{(B)} \qquad B_n^{o-1/2} A_n^o \sqrt{n} \left(\hat{\theta}_n^* (\cdot, \omega) - \hat{\theta}_n (\omega) \right) \Rightarrow^{d_{P_{n,\omega}^*}} N(0, I_p) \;\; prob - P.$$

A slightly stronger dependence assumption than Gallant and White (1988) use is imposed to satisfy condition b3) in our application of Lemma 3.3. Specifically, we require the elements of the scores to be $L_{2+\delta}-$NED on a mixing process (see Andrews (1988)), for small $\delta > 0$. We strengthen Assumptions A.1-A.10 as follows.

Assumption 3.1.b′) For some small $\delta > 0$ and some $r > 2$, the elements of $\left\{ s_{nt} \left(X^{nt}, \theta \right) \right\}$ are $L_{2+\delta} - NED$ on $\{V_t\}$ of size $-\frac{2(r-1)}{r-2}$ uniformly on (Θ, ρ); $\{V_t\}$ is $\alpha-$mixing with α_k of size $-\frac{(2+\delta)r}{r-2}$.

Assumption 3.2 $n^{-1} \sum_{t=1}^n E\left(s_{nt}^o\right) E\left(s_{nt}^o\right)' = o\left(\ell_n^{-1}\right)$, where $\ell_n = o(n)$ and $\ell_n \to \infty$.

Assumption 3.2 is satisfied under correct specification or stationarity of $\{X_{nt}\}$ and $f_{nt}(\cdot, \theta) = f(\cdot, \theta)$ for all n, t, since in these cases it follows that $E\left(s_{nt}^o\right) = 0$ for all $n, t = 1, 2, \ldots$.

Theorem 3.3. *Let Assumptions A.1-A.10 as strengthened by Assumptions 3.1.a), 3.1.b′) and 3.2 hold. If* $\ell_n \to \infty$ *and* $\ell_n = o\left(n^{1/2}\right)$, *then for any* $\varepsilon > 0$ *and for* $j = 1, 2$,

$$P\left\{ \omega : \sup_{x \in \mathbb{R}^p} \left| P_{n,\omega}^* \left[\sqrt{n} \left(\hat{\theta}_n^{*(j)} (\cdot, \omega) - \hat{\theta}_n (\omega) \right) \leq x \right] - P\left[\sqrt{n} \left(\hat{\theta}_n - \theta_n^o \right) \leq x \right] \right| > \varepsilon \right\} \to 0,$$

where " $\leq$ *" applies to each component of the relevant vectors.*

Theorem 3.3 establishes the consistency of the MBB and SB approximations, given by the conditional distribution of $\sqrt{n}\left(\hat{\theta}_n^{*(j)} - \hat{\theta}_n\right)$, for $j = 1, 2$, respectively, for the true sampling distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta_n^o\right)$. Thus, the order statistics of the bootstrap distribution can be used to construct confidence intervals for θ_n^o with the asymptotically correct coverage probabilities. Note nevertheless that Theorem 3.3 does not justify the use of the variance of the bootstrap distribution as a consistent estimator of the asymptotic variance of the QMLE without further conditions, for example that the sequence $\left\{ \sqrt{n}\left(\hat{\theta}_n^{*(j)} - \hat{\theta}_n\right) \right\}$ is uniformly integrable (see e.g.

Billingsley, 1995, p. 338). This point has been sometimes overlooked in the bootstrap literature. For instance, Fitzenberger (1997, p. 250) claims that his Theorem 3.2 (which shows that the discrepancy between the bootstrap and the normal distributions for the least squares (LS) estimator converges uniformly to zero in probability) "establishes that the MBB procedure provides a means for HAC inference in LS linear regression", without proving uniform integrability. Counterexamples of the inconsistency of the bootstrap variance of smooth functions of sample means in the i.i.d. context can be found in Ghosh et al. (1984) and Shao (1992). The consistency of the block bootstrap variance of $\sqrt{n}\left(\hat{\theta}_n^{*(j)} - \hat{\theta}_n\right)$ when $\hat{\theta}_n$ is a smooth function of sample means and, in particular, a nonlinear function of the LS estimator in linear dynamic models is studied in Gonçalves and White (2000b).

As remarked in Section 2, bootstrapping maximum likelihood estimators for nonlinear dynamic models may be computationally costly as it requires solving a QMLE optimization problem for each resample. In the context of testing nonlinear models, Davidson and MacKinnon (1999) have recently proposed approximate bootstrap methods that are based on a small number of iterative steps starting from the QMLE obtained for the original sample and that achieve the same level of accuracy as the fully-optimized bootstrap. For nonlinear regressions, Shao and Tu (1995) suggested a related one-step bootstrap estimator that further simplifies the computations by only requiring resampling the gradient of the nonlinear objective function. We will call this one-step bootstrap estimator Shao and Tu's one-step bootstrap estimator to distinguish it from Davidson and MacKinnon's (1999) one-step bootstrap estimator.

Let $\hat{A}_n = n^{-1}\sum_{t=1}^{n} \nabla^2 \log f_{nt}\left(X_n^t, \hat{\theta}_n\right)$ and let $\left\{\hat{s}_{nt}^{*(j)}\right\}$ be the resampled estimated scores obtained by the MBB and SB resampling schemes. The MBB and SB analogs to Shao and Tu's (1995) one-step bootstrap estimator are

$$\hat{\theta}_n^{*(j)} = \hat{\theta}_n - \hat{A}_n^{-1} n^{-1} \sum_{t=1}^{n} \hat{s}_{nt}^{*(j)}.$$

Davidson and MacKinnon's (1999) one-step procedure uses $\hat{A}_n^{*(j)-1}$ in place of $\hat{A}_n^{-1}$:

$$\theta_n^{\dagger*(j)} = \hat{\theta}_n - \hat{A}_n^{*(j)-1} n^{-1} \sum_{t=1}^{n} \hat{s}_{nt}^{*(j)}.$$

Corollary 3.1 states the first-order asymptotic validity of $\acute{\theta}_n^{*(j)}$ and $\theta_n^{\dagger*(j)}$ in the context of quasi-maximum likelihood estimation of nonlinear dynamic models with heterogeneous NED data.

Corollary 3.1. *Let Assumptions A.1-A.10 as strengthened by Assumptions 3.1.a), 3.1.b′) and 3.2 hold. If* $\ell_n \to \infty$ *and* $\ell_n = o\left(n^{1/2}\right)$, *then for* $j = 1, 2$, *and for any* $\varepsilon > 0$,

$$P\left\{\sup_{x\in\mathbb{R}^p}\left|P_{n,\omega}^*\left[\sqrt{n}\left(\hat{\theta}_n^{*(j)}(\cdot,\omega) - \hat{\theta}_n(\omega)\right) \le x\right] - P_{n,\omega}^*\left[\sqrt{n}\left(\hat{\theta}_{1n}^{*(j)}(\cdot,\omega) - \hat{\theta}_n(\omega)\right) \le x\right]\right| > \varepsilon\right\} \to 0,$$

where $\hat{\theta}_{1n}^{*(j)}$ *denotes the one-step bootstrap estimator* $\acute{\theta}_n^{*(j)}$ *or* $\theta_n^{\dagger*(j)}$.

Analogous results hold for the multi-step estimators.

4. Hypothesis Testing

Although the results of Section 3 justify the use of the bootstrap to approximate the distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta_n^o\right)$, they do not immediately justify testing hypotheses about θ_n^o based on studentized test statistics such as a t-statistic or a Wald statistic. Nevertheless, they are an important step towards proving the validity of the bootstrap to approximate the distribution of studentized statistics based on the QMLE $\hat{\theta}_n$, as we show in this section. In particular, we prove the first order asymptotic validity of a suitable bootstrap analog of the Wald test statistic for testing general nonlinear restrictions on θ_n^o. We focus on the moving blocks bootstrap, which is easier to analyze, but analogous results are expected to hold for the stationary bootstrap. Proving second-order optimality properties of block bootstrap tests in our framework is beyond our present scope. See Hall and Horowitz (1996) and Andrews (1999) for higher order improvements of bootstrap testing based on extremum estimators in a dependent stationary context.

Let $\{r_n : \Theta \to \mathbb{R}^q\}$, with $\Theta \subset \mathbb{R}^p$, $q \le p$, be a sequence of functions that satisfy the usual regularity conditions (see e.g. White, 1994, Assumption 8.2). In particular, assume it has elements continuously differentiable on Θ uniformly in n such that $\{R_n^o \equiv \nabla' r_n(\theta_n^o)\}$ is $O(1)$ and has full row rank q, uniformly in n. The Wald test statistic for testing $H_o : \sqrt{n} r_n(\theta_n^o) \to 0$ is

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

where $\hat{r}_n = r_n\left(\hat{\theta}_n\right)$, $\hat{R}_n = \nabla' r_n\left(\hat{\theta}_n\right)$ and $\hat{C}_n = \hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}$ is a consistent estimator of $C_n^o = A_n^{o-1}B_n^o A_n^{o-1}$. In particular, $\hat{A}_n = n^{-1}\sum_{t=1}^n \nabla^2 \log f_{nt}\left(X_n^t, \hat{\theta}_n\right)$ and $\hat{B}_n$ is such that $\hat{B}_n - B_n^o \xrightarrow{P} 0$. In our NED context, $\hat{B}_n$ is a kernel-type variance estimator, e.g. $\hat{B}_{n,1}$ or $\hat{B}_{n,2}$ in Section 3. For first order properties, the particular choice of $\hat{B}_n$ is not relevant as long as it is a consistent estimator of B_n^o. Nevertheless, this choice becomes important for second order properties (see remarks below). The bootstrap Wald statistic we consider is given by

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

where, with $\hat{\theta}_n^*$ the bootstrap QMLE, we set $\hat{r}_n^* = r_n\left(\hat{\theta}_n^*\right)$, $\hat{R}_n^* = \nabla' r_n\left(\hat{\theta}_n^*\right)$ and $\hat{C}_n^* = \hat{A}_n^{*-1}\hat{B}_n^*\hat{A}_n^{*-1}$. $\hat{A}_n^*$ is the bootstrap analog of $\hat{A}_n$, defined as $\hat{A}_n^* = n^{-1}\sum_{t=1}^n \nabla^2 \log f_{n,\tau_{nt}}\left(X_n^{\tau_{nt}}, \hat{\theta}_n^*\right)$, where $\{\tau_{nt}\}$ is a random array generated by the bootstrap resampling scheme. For the MBB, recall from Section 3 that $\{\tau_{nt}\} = \{I_{n1}+1, \ldots, I_{n1}+\ell, \ldots, I_{nk}+1, \ldots, I_{nk}+\ell\}$, where for each n, $\{I_{ni}\}$ are i.i.d. uniform random variables on $\{0, \ldots, n-\ell\}$. In particular, for any sequence $\{Z_{nt}\}$ we have that $Z_{n,(i-1)\ell+t}^* = Z_{n,I_{ni}+t}$ for $i = 1, \ldots, k$ and $t = 1, \ldots, \ell$, where $k = n/\ell$ is the number of blocks used to form the bootstrap sample. $\hat{B}_n^*$ is the bootstrap analog of $\hat{B}_n$, given by

$$\hat{B}_n^* = k^{-1}\sum_{i=1}^k \left(\ell^{-1/2}\sum_{t=1}^{\ell} s_{n,I_{ni}+t}\left(X_n^{I_{ni}+t}, \hat{\theta}_n^*\right)\right)\left(\ell^{-1/2}\sum_{t=1}^{\ell} s_{n,I_{ni}+t}\left(X_n^{I_{ni}+t}, \hat{\theta}_n^*\right)\right)'. \tag{4.1}$$

$\hat{B}_n^*$ is the multivariate analog of the estimator of the MBB variance proposed by Götze and Künsch (1996) for studentizing the bootstrap resampled mean, adapted to the QMLE context. To motivate this choice of $\hat{B}_n^*$ recall that $\hat{B}_n^*$ is the bootstrap analog of $\hat{B}_n$, which is an estimator of B_n^o, the variance matrix of the scaled scores evaluated at the "true parameter" θ_n^o. Analogously, $\hat{B}_n^*$ is an estimator of the bootstrap variance matrix of the scaled average of the resampled scores evaluated at the "bootstrap true parameter" $\hat{\theta}_n$, i.e. $\hat{B}_n^*$ is an *estimator* of $\hat{B}_{n,1}$ in Section 3. By definition,

$$(4.2)\quad \hat{B}_{n,1} = \text{var}^*\left(n^{-1/2}\sum_{t=1}^{n} s^*_{nt}\left(\hat{\theta}_n\right)\right) = \text{var}^*\left(k^{-1/2}\sum_{i=1}^{k}\left(\ell^{-1/2}\sum_{t=1}^{\ell} s_{n,I_{ni}+t}\left(X_n^{I_{ni}+t},\hat{\theta}_n\right)\right)\right).$$

Because the block bootstrap means $\ell^{-1}\sum_{t=1}^{\ell} s_{n,I_{ni}+t}\left(X_n^{I_{ni}+t},\hat{\theta}_n\right)$ are (conditionally) i.i.d., the estimator of the (bootstrap population) variance (4.2) is simply the sample variance of the bootstrap variables $\left\{\ell^{-1/2}\sum_{t=1}^{\ell} s_{n,I_{ni}+t}\left(X_n^{I_{ni}+t},\hat{\theta}_n\right) : i = 1,\ldots,k\right\}$, where $\hat{\theta}_n$ is replaced by its bootstrap estimator $\hat{\theta}^*_n$ to mimic the fact that θ^o_n is replaced with $\hat{\theta}_n$ when computing $\hat{B}_n$. Thus, (4.1) exploits the independence of the blocks bootstrap means and is a natural estimator of $\hat{B}_{n,1}$. Notice that we simplified (4.1) by using the first order conditions of the bootstrap problem to set $\overline{\hat{s}}{}^*_n \equiv n^{-1}\sum_{t=1}^{n} s^*_{nt}\left(\hat{\theta}^*_n\right) = 0$.

Götze and Künsch (1996) prove the second order correctness of the MBB distribution of a studentized statistic of smooth functions of sample means of stationary mixing data that uses (an equivalent version of) $\hat{B}^*_n$ to studentize the bootstrap statistic in the context of one-sided bootstrap-t intervals. As Götze and Künsch (1996) remark, $\hat{B}^*_n$ differs from $\hat{B}_{n,1}$ (or from any other HAC covariance estimator we might want to use) in a fundamental way: while for $\hat{B}_{n,1}$ we consider all pairs of observations at lag distance less than ℓ (with an appropriate lag weight), instead for $\hat{B}^*_n$ we only consider pairs in the same blocks. In order to have second order improvements, they note the need to choose the kernel variance estimator for studentizing the original statistic carefully. In particular, they claim that triangular weights should not be used, which suggests that $\hat{B}_{n,1}$ should not be used to studentize the Wald test statistic in the first place. Instead, rectangular or quadratic weights should be used in defining the HAC estimator of B^o_n. In addition, as we remarked in Section 3, in order to achieve second order improvements, $\hat{\theta}^*_n$ should be computed with recentering, as in Hall and Horowitz (1996) and Andrews (1999). For simplicity, we abstract from these considerations as they do not affect our first order asymptotic results, although they should be borne in mind in applications.

In order to obtain the first order asymptotic validity of the bootstrap Wald statistic $\mathcal{W}^*_n$ we strengthen Assumption 3.2 as follows.

Assumption 3.2$'$ $n^{-1}\sum_{t=1}^{n}|E(s_{nti}^{o})|^{2+\delta}=o\left(\ell_n^{-1-\delta/2}\right)$ for $i=1,\ldots,p$.

Theorem 4.1. *Let the assumptions of Theorem 3.3 hold as strengthened by Assumption 3.2$'$. Then, under H_o, for all $\varepsilon>0$, if $\ell=o\left(n^{1/2}\right)$,*

$$P\left[\sup_{x\in\mathbb{R}^q}\left|P_{n,\omega}^{*}\left(\mathcal{W}_n^{*}(\cdot,\omega)\leq x\right)-P\left(\mathcal{W}_n\leq x\right)\right|>\varepsilon\right]\to 0 \text{ as } n\to\infty.$$

Heuristically, by a mean value expansion of $r_n\left(\hat{\theta}_n^{*}\right)$ about $\hat{\theta}_n$ we have that with probability approaching one $\sqrt{n}\left(r_n\left(\hat{\theta}_n^{*}\right)-r_n\left(\hat{\theta}_n\right)\right)\Rightarrow^{d_{P^*}}N\left(0,R_n^oC_n^oR_n^{o\prime}\right)$, which implies that $n\left(\hat{r}_n^{*}-\hat{r}_n\right)'\left(R_n^oC_n^oR_n^{o\prime}\right)^{-1}\left(\hat{r}_n^{*}-\hat{r}_n\right)\Rightarrow^{d_{P^*}}\mathcal{X}_q^2$. In particular, the consistency of $\hat{B}_n^{*}$ for B_n^o implies the consistency of $\hat{R}_n^{*}\hat{C}_n^{*}\hat{R}_n^{*\prime}$ for $R_n^oC_n^oR_n^{o\prime}$, which delivers the first order asymptotic equivalence under the null between the bootstrap Wald statistic and the original Wald statistic.

The bootstrap can also be shown to work for an appropriate version of the Lagrange Multiplier (LM) statistic. Using notation analogous to that of Gallant and White (1988), the bootstrap LM statistic can be written

$$\mathcal{L}_n^{*}=n\nabla' L_n^{*}\left(\tilde{\theta}_n^{*}\right)\tilde{A}_n^{*-1}\tilde{R}_n^{*\prime}\left(\tilde{R}_n^{*}\tilde{C}_n^{*}\tilde{R}_n^{*\prime}\right)^{-1}\tilde{R}_n^{*}\tilde{A}_n^{*-1}\nabla L_n^{*}\left(\tilde{\theta}_n^{*}\right),$$

where, with $\tilde{\theta}_n^{*}$ the constrained bootstrap QMLE, we set $\nabla L_n^{*}\left(\tilde{\theta}_n^{*}\right)=n^{-1}\sum_{t=1}^{n}s_{nt}^{*}\left(\tilde{\theta}_n^{*}\right)$, $\tilde{R}_n^{*}=\nabla' r_n\left(\tilde{\theta}_n^{*}\right)$, $\tilde{C}_n^{*}=\tilde{A}_n^{*-1}\tilde{B}_n^{*}\tilde{A}_n^{*-1}$, and $\tilde{A}_n^{*}=n^{-1}\sum_{t=1}^{n}\nabla^2\log f_{n,\tau_{nt}}\left(X_n^{\tau_{nt}},\tilde{\theta}_n^{*}\right)$. Similarly, $\tilde{B}_n^{*}$ is defined as in (4.1) with $\tilde{\theta}_n^{*}$ instead of $\hat{\theta}_n^{*}$, where $\ell^{1/2}\nabla L_n^{*}\left(\tilde{\theta}_n^{*}\right)$ is subtracted off each term $\ell^{-1/2}\sum_{t=1}^{\ell}s_{n,I_{ni}+t}\left(X_n^{I_{ni}+t},\tilde{\theta}_n^{*}\right)$ to account for the fact that $\nabla L_n^{*}\left(\tilde{\theta}_n^{*}\right)$ is not zero for the constrained optimization problem. As the development is entirely parallel to that given in Gallant and White (1988), using arguments analogous to those of Theorem 4.1, we omit the formalities.

5. Conclusion

This paper gives conditions under which two commonly used block bootstrap procedures, the moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992) and the stationary bootstrap of Politis and Romano (1994a), provide valid tools for inference using maximum likelihood estimators of nonlinear dynamic models with heterogeneous dependent data. Our results apply to a wide class of data generating processes, the processes near epoch dependent on a mixing process, thus allowing for a considerable degree of heterogeneity and dependence in the data. We prove that the bootstrap works for the QMLE in that it gets the limiting distribution of the QMLE right. We introduce a new heteroskedasticity and autocorrelation consistent covariance matrix estimator for the QMLE. We also show the first order asymptotic validity of a suitable bootstrap analog of a Wald test statistic for testing general nonlinear restrictions on the pseudo-true parameters of the model.

A. Assumptions and Proofs for Section 3

Assumptions A.1 through A.10 are the doubly indexed counterparts of the regularity conditions used by Gallant and White (1988) to deliver the consistency and asymptotic normality of the QMLE.

Assumption A.1 Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. The observed data are a realization of a stochastic process $X \equiv \left\{X_{nt} : \Omega \to \mathbb{R}^l,\ l \in \mathbb{N},\ n, t = 1, 2, \ldots\right\}$, where $X_{nt}(\omega) = W_{nt}(\ldots, V_{t-1}(\omega), V_t(\omega), V_{t+1}(\omega), \ldots)$, $\omega \in \Omega$, where $V_t : \Omega \to \mathbb{R}^v$, $v \in \mathbb{N}$, and $W_{nt} : \times_{\tau=-\infty}^{\infty} \mathbb{R}^v \to \mathbb{R}^l$ are such that X_{nt} is measurable-$\mathcal{F}$, $n, t = 1, 2, \ldots$.

Assumption A.2 The functions $f_{nt} : \mathbb{R}^{lt} \times \Theta \to \mathbb{R}^+$ are such that $f_{nt}(\cdot, \theta)$ is measurable-$\mathcal{B}^{lt}$ for each $\theta \in \Theta$, a compact subset of $\mathbb{R}^p$, $p \in \mathbb{N}$, and $f_{nt}\left(X_n^t, \cdot\right) : \Theta \to \mathbb{R}^+$ is continuous on Θ $a.s. - P$, $n, t = 1, 2, \ldots$.

Assumption A.3 θ_n^o is identifiably unique with respect to $\bar{L}_n(\theta) \equiv E\left(L_n\left(X_n^n, \theta\right)\right)$, i.e. given $\varepsilon > 0$ there exists $N_o(\varepsilon) < \infty$ and $\delta(\varepsilon) < 0$ such that

$$\sup_{n \geq N_o(\varepsilon)} \left[\max_{\theta \in \eta_n^c(\varepsilon)} \bar{L}_n(\theta) - \bar{L}_n\left(\theta_n^o\right)\right] \equiv \delta(\varepsilon) < 0,$$

where $\eta_n^c(\varepsilon)$ is the compact complement of $\eta_n(\varepsilon) = \{\theta : |\theta - \theta_n^o| < \varepsilon\} \cap \Theta$.

Assumption A.4 θ_n^o is interior to Θ uniformly in n, i.e. there exists $\varepsilon > 0$ such that for all n sufficiently large $\{\theta \in \mathbb{R}^p : |\theta - \theta_n^o| < \varepsilon\} = \{\theta \in \Theta : |\theta - \theta_n^o| < \varepsilon\}$.

Assumption A.5

(i) $\left\{\log f_{nt}\left(X_n^t, \theta\right)\right\}$ is Lipschitz continuous on Θ, i.e. $\left|\log f_{nt}\left(X_n^t, \theta\right) - \log f_{nt}\left(X_n^t, \theta^o\right)\right| \leq L_{nt}$ $|\theta - \theta^o|$ $a.s. - P$, $\forall \theta, \theta^o \in \Theta$, where $\sup_n \left\{n^{-1} \sum_{t=1}^n E\left(L_{nt}\right)\right\} = O(1)$.

(ii) $\{\nabla' s_{nt}(X_n^t, \theta) \equiv \nabla^2 \log f_{nt}(X_n^t, \theta)\}$ is Lipschitz continuous on Θ.

Assumption A.6

(i) $\{\log f_{nt}(X_n^t, \theta)\}$ is $r-$dominated on Θ uniformly in n, t, i.e. there exists $D_{nt} : \mathbb{R}^{lt} \to \mathbb{R}$ such that $|\log f_{nt}(X_n^t, \theta)| \leq D_{nt}$ for all θ in Θ and D_{nt} is measurable-$\mathcal{B}^{lt}$ such that $\|D_{nt}\|_r \leq \Delta < \infty$ for $r > 2$ and all $n, t = 1, 2, \dots$.

(ii) $\{s_{nt}(X_n^t, \theta) \equiv \nabla \log f_{nt}(X_n^t, \theta)\}$ is r-dominated on Θ uniformly in $n, t = 1, 2, \dots, r > 2$.

(iii) $\{\nabla' s_{nt}(X_n^t, \theta) \equiv \nabla^2 \log f_{nt}(X_n^t, \theta)\}$is r-dominated on Θ uniformly in $n, t = 1, 2, \dots, r > 2$.

Assumption A.7 $\{V_t\}$ is an α-mixing sequence of size $-\frac{2r}{r-2}$, with $r > 2$.

Assumption A.8

(i) The elements of $\{\log f_{nt}(X_n^t, \theta)\}$ are near epoch dependent (NED) on $\{V_t\}$ of size $-\frac{1}{2}$.

(ii) The elements of $\{s_{nt}(X_n^t, \theta) \equiv \nabla \log f_{nt}(X_n^t, \theta)\}$ are NED on $\{V_t\}$ of size -1 uniformly on (Θ, ρ), where ρ is any convenient norm on $\mathbb{R}^p$.

(iii) The elements of $\{\nabla' s_{nt}(X_n^t, \theta) \equiv \nabla^2 \log f_{nt}(X_n^t, \theta)\}$ are NED on $\{V_t\}$ of size $-\frac{1}{2}$ uniformly on (Θ, ρ).

Assumption A.9 $\left\{B_n^o \equiv \operatorname{var}\left(n^{-\frac{1}{2}} \sum_{t=1}^n \nabla \log f_{nt}(X_n^t, \theta_n^o)\right)\right\}$ is uniformly positive definite, i.e. B_n^o is positive semi-definite for all n and $\det B_n^o > \kappa > 0$ for all n sufficiently large for some $\kappa > 0$.

Assumption A.10 $\{A_n^o \equiv E(n^{-1} \sum_{t=1}^n \nabla^2 \log f_{nt}(X_n^t, \theta_n^o))\}$ is uniformly nonsingular, i.e. $|\det A_n^o| \geq \kappa > 0$ for all n sufficiently large.

Lemma A.1. *Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, and let $X_{nt} : \Omega \to \mathbb{R}^l$, $l \in \mathbb{N}$, be measurable-$\mathcal{F}$, $n, t = 1, 2, \dots$. Let $(\Lambda, \mathcal{G})$ be a measurable space and for each $\omega \in \Omega$, let $P_{n,\omega}^*$ be the probability measure induced by the bootstrap on $(\Lambda, \mathcal{G})$, a complete measurable space. For each $n = 1, 2, \dots$, and each $\omega \in \Omega$, define $X_{n\omega}^* \equiv \{X_{nt}^*(\cdot, \omega)\} : \Lambda \to \mathbb{R}^{l\infty}$ as*

$$X_{nt}^*(\lambda, \omega) = X_{n,\tau_{nt}(\lambda)}(\omega), \qquad t = 1, 2, \dots ,$$

where τ_{nt} is a double array of random indexes on $(\Lambda, \mathcal{G}, P_{n,\omega}^*)$ taking values in $\{1, \dots, n\}$ for $n \in \mathbb{N}$. Then $X_n^*(\cdot, \omega)$ is a measurable mapping with respect to $\mathcal{G}/\mathcal{B}^{l\infty}$, i.e. $X_n^*(\cdot, \omega)$ is a stochastic process on $(\Lambda, \mathcal{G}, P_{n,\omega}^*)$ for each $n = 1, 2, \dots$.

Proof. Fix $\omega \in \Omega$ and let $l \in \mathbb{N}$, $n = 1, 2, \ldots$. For any Borel set $H \in \mathcal{B}\left(\mathbb{R}^l\right)$, consider

$$\begin{aligned}\left[\lambda \in \Lambda : X^*_{nt}(\lambda, \omega) \in H, \ \ t = 1, 2, \ldots\right] &= \left[\lambda \in \Lambda : X_{n, \tau_{nt}(\lambda)}(\omega) \in H, \ \ t = 1, 2, \ldots\right] \\ &= \cap_{t=1}^{\infty} \cup_{k=1}^{n} \left[\lambda \in \Lambda : \tau_{nt}(\lambda) = k\right] \cap \left[\lambda \in \Lambda : X_{n,k}(\omega) \in H\right].\end{aligned}$$

We claim this set is in $\mathcal{G}$. Indeed, for each t and k it is true that $[\lambda \in \Lambda : \tau_{nt}(\lambda) = k] \in \mathcal{G}$ because τ_{nt} is a simple random variable on $\mathcal{G}$. Because $X_{n,k}(\omega)$ is a constant for fixed ω, it trivially follows that $[\lambda \in \Lambda : X_{n,k}(\omega) \in H] \in \mathcal{G}$: $[\lambda \in \Lambda : X_{n,k}(\omega) \in H]$ is either Λ or $\varnothing$ depending on whether $X_{n,k}(\omega)$ is in H or not. Since a σ-field is closed under countable unions and complements the result follows.

Proof of Theorem 3.1. Theorem 3.1 of Gonçalves and White (2000a) implies $B_{n,j} - (B^o_n + U_{n,j}) \xrightarrow{P} 0$ for $j = 1, 2$. Thus, it suffices to show that $\hat{B}_{n,j} - B^o_{n,j} \xrightarrow{P} 0$ for $j = 1, 2$.

We start by proving this result for the MBB, i.e. for $j = 1$. We let $\alpha_{nt} \equiv \alpha_n(t)$ and $\beta_{n,t,\tau} \equiv \beta_n(t, \tau)$ throughout. Given that $\sum_{t=1}^{n} \alpha_{nt} = 1$, and that for each τ, $\sum_{t=1}^{n-\tau} \beta_{n,t,\tau} = 1$, by (3.1) and (3.6) one can write

$$\begin{aligned}
\hat{B}_{n,1} - B^o_{n,1} &= D_{n1} - D_{n2} + D_{n3} + D'_{n3} - D_{n4} - D'_{n4} + D_{n5}, \text{ where} \\
D_{n1} &\equiv \sum_{t=1}^{n} \alpha_{nt}\left(\hat{s}_{nt}\hat{s}'_{nt} - s^o_{nt}s^{o\prime}_{nt}\right) \\
D_{n2} &\equiv \overline{\hat{s}}_{\alpha,n}\overline{\hat{s}}'_{\alpha,n} - \overline{s}^o_{\alpha,n}\overline{s}^{o\prime}_{\alpha,n} \\
D_{n3} &\equiv \sum_{\tau=1}^{\ell-1}\left(1 - \frac{\tau}{\ell}\right)\sum_{t=1}^{n-\tau}\beta_{n,t,\tau}\left(\hat{s}_{nt}\hat{s}'_{n,t+\tau} - s^o_{nt}s^{o\prime}_{n,t+\tau}\right) \\
D_{n4} &\equiv \sum_{\tau=1}^{\ell-1}\left(1 - \frac{\tau}{\ell}\right)\sum_{t=1}^{n-\tau}\beta_{n,t,\tau}\left(\left(\hat{s}_{nt} + \hat{s}_{n,t+\tau}\right)\overline{\hat{s}}'_{\alpha,n} - \left(s^o_{nt} + s^o_{n,t+\tau}\right)\overline{s}^{o\prime}_{\alpha,n}\right) \\
D_{n5} &\equiv (\ell - 1)\left(\overline{\hat{s}}_{\alpha,n}\overline{\hat{s}}'_{\alpha,n} - \overline{s}^o_{\alpha,n}\overline{s}^{o\prime}_{\alpha,n}\right).
\end{aligned}$$

To prove that $D_{n1} \xrightarrow{P} 0$, we take a mean value expansion of a typical element of $\sum_{t=1}^{n} \alpha_{nt}\hat{s}_{nt}\hat{s}'_{nt}$ around θ^o_n. Let $\hat{s}_{nti}$ be a typical element of $\hat{s}_{nt}$. Then,

$$\sum_{t=1}^{n} \alpha_{nt}\left(\hat{s}_{nti}\hat{s}_{ntj} - s^o_{nti}s^o_{ntj}\right) = \bar{a}_n\left(\hat{\theta}_n - \theta^o_n\right),$$

where $\bar{a}_n \equiv \sum_{t=1}^{n} \alpha_{nt}\left(\bar{s}_{nti}\nabla'\bar{s}_{ntj} + \bar{s}_{ntj}\nabla'\bar{s}_{nti}\right)$, $\bar{s}_{nti} \equiv s_{nti}\left(\bar{\theta}_n\right)$, and $\bar{\theta}_n$ lies between $\hat{\theta}_n$ and θ^o_n. Given Assumptions A.6.(ii) and A.6.(iii) and the fact that $\sum_{t=1}^{n} \alpha_{nt} = 1$, an application of the Minkowski and Hölder inequalities implies that for some $r > 2$, $\|\bar{a}_n\|_{\frac{r}{2}} \leq 2\Delta^2 < \infty$. Thus, by Markov's inequality, $\bar{a}_n = O_P(1)$. It follows that $\bar{a}_n\left(\hat{\theta}_n - \theta^o_n\right) = o_P(1)$ because $\hat{\theta}_n - \theta^o_n = o_P(1)$ by Theorem 3.19 of Gallant and White (1988) under our assumptions.

Similarly, to prove that $D_{n3} \xrightarrow{P} 0$, we take a mean value expansion of a typical element of $\sum_{\tau=1}^{\ell-1}\left(1-\frac{\tau}{\ell}\right)\sum_{t=1}^{n-\tau}\beta_{n,t,\tau}\hat{s}_{nt}\hat{s}'_{n,t+\tau}$ around θ_n^o and get

$$\sum_{\tau=1}^{\ell-1}\left(1-\frac{\tau}{\ell}\right)\sum_{t=1}^{n-\tau}\beta_{n,t,\tau}\left(\hat{s}_{nti}\hat{s}_{n,t+\tau,j}-s_{nti}^o s_{n,t+\tau,j}^o\right)=n^{-1/2}\bar{a}_n\sqrt{n}\left(\hat{\theta}_n-\theta_n^o\right),$$

where $\bar{a}_n \equiv \sum_{\tau=1}^{\ell-1}\left(1-\frac{\tau}{\ell}\right)\sum_{t=1}^{n-\tau}\beta_{n,t,\tau}\left(\bar{s}_{nti}\nabla'\bar{s}_{n,t+\tau,j}+\bar{s}_{n,t+\tau,j}\nabla'\bar{s}_{nti}\right)$, with $\bar{s}_{nti} \equiv s_{nti}\left(\bar{\theta}_n\right)$ for some $\bar{\theta}_n$ lying between $\hat{\theta}_n$ and θ_n^o. Routine arguments deliver $n^{-1/2}\bar{a}_n = O_P\left(\frac{\ell}{n^{1/2}}\right)$ (apply the Minkowski and Hölder inequalities and notice that for fixed τ $\sum_{t=1}^{n-\tau}\beta_{n,t,\tau}=1$ to conclude that $\|\bar{a}_n\|_{\frac{r}{2}} \leq \Delta^2\ell$ for some $r>2$). The desired result follows because $\sqrt{n}\left(\hat{\theta}_n-\theta_n^o\right)=O_P(1)$ under our assumptions and because $\ell \equiv \ell_n = o\left(n^{1/2}\right)$ by assumption. A similar argument yields $D_{n4} \xrightarrow{P} 0$. To prove that D_{n2} and D_{n5} converge to zero in probability, it is first convenient to note (Fitzenberger, 1997, Lemma A.1) that if $\ell = o(n)$, $\bar{s}_{\alpha,n}^o = \bar{s}_n^o + O_P\left(\frac{\ell}{n}\right)$ and $\bar{\hat{s}}_{\alpha,n} = \bar{\hat{s}}_n + O_P\left(\frac{\ell}{n}\right)$. But $\bar{s}_n^o \equiv n^{-1}\sum_{t=1}^n s_{nt}^o = O_P\left(n^{-1/2}\right)$, because under our assumptions $B_n^{o-1/2}\sqrt{n}\bar{s}_n^o \Rightarrow N(0,I)$ and $B_n^{o-1/2}=O(1)$; moreover, the F.O.C. for $\hat{\theta}_n$ allow us to set $\bar{\hat{s}}_n=0$. Thus, $D_{n2}=O_P\left(n^{-1}\right)+O_P\left(\frac{\ell}{n^{3/2}}\right)+O_P\left(\frac{\ell^2}{n^2}\right)$ and $D_{n5}=O_P\left(\frac{\ell}{n}\right)+O_P\left(\frac{\ell^2}{n^{3/2}}\right)+O_P\left(\frac{\ell^3}{n^2}\right)$, which converge to zero in probability given our assumptions on ℓ.

The proof that $\hat{B}_{n,j}-B_{n,j}^o \xrightarrow{P} 0$ for $j=2$ follows similarly once we use (3.4) and (3.7) to write

$$\begin{aligned}
\hat{B}_{n,2}-B_{n,2}^o &= F_{n1}+F_{n2}+F'_{n2}-F_{n3}, \text{ where}\\
F_{n1} &\equiv n^{-1}\sum_{t=1}^n\left(\hat{s}_{nt}\hat{s}'_{nt}-s_{nt}^o s_{nt}^{o\prime}\right)\\
F_{n2} &\equiv \sum_{\tau=1}^{n-1}b_n(\tau)\,n^{-1}\sum_{t=1}^{n-\tau}\left(\hat{s}_{nt}\hat{s}'_{n,t+\tau}-s_{nt}^o s_{n,t+\tau}^{o\prime}\right)\\
F_{n3} &\equiv \left(\bar{s}_n^o\bar{s}_n^{o\prime}+2\bar{s}_n^o\bar{s}_n^{o\prime}\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)(1-p)^\tau\right),
\end{aligned}$$

and notice that $\sum_{\tau=1}^{n-1}b_n(\tau) \leq \frac{1}{p_n} = \ell_n$. ■

Proof of Lemma 3.1. Under our assumptions it follows by Theorem 3.4 of White (1994) that $\hat{\theta}_n-\theta_n^o \to 0$ as $n\to\infty$ $a.s.-P$. Let $F \equiv \left\{\omega : \hat{\theta}_n(\omega)-\theta_n^o \to 0\right\} \cap \left\{\omega : \sup_\Theta\left|Q_n(\omega,\theta)-\overline{Q}_n(\theta)\right| \to 0\right\}$. By hypothesis, $P(F)=1$. Fix $\varepsilon'>0$ and ω in F. Because $\hat{\theta}_n(\omega)-\theta_n^o \to 0$ there exists $N_0(\omega,\varepsilon')<\infty$ such that for all $n>N_0(\omega,\varepsilon')$, $\left|\hat{\theta}_n(\omega)-\theta_n^o\right|<\varepsilon'$. Because $\{\theta_n^o\}$ is identifiably unique on Θ, given $\varepsilon'>0$ there exists $N_1(\varepsilon')<\infty$ and $\delta'(\varepsilon')>0$ such that $\sup_{n\geq N_1(\varepsilon')}\left[\max_{\eta^c(\theta_n^o,\varepsilon')}\overline{Q}_n(\theta)-\overline{Q}_n(\theta_n^o)\right] \equiv -\delta'(\varepsilon')<0$, where $\eta(\theta_n^o,\varepsilon') \equiv \{\theta\in\Theta : |\theta-\theta_n^o|<\varepsilon'\}$. Next, by Corollary 3.8 of White (1994), for ω in F there exists $N_2\left(\omega,\delta'(\varepsilon')\right)<\infty$ such that for all $n>N_2\left(\omega,\delta'(\varepsilon')\right)$, $\left|Q_n\left(\omega,\hat{\theta}_n(\omega)\right)-\overline{Q}_n(\theta_n^o)\right|<\frac{\delta'(\varepsilon')}{4}$, or $-Q_n\left(\omega,\hat{\theta}_n(\omega)\right)<-\overline{Q}_n(\theta_n^o)+\frac{\delta'(\varepsilon')}{4}$.

For ω in F and all $n > N_2\left(\omega, \delta'\left(\varepsilon'\right)\right)$ it is also true that $\max_{\eta^c(\theta_n^o,\varepsilon')} Q_n\left(\omega,\theta\right) \leq \max_{\eta^c(\theta_n^o,\varepsilon')} \overline{Q}_n\left(\theta\right) + \frac{\delta'(\varepsilon')}{4}$ (by uniform convergence of $Q_n\left(.,\theta\right) - \overline{Q}_n\left(\theta\right)$ to zero). Let $N\left(\omega,\varepsilon'\right) = \max\left\{N_0\left(\omega,\varepsilon'\right), N_1\left(\varepsilon'\right), N_2\left(\omega,\delta'\left(\varepsilon'\right)\right)\right\}$ and notice that $\eta^c\left(\hat{\theta}_n\left(\omega\right), 2\varepsilon'\right) \subset \eta^c\left(\theta_n^o,\varepsilon'\right)$ for $n > N_0\left(\omega,\varepsilon'\right)$. It follows that

$$\begin{aligned}
&\sup_{n\geq N(\omega,\varepsilon')}\left[\max_{\eta^c(\hat{\theta}_n(\omega),2\varepsilon')} Q_n\left(\omega,\theta\right) - Q_n\left(\omega,\hat{\theta}_n\left(\omega\right)\right)\right] \\
&\leq \sup_{n\geq N(\omega,\varepsilon')}\left[\max_{\eta^c(\theta_n^o,\varepsilon')} Q_n\left(\omega,\theta\right) - \overline{Q}_n\left(\theta_n^o\right) + \frac{\delta'\left(\varepsilon'\right)}{4}\right] \\
&\leq \sup_{n\geq N(\omega,\varepsilon')}\left[\max_{\eta^c(\theta_n^o,\varepsilon')} \overline{Q}_n\left(\theta\right) - \overline{Q}_n\left(\theta_n^o\right) + \frac{2\delta'\left(\varepsilon'\right)}{4}\right] \leq -\frac{\delta'\left(\varepsilon'\right)}{2}.
\end{aligned}$$

Set $\varepsilon = 2\varepsilon'$ and $\delta\left(\varepsilon\right) = \frac{\delta'(\varepsilon/2)}{2} > 0$. Since ε' is arbitrary so is ε, and we have

$$\sup_{n\geq N(\omega,\varepsilon/2)}\left[\max_{\eta^c(\hat{\theta}_n(\omega),\varepsilon)} Q_n\left(\omega,\theta\right) - Q_n\left(\omega,\hat{\theta}_n\left(\omega\right)\right)\right] \leq -\delta\left(\varepsilon\right) < 0. \tag{A.1}$$

Because this holds for all ω in F and $P\left(F\right) = 1$, (A.1) holds $a.s. - P$.

If instead $\sup_\Theta \left|Q_n\left(\omega,\theta\right) - \overline{Q}_n\left(\theta\right)\right| = o_P\left(1\right)$, which implies $\hat{\theta}_n\left(\omega\right) - \theta_n^o = o_P\left(1\right)$, then for any subsequence $\{n'\}$ there exists a further subsequence $\{n''\}$ such that $\sup_\Theta \left|Q_{n''}\left(\omega,\theta\right) - \overline{Q}_{n''}\left(\theta\right)\right| = o\left(1\right)$ and $\hat{\theta}_{n''}\left(\omega\right) - \theta_{n''}^o = o\left(1\right)$ for all ω in a set with probability$-P$ equal to one. Now apply the above result to conclude that (A.1) holds for this subsequence $\{n''\}$. ■

Proof of Lemma 3.2. (A) follows by Theorem 3.4 of White (1994) under conditions a1)-a3). To prove (B), pick any subsequence $\{n'\}$. Given a1)-a3), by Lemma 3.1 there exists a further subsequence $\{n''\}$ such that $\left\{\hat{\theta}_{n''}\right\}$ is identifiably unique with respect to $\{Q_{n''}\}$ $a.s. - P$, i.e. for all ω in some $F \in \mathcal{F}$ with $P\left(F\right) = 1$. By b1) and b2) and for all ω in some $G \in \mathcal{F}$ with $P\left(G\right) = 1$, $\left\{Q_{n''}^*\left(\cdot,\omega,\theta\right)\right\}$ is a sequence of random functions on $\left(\Lambda, \mathcal{G}, P_{n,\omega}^*\right)$ continuous on Θ for all $\lambda \in \Lambda$. Hence, by Theorem 2.11 of White (1994) for fixed ω in G there exists $\hat{\theta}_{n''}^*\left(\cdot,\omega\right) : \Lambda \to \Theta$ measurable-$\mathcal{G}$ such that $\hat{\theta}_{n''}^*\left(\cdot,\omega\right) = \arg\max_\Theta Q_{n''}^*\left(\cdot,\omega,\theta\right)$. By b3), $Q_{n''}^*\left(\cdot,\omega,\theta\right) - Q_{n''}\left(\omega,\theta\right) \to 0$ as $n'' \to \infty$ $prob - P_{n,\omega}^*, prob - P$ uniformly on Θ. Hence, there exists a further subsequence $\{n'''\}$ such that $Q_{n'''}^*\left(\cdot,\omega,\theta\right) - Q_{n'''}\left(\omega,\theta\right) \to 0$ as $n''' \to \infty$ $prob - P_{n,\omega}^*$ for all ω in some $H \in \mathcal{F}$ with $P\left(H\right) = 1$. Choose ω in $\mathfrak{F} = F \cap G \cap H$, $\mathfrak{F} \in \mathcal{F}$. By theorem 3.4 of White (1994), $\hat{\theta}_{n'''}^*\left(\cdot,\omega\right) - \hat{\theta}_{n'''}\left(\omega\right) \to 0$ as $n''' \to \infty$ $prob - P_{n,\omega}^*$ for this fixed ω (note that $\left\{\hat{\theta}_{n'''}\right\}$ is identifiably unique for ω in F since $\left\{\hat{\theta}_{n''}\right\}$ is). Because this is true for any subsequence $\{n'\}$ and $P\left(\mathfrak{F}\right) = 1$, the result follows, i.e. $\hat{\theta}_n^*\left(\cdot,\omega\right) - \hat{\theta}_n\left(\omega\right) \to 0$ as $n \to \infty$ $prob - P_{n,\omega}^*, prob - P$. ■

The proof of our QMLE bootstrap consistency results below (Theorem 3.2 and Theorem 3.3) makes use of the following two lemmas. Lemma A.2 provides a convenient approach to establishing a bootstrap uniform weak law of large numbers. It requires a pointwise bootstrap weak law of large numbers, which we establish in Lemma A.3.

Lemma A.2 (Bootstrap Uniform WLLN). *For* $j = 1, 2$, *let* $\left\{q_{nt}^{*(j)}(\cdot, \omega, \theta)\right\}$ *be the MBB and SB resample obtained from* $\{q_{nt}(\omega, \theta), t = 1, \ldots, n\}$, *respectively, and assume the two following conditions hold:*

Bootstrap P-WLLN *For each* $\theta \in \Theta \subset \mathbb{R}^p$, Θ *a compact set,*

$$n^{-1} \sum_{t=1}^{n} \left(q_{nt}^{*(j)}(\cdot, \omega, \theta) - q_{nt}(\omega, \theta)\right) \to 0,\ prob - P_{n,\omega}^{*},\ prob - P.$$

Global Lipschitz $\forall \theta, \theta^{o} \in \Theta$, $|q_{nt}(\cdot, \theta) - q_{nt}(\cdot, \theta^{o})| \leq L_{nt} |\theta - \theta^{o}|$ *a.s.* $- P$, *where* $\sup_n \left\{n^{-1} \sum_{t=1}^{n} E(L_{nt})\right\} = O(1)$.

Then, if $\ell_n \to \infty$ *and* $\ell_n = o(n)$, *for any* $\delta > 0$ *and* $\xi > 0$,

$$\lim_{n \to \infty} P\left[P_{n,\omega}^{*}\left(\sup_{\theta \in \Theta} n^{-1}\left|\sum_{t=1}^{n}\left(q_{nt}^{*(j)}(\cdot, \omega, \theta) - q_{nt}(\omega, \theta)\right)\right| > \delta\right) > \xi\right] = 0.$$

Proof. We follow the idea of the proof of Lemma 8 of Hall and Horowitz (1996). So, given $\varepsilon > 0$ (to be chosen appropriately later), let $\{\eta(\theta_i, \varepsilon) : i = 1, \ldots, I\}$ be a finite subcover of Θ, where $\eta(\theta_i, \varepsilon) \equiv \{\theta \in \Theta : |\theta - \theta_i| < \varepsilon\}$. Then, for $j = 1, 2$,

$$\sup_{\Theta} n^{-1}\left|\sum_{t=1}^{n}\left(q_{nt}^{*(j)}(\theta) - q_{nt}(\theta)\right)\right| = \max_{i} \sup_{\theta \in \eta(\theta_i, \varepsilon)} n^{-1}\left|\sum_{t=1}^{n}\left(q_{nt}^{*(j)}(\theta) - q_{nt}(\theta)\right)\right|,$$

where for simplicity we omit the arguments ω and λ in the notations $q_{nt}^{*(j)}(\theta)$ and $q_{nt}(\theta)$. It follows that for any $\delta > 0$ (and any fixed ω),

$$P_{n,\omega}^{*}\left(\sup_{\theta \in \Theta} n^{-1}\left|\sum_{t=1}^{n}\left(q_{nt}^{*(j)}(\theta) - q_{nt}(\theta)\right)\right| > \delta\right) \leq \sum_{i=1}^{I} P_{n,\omega}^{*}\left(\sup_{\eta(\theta_i, \varepsilon)} n^{-1}\left|\sum_{t=1}^{n}\left(q_{nt}^{*(j)}(\theta) - q_{nt}(\theta)\right)\right| > \delta\right).$$

Now, for $\theta \in \eta(\theta_i, \varepsilon)$,

$$
\begin{aligned}
& n^{-1}\left|\sum_{t=1}^{n}\left(q_{nt}^{*(j)}(\theta)-q_{nt}(\theta)\right)\right| \\
\leq\ & n^{-1}\left|\sum_{t=1}^{n}\left(q_{nt}^{*(j)}(\theta_i)-q_{nt}(\theta_i)\right)\right|+n^{-1}\sum_{t=1}^{n}\left|q_{nt}^{*(j)}(\theta)-q_{nt}^{*(j)}(\theta_i)\right|+n^{-1}\sum_{t=1}^{n}\left|q_{nt}(\theta)-q_{nt}(\theta_i)\right| \\
\leq\ & n^{-1}\left|\sum_{t=1}^{n}\left(q_{nt}^{*(j)}(\theta_i)-q_{nt}(\theta_i)\right)\right|+n^{-1}\sum_{t=1}^{n}L_{nt}^{*(j)}\varepsilon+n^{-1}\sum_{t=1}^{n}L_{nt}\varepsilon,
\end{aligned}
$$

where $L_{nt}\left(L_{nt}^{*(j)}\right)$ is the Lipschitz (resampled Lipschitz) function. By Markov's inequality and because $\left\{E\left(n^{-1}\sum_{t=1}^{n}L_{nt}\right)\right\}=O(1)$, for any $\delta>0$ and $\xi>0$, $P\left[n^{-1}\sum_{t=1}^{n}L_{nt}\varepsilon>\frac{\delta}{3}\right]\leq\frac{3\varepsilon\Delta}{\delta}<\xi/3$, for all n sufficiently large, if we choose $\varepsilon<\frac{\xi\delta}{9\Delta}$, where Δ is a sufficiently large constant. Thus,

$$
\begin{aligned}
& P\left[P_{n,\omega}^{*}\left(\sup_{\theta\in\eta(\theta_i,\varepsilon)}n^{-1}\left|\sum_{t=1}^{n}\left(q_{nt}^{*(j)}(\theta)-q_{nt}(\theta)\right)\right|>\delta\right)>\xi\right] \\
\leq\ & P\left[P_{n,\omega}^{*}\left(n^{-1}\left|\sum_{t=1}^{n}\left(q_{nt}^{*(j)}(\theta_i)-q_{nt}(\theta_i)\right)\right|>\frac{\delta}{3}\right)>\frac{\xi}{3}\right] \\
& +P\left[P_{n,\omega}^{*}\left(n^{-1}\sum_{t=1}^{n}L_{nt}^{*(j)}>\frac{\delta}{3\varepsilon}\right)>\frac{\xi}{3}\right]+P\left[n^{-1}\sum_{t=1}^{n}L_{nt}\varepsilon>\frac{\delta}{3}\right].
\end{aligned}
$$

By the Bootstrap P-WLLN,

$$
P\left[P_{n,\omega}^{*}\left(n^{-1}\left|\sum_{t=1}^{n}\left(q_{nt}^{*(j)}(\theta_i)-q_{nt}(\theta_i)\right)\right|>\frac{\delta}{3}\right)>\frac{\xi}{3}\right]<\frac{\xi}{3},
$$

for all n sufficiently large. For fixed ω, Markov's inequality implies that $P_{n,\omega}^{*}\left(n^{-1}\sum_{t=1}^{n}L_{nt}^{*(j)}>\frac{\delta}{3\varepsilon}\right)\leq n^{-1}\sum_{t=1}^{n}E^{*}\left(L_{nt}^{*(j)}\right)/\frac{\delta}{3\varepsilon}$, since $L_{nt}^{*(j)}$is nonnegative. Now, for $j=1$, $n^{-1}\sum_{t=1}^{n}E^{*}\left(L_{nt}^{*(j)}\right)=\sum_{t=1}^{n}\alpha_n(t)L_{nt}\leq\frac{1}{n-\ell+1}\sum_{t=1}^{n}L_{nt}$, since $\alpha_n(t)\leq\frac{1}{n-\ell+1}$ by definition (3.2). For $j=2$, $n^{-1}\sum_{t=1}^{n}E^{*}\left(L_{nt}^{*(j)}\right)=n^{-1}\sum_{t=1}^{n}L_{nt}\leq\frac{1}{n-\ell+1}\sum_{t=1}^{n}L_{nt}$. Hence, for $j=1,2$, and for n sufficiently large (which implies $\frac{n}{n-\ell+1}\to 1$ given that $\ell=o(n)$), we can make $P\left[P_{n\omega}^{*}\left(n^{-1}\sum_{t=1}^{n}L_{nt}^{*(j)}>\frac{\delta}{3\varepsilon}\right)>\frac{\xi}{3}\right]<\frac{\xi}{3}$, if we choose $\varepsilon<\min\left\{\frac{\delta\xi}{9\Delta},\frac{\delta\xi^2}{27\Delta}\right\}$. This completes the proof. ■

Lemma A.3 (Bootstrap Pointwise WLLN). *For some $r>2$, let $\{q_{nt}:\Omega\times\Theta\to\mathbb{R}\}$ be such that for $n,t=1,2,\ldots$, there exists $D_{nt}:\Omega\to\mathbb{R}$ with $|q_{nt}(\cdot,\theta)|\leq D_{nt}$ for all $\theta\in\Theta$ and $\|D_{nt}\|_r\leq\Delta<\infty$. For each $\theta\in\Theta$ let $\left\{q_{nt}^{*(j)}(\cdot,\omega,\theta)\right\}$ be obtained from $\{q_{nt}(\omega,\theta)\}$ by the MBB and by the SB, for $j=1,2$, respectively. Then, if $\ell_n=o(n)$ and $\ell_n\to\infty$, it follows that for any*

$\delta > 0$, $\xi > 0$, *and for each* $\theta \in \Theta$,

$$\lim_{n\to\infty} P\left[P^*_{n,\omega}\left(n^{-1}\left|\sum_{t=1}^{n}\left(q^{*(j)}_{nt}(\cdot,\omega,\theta) - q_{nt}(\omega,\theta)\right)\right| > \delta\right) > \xi\right] = 0,\ \ j = 1, 2.$$

Proof. For $j = 1, 2$, write

$$n^{-1}\sum_{t=1}^{n}\left(q^{*(j)}_{nt}(\theta) - q_{nt}(\theta)\right) = n^{-1}\sum_{t=1}^{n}\left(q^{*(j)}_{nt}(\theta) - E^*\left(q^{*(j)}_{nt}(\theta)\right)\right)$$
$$+\left[E^*\left(n^{-1}\sum_{t=1}^{n} q^{*(j)}_{nt}(\theta)\right) - n^{-1}\sum_{t=1}^{n} q_{nt}(\theta)\right] \equiv Q^{(j)}_{1n} + Q^{(j)}_{2n}.$$

For $j = 1$, by Lemma A.1 of Fitzenberger (1997), $E^*\left(n^{-1}\sum_{t=1}^{n} q^{*(1)}_{nt}(\theta)\right) = n^{-1}\sum_{t=1}^{n} q_{nt}(\theta) + O_P\left(\frac{\ell}{n}\right)$, which implies $Q^{(1)}_{2n} \xrightarrow{P} 0$ since $\frac{\ell}{n} \to 0$ by assumption. For $j = 2$, $E^*\left(n^{-1}\sum_{t=1}^{n} q^{*(2)}_{nt}(\theta)\right) = n^{-1}\sum_{t=1}^{n} q_{nt}(\theta)$, so that $Q^{(2)}_{2n} \equiv 0$. Thus, it suffices to prove that for any $\delta > 0$ and $\xi > 0$ and n sufficiently large, $P\left[P^*_{n,\omega}\left(\left|Q^{(j)}_{1n}\right| > \delta\right) > \xi\right] < \xi$. But by Chebyshev's inequality, $P^*_{n,\omega}\left(\left|Q^{(j)}_{1n}\right| > \delta\right) \leq \frac{1}{\delta^2} n^{-1} \operatorname{var}^*\left(n^{-1/2}\sum_{t=1}^{n} q^{*(j)}_{nt}(\theta)\right)$, where $\operatorname{var}^*\left(n^{-1/2}\sum_{t=1}^{n} q^{*(j)}_{nt}(\theta)\right)$ has a closed form expression involving products of $q_{nt}(\theta)$ and $q_{n,t+\tau}(\theta)$ (the exact expressions are given by the univariate analogs of (3.1) and (3.4) for $j = 1, 2$, respectively, with s^o_{nt} replaced by $q_{nt}(\theta)$). Under the domination condition on $\{q_{nt}(\theta)\}$ and the properties of the sequences of weights (3.2), (3.3) and (3.5), repeated application of Minkowski's inequality and Hölder's inequality yields $\left\|\operatorname{var}^*\left(n^{-1/2}\sum_{t=1}^{n} q^{*(j)}_{nt}(\theta)\right)\right\|_{P^*_{\omega,n},\frac{r}{2}} = O(\ell)$, where $\|X\|_{P^*_{\omega,n},r} = E^*\left(|X|^r\right)^{1/r}$ for some $r > 2$. Thus, by Markov's inequality, for all n sufficiently large and for $j = 1, 2$, $P\left[P^*_{n,\omega}\left(\left|Q^{(j)}_{1n}\right| > \delta\right) > \xi\right] = O\left(\left(\frac{\ell}{n}\right)^{r/2}\right)$, which completes the proof, given that $\ell = o(n)$ by assumption. ■

Proof of Theorem 3.2. We apply Lemma 3.2 with $Q_n(\cdot,\theta) = n^{-1}\sum_{t=1}^{n} q_{nt}(\cdot,\theta)$ and $Q^{*(j)}_n(\cdot,\omega,\theta) = n^{-1}\sum_{t=1}^{n} q^{*(j)}_{nt}(\cdot,\omega,\theta)$, where we define $q_{nt}(\cdot,\theta) = \log f_{nt}\left(X^{nt}(\cdot),\theta\right)$ and, for $j = 1, 2$, we let $\left\{q^{*(j)}_{nt}(\cdot,\omega,\theta)\right\}$ denote the MBB and SB resamples, respectively. Assumptions A.1, A.2, A.3, A.5.(i), A.6.(i), A.7 and A.8.(i) ensure that conditions a1) through a3) in Lemma 3.2 are verified, which proves (A). The proof is identical to that of Theorem 3.19 of Gallant and White (1988), except that we rely on a generic uniform law of large numbers due to Andrews (1992, Theorem 3) to prove b3) instead of Gallant and White's (1988) Theorem 3.18. This requires that a pointwise law of large numbers and a global Lipschitz condition on Θ holds for $\{q_{nt}(\cdot,\theta)\}$. In particular, our assumptions guarantee that Andrews's (1988) weak law of large numbers for uniformly integrable L_1-mixingales applies to $\{q_{nt}(\cdot,\theta)\}$. To prove (B) we verify the additional conditions b1), b2) and b3) of Lemma 3.2. For each ω in Ω and $n = 1, 2, \ldots$, for $j = 1, 2$, the functions $q^{*(j)}_{nt}(\cdot,\omega,\cdot) : \Lambda \times \Theta \to \mathbb{R}$ are such that $q^{*(j)}_{nt}(\cdot,\omega,\theta)$ is measurable-$\mathcal{G}$ for each $\theta \in \Theta$ (by Lemma A.1) and $q_{nt}^{*(j)}(\lambda, \omega, \cdot)$ is continuous on Θ for all λ, $a.s.-P$ by Assumption A.2. For almost all $\omega \in \Omega$, the existence of $\hat{\theta}_n^{*(j)}(\cdot, \omega)$ as a measurable-$\mathcal{G}$ function follows then by Lemma 2 of Jennrich (1969), for each n. Condition b3) requires a bootstrap uniform weak law of large numbers to apply to $\left\{Q_n^{*(j)}(\cdot, \omega, \theta) - Q_n(\omega, \theta)\right\}$. This follows straightforwardly under our assumptions by Lemma A.2 and Lemma A.3, once we assume $\ell_n = o(n)$ and $\ell_n \to \infty$. ■

Proof of Lemma 3.3. (A) is a restatement of Theorem 6.2 of White (1994, p. 89) under conditions a1) through a4). To prove (B), by the method of subsequences it suffices to show that for an arbitrary subsequence indexed by $\{n'\}$ there exists a further subsequence indexed by $\{n''\}$ such that as $n'' \to \infty$,

$$B_{n''}^{*-1/2} A_{n''}^{*} \sqrt{n''}\left(\hat{\theta}_{n''}^{*}(\cdot, \omega) - \hat{\theta}_{n''}(\omega)\right) \Rightarrow^{d_{P^*_{n'',\omega}}} N(0, I_p), \ a.s. - P,$$

i.e. for all ω in some set F with $P(F) = 1$.

Pick an arbitrary subsequence $\{n'\}$. By a1) through a4), there exists a further subsequence $\{n''\}$, say $\left\{\left(\hat{\theta}_{n''}(\omega) - \theta_{n''}^{o}\right)', vec'\left[\nabla^2 Q_{n''}(\omega, \theta) - A_{n''}(\theta)\right]\right\} \to 0$ for all ω in some $F_1 \in \mathcal{F}$ with $P(F_1) = 1$ uniformly on Θ. Under b1) and b2), for all ω in some $F_2 \in \mathcal{F}$ with $P(F_2) = 1$, $\left\{Q_{n''}^{*}(\cdot, \omega, \cdot)\right\}$ is a sequence of random functions on $\left(\Lambda, \mathcal{G}, P_{n'',\omega}^{*}\right)$ continuously differentiable of order 2 on Θ. Thus, by Lemma 2 of Jennrich (1969) $\hat{\theta}_{n''}^{*}(\cdot, \omega) = \arg\max_{\Theta} Q_{n''}^{*}(\cdot, \omega, \theta)$ for $n'' = 1, 2, \ldots$, and ω in F_2. Under b1) through b4), and for all ω in some $F_3 \in \mathcal{F}$ with $P(F_3) = 1$, the sequence $\left\{\left(\hat{\theta}_{n''}^{*}(\cdot, \omega) - \hat{\theta}_{n''}(\omega)\right)', vec'\left[\nabla^2 Q_{n''}^{*}(\cdot, \omega, \theta) - \nabla^2 Q_{n''}(\omega, \theta)\right]\right\}$ contains a further subsequence, say $\left\{\left(\hat{\theta}_{n'''}^{*}(\cdot, \omega) - \hat{\theta}_{n'''}(\omega)\right)', vec'\left[\nabla^2 Q_{n'''}^{*}(\cdot, \omega, \theta) - \nabla^2 Q_{n'''}(\omega, \theta)\right]\right\}$, which converges to zero in probability-$P_{n''',\omega}^{*}$, and $B_{n'''}^{o-1/2} \sqrt{n'''} \nabla Q_{n'''}^{*}\left(\cdot, \omega, \hat{\theta}_{n'''}(\omega)\right)$ converges in distribution to $N(0, I_p)$ under $P_{n''',\omega}^{*}$.

Define $\mathfrak{F} = F_1 \cap F_2 \cap F_3$ so that $P(\mathfrak{F}) = 1$. For fixed $\omega \in \mathfrak{F}$ and n''' sufficiently large, it follows that $\hat{\theta}_{n'''}(\omega)$ is interior to Θ given that $\hat{\theta}_{n'''}(\omega) - \theta_{n'''}^{o} \to 0$ and given assumption A.3; and $\hat{A}_{n'''}(\omega) \equiv \nabla^2 Q_{n'''}\left(\omega, \hat{\theta}_{n'''}(\omega)\right)$ is $O(1)$ and nonsingular given that for $\omega \in F_1$, $\left|\det \hat{A}_{n'''}(\omega) - \det A_{n'''}^{o}\right| < \frac{\varepsilon}{2}$ for any $\varepsilon > 0$ and all n'''sufficiently large, which implies $\left|\det \hat{A}_{n'''}(\omega)\right| \geq \frac{\kappa}{2} > 0$ given assumption A.10.

For all such n''' and fixed ω in $\mathfrak{F}$, by Theorem 6.2 of White (1994) it follows that

$$\text{(A.2)} \quad \sqrt{n'''}\left(\hat{\theta}_{n'''}^{*}(\cdot, \omega) - \hat{\theta}_{n'''}(\omega)\right) = -\hat{A}_{n'''}(\omega)^{-1} \sqrt{n'''} \nabla Q_{n'''}^{*}\left(\cdot, \omega, \hat{\theta}_{n'''}(\omega)\right) + o_{P_{n''',\omega}^{*}}(1)$$

$$\text{(A.3)} \quad = -A_{n'''}^{o-1} \sqrt{n'''} \nabla Q_{n'''}^{*}\left(\cdot, \omega, \hat{\theta}_{n'''}(\omega)\right) + o_{P_{n''',\omega}^{*}}(1).$$

Given that $\left\{B_{n'''}^{o-1/2}A_{n'''}^{o}\right\}$ is $O(1)$ and that $B_{n'''}^{o-1/2}\sqrt{n'''}\nabla Q_{n'''}^{*}\left(\cdot,\omega,\hat{\theta}_{n'''}(\omega)\right)\Rightarrow^{d_{P^{*}_{n''',\omega}}} N(0,I_p)$, (A.3) implies that $B_{n'''}^{o-1/2}A_{n'''}^{o}\sqrt{n'''}\left(\hat{\theta}_{n'''}^{*}(\cdot,\omega)-\hat{\theta}_{n'''}(\omega)\right)\Rightarrow^{d_{P^{*}_{n''',\omega}}} N(0,I_p)$. The desired result follows because this holds for any subsequence $\{n'\}$ and all ω in $\mathfrak{F}$ with $P(\mathfrak{F})=1$. ■

The proof of our next results makes uses of the following lemma. Part (A) is just a restatement of Corollary 3.8 of White (1994) and part (B) is an extension of this result to the bootstrap context.

Lemma A.4. *Let $\{Q_n:\Omega\times\Theta\rightarrow\mathbb{R}\}$ be a sequence of functions continuous on Θ a.s. $-P$ and let $\left\{\hat{\theta}_n:\Omega\rightarrow\Theta\right\}$ be such that $\hat{\theta}_n-\theta_n^o\rightarrow 0$ prob $-P$ for some nonstochastic sequence $\{\theta_n^o\in\Theta\}$. Suppose that $\sup_{\theta\in\Theta}\left|Q_n(\cdot,\theta)-\overline{Q}_n(\theta)\right|\rightarrow 0$ prob $-P$ where $\left\{\overline{Q}_n:\Theta\rightarrow\mathbb{R}\right\}$ is continuous on Θ uniformly in n. Then,*

(A) $$Q_n\left(\cdot,\hat{\theta}_n(\cdot)\right)-\overline{Q}_n(\theta_n^o)\rightarrow 0 \text{ prob}-P.$$

Let $(\Lambda,\mathcal{G})$ be a complete probability space and for each $\omega\in\Omega$, let $P^{*}_{n,\omega}$ be a probability measure on $(\Lambda,\mathcal{G})$. If $\hat{\theta}_n^{*}(\cdot,\omega)-\hat{\theta}_n(\omega)\rightarrow 0$ prob $-P^{*}_{\omega,n}$, prob $-P$ and $\sup_{\theta\in\Theta}|Q_n^{*}(\cdot,\omega,\theta)-Q_n(\omega,\theta)|\rightarrow 0$ prob $-P^{*}_{n,\omega}$, prob $-P$, then

(B) $$Q_n^{*}\left(\cdot,\omega,\hat{\theta}_n^{*}(\cdot,\omega)\right)-Q_n\left(\omega,\hat{\theta}_n(\omega)\right)\rightarrow 0 \text{ prob}-P^{*}_{n,\omega},\ \text{prob}-P.$$

Proof. (A) is a restatement of Corollary 3.8 in White (1994, p. 32). To prove (B), by the continuity of $\bar{Q}_n$ on Θ uniformly in n, given $\varepsilon>0$ there exists a $\delta(\varepsilon)>0$ independent of n such that $\left|\bar{Q}_n(\theta)-\bar{Q}_n\left(\hat{\theta}_n\right)\right|>\varepsilon/3$ implies $\left|\theta-\hat{\theta}_n\right|>\delta(\varepsilon)$. It follows that

$$\begin{aligned}
&P\left[P^{*}_{n,\omega}\left(\left|Q_n^{*}\left(\cdot,\omega,\hat{\theta}_n^{*}\right)-Q_n\left(\omega,\hat{\theta}_n\right)\right|>\varepsilon\right)>\varepsilon\right]\\
&\leq P\left[P^{*}_{n,\omega}\left(\sup_{\Theta}|Q_n^{*}(\cdot,\omega,\theta)-Q_n(\omega,\theta)|>\varepsilon/3\right)>\varepsilon/3\right]\\
&+P\left[P^{*}_{n,\omega}\left(2\sup_{\Theta}\left|Q_n(\omega,\theta)-\bar{Q}_n(\theta)\right|>\varepsilon/3\right)>\varepsilon/3\right]\\
&+P\left[P^{*}_{n,\omega}\left(\left|\hat{\theta}_n^{*}-\hat{\theta}_n\right|>\delta(\varepsilon)\right)>\varepsilon/3\right]\equiv\xi_1+\xi_2+\xi_3,
\end{aligned}$$

with obvious definitions. By uniform convergence of $Q_n^{*}(\cdot,\omega,\theta)-Q_n(\omega,\theta)$ to zero $\xi_1\rightarrow 0$. Similarly, by uniform convergence of $Q_n(\cdot,\theta)-\bar{Q}_n(\theta)$ to zero, $\xi_2\rightarrow 0$ once we note that $\xi_2\leq P\left(2\sup_{\Theta}\left|Q_n(\omega,\theta)-\bar{Q}_n(\theta)\right|>\varepsilon/3\right)$. Finally, $\xi_3\rightarrow 0$ because $\hat{\theta}_n^{*}(\cdot,\omega)-\hat{\theta}_n(\omega)\rightarrow 0$ *prob* $-P^{*}_{n,\omega}$, *prob* $-P$. ■

Proof of Theorem 3.3. We show that (A) and (B) of Lemma 3.3 apply to the QMLE with the same choices of $Q_n(\cdot,\theta)$ and $Q_n^{*(j)}(\cdot,\omega,\theta)$, $j=1,2$, as in Theorem 3.2. The desired result follows then by Polya's theorem (see e.g. Serfling, 1980, p. 20) since $C_n^o=A_n^{o-1}B_n^oA_n^{o-1}$ is

$O(1)$ by assumption and the normal distribution is everywhere continuous. Part (A) of Lemma 3.3 follows in our particular application by Theorem 5.7 of Gallant and White (1988) given our assumptions A.1 through A.10, which imply conditions a1) through a4).

Next we verify conditions b1) through b4) to obtain the analog of (B) in Lemma 3.3 in the QMLE application. Given assumptions A.1 and A.2, by Lemma A.1 $Q_n^{*(j)}(\cdot,\omega,\theta)$ satisfies the measurability and continuity requirements (i.e. b1) and b2)) of Lemma 3.3. Given assumptions A.5(ii) and A.6(iii) and the conditions on ℓ_n, it follows from Lemmas A.2 and A.3 that a bootstrap uniform weak law of large numbers applies to $\left\{\nabla^2 Q_n^{*(j)}(\cdot,\omega,\theta) - \nabla^2 Q_n(\omega,\theta)\right\}$, which verifies b4). Lastly, we verify condition b3). Add and subtract appropriately to obtain (for any n and any ω)

$$n^{-1/2}\sum_{t=1}^{n} s_{nt}^{*(j)}\left(\cdot,\omega,\hat{\theta}_n\right) - n^{-1/2}\sum_{t=1}^{n} s_{nt}\left(\omega,\hat{\theta}_n\right) = \xi_{1n}^{(j)} + \xi_{2n} + \xi_{3n}^{(j)}, \text{ where}$$

$$\xi_{1n}^{(j)}(\cdot,\omega) = n^{-1/2}\sum_{t=1}^{n}\left(s_{nt}^{*(j)}(\cdot,\omega,\theta_n^o) - s_{nt}(\omega,\theta_n^o)\right);$$

$$\xi_{2n}(\omega) = -n^{-1/2}\sum_{t=1}^{n}\left(s_{nt}\left(\omega,\hat{\theta}_n\right) - s_{nt}(\omega,\theta_n^o)\right); \text{ and}$$

$$\xi_{3n}^{(j)}(\cdot,\omega) = n^{-1/2}\sum_{t=1}^{n}\left(s_{nt}^{*(j)}\left(\cdot,\omega,\hat{\theta}_n\right) - s_{nt}^{*(j)}(\cdot,\omega,\theta_n^o)\right).$$

Consider an arbitrary subsequence n'. Under assumptions A.1-A.10 strengthened by 3.1.a), 3.1b$'$) and 3.2, by Theorem 3.2 of Gonçalves and White (2000a) there exists a further subsequence n'' such that if $\ell_{n''} = o\left(n''^{1/2}\right)$, $B_{n''}^{o-1/2}\xi_{1n''}^{(j)}(\cdot,\omega) \overset{d_{P^*_{n'',\omega}}}{\Rightarrow} N(0,I_p)$, for $j=1,2$, for all ω in some set F_1 with $P(F_1)=1$. Because $\hat{\theta}_n - \theta_n^o \xrightarrow{P} 0$, it follows that $\hat{\theta}_{n''}(\omega) - \theta_{n''}^o \to 0$ and $n''^{-1}\sum_{t=1}^{n''} s_{n''t}\left(\omega,\hat{\theta}_{n''}\right) = 0$ for all n'' sufficiently large and for all ω in F_2 with $P(F_2)=1$ (since $\theta_{n''}^o$ is interior to Θ by assumption). We will show that $\xi_{2n''}(\omega) + \xi_{3n''}^{(j)}(\cdot,\omega) \to 0 \; prob - P^*_{n'',\omega}$ for all ω in some F with $P(F)=1$, which by Lemma 4.7 of White (2000) implies the desired result. F will be taken to be $F_1 \cap F_2 \cap F_3$, where F_3 is defined as the set of ω for which conditions a3), a4), b3) and b4) of Lemma 3.3 are satisfied. By definition, $P(F_3)=1$ along an appropriately defined subsequence, here indexed by n''. Thus, $P(F)=1$. For fixed ω in F we consider the following two mean value expansions

$$
\begin{aligned}
\xi_{2n''}(\omega) &= -n''^{-1}\sum_{t=1}^{n''}\nabla' s_{n''t}\left(\omega,\bar{\theta}_{n''}\right)\sqrt{n''}\left(\hat{\theta}_{n''}(\omega)-\theta^{o}_{n''}\right),\\
\xi^{(j)}_{3n''}(\cdot,\omega) &= n''^{-1}\sum_{t=1}^{n''}\nabla' s^{*(j)}_{n''t}\left(\cdot,\omega,\bar{\theta}^{*}_{n''}\right)\sqrt{n''}\left(\hat{\theta}_{n''}(\omega)-\theta^{o}_{n''}\right),
\end{aligned}
$$

where $\bar{\theta}_{n''}$ and $\bar{\theta}^{*}_{n''}$ are (possibly different) mean values lying between $\hat{\theta}_{n''}$ and $\theta^{o}_{n''}$. Hence

$$
\begin{aligned}
\xi_{2n''}(\omega)+\xi^{(j)}_{3n''}(\cdot,\omega) &= n''^{-1}\sum_{t=1}^{n''}\left(\nabla' s^{*(j)}_{n''t}\left(\cdot,\omega,\bar{\theta}^{*}_{n''}\right)-\nabla' s_{n''t}\left(\omega,\bar{\theta}_{n''}\right)\right)\sqrt{n''}\left(\hat{\theta}_{n''}(\omega)-\theta^{o}_{n''}\right)\\
&\equiv \zeta^{(j)}_{n''}(\cdot,\omega)\sqrt{n''}\left(\hat{\theta}_{n''}(\omega)-\theta^{o}_{n''}\right) \qquad \text{(A.4)}
\end{aligned}
$$

Now, for $j=1,2$, the bootstrap uniform convergence of $\left\{\nabla^2 Q^{*(j)}_{n''}(\cdot,\omega,\theta)-\nabla^2 Q_{n''}(\omega,\theta)\right\}$, the uniform convergence of $\left\{\nabla^2 Q_{n''}(\omega,\theta)-A_{n''}(\theta)\right\}$ and the convergences of $\bar{\theta}_{n''}-\theta^{o}_{n''}$ and $\bar{\theta}^{*}_{n''}-\theta^{o}_{n''}$ to zero together imply by Lemma A.4 that $\zeta^{(j)}_{n''}(\cdot,\omega)\rightarrow 0$ *prob* $-P^{*}_{n'',\omega}$ for all $\omega\in F$. Since $\sqrt{n''}\left(\hat{\theta}_{n''}(\omega)-\theta^{o}_{n''}\right)=O(1)$ on F, it follows that $\xi_{2n''}(\omega)+\xi^{(j)}_{3n''}(\cdot,\omega)\rightarrow 0$ *prob* $-P^{*}_{n'',\omega}$ for $\omega\in F$, $P(F)=1$. This delivers (B) in Lemma 3.3 and completes the proof. ■

Proof of Corollary 3.1. The definition of $\acute{\theta}^{*(j)}_{n}$ and equation (A.2) in the proof of Lemma 3.3 imply $\sqrt{n}\left(\acute{\theta}^{*(j)}_{n}-\hat{\theta}_{n}\right)-\sqrt{n}\left(\hat{\theta}^{*(j)}_{n}-\hat{\theta}_{n}\right)\rightarrow 0$ *prob* $-P^{*}_{n,\omega}$, *prob* $-P$, which delivers the result for $\acute{\theta}^{*(j)}_{n}$. A similar result can be established for $\theta^{\dagger*(j)}_{n}$ since by Lemma A.4 $\hat{A}^{*(j)}_{n}-\hat{A}_{n}\rightarrow 0$ *prob* $-P^{*}_{n,\omega}$, *prob* $-P$. ■

B. Proofs for Section 4

Throughout Appendix B, C will denote a generic constant that might change from one usage to the next. The dependence of the bootstrap variables on ω and on n will also be omitted as it is not relevant for the arguments made here. For instance, $P^{*}\equiv P^{*}_{n,\omega}$ and $X^{*}_{nt}\equiv X^{*}_{nt}(\cdot,\omega)$, where the star denotes resampling under the MBB.

Lemma B.1 (Studentization of the sample mean). *Let $\{X_{nt}\}$ satisfy Assumptions 2.1′ and 2.2 of Gonçalves and White (2000a), where Assumption 2.2′ is strengthened by*

Assumption 2.2′ $n^{-1}\sum_{t=1}^{n}\left|\mu_{nt}-\bar{\mu}_{n}\right|^{2+\delta}=o\left(\ell_n^{-1-\delta/2}\right)$ *for some small $\delta>0$ (i.e. $0<\delta\leq 2$).*

Then, if $\ell_n\rightarrow\infty$ with $\ell_n=o\left(n^{1/2}\right)$ we have that for any $\varepsilon>0$,

$$
\lim_{n\rightarrow\infty}P\left(P^{*}\left(\left|\hat{\sigma}^{*2}_{n,1}-\hat{\sigma}^{2}_{n,1}\right|>\varepsilon\right)>\varepsilon\right)=0,
$$

where $\hat{\sigma}^{2}_{n,1}=\operatorname{var}^{*}\left(\sqrt{n}\bar{X}^{*}_{n}\right)$ and $\hat{\sigma}^{*2}_{n,1}=k^{-1}\sum_{i=1}^{k}\left(\ell^{-1/2}\sum_{t=1}^{\ell}\left(X_{I_i+t}-\bar{X}^{*}_{n}\right)\right)^{2}$.

Proof. The proof consists of two steps:

Step 1: Show that $\tilde{\sigma}^{*2}_{n,1} - \hat{\sigma}^{2}_{n,1} \to 0$ $prob-P^*, prob-P$, where

$$\tilde{\sigma}^{*2}_{n,1} = k^{-1} \sum_{i=1}^{k} \left(\ell^{-1/2} \sum_{t=1}^{\ell} \left(X_{I_i+t} - \bar{X}_{\alpha,n} \right) \right)^2, \text{ with } \bar{X}_{\alpha,n} = E^* \left(\bar{X}^*_n \right); \text{ and}$$

Step 2: Show that $\hat{\sigma}^{*2}_{n,1} - \tilde{\sigma}^{*2}_{n,1} \to 0$ $prob-P^*, prob-P$.

Define $\hat{A}_i = \ell^{-1/2} \sum_{t=1}^{\ell} \left(X_{i+t} - \bar{X}^*_n \right)$ and $A_i = \ell^{-1/2} \sum_{t=1}^{\ell} \left(X_{i+t} - \bar{X}_{\alpha,n} \right)$ so that $\hat{\sigma}^{*2}_{n,1} = k^{-1} \sum_{i=1}^{k} \hat{A}^2_{I_i}$ and $\tilde{\sigma}^{*2}_{n,1} = k^{-1} \sum_{i=1}^{k} A^2_{I_i}$. By Künsch (1989, Theorems 3.1 and 3.4), we can also write $\hat{\sigma}^{2}_{n,1} = (n-\ell+1)^{-1} \sum_{i=0}^{n-\ell} A^2_i$.

To prove step 1, use Markov's inequality to write

$$P^* \left(\left| \hat{\sigma}^{*2}_{n,1} - \hat{\sigma}^{2}_{n,1} \right| > \varepsilon \right) \leq \varepsilon^{-p} E^* \left| \hat{\sigma}^{*2}_{n,1} - \hat{\sigma}^{2}_{n,1} \right|^p \text{ for any } \varepsilon > 0 \text{ and some } p > 1.$$

Given that the I_i are i.i.d. uniform on $\{0, \ldots, n-\ell\}$, $E^* \left(\tilde{\sigma}^{*2}_{n,1} \right) = E^* \left(A^2_{I_1} \right) = (n-\ell+1)^{-1} \sum_{i=0}^{n-\ell} A^2_i \equiv \hat{\sigma}^{2}_{n,1}$, which implies that

$$E^* \left| \tilde{\sigma}^{*2}_{n,1} - \hat{\sigma}^{2}_{n,1} \right|^p = E^* \left| k^{-1} \sum_{i=1}^{k} \left(A^2_{I_i} - E^* \left(A^2_{I_1} \right) \right) \right|^p \leq k^{-p} C E^* \left| \sum_{i=1}^{k} \left(A^2_{I_i} - E^* \left(A^2_{I_1} \right) \right)^2 \right|^{p/2},$$

for some $C < \infty$ by Burkholder's inequality, given that $\left\{ A^2_{I_i} - E^* \left(A^2_{I_1} \right) \right\}$ are i.i.d. zero mean, conditional on the original sample[4]. For $1 < p \leq 2$, $x \geq 0$ and $y \geq 0$, the inequality $(x+y)^{p/2} \leq x^{p/2} + y^{p/2}$ (this is the c_r-inequality with $r \leq 1$, e.g. Davidson, 1994, p. 140) implies that $E^* \left| \sum_{i=1}^{k} \left(A^2_{I_i} - E^* \left(A^2_{I_1} \right) \right)^2 \right|^{p/2} \leq k E^* \left| A^2_{I_1} - E^* \left(A^2_{I_1} \right) \right|^p$ so that

$$E^* \left| \tilde{\sigma}^{*2}_{n,1} - \hat{\sigma}^{2}_{n,1} \right|^p \leq k^{-(p-1)} C E^* \left| A^2_{I_1} - E^* \left(A^2_{I_1} \right) \right|^p \leq 2^p C k^{-(p-1)} E^* \left| A_{I_1} \right|^{2p}.$$

Take $p = 1 + \delta/2$ with $0 < \delta \leq 2$. Then, by another application of Markov's inequality,

$$\begin{aligned} P \left(P^* \left(\left| \hat{\sigma}^{*2}_{n,1} - \hat{\sigma}^{2}_{n,1} \right| > \varepsilon \right) > \varepsilon \right) &\leq P \left(\varepsilon^{-(1+\delta/2)} C k^{-\delta/2} E^* \left| A_{I_1} \right|^{2+\delta} > \varepsilon \right) \\ &\leq \varepsilon^{-(2+\delta/2)} C k^{-\delta/2} E \left(E^* \left| A_{I_1} \right|^{2+\delta} \right). \end{aligned}$$

Next, we show that $k^{-\delta/2} E \left(E^* \left| A_{I_1} \right|^{2+\delta} \right) = o(1)$. We have

[4] More precisely, we use an extension of Burkholder's inequality to m.d.s. triangular arrays due to Chen and White (1994) since $\left\{ A^2_{I_i} - E^* \left(A^2_{I_1} \right) \right\}$ is in fact a m.d.s array with respect to $\mathcal{F}^i = \sigma \left(I_1, \ldots, I_i \right)$.

$$E\left(E^*\left|A_{I_1}\right|^{2+\delta}\right) = (n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}\ell^{-1-\delta/2}E\left(\left|\sum_{t=1}^{\ell}X_{n,i+t}-\ell\bar{X}_{\alpha,n}\right|^{2+\delta}\right)$$

$$= (n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}\ell^{-1-\delta/2}E\left(\left|\sum_{t=1}^{\ell}Z_{n,i+t}+\sum_{t=1}^{\ell}\left(\mu_{n,i+t}-\bar{\mu}_{\alpha,n}\right)-\ell\bar{Z}_{\alpha,n}\right|^{2+\delta}\right) \leq F_1+F_2+F_3,$$

where $Z_{nt} \equiv X_{nt}-\mu_{nt}$, $\bar{\mathcal{Y}}_{\alpha,n} = (n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}\ell^{-1}\sum_{t=1}^{\ell}\mathcal{Y}_{n,i+t} = \sum_{t=1}^{n}\alpha_{nt}\mathcal{Y}_{n,t}$ for any $\{\mathcal{Y}_t\}$ and where, apart from a multiplicative constant,

$$\begin{aligned}
F_1 &= (n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}\ell^{-1-\delta/2}E\left(\left|\sum_{t=1}^{\ell}Z_{n,i+t}\right|^{2+\delta}\right); \\
F_2 &= (n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}\ell^{-1-\delta/2}\left|\sum_{t=1}^{\ell}\left(\mu_{n,i+t}-\bar{\mu}_{\alpha,n}\right)\right|^{2+\delta}; \text{ and} \\
F_3 &= (n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}\ell^{-1-\delta/2}E\left(\left|\ell\bar{Z}_{\alpha,n}\right|^{2+\delta}\right).
\end{aligned}$$

By an extension of a maximal inequality for $L_{2+\delta}-$mixingales to a triangular array setting (e.g. Lemma 1 of Hansen, 1991; see also Hansen, 1992b), $E\left|\sum_{t=1}^{\ell}Z_{n,i+t}\right|^{2+\delta} < C\ell^{1+\delta/2}$ under Assumption 2.1$'$, which implies that $\left|k^{-\delta/2}F_1\right| \leq Ck^{-\delta/2}\ell^{-1-\delta/2}\ell^{1+\delta/2} = O\left(\left(\frac{\ell}{n}\right)^{\delta/2}\right) = o(1)$ given that $\ell = o\left(n^{1/2}\right)$. Similarly, $E\left(\left|\ell\bar{Z}_{\alpha,n}\right|^{2+\delta}\right) \leq (n-\ell+1)^{-(1+\delta)}C\ell^{1+\delta/2} \leq C\ell^{1+\delta/2}$ implying that $k^{-\delta/2}F_3 = O\left(\left(\frac{\ell}{n}\right)^{\delta/2}\right) = o(1)$. If $\mu_{nt} = \mu$ for all t, $F_2 = 0$ because $\bar{\mu}_{\alpha,n} = \sum_{t=1}^{n}\alpha_{nt}\mu = \mu$ since $\sum_{t=1}^{n}\alpha_{nt} = 1$. Otherwise, we can show that

$$\begin{aligned}
F_2 &\leq \ell^{1+\delta/2}(n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}\ell^{-1}\sum_{t=1}^{\ell}\left|\mu_{n,i+t}-\bar{\mu}_{\alpha,n}\right|^{2+\delta} = \ell^{1+\delta/2}\sum_{t=1}^{n}\alpha_{nt}\left|\mu_{n,t}-\bar{\mu}_{\alpha,n}\right|^{2+\delta} \\
&\leq \frac{n}{n-\ell+1}\ell^{1+\delta/2}n^{-1}\sum_{t=1}^{n}\left|\mu_{n,t}-\bar{\mu}_{\alpha,n}\right|^{2+\delta},
\end{aligned}$$

since $0 < \alpha_{nt} \leq \frac{1}{n-\ell+1}$. Routine calculations show that Assumption 2.2$'$ is sufficient to yield $F_2 = o(1)$, and thus $k^{-\delta/2}F_2 = o(1)$, given that $\ell = o\left(n^{1/2}\right)$. In fact, the weaker condition $n^{-1}\sum_t\left|\mu_{nt}-\bar{\mu}_n\right|^{2+\delta} = o\left(\ell_n^{-1}\right)$ is sufficient for the result. This condition may be stronger or weaker than Assumption 2.2 depending on whether $\left|\mu_{nt}-\bar{\mu}_n\right| > 1$ or < 1. Assumption 2.2$'$ is slightly stronger than both of these other conditions and it is adopted for simplicity of the presentation of the results.

To prove step 2, notice that we can write $\hat{A}_{I_i} = \sqrt{\ell}\left(\bar{X}_{I_i}-\bar{X}_n^*\right)$, where $\bar{X}_{I_i} = \ell^{-1}\sum_{t=1}^{\ell}X_{I_i+t}$, and $A_i = \sqrt{\ell}\left(\bar{X}_{I_i}-\bar{X}_{\alpha,n}\right)$. Thus, with probability approaching one,

$$\hat{\sigma}_{n,1}^{*2} - \tilde{\sigma}_{n,1}^{*2} = k^{-1} \sum_{i=1}^{k} \left(\ell \left(\bar{X}_{I_i} - \bar{X}_n^* \right)^2 - \ell \left(\bar{X}_{I_i} - \bar{X}_{\alpha,n} \right)^2 \right) = -\ell \left(\bar{X}_n^* - \bar{X}_{\alpha,n} \right)^2 = O_{P^*} \left(\frac{\ell}{n} \right) \to 0,$$

since under our assumptions $\sqrt{n} \left(\bar{X}_n^* - \bar{X}_{\alpha,n} \right) \Rightarrow^{d_{P^*}} N(0,1)$ *prob*$-P$ by Theorem 2.2 of Gonçalves and White (2000a). This completes the proof of Lemma B.1. ■

Lemma B.2. *Let* $\{X_{nt}\}$ *and* $\{Z_{nt}\}$ *satisfy* $\|X_{nt}\|_{2+\delta} \leq \Delta$ *and* $\|Z_{nt}\|_{2+\delta} \leq \Delta$, $t = 1, \ldots, n$, $n = 1, 2, \ldots$, *for any* $0 < \delta \leq 2$ *and* $\Delta < \infty$. *Let* $k = n/\ell$. *If* $\{I_i\}_{i=1}^k$ *are i.i.d. uniform on* $\{0, \ldots, n - \ell\}$ *and if* $\ell_n \to \infty$ *and* $\ell_n = o\left(n^{1/2}\right)$, *then for any* $\varepsilon > 0$,

$$\lim_{n\to\infty} P \left(P^* \left(\left| k^{-1} \sum_{i=1}^{k} \ell^{-1} \sum_{t=1}^{\ell} X_{n,I_i+t} \sum_{t=1}^{\ell} Z_{n,I_i+t} \right| > n^{1/2} \varepsilon \right) > \varepsilon \right) = 0.$$

Proof. Let $S_{n,i}^1 = \sum_{t=1}^{\ell} X_{n,i+t}$ and $S_{n,i}^2 = \sum_{t=1}^{\ell} Z_{n,i+t}$. By Markov's inequality, for some $1 < p \leq 2$

$$P^* \left(\left| k^{-1} \sum_{i=1}^{k} \ell^{-1} S_{n,I_i}^1 S_{n,I_i}^2 \right| > n^{1/2} \varepsilon \right) \leq \varepsilon^{-p} n^{-p/2} E^* \left(\left| k^{-1} \sum_{i=1}^{k} \ell^{-1} S_{n,I_i}^1 S_{n,I_i}^2 \right|^p \right).$$

Adding and subtracting appropriately yields

$$\begin{aligned} E^* \left(\left| k^{-1} \sum_{i=1}^{k} \ell^{-1} S_{n,I_i}^1 S_{n,I_i}^2 \right|^p \right) \leq C \Bigg\{ & E^* \left(\left| k^{-1} \sum_{i=1}^{k} \ell^{-1} \left(S_{n,I_i}^1 S_{n,I_i}^2 - E^* \left(S_{n,I_i}^1 S_{n,I_i}^2 \right) \right) \right|^p \right) \\ & + E^* \left(\left| k^{-1} \sum_{i=1}^{k} \ell^{-1} E^* \left(S_{n,I_i}^1 S_{n,I_i}^2 \right) \right|^p \right) \Bigg\} \equiv F_1 + F_2. \end{aligned}$$

By the Burkholder and c_r-inequalities (with $r = p/2 \leq 1$)

$$F_1 \leq C \ell^{-p} k^{-p} E^* \left[\left(\sum_{i=1}^{k} \left| S_{n,I_i}^1 S_{n,I_i}^2 - E^* \left(S_{n,I_i}^1 S_{n,I_i}^2 \right) \right|^2 \right)^{p/2} \right] \leq C \ell^{-p} k^{-(p-1)} E^* \left| S_{n,I_1}^1 S_{n,I_1}^2 \right|^p,$$

given that $\left\{ S_{n,I_i}^1 S_{n,I_i}^2 - E^* \left(S_{n,I_i}^1 S_{n,I_i}^2 \right) \right\}$ are i.i.d. with zero mean. Repeated applications of the c_r-inequality and Jensen's inequality yield $F_2 \leq C \ell^{-p} E^* \left| S_{n,I_1}^1 S_{n,I_1}^2 \right|^p$. Take $p = 1 + \delta/2$ with $0 < \delta \leq 2$. Then,

$$\begin{aligned} E^* \left(\left| k^{-1} \sum_{i=0}^{k} \ell^{-1} S_{n,I_i}^1 S_{n,I_i}^2 \right|^{1+\delta/2} \right) & \leq C \ell^{-(1+\delta/2)} \left(k^{-\delta/2} E^* \left| S_{n,I_1}^1 S_{n,I_1}^2 \right|^{1+\delta/2} + E^* \left| S_{n,I_1}^1 S_{n,I_1}^2 \right|^{1+\delta/2} \right) \\ & \leq C \ell^{-(1+\delta/2)} E^* \left| S_{n,I_1}^1 S_{n,I_1}^2 \right|^{1+\delta/2}, \end{aligned}$$

because $k^{-\delta/2} = \left(\frac{\ell}{n}\right)^{\delta/2} \leq 1$. By the implication rule and Markov's inequality, it follows that

$$P\left(P^*\left(\left|k^{-1}\sum_{i=0}^{k}\ell^{-1}S^1_{n,I_i}S^2_{n,I_i}\right|>n^{1/2}\varepsilon\right)>\varepsilon\right)\leq\varepsilon^{-(2+\delta/2)}n^{-(1+\delta/2)/2}\ell^{-(1+\delta/2)}E\left(E^*\left|S^1_{n,I_1}S^2_{n,I_1}\right|^{1+\delta/2}\right).$$

An application of the Cauchy-Schwartz and Minkowski inequalities implies

$$\begin{aligned}E\left(E^*\left|S^1_{n,I_1}S^2_{n,I_1}\right|^{1+\delta/2}\right) &= (n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}E\left(\left|S^1_iS^2_i\right|^{1+\delta/2}\right)\\ &= (n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}\left\|S^1_i\right\|^{1+\delta/2}_{2+\delta}\left\|S^2_i\right\|^{1+\delta/2}_{2+\delta}\leq\Delta^{2+\delta}\ell^{2+\delta}.\end{aligned}$$

Thus, $P\left(P^*\left(\left|k^{-1}\sum_{i=0}^{k}\ell^{-1}S^1_{n,I_i}S^2_{n,I_i}\right|>n^{1/2}\varepsilon\right)>\varepsilon\right)=O\left(\left(\frac{\ell}{n^{1/2}}\right)^{1+\delta/2}\right)\rightarrow 0$ given that $\ell = o\left(n^{1/2}\right)$. ■

Proof of Theorem 4.1. By Theorem 7.5 of Gallant and White (1988), $\mathcal{W}_n\Rightarrow\mathcal{X}^2_q$ under H_o. Thus, it suffices to prove that $\mathcal{W}^*_n\Rightarrow^{d_{P^*}}\mathcal{X}^2_q$ with probability approaching one. By a mean value expansion of $r_n\left(\hat{\theta}^*_n\right)$ around $\hat{\theta}_n$, it follows that $\sqrt{n}\left(r_n\left(\hat{\theta}^*_n\right)-r_n\left(\hat{\theta}_n\right)\right)\Rightarrow^{d_{P^*}}N\left(0,R^o_nC^o_nR^{o\prime}_n\right)$ $prob-P$, which implies that $n\left(\hat{r}^*_n-\hat{r}_n\right)'\left(R^o_nC^o_nR^{o\prime}_n\right)^{-1}\left(\hat{r}^*_n-\hat{r}_n\right)\Rightarrow^{d_{P^*}}\mathcal{X}^2_q$ $prob-P$. Thus, it suffices to prove: (i) $\hat{R}^*_n-R^o_n\rightarrow 0$ $prob-P^*$, $prob-P$; (ii) $\hat{A}^*_n-A^o_n\rightarrow 0$ $prob-P^*$, $prob-P$; and (iii) $\hat{B}^*_n-B^o_n\rightarrow 0$ $prob-P^*$, $prob-P$.

(i) follows by continuity of r_n on Θ (uniformly in n) and because $\hat{\theta}^*_n-\theta^o_n\rightarrow 0$ $prob-P^*$, $prob-P$ by Theorem 3.2; similarly, by Theorem 3.2 and Lemma A.4, we have that $\hat{A}^*_n-\hat{A}_n\rightarrow 0$ $prob-P^*$, $prob-P$, which together with the fact that $\hat{A}_n-A^o_n\rightarrow 0$ $prob-P$ implies (ii).

To prove (iii), consider the infeasible estimator based on θ^o_n,

$$\tilde{B}^{*o}_n=k^{-1}\sum_{i=1}^{k}\left(\ell^{-1/2}\sum_{t=1}^{\ell}\left(s_{n,I_i+t}\left(X^{I_i+t}_n,\theta^o_n\right)-\bar{s}^{*o}_n\right)\right)\left(\ell^{-1/2}\sum_{t=1}^{\ell}\left(s_{n,I_i+t}\left(X^{I_i+t}_n,\theta^o_n\right)-\bar{s}^{*o}_n\right)\right)'$$

$$=k^{-1}\sum_{i=1}^{k}\ell^{-1}\sum_{t=1}^{\ell}s_{n,I_i+t}\left(X^{I_i+t}_n,\theta^o_n\right)\sum_{t=1}^{\ell}s'_{n,I_i+t}\left(X^{I_i+t}_n,\theta^o_n\right)-\ell\bar{s}^{*o}_n\bar{s}^{*o\prime}_n, \tag{B.1}$$

where $\bar{s}^{*o}_n=n^{-1}\sum_{t=1}^{n}s^*_{nt}\left(\theta^o_n\right)$. By Lemma B.1 applied to $\left\{\lambda's_{nt}\left(X^t_n,\theta^o_n\right)\right\}$ with an arbitrary nonzero fixed $\lambda\in\mathbb{R}^p$, it follows that $\tilde{B}^*_n-B^o_{n,1}\stackrel{P^*}{\rightarrow}0$, $prob-P$, where $B^o_{n,1}=\text{var}^*\left(n^{-1/2}\sum_{t=1}^{n}s^*_{nt}\left(\theta^o_n\right)\right)$. By Theorem 3.1 of Gonçalves and White (2000a), $B^o_{n,1}-B^o_n\rightarrow 0$, $prob-P$, which implies $\tilde{B}^*_n-B^o_n\stackrel{P^*}{\rightarrow}0$, $prob-P$. Therefore, it suffices to show that $\hat{B}^*_n-\tilde{B}^*_n\stackrel{P^*}{\rightarrow}0$, $prob-P$. From (4.1) and (B.1) we can write

$$\hat{B}^*_n-\tilde{B}^*_n=D_1+D_2,\text{ where}$$

$$D_1 \equiv k^{-1} \sum_{i=1}^{k} \ell^{-1} \left[\sum_{t=1}^{\ell} s_{n,I_i+t} \left(X_n^{I_i+t}, \hat{\theta}_n^* \right) \sum_{t=1}^{\ell} s'_{n,I_i+t} \left(X_n^{I_i+t}, \hat{\theta}_n^* \right) - \sum_{t=1}^{\ell} s_{n,I_i+t} \left(X_n^{I_i+t}, \theta_n^o \right) \sum_{t=1}^{\ell} s'_{n,I_i+t} \left(X_n^{I_i+t}, \theta_n^o \right) \right], \quad \text{and}$$

$$D_2 \equiv \ell \bar{s}_n^{*o} \bar{s}_n^{*o\prime}.$$

Consider D_2 first. We can write $\bar{s}_n^{*o} = B_n^{o1/2} B_n^{o-1/2} \left(\bar{s}_n^{*o} - \bar{s}_n^o \right) + B_n^{o1/2} B_n^{o-1/2} \bar{s}_n^o \equiv E_1 + E_2$, where $\bar{s}_n^o = n^{-1} \sum_{t=1}^n s_{nt} \left(\theta_n^o \right)$. By Theorem 3.2 of Gonçalves and White (2000a), $B_n^{o-1/2} \sqrt{n} \left(\bar{s}_n^{*o} - \bar{s}_n^o \right) \overset{d_{P^*}}{\Rightarrow} N \left(0, I_p \right)$ $prob - P$ so that $E_1 = O_{P^*} \left(n^{-1/2} \right)$ with probability approaching one. By the CLT for $\{ s_{nt}^o \}$ and noticing that $E \left(\bar{s}_n^o \right) = 0$ by the F.O.C. that define θ_n^o, it follows that $E_2 = O_P \left(n^{-1/2} \right)$. This implies that $D_2 \overset{P^*}{\to} 0$ $prob - P$, since $\ell = o \left(n^{1/2} \right)$.

Next, consider D_1. For $J = 1, \ldots, p$, let $S_{n,I_i}^J \left(\theta \right) \equiv \sum_{t=1}^{\ell} s_{n,I_i+t,J} \left(X_n^{I_i+t}, \theta \right)$, where J indexes the J^{th} element of the score. A mean value expansion of $k^{-1} \sum_{i=1}^k \ell^{-1} S_{n,I_i}^{j_1} \left(\hat{\theta}_n^* \right) S_{n,I_i}^{j_2} \left(\hat{\theta}_n^* \right)$ around θ_n^o yields for a typical element (j_1, j_2) of D_1

$$D_1^{(j_1,j_2)} = k^{-1} \sum_{i=1}^{k} \ell^{-1} \frac{\partial}{\partial \theta'} \left(S_{n,I_i}^{j_1} \left(\ddot{\theta}_n^* \right) S_{n,I_i}^{j_2} \left(\ddot{\theta}_n^* \right) \right) \left(\hat{\theta}_n^* - \theta_n^o \right),$$

where $\ddot{\theta}_n^*$ lies between $\hat{\theta}_n^*$ and θ_n^o. It follows that

$$\left| D_1^{(d_1,d_2)} \right| \leq n^{-1/2} k^{-1} \sum_{i=1}^{k} \ell^{-1} \left\{ \sum_{t=1}^{\ell} \mathcal{Z}_{n,I_i+t}^{j_1} \sum_{t=1}^{\ell} \mathcal{Y}_{n,I_i+t}^{j_2} + \sum_{t=1}^{\ell} \mathcal{Z}_{n,I_i+t}^{j_2} \sum_{t=1}^{\ell} \mathcal{Y}_{n,I_i+t}^{j_1} \right\} \left| \sqrt{n} \left(\hat{\theta}_n^* - \theta_n^o \right) \right|,$$

where $|x| = \left(x'x \right)^{1/2}$, $\mathcal{Z}_{n,I_i+t}^{j} = \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta'} s_{n,I_i+t,j} \left(X_n^{I_i+t}, \theta \right) \right|$, and $\mathcal{Y}_{n,I_i+t}^{j} = \sup_{\theta \in \Theta} \left| s_{n,I_i+t,j} \left(X_n^{I_i+t}, \theta \right) \right|$, $j = \{ j_1, j_2 \}$. Given Assumption A.6 we can apply Lemma B.2 twice with $X_{nt} = \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta'} s_{nt,j} \left(X_n^t, \theta \right) \right|$ and $Z_{nt} = \sup_{\theta \in \Theta} \left| s_{nt,j} \left(X_n^t, \theta \right) \right|$ to obtain that

$$n^{-1/2} k^{-1} \sum_{i=1}^{k} \ell^{-1} \left\{ \sum_{t=1}^{\ell} \mathcal{Z}_{n,I_i+t}^{j_1} \sum_{t=1}^{\ell} \mathcal{Y}_{n,I_i+t}^{j_2} + \sum_{t=1}^{\ell} \mathcal{Z}_{n,I_i+t}^{j_2} \sum_{t=1}^{\ell} \mathcal{Y}_{n,I_i+t}^{j_1} \right\} \to 0 \ prob - P^*, \ prob - P.$$

By Theorem 3.2, with probability approaching one $\sqrt{n} \left(\hat{\theta}_n^* - \theta_n^o \right) = O_{P^*}(1)$, which delivers the result. ■

References

[1] ANDREWS, D.W.K. (1988): "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables," *Econometric Theory,* 4, 458-467.

[2] ———— (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica,* 59, 817-858.

[3] ———— (1999): "Higher-order Improvements of a Computationally Attractive k-step Bootstrap for Extremum Estimators," *Cowles Foundation for Research in Economics*, Yale University.

[4] Andrews, D.W.K., and J.C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica,* 60, 953-966.

[5] Bickel, P.J., and D.A. Freedman (1981): "Some Asymptotic Theory for the Bootstrap," *Annals of Statistics,* 9, 1196-1217.

[6] Billingsley, P. (1995): *Probability and Measure*, New York: Wiley.

[7] Chen, X., and H. White (1994): "Laws of Large Numbers for Hilbert Space-Valued Mixingales with Application to Recursive Nonparametric Estimation," UCSD Discussion Paper.

[8] Davison, A.C., and P. Hall (1993): "On Studentizing and Blocking Methods for Implementing the Bootstrap with Dependent Data," *Australian Journal of Statistics*, 35, 215-224.

[9] Davidson, J. (1994): *Stochastic limit theory.* Oxford: Oxford University Press.

[10] Davidson, R., and J.G. MacKinnon (1999): "Bootstrap Testing in Nonlinear Models," *International Economic Review,* 40, 487-508.

[11] de Jong, R.M., and J. Davidson (2000): "Consistency of Kernel Estimators of Heteroskedastic and Autocorrelation Covariance Matrices," *Econometrica,* 68, 407-423.

[12] Den Haan, W.J., and A. Levin (1997): "A Practitioner's Guide to Robust Covariance Matrix Estimation," *Handbook of Statistics,* 15, Chapter 12, 291-341.

[13] Domowitz, I., and H. White (1982): "Misspecified Models with Dependent Observations," *Journals of Econometrics,* 20, 35-58.

[14] Efron, B. (1979): "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics,* 7, 1-26.

[15] Efron, B., and R.J. Tibshirani (1986): "Bootstrap Methods for Standard errors, Confidence Intervals and Other Measures of Statistical Accuracy (with discussion)," *Statistica Sinica,* 1, 54-77.

[16] Eicker, F. (1967): "Limit Theorems for Regressions with Unequal and Dependent Errors", in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1,* ed. by LeCam, L.M. and J. Neyman. Berkeley: University of California, Berkeley.

[17] Fitzenberger, B. (1997): "The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions," *Journal of Econometrics,* 82, 235-287.

[18] Freedman, D.A. (1981): "Bootstrapping Regression Models," *Annals of Statistics,* 9, 1218-1228.

[19] ———— (1984): "On Bootstrapping Two-stage Least Squares Estimates in Stationary Linear Models," *Annals of statistics,* 12, 827-842.

[20] Gallant, A.R., and H. White (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. New York: Basil Blackwell.

[21] GINÉ, E., and J. Zinn (1989): "Necessary Conditions for the Bootstrap of the Mean," *Annals of Statistics,* 17, 684-691.

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

[23] GONÇALVES, S., and H. White (2000a): "The Bootstrap of the Mean for Dependent Heterogeneous Arrays," UCSD, manuscript.

[24] ——— (2000b): "Bootstrap Variance Estimation for Smooth Functions of Sample Means and for Linear Regressions," UCSD, manuscript.

[25] GÖTZE, F., and H.R. KÜNSCH (1996): "Second-order Correctness of the Blockwise Bootstrap for Stationary Observations," *Annals of Statistics*, 24, 1914-1933.

[26] Hahn, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators," *Econometric Theory,* 12, 187-197.

[27] Hall, P., and J. Horowitz (1996): "Bootstrap Critical Values for Tests based on Generalized-Method-of-Moments Estimators," *Econometrica,* 64, 891-916.

[28] Hansen, B.E. (1991): "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory,* 7, 213-221.

[29] ——— (1992a): "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica,* 60, 967-972.

[30] ——— (1992b): "Strong Laws for Dependent Heterogeneous Processes. Erratum," *Econometric Theory, 8, 421-422.*

[31] Jennrich, R.I. (1969): "Asymptotic Properties of Nonlinear Least Squares Estimators," *Annals of Mathematical Statistics,* 40, 633-643.

[32] KÜNSCH, H.R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics,* 17, 1217-1241.

[33] Lahiri, S.N. (1996): "On Edgeworth Expansion and Moving Block Bootstrap for Studentized M-Estimators in Multiple Linear Regression Models," *Journal of Multivariate Analysis,* 56, 42-59.

[34] ——— (1999a): "Theoretical Comparisons of Block Bootstrap Methods," *Annals of Statistics,* 27, 386-404.

[35] ——— (1999b): "On Second-order Properties of the Stationary Bootstrap Method for Studentized Statistics," in *Asymptotics, Nonparametrics, and Time Series*, ed. by S. Ghosh. Mercel-Dekker.

[36] Liu, R.Y. (1988): "Bootstrap Procedures Under Some Non-iid Models," *Annals of Statistics,* 16, 1696-1708.

[37] Liu, R.Y., and K. Singh (1992) "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of the Bootstrap,* ed. by R. LePage and L. Billiard. New York: Wiley.

[38] Newey, W.K., and K.D. West (1987): "A Simple Positive Semi-definite, Heteroskedastic and Autocorrelation Consistent Covariance Matrix," *Econometrica,* 55, 703-708.

[39] Politis, D., and J. Romano (1994a): "The Stationary Bootstrap," *Journal of American Statistics Association*, 89, 1303-1313.

[40] ———— (1994b): "Limit Theorems for Weakly Dependent Hilbert Space Valued Random Variables with Application to the Stationary Bootstrap," *Statistica Sinica,* 4, 461-476.

[41] PÖTSHER, B.M., and I.R. Prucha (1991): "Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, part I: Consistency and Approximation Concepts," *Econometric Reviews,* 10, 125-216.

[42] Serfling, R.J. (1980): *Approximation theorems of mathematical statistics.* New York: Wiley.

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

[44] Shao, J., and D. Tu (1995): *The Jackknife and Bootstrap.* New York: Springer-Verlag.

[45] Singh, K. (1981): "On the Asymptotic Accuracy of Efron's Bootstrap," *Annals of Statistics,* 9, 1187-1195.

[46] White, H. (1980) "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica,* 48, 817-838.

[47] ———— (1994) *Estimation, Inference and Specification Analysis.* Cambridge: Cambridge University Press.

[48] ———— (2000) *Asymptotic Theory for Econometricians.* Orlando: Academic Press.

[49] White, H., and I. Domowitz (1984) "Nonlinear Regression with Dependent Observations," *Econometrica,* 52, 143-161.