Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models*

Sílvia Gonçalves
CIRANO, CIREQ and Département de sciences économiques, Université de Montréal†

Halbert White
University of California, San Diego‡

July 26, 2003

Abstract

We provide a unified framework for analyzing bootstrapped extremum estimators of nonlinear dynamic models for heterogeneous dependent stochastic processes. We apply our results to the moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992) and prove the first order asymptotic validity of the bootstrap approximation to the true distribution of quasi-maximum likelihood estimators. We also consider bootstrap testing. In particular, we prove the first order asymptotic validity of the bootstrap distribution of suitable bootstrap analogs of Wald and Lagrange Multiplier statistics for testing hypotheses.

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

JEL codes: C15, C22.

*We would like to thank Rui Castro and Nour Meddahi for many helpful discussions, and Jeff Wooldridge for suggesting the logit example. In addition, we are grateful to numerous seminar participants and to two anonymous referees for many valuable suggestions.

†C.P.6128, succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada. Tel: (514) 343 6556. Email: silvia.goncalves@umontreal.ca. Gonçalves acknowledges financial support from Praxis XXI and a Sloan Dissertation fellowship.

‡9500 Gilman Drive, La Jolla, California, 92093-0508. USA . Tel: (858) 534-3502. Email: hwhite@weber.ucsd.edu. White's participation has been supported by the National Science Foundation, grants SBR-9811562 and SES 0111238.

1. Introduction

The bootstrap is a powerful and increasingly utilized method for obtaining confidence intervals and performing statistical inference. Despite this, results validating the bootstrap for the quasi-maximum likelihood estimator (QMLE) or generalized method of moments (GMM) estimator have previously been available only under restrictive assumptions, such as stationarity and limited memory. A main goal of this paper is thus to establish the bootstrap's first order asymptotic validity in the framework of 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 treat primarily QML estimators for concreteness and because there are fewer results in this area.

We apply our results to the moving blocks bootstrap (MBB) of Künsch (1989) and Liu and Singh (1992). Similar results hold for the stationary bootstrap of Politis and Romano (1994) (see Gonçalves and White, 2000). In the context of (possibly misspecified) nonlinear dynamic models, the bootstrap QMLE involves resampling blocks of the quasi-log-likelihood values. With misspecified models, the associated scores are generally dependent, justifying our use of block bootstrap methods.

Results for bootstrapping extremum estimators are available for special cases. For example, Hahn (1996) shows first order asymptotic validity of Efron's bootstrap for GMM with i.i.d. data. Hall and Horowitz (1996) give asymptotic refinements for bootstrapped GMM estimators with stationary ergodic data. Andrews (2002) extends their results, establishing higher-order improvements of k-step bootstrap estimators (see Davidson and MacKinnon (1999)) for nonlinear extremum estimators, including GMM and ML. Both Hall and Horowitz (1996) and Andrews (2002) take the moment conditions defining the estimator to be uncorrelated after finitely many lags, obviating use of HAC covariance estimators. For stationary mixing processes, Inoue and Shintani (2001) prove asymptotic refinements for GMM applied to linear models where the defining moment conditions have unknown covariance.

Here, we do not attempt to prove asymptotic refinements. Instead, we prove the consistency of the block bootstrap estimator of the QMLE sampling distribution for a broad class of models and data generating processes. Specifically, we avoid stationarity and restrictive memory conditions, and show that the block bootstrap distribution of the QMLE converges weakly to the distribution of the QMLE. Thus, bootstrap confidence intervals have correct asymptotic coverage probability.

An important bootstrap application is hypothesis testing. We show first order asymptotic validity for new bootstrap Wald and LM tests. The asymptotic validity of the percentile-t test follows from that of the Wald test, justifying use of MBB to construct percentile-t confidence intervals.

Another important application of the bootstrap is to Kolmogorov type conditional distribution tests in the presence of dynamic misspecification and parameter estimation error. The limiting distribution of such tests is typically not nuisance free, implying that critical values cannot be tabulated. The bootstrap offers a suitable alternative for computing critical values. Recently, Corradi and Swanson (2003a,b) have established the first-order asymptotic validity of the block bootstrap in this context, relying on the consistency results presented here.

We illustrate MBB finite sample performance for confidence intervals via two Monte Carlo experiments. Specifically, we compute confidence intervals for 1) a logit model with neglected autocorrelation, and 2) a possibly misspecified ARCH(1) model. In both cases the MBB outperforms standard asymptotics, especially when robustness to autocorrelated scores is needed.

2. Consistency of the Bootstrap QMLE

We adopt the framework of Gallant and White (1988). The goal is to conduct inference on a parameter of interest θ_n^o from data $X_{n1}, \ldots, X_{nn}$ near epoch dependent (NED) on an underlying mixing process. Here, X_{nt} is a vector containing both explanatory and dependent variables. We define $\{X_{nt}\}$ to be NED on a mixing process $\{V_t\}$ if $E\left(X_{nt}^2\right) < \infty$ and $v_k \equiv \sup_{n,t} \left\| X_{nt} - E_{t-k}^{t+k}\left(X_{nt}\right) \right\|_2 \to 0$ as $k \to \infty$. Here, $\|X_{nt}\|_p \equiv \left(E\left|X_{nt}\right|^p\right)^{1/p}$ is the L_p norm and $E_{t-k}^{t+k}(\cdot) \equiv E\left(\cdot | \mathcal{F}_{t-k}^{t+k}\right)$, where $\mathcal{F}_{t-k}^{t+k} \equiv \sigma\left(V_{t-k}, \ldots, V_{t+k}\right)$ is the σ-field generated by $V_{t-k}, \ldots, V_{t+k}$. If $v_k = O\left(k^{-a-\delta}\right)$ for some $\delta > 0$, we say $\{X_{nt}\}$ is NED of size $-a$. We assume $\{V_t\}$ is strong mixing. The strong mixing coefficients are $\alpha_k \equiv \sup_m \sup_{\{A \in \mathcal{F}_{-\infty}^m, B \in \mathcal{F}_{m+k}^\infty\}} |P(A \cap B) - P(A) P(B)|$; we require $\alpha_k \to 0$ as $k \to \infty$ suitably fast.

Our methods involve using the MBB to resample certain functions of the data. Thus, consider a generic array of random variables $\{Z_{nt} : t = 1, \ldots, n\}$. Let $\ell = \ell_n \in \mathbb{N}$ $(1 \leq \ell < n)$ be a block length, and let $B_{t,\ell} = \{Z_{nt}, Z_{n,t+1}, \ldots, Z_{n,t+\ell-1}\}$ be the block of ℓ consecutive observations starting at Z_{nt} ($\ell = 1$ gives the standard bootstrap). For simplicity take $n = k\ell$. The MBB draws $k = n/\ell$ blocks randomly with replacement from the set of overlapping blocks $\{B_{1,\ell}, \ldots, B_{n-\ell+1,\ell}\}$. Letting $I_{n1}, \ldots, I_{nk}$ be i.i.d. random variables distributed uniformly on $\{0, \ldots, n-\ell\}$, we have $\{Z_{nt}^* = Z_{n,\tau_{nt}}, t = 1, \ldots, n\}$, where τ_{nt}

defines a random array $\{\tau_{nt}\} \equiv \{I_{n1}+1,\ldots,I_{n1}+\ell,\ldots,I_{nk}+1,\ldots,I_{nk}+\ell\}$.

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

$$\max_{\Theta} L_n(\theta), \quad n=1,2,\ldots,$$

where $L_n(\theta) \equiv n^{-1}\sum_{t=1}^{n} \log f_{nt}\left(X_n^t,\theta\right)$, $X_n^t \equiv (X_{n1}',\ldots,X_{nt}')'$, $t=1,2,\ldots,n$, and θ belongs to Θ, a compact subset of $\mathbb{R}^p$, $p \in \mathbb{N}$. Thus, X_n^t contains all explanatory and dependent variables entering f_{nt}, the "quasi-likelihood" for observation t. The function L_n is the "quasi-log-likelihood function".

Gallant and White (1988) study the properties of the QMLE $\hat{\theta}_n$ (consistency and asymptotic normality) under certain regularity assumptions, collected in Appendix A for convenience. In particular, Assumption A in Appendix A is our doubly indexed version of Gallant and White's (1988) regularity assumptions. Under these assumptions, Theorem 5.7 of Gallant and White (1988) shows that $B_n^{o-1/2}A_n^o\sqrt{n}\left(\hat{\theta}_n-\theta_n^o\right) \Rightarrow N(0,I_p)$, where $\Rightarrow$ denotes convergence in distribution. The asymptotic covariance matrix is thus $C_n^o \equiv A_n^{o-1}B_n^oA_n^{o-1}$, where $A_n^o \equiv E\left(n^{-1}\sum_{t=1}^{n}\nabla^2\log f_{nt}\left(X_n^t,\theta_n^o\right)\right)$ and $B_n^o \equiv \operatorname{var}\left(n^{-1/2}\sum_{t=1}^{n}\nabla\log f_{nt}\left(X_n^t,\theta_n^o\right)\right)$. Here and throughout the paper, for any function $g:\Omega\times\Theta\to\mathbb{R}$, we let $\nabla g(\cdot,\theta)=\partial g/\partial\theta$ and $\nabla^2 g(\cdot,\theta)=\partial^2 g/\partial\theta\partial\theta'$.

Given the original sample $X_{n1},\ldots,X_{nn}$, let $\hat{\theta}_n^*$ be a bootstrap version of $\hat{\theta}_n$, solving

$$\max_{\Theta} L_n^*(\theta), \quad n=1,2,\ldots,$$

where $L_n^*(\theta) \equiv n^{-1}\sum_{t=1}^{n}\log f_{nt}^*(\theta)$, and for $n=1,2,\ldots$ and each $\theta\in\Theta$, $\{f_{nt}^*(\theta), t=1,\ldots,n\}$ is given by $f_{nt}^*(\theta)=f_{n,\tau_{nt}}\left(X_n^{\tau_{nt}},\theta\right)$, with τ_{nt} chosen by the MBB. Thus, the bootstrap QMLE resamples the contributions $\log f_{nt}\left(X_n^t,\theta\right)$ to $L_n(\theta)$. This is often equivalent to directly resampling the data, for example in linear regression where f_{nt} depends only upon $X_{nt}=(y_{nt},W_{nt}')'$ (here y_{nt} is the dependent variable at time t and W_{nt} is a vector of explanatory variables at time t that may include lagged dependent variables). In this case, resampling blocks of $f_{nt}\left(X_n^t,\theta\right)$ is equivalent to resampling blocks of $X_{nt}=(y_{nt},W_{nt}')'$, the "blocks of blocks bootstrap" (Politis and Romano, 1992). But if f_{nt} depends on the entire past history X_n^t, it may not be possible to define "tuples" of observables on which to apply the MBB. This is the case for GARCH models, for which bootstrapping the QMLE does not involve directly bootstrapping the data. Instead, for GARCH models we resample the contributions $\log f_{nt}\left(X_n^t,\theta\right)$ to the quasi-log-likelihood function $L_n(\theta)$.

We first show that $\hat{\theta}_n^*$ converges in probability to $\hat{\theta}_n$, conditional on all samples with probability tending to one. Conventionally, P^* is the probability measure induced by the MBB. For a bootstrap statistic T_n^* we write $T_n^* \to 0$ $prob - P^*$, $prob - P$ if for any $\varepsilon > 0$ and any $\delta > 0$, $\lim_{n\to\infty} P\left[P^*\left[|T_n^*| > \varepsilon\right] > \delta\right] = 0$.

Theorem 2.1. *Let Assumption A in Appendix A hold. Then,* $\hat{\theta}_n - \theta_n^o \to 0$ $prob - P$. *If also* $\ell_n \to \infty$, *and* $\ell_n = o(n)$, *then* $\hat{\theta}_n^* - \hat{\theta}_n \to 0$ $prob - P^*$, $prob - P$.

Thus, $\hat{\theta}_n$ is asymptotically the bootstrap "pseudo-true parameter". Nevertheless, for given n, the MBB population first-order conditions evaluated at $\hat{\theta}_n$ are not generally zero. That is, $E^*\left[n^{-1}\sum_{t=1}^n s_{nt}^*\left(\hat{\theta}_n\right)\right] \neq 0$, where $\left\{s_{nt}^*\left(\hat{\theta}_n\right) = \nabla \log f_{n,\tau_{nt}}\left(X_n^{\tau_{nt}}, \hat{\theta}_n\right)\right\}$. In order to obtain asymptotic refinements of the bootstrap, Andrews (2002) recenters the bootstrap objective function to $L_n^*(\theta) - n^{-1}\sum_{t=1}^n E^*\left(s_{nt}^*\left(\hat{\theta}_n\right)\right)'\theta$, ensuring that the bootstrap population first-order conditions equal zero at $\hat{\theta}_n$. A similar recentering of the criterion function is used by Hall and Horowitz (1996) in the GMM context. Recentering is not needed for establishing the first-order properties of the bootstrap QMLE and we therefore leave this aside here.

Next we show that the sampling distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta_n^o\right)$ is well-approximated by the distribution of $\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$, conditional on $X_{n1}, \ldots, X_{nn}$. For this, we strengthen Assumption A as follows:

Assumption 2.1

2.1.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 $n, t = 1, 2, \ldots$, $r > 2$.

2.1.b) For some small $\delta > 0$ and some $r > 2$, the elements of $\left\{s_{nt}\left(X_n^t, \theta\right)\right\}$ 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 2.2 $n^{-1}\sum_{t=1}^n E\left(s_{nt}^o\right) E\left(s_{nt}^o\right)' = o\left(\ell_n^{-1}\right)$, where $\ell_n = o(n)$ and $\ell_n \to \infty$.

The consistency of the MBB distribution depends crucially on the consistency of the MBB covariance matrix of the scaled average of the MBB-resampled scores $\{s_{nt}^{*o} \equiv \nabla \log f_{n,\tau_{nt}}(X_n^{\tau_{nt}}, \theta_n^o)\}$. With misspecification, $\{s_{nt}^o\}$ is dependent and possibly heterogeneous. Accordingly, Assumption 2.1.b) takes $\{s_{nt}^o\}$ to be $L_{2+\delta}$-NED on a mixing process (see Andrews (1988)), for small $\delta > 0$. An application of Gonçalves and White (2002) Theorem 2.1 shows that the MBB covariance matrix of the scaled average of $\{s_{nt}^{*o}\}$ is consistent under this NED condition for $B_n^o + U_n^o$, where $B_n^o \equiv \text{var}\left(n^{-1/2}\sum_{t=1}^n s_{nt}^o\right)$, and

$U_n^o \equiv \text{var}^*\left(n^{-1/2}\sum_{t=1}^n [E(s_{nt}^o)]^*\right)$, with $\{[E(s_{nt}^o)]^*\}$ a MBB resample of $\{E(s_{nt}^o)\}$. Assumption 2.2 eliminates the bias U_n^o asymptotically, ensuring that $\hat{\theta}_n^*$ converges to a normal distribution with the correct covariance matrix. A sufficient condition for Assumption 2.2 is that $E(s_{nt}^o) = 0$ be zero for all t, n. As Gallant and White (1988, p. 102) remark, 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. $f_{nt}(\cdot,\theta) = f(\cdot,\theta)$ for all t, n).

Theorem 2.2. *Let Assumption A as strengthened by Assumptions 2.1 and 2.2 hold. If* $\ell_n \to \infty$ *and* $\ell_n = o(n^{1/2})$, *then for any* $\varepsilon > 0$, $P\left\{\sup_{x\in\mathbb{R}^p}\left|P^*\left[\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \le x\right] - P\left[\sqrt{n}\left(\hat{\theta}_n - \theta_n^o\right) \le x\right]\right| > \varepsilon\right\} \to 0.$

Theorem 2.2 justifies using order statistics of the bootstrap distribution to form percentile confidence intervals for θ_n^o with asymptotically correct coverage probabilities. Note that this does not justify using the variance of the bootstrap distribution, $\text{var}^*\left(\sqrt{n}\hat{\theta}_n^*\right) = \lim_{B\to\infty}\frac{1}{B}\sum_{b=1}^B n\left(\hat{\theta}_n^{*(b)} - \overline{\hat{\theta}_n^*}\right)\left(\hat{\theta}_n^{*(b)} - \overline{\hat{\theta}_n^*}\right)'$, with B the number of bootstrap replications and $\overline{\hat{\theta}_n^*} = \frac{1}{B}\sum_{b=1}^B \hat{\theta}_n^{*(b)}$, to consistently estimate the QMLE asymptotic variance without further conditions, e.g. that $\left\{n\left(\hat{\theta}_n^* - \hat{\theta}_n\right)\left(\hat{\theta}_n^* - \hat{\theta}_n\right)'\right\}$ is uniformly integrable (e.g. Billingsley, 1995, p. 338). This has been sometimes overlooked in the literature. Counterexamples to the consistency 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). See also Gonçalves and White (2003).

Bootstrapping the QMLE may be computationally costly as it requires an optimization for each resample. Davidson and MacKinnon (1999) have proposed approximate bootstrap methods based on a few iterations starting from the original QMLE, achieving the same accuracy as the fully-optimized bootstrap. Let $A_n^*\left(\hat{\theta}_n\right) = n^{-1}\sum_{t=1}^n \nabla^2 \log f_{nt}^*\left(\hat{\theta}_n\right)$ be the MBB resampled estimated Hessian, and let $\left\{s_{nt}^*\left(\hat{\theta}_n\right)\right\}$ be the MBB resampled estimated scores. The one-step MBB QMLE is:

$$\hat{\theta}_{1n}^* = \hat{\theta}_n - A_n^*\left(\hat{\theta}_n\right)^{-1} n^{-1}\sum_{t=1}^n s_{nt}^*\left(\hat{\theta}_n\right).$$

Corollary 2.1. *Let Assumption A as strengthened by Assumptions 2.1 and 2.2 hold. If* $\ell_n = o(n^{1/2})$, *then for any* $\varepsilon > 0$, $P\left[\sup_{x\in\mathbb{R}^p}\left|P^*\left[\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \le x\right] - P^*\left[\sqrt{n}\left(\hat{\theta}_{1n}^* - \hat{\theta}_n\right) \le x\right]\right| > \varepsilon\right] \to 0.$

Following Andrews (2002), we define the multi-step bootstrap estimators recursively as follows:

$$\hat{\theta}_{j,n}^* = \hat{\theta}_{j-1,n}^* - A_n^*\left(\hat{\theta}_{j-1,n}^*\right)^{-1} n^{-1}\sum_{t=1}^n s_{nt}^*\left(\hat{\theta}_{j-1,n}^*\right) \text{ for } 1 \le j \le m,$$

where $\hat{\theta}^*_{0,n} = \hat{\theta}_n$. Here, $A^*_n\left(\hat{\theta}^*_{j-1,n}\right) = n^{-1}\sum_{t=1}^n \nabla^2 \log f^*_{nt}\left(\hat{\theta}^*_{j-1,n}\right)$ is the MBB resampled Hessian and $\left\{s^*_{nt}\left(\hat{\theta}^*_{j-1,n}\right)\right\}$ are the MBB resampled scores, both evaluated at $\hat{\theta}^*_{j-1,n}$. Results analogous to Corollary 2.1 hold for the multi-step estimators under the same conditions.

3. Hypothesis Testing

The results of Section 2 do not immediately justify testing hypotheses about θ^o_n based on studentized statistics such as t- or Wald statistics. Nevertheless, they are the key to proving the ability of the bootstrap to approximate the distribution of studentized statistics, as we now show.

Let $\{r_n : \Theta \to \mathbb{R}^q\}$, with $\Theta \subset \mathbb{R}^p$, $q \leq p$, be a sequence of functions that have elements continuously differentiable on Θ uniformly in n such that $\{R^o_n \equiv \nabla' r_n(\theta^o_n)\}$ is $O(1)$ with full row rank q, uniformly in n. The Wald statistic for testing $H_o : \sqrt{n} r_n(\theta^o_n) \to 0$ 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 consistent for $C^o_n = A^{o-1}_n B^o_n A^{o-1}_n$. In particular, $\hat{A}_n = n^{-1}\sum_{t=1}^n \nabla^2 \log f_{nt}\left(X^t_n, \hat{\theta}_n\right)$ is an estimator of $A^o_n \equiv E\left(n^{-1}\sum_{t=1}^n \nabla^2 \log f_{nt}\left(X^t_n, \theta^o_n\right)\right)$ and $\hat{B}_n$ is such that $\hat{B}_n - B^o_n \xrightarrow{P} 0$. For our context, $\hat{B}_n$ is a kernel-type variance estimator, e.g. a Bartlett (Newey-West, 1987) or a Quadratic Spectral (Andrews, 1991) estimator. For first order properties, we just need $\hat{B}_n$ to be consistent for B^o_n. Our bootstrap Wald statistic is

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

where we set $\hat{r}^*_n = r_n\left(\hat{\theta}^*_n\right)$, $\hat{R}^*_n = \nabla' r_n\left(\hat{\theta}^*_n\right)$ and $\hat{C}^*_n = \hat{A}^{*-1}_n\hat{B}^*_n\hat{A}^{*-1}_n$. Here,
$\hat{A}^*_n = n^{-1}\sum_{t=1}^n \nabla^2 \log f_{n,\tau_{nt}}\left(X^{\tau_{nt}}_n, \hat{\theta}^*_n\right)$ and $\hat{B}^*_n$ is

$$\hat{B}^*_n = k^{-1}\sum_{i=1}^k\left(\ell^{-1/2}\sum_{t=1}^{\ell} s_{n,I_{ni}+t}\left(X^{I_{ni}+t}_n, \hat{\theta}^*_n\right)\right)\left(\ell^{-1/2}\sum_{t=1}^{\ell} s_{n,I_{ni}+t}\left(X^{I_{ni}+t}_n, \hat{\theta}^*_n\right)\right)'. \tag{3.1}$$

$\hat{B}^*_n$ is the multivariate QMLE analog of the MBB variance estimator of Davison and Hall (1993) and Götze and Künsch (1996). To motivate this, recall that $\hat{B}^*_n$ is the bootstrap analog of $\hat{B}_n$, which estimates B^o_n, the covariance of the scaled average of the scores at θ^o_n. Analogously, $\hat{B}^*_n$ estimates the bootstrap covariance of the scaled average of the resampled scores at $\hat{\theta}_n$, i.e. $\hat{B}_n^*$ is an *estimator* of

$$\operatorname{var}^*\left(n^{-1/2}\sum_{t=1}^{n} s_{nt}^*\left(\hat{\theta}_n\right)\right) = \operatorname{var}^*\left(k^{-1/2}\sum_{i=1}^{k}\left(\ell^{-1/2}\sum_{t=1}^{\ell} s_{n,I_{ni}+t}\left(X_n^{I_{ni}+t},\hat{\theta}_n\right)\right)\right). \tag{3.2}$$

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 (3.1) of the (bootstrap population) variance (3.2) is just the sample variance of these means, with $\hat{\theta}_n$ replaced by $\hat{\theta}_n^*$ to mimic the replacement of θ_n^o with $\hat{\theta}_n$ when computing $\hat{B}_n$. Note that in (3.1) we use the bootstrap optimization first order conditions to set $\bar{\hat{s}}_n^* \equiv n^{-1}\sum_{t=1}^{n} s_{nt}^*\left(\hat{\theta}_n^*\right) = 0$.

For proving asymptotic refinements of the bootstrap, Götze and Künsch (1996) note that one must carefully choose the studentizing kernel variance estimator. Instead of triangular weights, rectangular or quadratic weights should be used in estimating B_n^o. For our first-order results, consistency of $\hat{B}_n$ for B_n^o suffices; for instance, our results hold for the Bartlett and the Quadratic Spectral kernel variance estimators used in our simulations below, provided these are consistent for B_n^o.

To analyze the bootstrap Wald statistic $\mathcal{W}_n^*$ we strengthen Assumption 2.2:

Assumption 2.2′ $n^{-1}\sum_{t=1}^{n}\left|E\left(s_{nti}^o\right)\right|^{2+\delta} = o\left(\ell_n^{-1-\delta/2}\right)$ for $i = 1,\ldots,p$.

Theorem 3.1. *Let the assumptions of Theorem 2.2 hold as strengthened by Assumption 2.2′. Suppose further that $\mathcal{W}_n$ uses a consistent estimator of B_n^o. Then, under H_o, for all $\varepsilon > 0$, if $\ell = o\left(n^{1/2}\right)$, $P\left[\sup_{x\in\mathbb{R}}\left|P^*\left(\mathcal{W}_n^* \le x\right) - P\left(\mathcal{W}_n \le x\right)\right| > \varepsilon\right] \to 0$.*

This proves the first order asymptotic equivalence under the null of the bootstrap Wald and the original Wald statistic. Consistency of a bootstrap t-statistic studentized with $\hat{C}_n^*$ follows by almost identical arguments, justifying the construction of MBB percentile-t confidence intervals.

The bootstrap also works for the Lagrange Multiplier (LM) statistic. Using notation analogous to Gallant and White (1988), the LM statistic is $\mathcal{L}_n$ and its bootstrap analog is

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

where, with $\tilde{\theta}_n^*$ the constrained bootstrap QMLE, $\nabla L_n^*\left(\tilde{\theta}_n^*\right) \equiv n^{-1}\sum_{t=1}^{n} s_{nt}^*\left(\tilde{\theta}_n^*\right)$, $\tilde{R}_n^* \equiv \nabla' r_n\left(\tilde{\theta}_n^*\right)$, $\tilde{C}_n^* \equiv \tilde{A}_n^{*-1}\tilde{B}_n^*\tilde{A}_n^{*-1}$, and $\tilde{A}_n^* \equiv n^{-1}\sum_{t=1}^{n}\nabla^2\log f_{n,\tau_{nt}}\left(X_n^{\tau_{nt}},\tilde{\theta}_n^*\right)$. Similarly, $\tilde{B}_n^*$ is as in (3.1) using $\tilde{\theta}_n^*$ instead of $\hat{\theta}_n^*$, with $\ell^{1/2}\nabla L_n^*\left(\tilde{\theta}_n^*\right)$ subtracted off each term $\ell^{-1/2}\sum_{t=1}^{\ell} s_{n,I_{ni}+t}\left(X_n^{I_{ni}+t},\tilde{\theta}_n^*\right)$ because $\nabla L_n^*\left(\tilde{\theta}_n^*\right)$ is not generally zero.

Theorem 3.2. *Let the assumptions of Theorem 2.2 hold as strengthened by Assumption 2.2′. Suppose further that* $\mathcal{L}_n$ *uses a consistent estimator of* B_n^o. *Then, under* H_o, *for all* $\varepsilon > 0$, *if* $\ell = o\left(n^{1/2}\right)$, $P\left[\sup_{x\in\mathbb{R}}\left|P^*\left(\mathcal{L}_n^* \leq x\right) - P\left(\mathcal{L}_n \leq x\right)\right| > \varepsilon\right] \to 0$.

4. Monte Carlo Results

This section provides Monte Carlo evidence on the relative finite sample performance of the MBB and the asymptotic normal approximation for confidence intervals. We consider two practical examples of nonlinear models that are typically estimated by QML. The first examines the MBB percentile-t and asymptotic normal coverage probabilities of confidence intervals in the context of logit models with neglected autocorrelation. Next we compare the MBB to asymptotic normal confidence intervals for possibly misspecified ARCH models.

Confidence Intervals for Logit models

Let a dependent variable y_t take the value 0 or 1, whenever the unobserved $y_t^* = W_t'\beta + \varepsilon_t$ is positive or negative, respectively. W_t is a $k \times 1$ vector of explanatory variables and β a vector of parameters. We generate ε_t as AR(1):

$$\varepsilon_t = \rho\varepsilon_{t-1} + \sqrt{1-\rho^2}v_t,$$

with $\text{Prob}(v_t \leq a) = \frac{\exp(a)}{1+\exp(a)}$ for any $a \in \mathbb{R}$. Thus, the DGP is logit with autocorrelated errors whenever $\rho \neq 0$. We estimate an ordinary logit model by QMLE ignoring the autocorrelation. The QMLE $\hat{\beta}_n$ remains consistent for β and asymptotically normal (cf. Gourieroux, Monfort and Trognon (1984) for the related probit model). Nevertheless, confidence intervals for β require an HAC covariance estimator using asymptotic normality, or a bootstrap confidence interval (e.g. a MBB with $\ell > 1$).

Asymptotic normal intervals rely on $t_{\hat{\beta}_{ni}} = \frac{\sqrt{n}(\hat{\beta}_{ni}-\beta_i)}{\sqrt{\hat{C}_{ni,i}}}$, where $\hat{C}_n = \hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}$. We consider three choices for $\hat{B}_n$: the outer product of the gradient (OP), $\hat{B}_n = n^{-1}\sum_{t=1}^n \hat{s}_{nt}\hat{s}_{nt}'$, and two HAC estimators, using either the Bartlett (BT) or the Quadratic Spectral (QS) kernel. As we remarked before, other choices of $\hat{B}_n$ (not considered here) may be expected to provide better finite sample properties, as suggested by the asymptotic refinements arguments of Götze and Künsch (1996). The MBB intervals are based on $t_{\hat{\beta}_{ni}^*} = \frac{\sqrt{n}(\hat{\beta}_{ni}^*-\hat{\beta}_{ni})}{\sqrt{\hat{C}_{ni,i}^*}}$, where $\hat{C}_n^* = \hat{A}_n^{*-1}\hat{B}_n^*\hat{A}_n^{*-1}$, with $\hat{B}_n^*$ as in (3.1). In particular, $\hat{\beta}_n^*$ is computed without recentering. Further finite-sample improvements may be expected with recentering of the criterion function, as in Andrews (2002). We note that the BT, QS, and MBB intervals are robust to neglected autocorrelation, whereas the OP intervals are not.

The choice of the block size/bandwidth is critical. In our simulations, we use Andrews' (1991) automatic procedure to compute a data-driven bandwidth for BT and QS based on approximating AR(1) models for the elements of the score. For the MBB, we use the (integer part of the) automatic bandwidth chosen by Andrews' (1991) method for the BT kernel. The asymptotic equivalence of the MBB variance estimator to the Bartlett kernel variance estimator motivates our choice of the block length by the Andrews procedure. Although the block length that results from this approach may not be optimal in the context of estimating distributions of studentized test statistics, as we do in our application, it is readily accessible to practitioners, typically performs well, and allows meaningful comparisons of the bootstrap and asymptotic normal theory-based methods. In addition, it is computationally very inexpensive (as opposed to other methods that have been proposed in the literature such as the calibration method of Politis, Romano and Wolf (1997) or the method of Hall, Horowitz and Jing (1995)).

In the experiments, W contains a constant, and either one, two, three, or four random regressors, independently generated as AR(1) with autocorrelation coefficient equal to 0.5. The intercept is always 0, so on average half the y_t's are 0 and half are 1. The slope parameters are all set to 0.25. For each experiment we let $\rho \in \{0, 0.5, 0.9\}$, and use 10,000 Monte Carlo trials with 999 bootstrap replications. We discarded 27 out of the 10,000 trials due to nonconvergence of the logit routine with $k = 5$, $n = 50$, $\rho = 0.9$. Nonconvergence in the bootstrap resamples occurred on average less than 0.08% per Monte Carlo trial, for all experiments, except when $k = 5$, $n = 50$, $\rho = 0.9$, in which case this rate was 1.07%. When bootstrap optimization failed, we redrew new bootstrap indices. Table 1 reports coverage rates for the first slope parameter.

The main results are: a) when $\rho = 0$ all methods work well, even for $n = 50$; b) when $\rho \neq 0$ all intervals undercover, but the robust methods (BT, QS and MBB) outperform OP, as expected. The undercoverage is worse the larger is k and the smaller is n (an exception is MBB when $k = 5$, $\rho = 0.9$, perhaps due to the larger rate of non-convergence); c) the MBB always outperforms BT or QS, especially for small n and large k, and d) the average block size is larger for larger ρ and n, as expected.

Confidence Intervals for ARCH models

We assume the following DGP:

$$y_t = \bar{\gamma} + \varepsilon_t, \qquad \varepsilon_t = v_t h_t^{1/2}, \qquad h_t = \bar{\omega} + \bar{\alpha}\varepsilon_{t-1}^2. \tag{4.1}$$

A bar denotes true parameters and a superscript o denotes pseudo-true parameters throughout. Usually $\{v_t\}$ is assumed i.i.d. Here, we generate v_t as AR(1):

$$v_t = \bar{\rho} v_{t-1} + u_t, \quad |\bar{\rho}| < 1, \quad u_t \sim \text{i.i.d. } N\left(0, \sigma_u^2\right), \ \sigma_u^2 = 1 - \bar{\rho}^2. \tag{4.2}$$

We can write (4.1)-(4.2) as $y_t = \bar{\gamma} + \bar{\rho} h_t^{1/2} h_{t-1}^{-1/2} (y_{t-1} - \bar{\gamma}) + h_t^{1/2} u_t$, with $u_t \sim$ i.i.d. $N\left(0, \sigma_u^2\right)$. For $\bar{\rho} = 0$, this is the usual ARCH(1). For $\bar{\rho} \neq 0$ an extra term appears. Letting $\mathcal{F}^{t-1} = \sigma(\ldots y_{t-2}, y_{t-1})$ and using $v_{t-1} = \varepsilon_{t-1}/h_{t-1}^{1/2}$, we have $E\left(\varepsilon_t | \mathcal{F}^{t-1}\right) = \bar{\rho} h_t^{1/2} v_{t-1}$ and $E\left(\varepsilon_t^2 | \mathcal{F}^{t-1}\right) = h_t\left(1 - \bar{\rho}^2 + \bar{\rho}^2 v_{t-1}^2\right)$, so $E\left(y_t | \mathcal{F}^{t-1}\right) = \bar{\gamma} + \bar{\rho} h_t^{1/2} h_{t-1}^{-1/2} (y_{t-1} - \bar{\gamma})$ and var $\left(y_t | \mathcal{F}^{t-1}\right) = \sigma_u^2 h_t$. We (mis)specify a Gaussian ARCH(1) model parameterized by $\theta = (\gamma, \omega, \alpha)'$:

$$y_t = \gamma + e_t, \quad e_t | \mathcal{F}^{t-1} \sim N(0, h_t(\theta)), \ t = 1, \ldots, n,$$

with $h_t(\theta) = \omega + \alpha e_{t-1}^2$. The QMLE $\hat{\theta}_n$ maximizes the log-likelihood

$$L_n(\theta) = \frac{1}{2n} \sum_{t=1}^{n} \log f_t(\theta), \text{ where } \log f_t(\theta) = -\left(\ln h_t(\theta) + \frac{e_t^2}{h_t(\theta)}\right). \tag{4.3}$$

The model is correctly specified if and only if $\bar{\rho} = 0$. With misspecification the QMLE is generally inconsistent for $\bar{\theta} = (\bar{\gamma}, \bar{\omega}, \bar{\alpha})$; instead confidence intervals for pseudo-true parameters $\theta^o = (\gamma^o, \omega^o, \alpha^o)$ pertain. We evaluate θ^o by simulation, as the value maximizing the expectation of (4.3), computed using 50,000 simulations. Considering the expected score corresponding to γ, we have

$$E\left(s_{1t}\left(\bar{\theta}^o\right)\right) = E\left(\frac{\alpha^o \varepsilon_{t-1}}{h_t\left(\bar{\theta}^o\right)} - \frac{\varepsilon_t^2}{h_t^2\left(\bar{\theta}^o\right)}(\alpha^o \varepsilon_{t-1}) + \frac{\varepsilon_t}{h_t\left(\bar{\theta}^o\right)}\right), \tag{4.4}$$

where $\bar{\theta}^o = (\bar{\gamma}, \omega^o, \alpha^o)'$, $h_t\left(\bar{\theta}^o\right) = \omega^o + \alpha^o \varepsilon_{t-1}^2$ and $\varepsilon_t = y_t - \bar{\gamma}$. For suitably symmetric joint distributions of $(\varepsilon_t, \varepsilon_{t-1})$ centered at zero, it is plausible that this expectation equals zero, implying that $\gamma^o = \bar{\gamma}$, despite the misspecification. Proving this conjecture would distract us from our purpose here, but our simulations (with normal errors) always delivered $\gamma^o = \bar{\gamma}$. Accordingly we set $\gamma^o = \bar{\gamma}$ in what follows.

With misspecification, the scores are generally not a martingale difference sequence, justifying the use of robust inference on θ^o. As before, we consider OP, BT, QS and MBB. All but OP are robust to misspecification. The OP interval is valid if the first two conditional moments of y_t are not misspecified.

Data on $\{y_t\}$ were generated by (4.1)-(4.2) with $\bar{\gamma} = 1.0$, $\bar{\omega} = 0.1$ and six combinations of $\bar{\alpha}$ and $\bar{\rho}$

taken from $\bar{\alpha} \in \{0.5, 0.9\}$ and $\bar{\rho} \in \{0.0, 0.5, 0.9\}$. Table 2 contains results. We summarize as follows. When $\bar{\rho} = 0$ all methods tend to perform well, though the coverage of the BT and QS intervals tends to slightly understate the true levels for $\bar{\omega}$ and $\bar{\alpha}$. In contrast, the MBB intervals achieve almost correct coverage for $\bar{\theta}$, with slight overstatement for $\bar{\omega}$. When $\bar{\rho} = 0$, the scores are a martingale difference sequence, and this is reflected in the block size parameter. When $\bar{\rho} \neq 0$, major findings are: (i) the OP intervals fail dramatically for $\bar{\gamma}$, exhibiting severe undercoverage which worsens as $\bar{\rho}$ increases; (ii) the coverages of BT and QS for $\bar{\gamma}$ are also well below the 95% nominal level, but we see clear improvement as n increases; (iii) the MBB outperforms the HAC methods; and (iv) the average chosen block size exceeds one, and tends to increase with n, as we expect.

5. Conclusion

The results presented here justify routine use of MBB methods for the QMLE in a general context. Further results in our setting establishing higher order improvements for the MBB (with recentering) are a logical next step and a promising subject for future work.

Appendix A: Assumptions and Proofs for Section 2

Throughout Appendix A, P is the probability measure governing the behavior of the original time series while $P^*_{n,\omega}$ denotes the probability measure induced by the bootstrap. For the arguments presented in this section, it is important to make explicit the dependence of $P^*_{n,\omega}$ on n and ω. For any bootstrap statistic $T^*_n(\cdot,\omega)$ we write $T^*_n(\cdot,\omega) \to 0$ $prob - P^*_{n,\omega}$, $a.s. - P$ if for any $\varepsilon > 0$ there exists $F \in \mathcal{F}$ with $P(F) = 1$ such that for all ω in F, $\lim_{n\to\infty} P^*_{n,\omega}[\lambda : |T^*_n(\lambda,\omega)| > \varepsilon] = 0$. We write $T^*_n(\cdot,\omega) \to 0$ $prob - P^*_{n,\omega}$, $prob - P$ if for any $\varepsilon > 0$ and for any $\delta > 0$, $\lim_{n\to\infty} P\left[\omega : P^*_{n,\omega}[\lambda : |T^*_n(\lambda,\omega)| > \varepsilon] > \delta\right] = 0$. Using a subsequence argument (e.g. Billingsley, 1995, Theorem 20.5), $T^*_n(\cdot,\omega) \to 0$ $prob - P^*_{n,\omega}$, $prob - P$ is equivalent to having that for any subsequence $\{n'\}$ there exists a further subsequence $\{n''\}$ such that $T^*_{n''}(\cdot,\omega) \to 0$ $prob - P^*_{n'',\omega}$, $a.s. - P$. For any distribution D we write $T^*_n(\cdot,\omega) \Rightarrow^{d_{P^*_{n,\omega}}} D$ $prob - P$ when weak convergence under $P^*_{n,\omega}$ occurs in a set with probability converging to one, or equivalently, when for every subsequence there exists a further subsequence for which weak convergence under $P^*_{n,\omega}$ takes place almost surely $-P$.

Assumption A is the doubly indexed counterpart of the regularity conditions used by Gallant and White (1988).

Assumption A

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

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

A.3: (i) θ_n^o is identifiably unique with respect to $E(L_n(X_n^n, \theta))$. (ii) θ_n^o is interior to Θ uniformly in n.

A.4: (i)$\{\log f_{nt}(X_n^t, \theta)\}$ is Lipschitz continuous on Θ, i.e. $|\log f_{nt}(X_n^t, \theta) - \log f_{nt}(X_n^t, \theta^o)| \leq L_{nt}$ $|\theta - \theta^o|$ *a.s.* $- P$, $\forall \theta, \theta^o \in \Theta$, where $\sup_n \{n^{-1} \sum_{t=1}^n E(L_{nt})\} = O(1)$. (ii) $\{\nabla^2 \log f_{nt}(X_n^t, \theta)\}$ is Lipschitz continuous on Θ.

A.5: For some $r > 2$: (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 such that $\|D_{nt}\|_r \leq \Delta < \infty$ for all n, t. (ii) $\{\nabla \log f_{nt}(X_n^t, \theta)\}$ is r-dominated on Θ uniformly in n, t. (iii) $\{\nabla^2 \log f_{nt}(X_n^t, \theta)\}$ is r-dominated on Θ uniformly in n, t.

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

A.7: The elements of (i) $\{\log f_{nt}(X_n^t, \theta)\}$ are NED on $\{V_t\}$ of size $-\frac{1}{2}$; (ii) $\{\nabla \log f_{nt}(X_n^t, \theta)\}$ are NED on $\{V_t\}$ of size -1 uniformly on (Θ, ρ), where ρ is any convenient norm on $\mathbb{R}^p$, and (iii) $\{\nabla^2 \log f_{nt}(X_n^t, \theta)\}$ are NED on $\{V_t\}$ of size $-\frac{1}{2}$ uniformly on (Θ, ρ).

A.8: (i) $\left\{B_n^o \equiv \text{var}\left(n^{-\frac{1}{2}} \sum_{t=1}^n \nabla \log f_{nt}(X_n^t, \theta_n^o)\right)\right\}$ is uniformly positive definite. (ii) $\{A_n^o \equiv E(n^{-1} \sum_{t=1}^n \nabla^2 \log f_{nt}(X_n^t, \theta_n^o))\}$ is uniformly nonsingular.

The usefulness of the following lemmas extends beyond the QMLE as they apply to prove the validity of bootstrap methods for other extremum estimators, such as GMM.

Lemma A.1 (Identifiable uniqueness of $\hat{\theta}_n$). *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 $\{Q_n : \Omega \times \Theta \to \overline{\mathbb{R}}\}$ 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 $\sup_{\theta \in \Theta} |Q_n(\cdot, \theta) - \overline{Q}_n(\theta)| \to 0$ a.s.-P and if $\{\overline{Q}_n : \Theta \to \overline{\mathbb{R}}\}$ has identifiably unique maximizers $\{\theta_n^o\}$ on Θ, then $\{\hat{\theta}_n\}$ 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*

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

where $\eta^c\left(\hat{\theta}_n, \varepsilon\right)$ is the compact complement of $\eta\left(\hat{\theta}_n, \varepsilon\right) \equiv \left\{\theta \in \Theta : \left|\theta - \hat{\theta}_n\right| < \varepsilon\right\}$. If instead $\sup_{\theta\in\Theta}\left|Q_n(\cdot,\theta) - \overline{Q}_n(\theta)\right| \to 0$ *prob* $- P$ then for any subsequence $\left\{\hat{\theta}_{n'}\right\}$ of $\left\{\hat{\theta}_n\right\}$, there exists a further subsequence $\left\{\hat{\theta}_{n''}\right\}$ such that $\left\{\hat{\theta}_{n''}\right\}$ is identifiably unique with respect to $\{Q_{n''}\}$ $a.s. - P$.

Lemma A.2 (Consistency of $\hat{\theta}_n^*$). *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 such that (a1) $Q_n(\cdot,\theta) : \Omega \to \overline{\mathbb{R}}$ is measurable-$\mathcal{F}$ for each $\theta \in \Theta$; (a2) $Q_n(\omega,\cdot) : \Theta \to \overline{\mathbb{R}}$ is continuous on Θ $a.s. - P$. Let $\hat{\theta}_n = \arg\max_\Theta Q_n(\cdot,\theta)$ $a.s. - P$ be measurable and assume there exists $\left\{\overline{Q}_n : \Theta \to \overline{\mathbb{R}}\right\}$ with identifiably unique maximizers $\{\theta_n^o\}$ such that (a3) $\sup_{\theta\in\Theta}\left|Q_n(\cdot,\theta) - \overline{Q}_n(\theta)\right| \to 0$ prob $- P$. Then*

$$\hat{\theta}_n - \theta_n^o \to 0 \; prob - P. \tag{A}$$

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 such that (b1) $Q_n^*(\cdot,\omega,\theta) : \Lambda \to \overline{\mathbb{R}}$ is measurable-$\mathcal{G}$ for each (ω,θ) in $\Omega \times \Theta$; (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). Let $\left\{\hat{\theta}_n^* : \Lambda \times \Omega \to \Theta\right\}$ be 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 further that (b3) $\sup_{\theta\in\Theta}\left|Q_n^*(\cdot,\omega,\theta) - Q_n(\omega,\theta)\right| \to 0$ prob $- P_{n,\omega}^*$, prob $- P$. Then

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

Lemma A.3 (Asymptotic Normality of $\hat{\theta}_n^*$). *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 such that (a1) $Q_n(\cdot,\theta) : \Omega \to \overline{\mathbb{R}}$ is measurable-$\mathcal{F}$ for each $\theta \in \Theta$; (a2) $Q_n(\omega,\cdot) : \Theta \to \overline{\mathbb{R}}$ is continuously differentiable of order 2 on Θ $a.s. - P$. Let $\hat{\theta}_n = \arg\max_\Theta Q_n(\cdot,\theta)$ $a.s. - P$ be measurable such that $\hat{\theta}_n - \theta_n^o \to 0$ 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) $\sup_{\theta\in\Theta}\left|\nabla^2 Q_n(\cdot,\theta) - A_n(\theta)\right| \to 0$ prob $- P$, 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 such that (b1) $Q_n^*(\cdot,\omega,\theta) : \Lambda \to \overline{\mathbb{R}}$ is measurable-$\mathcal{G}$ for each (ω,θ) in $\Omega \times \Theta$; (b2) $Q_n^*(\lambda,\omega,\cdot) : \Theta \to \overline{\mathbb{R}}$ is continuously differentiable of order 2 on Θ $a.s. - P$. For each $n = 1, 2, \ldots$, let $\hat{\theta}_n^*(\cdot,\omega) = \arg\max_\Theta Q_n^*(\cdot,\omega,\theta)$ $a.s. - P$ be measurable such that $\hat{\theta}_n^*(\cdot,\omega) - \hat{\theta}_n(\omega) \to 0$, 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)* $\sup_{\theta\in\Theta} \left|\nabla^2 Q_n^*(\cdot,\omega,\theta) - \nabla^2 Q_n(\omega,\theta)\right| \to 0\ prob - P_{n,\omega}^*,\ prob - P$. *Then*

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

Lemma A.4 (Bootstrap Uniform WLLN). *Let* $\{q_{nt}^*(\cdot,\omega,\theta)\}$ *be a MBB resample of* $\{q_{nt}(\omega,\theta)\}$ *and assume: (a) For each* $\theta \in \Theta \subset \mathbb{R}^p$, Θ *a compact set,* $n^{-1}\sum_{t=1}^n (q_{nt}^*(\cdot,\omega,\theta) - q_{nt}(\omega,\theta)) \to 0,\ prob - P_{n,\omega}^*,\ prob - P$; *and (b)* $\forall \theta, \theta^o \in \Theta$, $|q_{nt}(\cdot,\theta) - q_{nt}(\cdot,\theta^o)| \le L_{nt}|\theta - \theta^o|\ a.s. - P$, *where* $\sup_n \left\{n^{-1}\sum_{t=1}^n E(L_{nt})\right\} = O(1)$. *Then, if* $\ell_n = o(n)$, *for any* $\delta > 0$ *and* $\xi > 0$,

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

Lemma A.5 (Bootstrap Pointwise WLLN). *For some* $r > 2$, *let* $\{q_{nt} : \Omega \times \Theta \to \mathbb{R}\}$ *be such that for all* n, t, *there exists* $D_{nt} : \Omega \to \mathbb{R}$ *with* $|q_{nt}(\cdot,\theta)| \le D_{nt}$ *for all* $\theta \in \Theta$ *and* $\|D_{nt}\|_r \le \Delta < \infty$. *For each* $\theta \in \Theta$ *let* $\{q_{nt}^*(\cdot,\omega,\theta)\}$ *be a MBB resample of* $\{q_{nt}(\omega,\theta)\}$. *If* $\ell_n = o(n)$, *then for any* $\delta > 0$, $\xi > 0$, *and for each* $\theta \in \Theta$,

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

Lemma A.6. *Let* $\{Q_n : \Omega \times \Theta \to \mathbb{R}\}$ *be a sequence of functions continuous on* Θ *a.s.* $- P$ *and let* $\left\{\hat{\theta}_n : \Omega \to \Theta\right\}$ *be such that* $\hat{\theta}_n - \theta_n^o \to 0\ prob - P$. *Suppose* $\sup_{\theta\in\Theta}\left|Q_n(\cdot,\theta) - \overline{Q}_n(\theta)\right| \to 0\ prob - P$ *where* $\left\{\overline{Q}_n : \Theta \to \mathbb{R}\right\}$ *is continuous on* Θ *uniformly in* n. *Then*

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

For each $\omega \in \Omega$, *let* $\left(\Lambda, \mathcal{G}, P_{n,\omega}^*\right)$ *be a complete probability space. If* $\hat{\theta}_n^*(\cdot,\omega) - \hat{\theta}_n(\omega) \to 0\ prob - P_{\omega,n}^*$, $prob - P$ *and* $\sup_{\theta\in\Theta}|Q_n^*(\cdot,\omega,\theta) - Q_n(\omega,\theta)| \to 0\ prob - P_{n,\omega}^*,\ prob - P$, *then*

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

Proof of Theorem 2.1. We apply Lemma A.2 with $Q_n(\cdot,\theta) = n^{-1}\sum_{t=1}^n q_{nt}(\cdot,\theta)$ and $Q_n^*(\cdot,\omega,\theta) = n^{-1}\sum_{t=1}^n q_{nt}^*(\cdot,\omega,\theta)$, where $q_{nt}(\cdot,\theta) \equiv \log f_{nt}\left(X_n^t(\cdot),\theta\right)$, and $\{q_{nt}^*(\cdot,\omega,\theta)\}$ is the MBB resample. Conditions (a1)-(a3) are readily verified under Assumption A. Assumption A.2. implies (b1) and (b2). To verify (b3) apply Lemmas A.4 and A.5, noting that $\ell_n = o(n)$. ■

Proof of Theorem 2.2. We apply Lemma A.3 with the same choices of $Q_n(\cdot,\theta)$ and $Q_n^*(\cdot,\omega,\theta)$ as in Theorem 2.1. The result follows then by Polya's theorem (e.g. Serfling, 1980, p. 20) since $C_n^o = A_n^{o-1} B_n^o A_n^{o-1}$ is $O(1)$ and the normal distribution is everywhere continuous. (a1)-(a4) can be verified as in Theorem 5.7 of Gallant and White (1988). (b1) and (b2) follow from assumption A.2. Lemmas A.4 and A.5 imply (b4) given A.4(ii) and A.5(iii) and the conditions on ℓ_n. Lastly, we verify (b3). We have that (for any n and any ω)

$$n^{-1/2}\sum_{t=1}^{n}s_{nt}^{*}\left(\cdot,\omega,\hat{\theta}_n\right)=\xi_{1n}\left(\cdot,\omega\right)+\xi_{2n}\left(\omega\right)+\xi_{3n}\left(\cdot,\omega\right)+\xi_{4n}\left(\omega\right),$$

where $\xi_{1n}(\cdot,\omega)\equiv n^{-1/2}\sum_{t=1}^{n}\left(s_{nt}^{*}\left(\cdot,\omega,\theta_n^o\right)-s_{nt}\left(\omega,\theta_n^o\right)\right);\xi_{2n}(\omega)\equiv -n^{-1/2}\sum_{t=1}^{n}\left(s_{nt}\left(\omega,\hat{\theta}_n\right)-s_{nt}\left(\omega,\theta_n^o\right)\right)$; $\xi_{3n}(\cdot,\omega)\equiv n^{-1/2}\sum_{t=1}^{n}\left(s_{nt}^{*}\left(\cdot,\omega,\hat{\theta}_n\right)-s_{nt}^{*}\left(\cdot,\omega,\theta_n^o\right)\right)$; and $\xi_{4n}(\omega)\equiv n^{-1/2}\sum_{t=1}^{n}s_{nt}\left(\omega,\hat{\theta}_n\right)$. The F.O.C. for $\hat{\theta}_n$ and assumption A.3(ii) imply that $\xi_{4n}(\omega)=0$ for all n sufficiently large $a.s.-P$ (see Theorem 3.13 of White, 1994). Also, Theorem 2.2 of Gonçalves and White (2002) implies that $B_n^{o-1/2}\xi_{1n}(\cdot,\omega)\Rightarrow^{d_{P^*_{n,\omega}}} N(0,I_p)$ $prob-P$ under Assumption A strengthened by 2.1 and 2.2. Thus, it suffices to show that $\xi_{2n}(\omega)+\xi_{3n}(\cdot,\omega)\to 0$ $prob-P^*_{n,\omega}, prob-P$. Two mean value expansions yield

$$\xi_{2n}(\omega)+\xi_{3n}(\cdot,\omega)=\zeta_n(\cdot,\omega)\sqrt{n}\left(\hat{\theta}_n(\omega)-\theta_n^o\right),$$

where $\zeta_n(\cdot,\omega)=n^{-1}\sum_{t=1}^{n}\left(\nabla' s_{nt}^{*}\left(\cdot,\omega,\bar{\theta}_n^{*}\right)-\nabla' s_{nt}\left(\omega,\bar{\theta}_n\right)\right)$, with $\bar{\theta}_n$ and $\bar{\theta}_n^{*}$ (possibly different) mean values lying between $\hat{\theta}_n$ and θ_n^o. Lemma A.6 implies $\zeta_n(\cdot,\omega)\to 0$ $prob-P^*_{n,\omega}, prob-P$, given the uniform convergence of $\left\{\nabla^2 Q_n^{*}(\cdot,\omega,\theta)-\nabla^2 Q_n(\omega,\theta)\right\}$ and $\left\{\nabla^2 Q_n(\omega,\theta)-A_n(\theta)\right\}$, and the convergences of $\bar{\theta}_n-\theta_n^o$ and $\bar{\theta}_n^{*}-\theta_n^o$ to zero. Since $\sqrt{n}\left(\hat{\theta}_n(\omega)-\theta_n^o\right)=O_P(1)$, it follows that $\xi_{2n}(\omega)+\xi_{3n}(\cdot,\omega)\to 0$ $prob-P^*_{n,\omega}, prob-P$. ■

Proof of Corollary 2.1. We can show $\sqrt{n}\left(\hat{\theta}_{1n}^{*}-\hat{\theta}_n\right)-\sqrt{n}\left(\hat{\theta}_n^{*}-\hat{\theta}_n\right)\to 0$ $prob-P^*_{n,\omega}$, $prob-P$, given the definition of $\hat{\theta}_{1n}^{*}$ and the fact that $A_n^{*}\left(\hat{\theta}_n\right)-\hat{A}_n\to 0$ $prob-P^*_{n,\omega}$, $prob-P$, by Lemma A.6. ■

Proof of Lemma A.1. 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 Theorem 3.4 of White (1994), $P(F)=1$. Fix $\varepsilon'>0$ and ω in F. Then, 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 $\eta(\theta_n^o,\varepsilon')\equiv\{\theta\in\Theta:|\theta-\theta_n^o|<\varepsilon'\}$. By Corollary 3.8 of White (1994), there exists $N_2\left(\omega,\delta'(\varepsilon')\right)<\infty$ such that for all $n>N_2\left(\omega,\delta'(\varepsilon')\right)$, $\left|Q_n\left(\omega,\hat{\theta}_n(\omega)\right)-\overline{Q}_n(\theta_n^o)\right|<\frac{\delta'(\varepsilon')}{4}$. Also, for all $n>N_2\left(\omega,\delta'(\varepsilon')\right)$, $\max_{\eta^c(\theta_n^o,\varepsilon')}Q_n(\omega,\theta)\leq\max_{\eta^c(\theta_n^o,\varepsilon')}\overline{Q}_n(\theta)+\frac{\delta'(\varepsilon')}{4}$. Let $N(\omega,\varepsilon')=\max\left\{N_0(\omega,\varepsilon'),N_1(\varepsilon'),N_2\left(\omega,\delta'(\varepsilon')\right)\right\}$. Hence

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

Set $\varepsilon=2\varepsilon'$ and $\delta(\varepsilon)=\frac{\delta'(\varepsilon/2)}{2}>0$ to obtain the result for all ω in F and $P(F)=1$. If instead $\sup_{\Theta}\left|Q_n(\omega,\theta)-\overline{Q}_n(\theta)\right|\to 0$ $prob-P$, then for any $\{n'\}$ there exists $\{n''\}$ such that $\sup_{\Theta}\left|Q_{n''}(\omega,\theta)-\overline{Q}_{n''}(\theta)\right|=o(1)$ and $\hat{\theta}_{n''}(\omega)-\theta_{n''}^o=o(1)$ $a.s.-P$. The result thus holds for $\{n''\}$ $a.s.-P$. ■

Proof of Lemma A.2. (A) follows by Theorem 3.4 of White (1994) under (a1)-(a3). To prove (B), note that for any subsequence $\{n'\}$, by Lemma A.1 there exists a further subsequence $\{n''\}$ such that $\left\{\hat{\theta}_{n''}\right\}$ is identifiably unique $a.s. - P$, given (a1)-(a3). Now apply Theorem 3.4 of White (1994). ■

Proof of Lemma A.3. (A) follows by White's (1994) Theorem 6.2 under (a1)-(a4). To prove (B), it suffices to show that for any subsequence $\{n'\}$ there exists a further subsequence $\{n''\}$ such that $B_{n''}^{*-1/2} A_{n''}^{*} \sqrt{n''}\left(\hat{\theta}_{n''}^{*}(\cdot,\omega) - \hat{\theta}_{n''}(\omega)\right) \Rightarrow^{d_{P^{*}_{n'',\omega}}} N(0, I_p)$, $a.s. - P$. This follows by applying White's (1994) Theorem 6.2 to an appropriately chosen subsequence, for given ω in some appropriate $\mathfrak{F}$ such that $P(\mathfrak{F}) = 1$. ■

Proof of Lemma A.4. The proof closely follows that of Lemma 8 of Hall and Horowitz (1996).

Proof of Lemma A.5. Fix $\theta \in \Theta$, and write $n^{-1}\sum_{t=1}^{n}\left(q_{nt}^{*}(\theta) - q_{nt}(\theta)\right) = Q_{1n} + Q_{2n}$, with

$$Q_{1n} \equiv n^{-1}\sum_{t=1}^{n}\left(q_{nt}^{*}(\theta) - E^{*}\left(q_{nt}^{*}(\theta)\right)\right), \text{ and } Q_{2n} \equiv E^{*}\left(n^{-1}\sum_{t=1}^{n} q_{nt}^{*}(\theta)\right) - n^{-1}\sum_{t=1}^{n} q_{nt}(\theta),$$

where we omit ω to conserve space. $Q_{2n} \to 0\ prob - P$ since $E^{*}\left(n^{-1}\sum_{t=1}^{n} q_{nt}^{*}(\theta)\right) = n^{-1}\sum_{t=1}^{n} q_{nt}(\theta) + O_P\left(\frac{\ell}{n}\right)$ (cf. Lemma A.1 of Fitzenberger (1997)) and $\frac{\ell}{n} \to 0$. By Chebyshev's inequality, for any $\delta > 0$, $P_{n,\omega}^{*}\left(|Q_{1n}| > \delta\right) \leq \frac{1}{\delta^2} n^{-1} \text{var}^{*}\left(n^{-1/2}\sum_{t=1}^{n} q_{nt}^{*}(\theta)\right)$, where $\text{var}^{*}\left(n^{-1/2}\sum_{t=1}^{n} q_{nt}^{*}(\theta)\right)$ has a closed form expression involving products of $q_{nt}(\theta)$ and $q_{n,t+\tau}(\theta)$ (cf. Gonçalves and White (2002), equation (2.1)). Under the domination condition on $\{q_{nt}(\theta)\}$ and the properties of the MBB, repeated application of Minkowski and Hölder's inequalities yields $\left\|\text{var}^{*}\left(n^{-1/2}\sum_{t=1}^{n} q_{nt}^{*}(\theta)\right)\right\|_{\frac{r}{2}} = O(\ell)$ for some $r > 2$. Thus, by Markov's inequality, $P\left[P_{n,\omega}^{*}\left(|Q_{1n}| > \delta\right) > \xi\right] = O\left(\left(\frac{\ell}{n}\right)^{r/2}\right) \to 0$ given $\ell = o(n)$. ■

Proof of Lemma A.6. (A) holds by White (1994, Corollary 3.8). To prove (B), by the continuity of $\bar{Q}_n$ on Θ uniformly in n, given $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$ independent of n such that $\left|\bar{Q}_n(\theta) - \bar{Q}_n\left(\hat{\theta}_n\right)\right| > \varepsilon/3$ implies $\left|\theta - \hat{\theta}_n\right| > \delta(\varepsilon)$. So

$$P\left[P_{n,\omega}^{*}\left(\left|Q_n^{*}\left(\cdot,\omega,\hat{\theta}_n^{*}\right) - Q_n\left(\omega,\hat{\theta}_n\right)\right| > \varepsilon\right) > \varepsilon\right] \leq P\left[P_{n,\omega}^{*}\left(\sup_{\Theta}\left|Q_n^{*}(\cdot,\omega,\theta) - Q_n(\omega,\theta)\right| > \varepsilon/3\right) > \varepsilon/3\right]$$

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

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

Appendix B: Proofs for Section 3

Throughout Appendix B, C denotes a generic constant. 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.

Lemma B.1 (Studentization of the sample mean). *Let $\{X_{nt}\}$ satisfy Assumptions 2.1′ and 2.2 of Gonçalves and White (2002), with Assumption 2.2 strengthened by*

A.2.2′ $n^{-1}\sum_{t=1}^{n}|\mu_{nt}-\bar{\mu}_n|^{2+\delta}=o\left(\ell_n^{-1-\delta/2}\right)$ *for some δ such that $0<\delta\leq 2$,*

where $\mu_{nt}\equiv E(X_{nt})$ and $\bar{\mu}_n\equiv n^{-1}\sum_{t=1}^{n}\mu_{nt}$.

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

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

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

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 some $\Delta<\infty$. 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 of Theorem 3.1. By Gallant and White's (1988) Theorem 7.5, if $\hat{B}_n$ used in forming $\mathcal{W}_n$ is consistent for B_n^o, then $\mathcal{W}_n\Rightarrow\mathcal{X}_q^2$ under H_o. Next, we prove $\mathcal{W}_n^*\Rightarrow^{d_{P^*}}\mathcal{X}_q^2$ *prob* $-P$. A mean value expansion of $r_n\left(\hat{\theta}_n^*\right)$ around $\hat{\theta}_n$ yields $\sqrt{n}\left(r_n\left(\hat{\theta}_n^*\right)-r_n\left(\hat{\theta}_n\right)\right)\Rightarrow^{d_{P^*}}N\left(0,R_n^oC_n^oR_n^{o\prime}\right)$ *prob* $-P$, implying $n\left(\hat{r}_n^*-\hat{r}_n\right)'\left(R_n^oC_n^oR_n^{o\prime}\right)^{-1}\left(\hat{r}_n^*-\hat{r}_n\right)\Rightarrow^{d_{P^*}}\mathcal{X}_q^2$ *prob*$-P$. Thus, it suffices to prove: (i) $\hat{R}_n^*-R_n^o\to 0$ *prob* $-P^*$, *prob* $-P$; (ii) $\hat{A}_n^*-A_n^o\to 0$ *prob* $-P^*$, *prob* $-P$; and (iii) $\hat{B}_n^*-B_n^o\to 0$ *prob* $-P^*$, *prob* $-P$. (i) follows by continuity of r_n on Θ (uniformly in n) and because $\hat{\theta}_n^*-\theta_n^o\to 0$ *prob* $-P^*$, *prob* $-P$ by Theorem 2.1; similarly, by Theorem 2.1 and Lemma A.6, we have $\hat{A}_n^*-\hat{A}_n\to 0$ *prob* $-P^*$, *prob* $-P$, which implies (ii) since $\hat{A}_n-A_n^o\to 0$ *prob* $-P$. To prove (iii), consider

$$\begin{aligned}\tilde{B}_n^{*o}&=k^{-1}\sum_{i=1}^{k}\left(\ell^{-1/2}\sum_{t=1}^{\ell}\left(s_{n,I_i+t}\left(X_n^{I_i+t},\theta_n^o\right)-\bar{s}_n^{*o}\right)\right)\left(\ell^{-1/2}\sum_{t=1}^{\ell}\left(s_{n,I_i+t}\left(X_n^{I_i+t},\theta_n^o\right)-\bar{s}_n^{*o}\right)\right)'\\&=k^{-1}\sum_{i=1}^{k}\ell^{-1}\sum_{t=1}^{\ell}s_{n,I_i+t}\left(X_n^{I_i+t},\theta_n^o\right)\sum_{t=1}^{\ell}s'_{n,I_i+t}\left(X_n^{I_i+t},\theta_n^o\right)-\ell\bar{s}_n^{*o}\bar{s}_n^{*o\prime},\end{aligned}\tag{B.1}$$

where $\bar{s}_n^{*o}=n^{-1}\sum_{t=1}^{n}s_{nt}^*\left(\theta_n^o\right)$. By Lemma B.1 $\tilde{B}_n^{*o}-B_{n,1}^o\to 0$, *prob* $-P^*$,*prob* $-P$, where $B_{n,1}^o=\operatorname{var}^*\left(n^{-1/2}\sum_{t=1}^{n}s_{nt}^*\left(\theta_n^o\right)\right)$. By Gonçalves and White's (2002) Corollary 2.1, $B_{n,1}^o-B_n^o\to 0$, *prob* $-P$, implying $\tilde{B}_n^{*o}-B_n^o\to 0$, *prob* $-P^*$,*prob* $-P$. Thus, it suffices that $\hat{B}_n^*-\tilde{B}_n^{*o}\to 0$, *prob* $-P^*$,*prob* $-P$.

From (3.1) and (B.1) we can write $\hat{B}_n^* - \tilde{B}_n^{*o} = D_1 + D_2$, where

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

and $D_2 \equiv \ell\bar{s}_n^{*o}\bar{s}_n^{*o\prime}$. Note that $\bar{s}_n^{*o} = B_n^{o1/2}B_n^{o-1/2}\left(\bar{s}_n^{*o}-\bar{s}_n^o\right) + B_n^{o1/2}B_n^{o-1/2}\bar{s}_n^o \equiv E_1 + E_2$, with $\bar{s}_n^o = n^{-1}\sum_{t=1}^n s_{nt}^o$. By Theorem 2.2 of Gonçalves and White's (2002), $B_n^{o-1/2}\sqrt{n}\left(\bar{s}_n^{*o}-\bar{s}_n^o\right) \Rightarrow^{d_{P^*}} N\left(0,I_p\right)$ $prob-P$ 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 $E\left(\bar{s}_n^0\right)=0$ by the F.O.C. that define θ_n^o, it follows that $E_2 = O_P\left(n^{-1/2}\right)$. Thus, $D_2 \to 0$ $prob-P^*, prob-P$. To show that $D_1 \to 0$ $prob-P^*, prob-P$ we take a mean value expansion about θ_n^o of a typical element of D_1 and apply Lemma B.2 twice with $X_{nt} = \sup_{\theta\in\Theta}\left|\frac{\partial}{\partial\theta'}s_{n,t,j}\left(X_n^t,\theta\right)\right|$ and $Z_{nt} = \sup_{\theta\in\Theta}\left|s_{n,t,j}\left(X_n^t,\theta\right)\right|$, where j indexes the j^{th} element of the score. ■

Proof of Theorem 3.2. The proof follows Gallant and White (1988), Theorem 7.9 (p. 128) using Lemmas B.1 and B.2.

Proof of Lemma B.1. The proof consists of two steps. *Step 1*: show $\tilde{\sigma}_n^{*2} - \hat{\sigma}_n^2 \to 0$ $prob-P^*, prob-P$, where $\tilde{\sigma}_n^{*2} = k^{-1}\sum_{i=1}^k\left(\ell^{-1/2}\sum_{t=1}^{\ell}\left(X_{I_i+t}-\bar{X}_{\alpha,n}\right)\right)^2$, with $\bar{X}_{\alpha,n} = E^*\left(\bar{X}_n^*\right)$; *Step 2*: show $\hat{\sigma}_n^{*2} - \tilde{\sigma}_n^{*2} \to 0$ $prob-P^*, prob-P$. 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 $\hat{\sigma}_n^{*2} = k^{-1}\sum_{i=1}^k \hat{A}_{I_i}^2$ and $\tilde{\sigma}_n^{*2} = k^{-1}\sum_{i=1}^k A_{I_i}^2$. To prove step 1, by two applications of Markov's inequality it suffices to show $E\left(E^*\left|\tilde{\sigma}_n^{*2}-\hat{\sigma}_n^2\right|^p\right) = o(1)$ for some $p>1$. We take $p = 1+\delta/2$ with $0<\delta\leq 2$. Since $E^*\left(\tilde{\sigma}_n^{*2}\right) = E^*\left(A_{I_1}^2\right) = (n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}A_i^2 \equiv \hat{\sigma}_n^2$ (cf. Künsch (1989, Theorems 3.1 and 3.4)), we have

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

by Burkholder's inequality, because $\left\{A_{I_i}^2 - E^*\left(A_{I_1}^2\right)\right\}$ are (conditionally) i.i.d. zero mean. For $1<p\leq 2$, $x\geq 0$ and $y \geq 0$, the inequality $(x+y)^{p/2} \leq x^{p/2}+y^{p/2}$ implies $E^*\left|\sum_{i=1}^k\left(A_{I_i}^2 - E^*\left(A_{I_1}^2\right)\right)^2\right|^{p/2} \leq kE^*\left|A_{I_1}^2 - E^*\left(A_{I_1}^2\right)\right|^p$ so that $E^*\left|\tilde{\sigma}_n^{*2}-\hat{\sigma}_n^2\right|^p \leq 2^pCk^{-(p-1)}E^*\left|A_{I_1}\right|^{2p}$. Thus, it suffices that $k^{-(p-1)}E\left(E^*\left|A_{I_1}\right|^{2p}\right) = o(1)$. Some algebra yields $E\left(E^*\left|A_{I_1}\right|^{2p}\right) \leq C\left(F_1+F_2+F_3\right)$, where

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

and $F_3 = (n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}\ell^{-p}E\left(\left|\ell\bar{Z}_{\alpha,n}\right|^{2p}\right)$, with $Z_{nt} \equiv X_{nt}-\mu_{nt}$, and for any $\{\mathcal{Y}_{nt}\}$ $\bar{\mathcal{Y}}_{\alpha,n} = (n-\ell+1)^{-1}\sum_{i=0}^{n-\ell}\ell^{-1}\sum_{t=1}^{\ell}\mathcal{Y}_{n,i+t} = \sum_{t=1}^n\alpha_{nt}\mathcal{Y}_{nt}$ with $\alpha_{nt} = \frac{1}{(n-\ell+1)\ell}\min\{t,\ell,n-t+1\}$. Under Assumption 2.1$'$, $E\left|\sum_{t=1}^{\ell}Z_{n,i+t}\right|^{2p} < C\ell^p$ (cf. Gonçalves and White (2002), p. 1384 for a similar argument), implying $k^{-(p-1)}F_1 = O\left(\left(\frac{\ell}{n}\right)^{p-1}\right) = o(1)$, and similarly for $k^{-(p-1)}F_3$. If $\mu_{nt} = \mu$ for all t,

$F_2 = 0$ because $\bar{\mu}_{\alpha,n} = \sum_{t=1}^{n} \alpha_{nt}\mu = \mu$ as $\sum_{t=1}^{n} \alpha_{nt} = 1$. Otherwise, we can show that assumption A.2.2$'$ implies $F_2 = o(1)$, so $k^{-(p-1)}F_2 = o(1)$. To prove step 2, note that $\hat{A}_{I_i} = \sqrt{\ell}\left(\bar{X}_{I_i} - \bar{X}_n^*\right)$, where $\bar{X}_{I_i} = \ell^{-1}\sum_{t=1}^{\ell} X_{I_i+t}$, and $A_i = \sqrt{\ell}\left(\bar{X}_{I_i} - \bar{X}_{\alpha,n}\right)$, implying $\hat{\sigma}_n^{*2} - \tilde{\sigma}_n^{*2} = -\ell\left(\bar{X}_n^* - \bar{X}_{\alpha,n}\right)^2 = O_{P^*}\left(\frac{\ell}{n}\right) \to 0$ $prob - P$, since $\sqrt{n}\left(\bar{X}_n^* - \bar{X}_{\alpha,n}\right)$ converges in distribution under P^* with probability P approaching one, by Theorem 2.2 of Gonçalves and White (2002). ■

Proof of Lemma B.2. 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 $1 < p \leq 2$, it suffices to show $n^{-p/2}E^*\left(\left|k^{-1}\sum_{i=1}^{k}\ell^{-1}S_{n,I_i}^1 S_{n,I_i}^2\right|^p\right) \to 0$ $prob - P$. Let $p = 1 + \delta/2$, $0 < \delta \leq 2$, and note $E^*\left(\left|k^{-1}\sum_{i=1}^{k}\ell^{-1}S_{n,I_i}^1 S_{n,I_i}^2\right|^p\right) \leq C\left(F_1 + F_2\right)$, with

$$F_1 = E^*\left(\left|k^{-1}\sum_{i=1}^{k}\ell^{-1}\left(S_{n,I_i}^1 S_{n,I_i}^2 - E^*\left(S_{n,I_i}^1 S_{n,I_i}^2\right)\right)\right|^p\right) \text{ and } F_2 = E^*\left|k^{-1}\sum_{i=1}^{k}\ell^{-1}E^*\left(S_{n,I_i}^1 S_{n,I_i}^2\right)\right|^p.$$

By the Burkholder and c_r-inequalities $F_1 \leq C\ell^{-p}k^{-(p-1)}E^*\left|S_{n,I_1}^1 S_{n,I_1}^2\right|^p \leq C\ell^{-p}E^*\left|S_{n,I_1}^1 S_{n,I_1}^2\right|^p$, since $\left\{S_{n,I_i}^1 S_{n,I_i}^2 - E^*\left(S_{n,I_i}^1 S_{n,I_i}^2\right)\right\}$ are i.i.d. zero mean, and $k^{-(p-1)} \leq 1$ for $p > 1$. Similarly, $F_2 \leq C\ell^{-p}E^*\left|S_{n,I_1}^1 S_{n,I_1}^2\right|^p$. By the Cauchy-Schwarz and Minkowski inequalities, $E\left(E^*\left|S_{n,I_1}^1 S_{n,I_1}^2\right|^p\right) \leq C\ell^{2p}$. Thus, $F_1 + F_2 = O_P\left(\ell^p\right)$, and so $n^{-p/2}\left(F_1 + F_2\right) = O_P\left(\left(\frac{\ell}{n^{1/2}}\right)^p\right) = o_p(1)$, since $\ell = o\left(n^{1/2}\right)$. ■

References

Andrews, D.W.K., 1988. Laws of large numbers for dependent nonidentically distributed random variables, Econometric Theory 4, 458-467.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica, 59, 817-858.

Andrews, D.W.K., 2002. Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators, Econometrica, 69, 119-162.

Billingsley, P., 1995. Probability and Measure. New York: Wiley.

Corradi, V., and N. R. Swanson, 2003a. Bootstrap conditional distribution tests in the presence of dynamic misspecification, Rutgers University, manuscript.

Corradi, V., and N. R. Swanson, 2003b. A test of comparing multiple misspecified conditional distributions, Rutgers University, manuscript.

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.

Davidson, R., and J.G. MacKinnon, 1999. Bootstrap testing in nonlinear models, International Economic Review, 40, 487-508.

Fitzenberger, B., 1997. The moving blocks bootstrap and robust inference for linear least squares and quantile regressions, Journal of Econometrics, 82, 235-287.

Gallant, A.R., and H. White, 1988. A unified theory of estimation and inference for nonlinear dynamic models. New York: Basil Blackwell.

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.

Gonçalves, S., and H. White, 2000. Maximum likelihood and the bootstrap for nonlinear dynamic models, UCSD discussion paper 2000-32.

Gonçalves, S., and H. White, 2003. Bootstrap Standard Error Estimates for Linear Regression, Université de Montréal, manuscript.

Gonçalves, S., and H. White, 2002. The bootstrap of the mean for dependent heterogeneous arrays, Econometric Theory, 18, 1367-1384.

Götze, F., and H.R. Künsch, 1996. Second-order correctness of the blockwise bootstrap for stationary observations, Annals of Statistics, 24, 1914-1933.

Gourieroux, C., Monfort, A. and A. Trognon, 1984. Estimation and test in probit models with serial correlation, in Alternative Approaches to Time Series Analysis, ed. Florens, J.P., Mouchart, M., Raoult, J.P. and L. Simar. Brussels: Publications des Facultés Universitaires Saint-Louis.

Hahn, J., 1996. A note on bootstrapping generalized method of moments estimators, Econometric Theory, 12, 187-197.

Hall, P, J. Horowitz, and B. Jing, 1995. On blocking rules for the bootstrap with dependent data, Biometrika 82, 561-74.

Hall, P., and J. Horowitz, 1996. Bootstrap critical values for tests based on generalized-method-of-moments estimators, Econometrica, 64, 891-916.

Inoue, A. and Shintani, M., 2001. Bootstrapping GMM estimators for time series. Forthcoming Journal of Econometrics.

Künsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations, Annals of Statistics, 17, 1217-1241.

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.

Newey, W.K., and K.D. West, 1987. A simple positive semi-definite, heteroskedastic and autocorrelation consistent covariance matrix, Econometrica, 55, 703-708.

Politis, D., and J. Romano, 1992. A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation, Annals of Statistics, 20, 1985-2007.

Politis, D. and J. Romano, 1994. The stationary bootstrap, Journal of American Statistics Association 89, 1303-1313.

Politis, D., Romano, J. and M. Wolf, 1997. Subsampling for heteroskedastic time series, Journal of Econometrics 81, 281-317.

Pötscher, 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.

Serfling, R.J., 1980. Approximation theorems of mathematical statistics. New York: Wiley.

Shao, J., 1992. Bootstrap variance estimators with truncation, Statistics and Probability Letters, 15, 95-101.

White, H., 1994. Estimation, inference and specification analysis. Cambridge: Cambridge University Press.

Table 1. Coverage Rates of Nominal 95% symmetric Percentile-*t* Intervals: Logit[a]

k	n	ρ	OP	BT	QS	MBB	Avg. block size
2	50	0.0	95.3	94.4	94.4	94.9	1.7
		0.5	90.2	91.5	91.9	92.5	2.0
		0.9	80.7	85.2	85.7	90.8	3.0
	100	0.0	95.2	94.6	94.6	94.9	1.7
		0.5	89.6	92.0	92.5	92.6	2.4
		0.9	82.0	89.0	90.0	91.9	4.1
3	50	0.0	94.7	93.9	93.9	94.7	1.8
		0.5	90.4	91.4	91.4	92.8	2.2
		0.9	81.1	84.8	84.8	91.8	3.3
	100	0.0	95.2	94.8	94.7	95.1	1.8
		0.5	90.2	92.0	92.4	93.2	2.7
		0.9	81.5	88.6	89.5	92.1	4.4
4	50	0.0	94.9	94.1	94.0	95.0	1.8
		0.5	90.2	90.6	90.4	92.8	2.3
		0.9	81.3	83.6	83.2	92.2	3.3
	100	0.0	95.3	94.8	94.6	95.2	1.9
		0.5	90.1	92.2	92.4	93.3	2.8
		0.9	82.1	88.2	88.8	92.3	4.4
5	50	0.0	94.2	93.5	93.1	94.4	1.8
		0.5	89.2	89.6	89.2	92.6	2.4
		0.9	80.8	82.7	82.3	94.2	3.2
	100	0.0	94.7	94.3	94.2	95.0	1.9
		0.5	89.7	91.5	91.8	93.0	2.8
		0.9	81.3	86.9	87.5	91.8	4.4

[a]10,000 Monte Carlo trials with 999 bootstrap replications each.

Table 2. Coverage Rates of Nominal 95% symmetric Percentile-t Intervals: ARCH[a]

n	$\bar{\alpha}$	$\bar{\rho}$	C.I. for θ	OP	BT	QS	MBB	Avg. block size
200	0.5	0.0	$\bar{\gamma}$	94.4	94.4	94.4	95.0	1.37
			$\bar{\omega}$	93.6	93.5	93.6	95.4	
			$\bar{\alpha}$	91.8	91.5	91.4	94.9	
		0.5	$\bar{\gamma}$	83.8	90.9	91.7	93.0	4.50
			ω^o	93.8	93.2	93.2	95.6	
			α^o	92.6	92.6	92.6	95.4	
		0.9	$\bar{\gamma}$	60.3	76.0	77.0	86.4	6.49
			ω^o	91.6	91.9	91.9	95.7	
			α^o	82.4	92.4	93.0	94.6	
500	0.5	0.0	$\bar{\gamma}$	94.8	94.8	94.8	95.0	1.35
			$\bar{\omega}$	94.5	94.6	94.6	95.5	
			$\bar{\alpha}$	93.6	93.5	93.4	94.7	
		0.5	$\bar{\gamma}$	85.0	92.6	93.0	93.5	6.11
			ω^o	94.9	94.3	94.3	95.4	
			α^o	93.3	94.0	94.1	95.4	
		0.9	$\bar{\gamma}$	62.9	83.2	83.7	87.6	9.16
			ω^o	94.0	94.0	94.0	95.5	
			α^o	80.0	93.5	94.1	94.7	
200	0.9	0.0	$\bar{\gamma}$	94.5	94.4	94.4	95.1	1.34
			$\bar{\omega}$	93.2	93.2	93.2	95.3	
			$\bar{\alpha}$	92.3	92.2	92.2	94.9	
		0.5	$\bar{\gamma}$	86.7	91.3	92.0	93.2	3.54
			ω^o	93.1	92.9	92.9	95.4	
			α^o	92.0	93.0	93.2	95.4	
		0.9	$\bar{\gamma}$	65.1	77.0	78.0	88.3	5.39
			ω^o	90.3	90.6	90.6	95.4	
			α^o	69.6	86.6	88.0	89.6	
500	0.9	0.0	$\bar{\gamma}$	94.8	94.8	94.9	95.1	1.32
			$\bar{\omega}$	94.3	94.3	94.3	95.4	
			$\bar{\alpha}$	93.4	93.4	93.4	94.3	
		0.5	$\bar{\gamma}$	88.0	92.6	93.0	93.5	4.82
			ω^o	94.4	94.0	94.0	95.2	
			α^o	92.4	94.6	94.9	95.8	
		0.9	$\bar{\gamma}$	69.2	84.4	84.8	88.8	7.66
			ω^o	93.6	93.8	93.8	95.5	
			α^o	68.7	90.4	91.1	92.2	

[a]10,000 Monte Carlo trials with 999 bootstrap replications each. Pseudo-true parameters were calculated by 50,000 simulations: for $\bar{\alpha} = 0.5, (\omega^o, \alpha^o) = (0.07, 0.798)$ when $\bar{\rho} = 0.5$ and $(\omega^o, \alpha^o) = (0.017, 1.130)$ when $\bar{\rho} = 0.9$; for $\bar{\alpha} = 0.9$, $(\omega^o, \alpha^o) = (0.069, 1.192)$ when $\bar{\rho} = 0.5$ and $(\omega^o, \alpha^o) = (0.015, 1.480)$ when $\bar{\rho} = 0.9$. $\bar{\gamma} = 1.0$ and $\bar{\omega} = 0.1$ were set throughout.