Discussion Paper #2010–01

EMPIRICAL LIKELIHOOD BLOCK BOOTSTRAPPING

Jason Allen[†], Allan W. Gregory[‡], and Katsumi Shimotsu[‡♭*]
Financial Stability Department, Bank of Canada[†]
Department of Economics, Queen's University[‡]
Department of Economics, Hitotsubashi University[♭]

March 13, 2010

Abstract

Monte Carlo evidence has made it clear that asymptotic tests based on generalized method of moments (GMM) estimation have disappointing size. The problem is exacerbated when the moment conditions are serially correlated. Several block bootstrap techniques have been proposed to correct the problem, including Hall and Horowitz (1996) and Inoue and Shintani (2006). We propose an empirical likelihood block bootstrap procedure to improve inference where models are characterized by nonlinear moment conditions that are serially correlated of possibly infinite order. Combining the ideas of Kitamura (1997) and Brown and Newey (2002), the parameters of a model are initially estimated by GMM which are then used to compute the empirical likelihood probability weights of the blocks of moment conditions. The probability weights serve as the multinomial distribution used in resampling. The first-order asymptotic validity of the proposed procedure is proven, and a series of Monte Carlo experiments show it may improve test sizes over conventional block bootstrapping.

Keywords: generalized methods of moments, empirical likelihood, block bootstrap

JEL classification: C14, C22

[*]Correspondence can be sent to J. Allen (JAllen@bankofcanada.ca). We thank Don Andrews, Geoffrey Dunbar, Atushi Inoue, Gregor Smith, Silvia Gonçalves, Sharon Kozicki, Thanasis Stengos, and Tim Vogelsang for helpful comments and insightful discussion. We also thank seminar participants at the Bank of Canada, Indiana University, Queen's University, Econometric Society World Congress in London (2005), Canadian Econometric Study Group in Vancouver (2005), and Far East Meetings of the Econometric Society in Taiwan (2007). We also thank the referee for insightful comments. We acknowledge the Social Sciences and Humanities Research Council of Canada for support of this research. The views in this paper do not necessarily reflect those of the Bank of Canada. All errors are our own.

1 Introduction

Generalized method of moments (GMM, Hansen (1982)) has been an essential tool for econometricians, partly because of its straightforward application and fairly weak restrictions on the data generating process. GMM estimation is widely used in applied economics to estimate and test asset pricing models (Hansen and Singleton (1982), Kocherlakota (1990), Altonji and Segal (1996)), business cycle models (Christiano and Haan (1996)), models that use longitudinal data (Arellano and Bond (1991), Ahn and Schmidt (1995)), as well as stochastic dynamic general equilibrium models (Ruge-Murcia (2007)).

Despite the widespread use of GMM, there is ample evidence that the finite sample properties for inference have been disappointing (e.g. the 1996 special issue of JBES); t-tests on parameters and Hansen's test of overidentifying restrictions (J-test, or Sargan test) for model specification perform poorly and tend to be biased away from the null hypothesis. The situation is especially severe for dependent data (see Clark (1996)). Consequently, inferences based on asymptotic critical values can often be very misleading. From an applied perspective, this means that theoretical models may be more frequently rejected than necessary due to poor inference rather than poor modeling.

Various attempts have been made to address finite sample size problems while allowing for dependence in the data. Berkowitz and Kilian (2000), Ruiz and Pascual (2002), and Härdle, Horowitz, and Kreiss (2003) review some of the techniques developed for bootstrapping time-series models, including financial time series. Lahiri (2003) is an excellent monograph on resampling methods for dependent data. Hall and Horowitz (1996) apply the block bootstrap approach to GMM and establish the asymptotic refinements of their procedure when the moment conditions are uncorrelated after finitely many lags. Andrews (2002) provides similar results for the k-step bootstrap procedure first proposed by Davidson and Mackinnon (1999).

Limited Monte Carlo results indicate the block-bootstrap has some success at improving inference in GMM. More recent papers by Zvingelis (2003) and Inoue and Shintani (2006) attempt refinements to Hall and Horowitz (1996) and Andrews (2002). The main requirement of these earlier papers is that the data is serially uncorrelated after a finite number of lags. In contrast, Inoue and Shintani (2006) prove that the block bootstrap provides asymptotic refinements for the GMM estimator of linear models when the moment conditions are serially correlated of possibly infinite order. Zvingelis (2003) derives the optimal block length for coverage probabilities of normalized and Studentized statistics.

A complementary line of research has examined empirical likelihood (EL) estimators, or their generalization (GEL). Rather than try to improve the finite properties of the GMM estimator directly, researchers such as Kitamura (1997), Kitamura and Stutzer (1997), Smith (1997), and Imbens, Spady, and Johnson (1998) have proposed and/or tested new statistics, ones based on GEL-estimators.[1] A GEL estimator minimizes the distance between the empirical density and a synthetic density subject to the restriction that all the moment conditions are satisfied. GEL estimators have the same first-order asymptotic properties as GMM but have smaller bias than GMM in finite samples. Furthermore, these biases do not increase in the number of overidentifying restrictions in the case of GEL. Newey and Smith (2004) provide theoretical evidence of the higher-order efficiency of GEL estimators. Gregory, Lamarche, and Smith (2002) have shown, however, that these alternatives to GMM do not solve the over-rejection problem in finite samples.

Brown and Newey (2002) introduce the empirical likelihood bootstrap technique for *iid* data. Rather than resampling from the empirical distribution function, the empirical likelihood bootstrap resamples from a multinomial distribution function, where the probability weights are computed by empirical likelihood. Brown and Newey (2002) show that empirical likelihood bootstrap provides an asymptotically efficient estimator of the distribution of t ratios and overidentification test-statistics. The author's Monte Carlo design features a dynamic panel model with persistence and *iid* error structure. The results suggest that the empirical likelihood bootstrap is more accurate than the asymptotic approximation, and not dissimilar to the Hall and Horowitz (1996) bootstrap.

In this paper, the approach of Brown and Newey (2002) is extended to the case of dependent data, using the empirical likelihood (Owen (1990)). A number of researchers have implemented this approach with some success in linear time-series models (Ramalho (2006)) as well as dynamic panel data models (Gonzalez (2007)). With serially correlated data the idea is that parameters of a model are initially estimated by GMM and then used to compute the empirical likelihood probability weights of the *blocks* of moment conditions, which serve as the multinomial distribution for resampling. In this paper the first-order asymptotic validity of the proposed empirical likelihood block bootstrap is proven using the results in Gonçalves and White (2004) and the approach of Mason and Newton (1992), who analyze the consistency of generalized bootstrap (weighted bootstrap) procedures. Our consistency results may be viewed as an extension of Mason and Newton (1992) to block bootstrapping. We report on the finite-sample properties of t-ratios and overidentification test-statistics. A series of Monte Carlo experiments show that the empirical likelihood block bootstrap can reduce size distortions considerably and improve test sizes over first-order asymptotic theory and frequently outperforms conventional block bootstrapping approaches.[2] Furthermore, the empir-

[1] See Kitamura (2007) for a review of recent research on empirical likelihood methods.

[2] In addition to bootstrapping using empirical likelihood estimated weights it would seem natural to consider subsampling using the same weights. Subsampling (Politis and Romano (1994), Politis, Romano, and Wolf (1999), and Hong and Scaillet (2006)) is an alternative to bootstrapping where each block is treated as it's own series and test-statistics are calculated for each sub-series. This is left as future work.

ical likelihood block bootstrap does not require solving the difficult saddle point problem associated with GEL estimators. This is because estimation of the probability weights can be conducted by plugging-in first-stage GMM estimates. Difficulties with solving the saddle point problem is a common argument amongst applied researchers for not switching from GMM to EL, even though the latter is higher-order efficient.

In related work, Hall and Horowitz (1996) analyze an application of the block bootstrap to GMM. Hall and Horowitz (1996) assume that the moment conditions are uncorrelated after finitely many lags, and derive the higher-order improvements of the block bootstrap. The key insight of Hall and Horowitz (1996) is that, when the number of moment conditions exceeds the number of parameters, one needs to re-center the moment conditions because there is in general no parameter value such that the resampled moment conditions will be exactly equal to zero in expectation. One difference between our paper and Hall and Horowitz (1996) is that we do not assume that the moment conditions are uncorrelated after finitely many lags. Further, in the empirical likelihood block bootstrap one does not need to re-center the moment conditions by virtue of the EL weights. However, we only derive the consistency of our proposed procedure, and do not derive its higher-order properties.

The paper is organized as follows. Section 2 provides an overview of GMM and EL. Section 3 presents a discussion of how resampling methods might improve inference in GMM. Section 4 presents the asymptotic results. Section 5 presents the Monte Carlo design for both linear and nonlinear models. Section 6 concludes. The technical assumption and proofs are collected at the end of the paper in the mathematical appendix.

2 Overview of GMM and GEL

In this section we present an overview of GMM and EL to establish notation and framework.

2.1 GMM

Let $X_t \in \mathbb{R}^k, t = 1, \ldots n$, be a set of observations from a stochastic sequence. Suppose for some true parameter value θ_0 ($p \times 1$) the following moment conditions (m equations) hold and $p \leq m < n$:

$$E\left[g(X_t, \theta_0)\right] = 0, \tag{1}$$

where $g : \mathbb{R}^k \times \Theta \to \mathbb{R}^m$. The GMM estimator is defined as:

$$\hat{\theta} = \arg\min Q_n(\theta), \quad Q_n(\theta) = \left(n^{-1}\sum_{t=1}^{n} g(X_t,\theta)\right)' W_n \left(n^{-1}\sum_{t=1}^{n} g(X_t,\theta)\right), \tag{2}$$

where the weighting matrix $W_n \to_p W$. Hansen (1982) shows that the GMM estimator $\hat{\theta}$ is consistent and asymptotically normally distributed subject to some regularity conditions. The elements of $\{g(X_t,\theta)\}$ and $\{\nabla g(x,\theta)\}$ are assumed to be near epoch dependent (NED) on the α-mixing sequence $\{V_t\}$ of size -1 uniformly on (Θ,ρ) where ρ is any convenient norm on $\mathbb{R}^p$.

Define $\Sigma = \lim_{n\to\infty}\mathrm{var}(n^{-1/2}\sum_{t=1}^{n} g(X_t,\theta_0))$. The standard kernel estimate of Σ is:

$$S_n(\theta) = \sum_{h=-n}^{n} k\left(\frac{h}{m}\right)\hat{\Gamma}(h,\theta), \tag{3}$$

where $k(\cdot)$ is a kernel and $\hat{\Gamma}(h,\theta) = n^{-1}\sum_{t=h+1}^{n} g(X_t,\theta)g(X_{t+h},\theta)'$ for $h \geq 0$ and $n^{-1}\sum_{t=1}^{n-h} g(X_t,\theta)g(X_{t-h},\theta)'$ for $h < 0$. It is known that $S_n(\tilde{\theta}) \to_p \Sigma$ if $\tilde{\theta} \to_p \theta_0$ under weak conditions on the kernel and bandwidth; see de Jong and Davidson (2000).

The optimal weighting matrix is given by $S_n(\tilde{\theta})^{-1}$ with $\tilde{\theta} \to_p \theta_0$. When the optimal weighting matrix is used, the asymptotic covariance matrix of $\hat{\theta}$ is $(G'\Sigma^{-1}G)^{-1}$, where $G = \lim_{n\to\infty} E(n^{-1}\sum_{t=1}^{n}\nabla g(X_t,\theta_0))$ with $\nabla g(x,\theta) = \partial g(x,\theta)/\partial\theta'$.

In terms of testing for model misspecification, the most popular test is Hansen's J-test for overidentifying restrictions:

$$\mathcal{J}_n = K_n(\hat{\theta}_n)' K_n(\hat{\theta}_n) \to_d \chi_{m-r}, \tag{4}$$

where

$$K_n(\theta) = S_n^{-1/2} n^{-1/2} \sum_{t=1}^{n} g(X_t,\theta),$$

and S_n is a consistent estimate of Σ. Let θ_r denote the rth element of θ, and let θ_{0r} denote the rth element of θ_0. The t-statistic for testing the null hypothesis $H_0 : \theta_r = \theta_{0r}$ is:

$$T_{nr} = \frac{\sqrt{n}(\hat{\theta}_{nr} - \theta_{0r})}{\hat{\sigma}_{nr}} \to_d N(0,1), \tag{5}$$

where $\hat{\theta}_{nr}$ is the rth element of $\hat{\theta}_n$, and $\hat{\sigma}_{nr}^2$ is a consistent estimate of the asymptotic variance of $\hat{\theta}_{nr}$.

2.2 Empirical Likelihood

Empirical Likelihood (EL) estimation has some history in the statistical literature but has only recently been explored by econometricians. One attractive feature is that while its first-order asymptotic properties are the same as GMM, there is an improvement for EL at the second-order (see Qin and Lawless (1994) and Newey and Smith (2004)). For time-series models see Anatolyev (2005). This suggests that there might be some gain for EL over GMM in finite sample performance. At present, limited Monte Carlo evidence (see Gregory, Lamarche, and Smith (2002)) has provided mixed results.

The idea of EL is to use likelihood methods for model estimation and inference without having to choose a specific parametric family or probability densities. The parameters are estimated by minimizing the distance between the empirical density and a density that identically satisfies all of the moment conditions. The main advantages over GMM are that it is invariant to linear transformations of the moment functions and does not require the calculation of the optimal weighting matrix for asymptotic efficiency (although smoothing or blocking of the moment condition is necessary for dependent data). The main disadvantage is that it is computationally more demanding than GMM in that a saddle point problem needs to be solved.

The Generalized Empirical Likelihood Estimator solves the following Lagrangian:

$$\max L = \frac{1}{n}\sum_{t=1}^{n} h(\cdot) - \mu(\sum_{t=1}^{n} \pi_t - 1) - \gamma' \sum_{t=1}^{n} \pi_t g(x_t, \theta). \tag{6}$$

Solving for π_t gives

$$\pi_t = \frac{h_1(\delta' g(x_t, \theta))}{\sum h_1(\delta' g(x_t, \theta))}, \qquad h_1(v) = \partial h(v)/\partial v. \tag{7}$$

In the case of EL, $h(\cdot) = \log(\pi_t)$. The presence of serially correlated observations necessitates a modification of equation (6). Kitamura and Stutzer (1997) address the data dependency problem by smoothing the moment conditions. Anatolyev (2005) provides conditions on the amount of smoothing necessary for the bias of the GEL estimator to be less than the GMM estimator. Kitamura (1997) and Bravo (2005) address serial correlation in the moment conditions by using averages across blocks of data.

3 Improving Inference: Resampling Methods

This section presents an overview of block bootstrap methods typically used to improve inference in models estimated by GMM and follows up with a detailed proposal of a new method based on empirical likelihood.

3.1 The Block Bootstrap

The bootstrap amounts to treating the estimation data as if they were the population and generating bootstrap observations by resampling the estimation data. If the estimation data is serially correlated, then blocks of data are resampled and the blocks are treated as the *iid* sample.

We implement two forms of the block bootstrap. The first approach implements the overlapping bootstrap (MBB, Künsch (1989)). Let b be the number of blocks and ℓ the block length, such that $n = b\ell$. The ith overlapping block is $\tilde{X}_i = \{X_i, ..., X_{i+\ell-1}\}$, $i = 1, ..., n-\ell+1$. The MBB resample is $\{X_t^*\}_{t=1}^n = \{\tilde{X}_1^*, ..., \tilde{X}_b^*\}$, where $\tilde{X}_i^* \sim iid(\tilde{X}_1, ..., \tilde{X}_{n-\ell+1})$. The GMM estimator is therefore:

$$\theta_{MBB}^{**} = \arg\min Q_{MBB,n}^{**}(\theta),$$
$$Q_{MBB,n}^{**}(\theta) = \left(n^{-1}\textstyle\sum_{t=1}^n g^*(X_t^*, \theta)\right)' W_n^{**} \left(n^{-1}\textstyle\sum_{t=1}^n g^*(X_t^*, \theta)\right),$$

where $g^*(X_t^*, \theta) = g(X_t^*, \theta) - n^{-1}\sum_{t=1}^n g(X_t, \hat{\theta}_n)$ and W_n^{**} is a weighting matrix. That is, given a weighting matrix W_n^{**}, the GMM estimator that minimizes the quadratic form of the demeaned block-resampled moment conditions is θ_{MBB}^{**}.

Hall and Horowitz (1996) implement the nonoverlapping block bootstrap (NBB, Carlstein (1986)). This approach is also considered (in addition to the MBB). Let b be the number of blocks and ℓ the block length, and assume $b\ell = n$. We resample b blocks with replacement from $\{\tilde{X}_i : i = 1, \ldots, b\}$ where $\tilde{X}_i = (X_{(i-1)\ell+1}, \ldots, X_{(i-1)\ell+\ell})$. The NBB resample is $\{X_t^*\}_{t=1}^n$. The NBB version of the GMM problem is identical to the MBB version, except for the way one resamples the data. We consider both MBB and NBB approaches because there is little known about the superiority of either method in finite samples.[3]

As shown in Gonçalves and White (2004) (hereafter GW04), because the resampled b blocks are (conditionally) *iid*, the bootstrap version of the long-run autocovariance matrix estimate takes the form (cf. equation (3.1) of GW04):

$$S_n^{**}(\theta^{**}) = \ell b^{-1} \sum_{i=1}^{b} \left(\ell^{-1} \sum_{t=1}^{\ell} g^*(X_{(i-1)\ell+t}^*, \theta^{**})\right) \left(\ell^{-1} \sum_{t=1}^{\ell} g^*(X_{(i-1)\ell+t}^*, \theta^{**})\right)', \tag{8}$$

where θ^{**} denotes either θ_{MBB}^{**} or θ_{NBB}^{**}. The optimal weighting matrix is given by $(S_n^{**}(\tilde{\theta}^{**}))^{-1}$, where $\tilde{\theta}^{**}$ is the first-stage MBB/NBB estimator. The bootstrap version of the J-statistic, $\mathcal{J}_{MBB,n}^{**}$ and

[3]It is only known that the MBB is more efficient than the NBB in estimating the variance (Lahiri (1999)).

$\mathcal{J}^{**}_{NBB,n}$, is defined analogously to $\mathcal{J}_n$ but using $(S^{**}_n(\tilde{\theta}^{**}))^{-1/2}$ and $n^{-1/2}\sum_{t=1}^{n} g^*(X_t^*, \theta)$.

Note that in Hall and Horowitz (1996), the recentering of the sample moment condition is necessary in order to establish the asymptotic refinements of the bootstrap. This is because in general there is no θ such that $E^* g(x,\theta) = 0$ when there are more moments than parameters and the resampling schemes must impose the null hypothesis. Recentering is not necessary for establishing the first-order validity of the bootstrap version of $\hat{\theta}_n$ (see Hahn (1996)), but is necessary for the first-order validity of the bootstrap J-test.

Operationally one needs to choose a block size when implementing the block-bootstrap. Härdle, Horowitz, and Kreiss (2003) point out that the optimal block length depends on the objective of bootstrapping. That is, the block length depends on whether or not one is interested in bootstrapping one-sided or two-sided tests or whether one is concerned with estimating a distribution function. Among others, Zvingelis (2003) solves for optimal block lengths given different scenarios. Practically, the optimal block lengths for each different hypothesis test are unlikely to be implemented since practitioner's are interested in a variety of problems across various hypotheses. Experimentation is done with fixed block lengths as well as data-dependent methods.

Following the literature we recommend using a data-dependent approach for selecting a block length. We set the block length equal to the data-driven lag length for the Bartlett kernel using the method proposed by Newey and West (1994). This is motivated by the asymptotic equivalence of the bootstrap variance to a Bartlett kernel variance estimator (see Bühlmann and Künsch (1999), equation (2.5)). Gonçalves and White (2004) use the automatic bandwidth selection procedure proposed by Andrews (1991) in their simulation study for similar reasons. There may be some gain in using a more advanced algorithm than the one we currently employ but given its simplicity and availability in pre-packaged GMM software, we believe that most practitioners are likely to continue using a Newey-West type lag-selection procedure.[4] A number of approaches that are particular to block bootstrapping, but under different conditions than our model, have been suggested. Berkowitz and Kilian (2000) propose a two-step parametric approach for linear models and Politis and White (2004) propose an automatic block-length selection procedure based on spectral estimation (Politis and Romano (1995)), and which is appropriate for the circular and stationary bootstrap (Politis and Romano (1994)).

[4]Note that in the case of covariance matrix estimation there is also the issue of smoothing, and therefore the choice of the appropriate kernel. The block samples in our approach, however, are (conditionally) *iid*, therefore the choice of kernel does not arise.

3.2 Empirical Likelihood Bootstrap

In this section we develop the empirical likelihood (EL) approach to bootstrapping time-series models. Two cases are considered: (i) the overlapping empirical likelihood block bootstrap (EMB), and (ii) the non-overlapping empirical likelihood block bootstrap (ENB). The procedure for implementing the empirical block bootstrap is straightforward and outlined in Section 7.

An advantage of the EL block bootstrap over the standard block bootstrap is that EL weighted observations estimate the distribution function of the data more efficiently than non-weighted observations. We think this provides the EL block bootstrap with an improvement in test level accuracy over the standard block bootstrap, although a rigorous proof by an Edgeworth expansion is beyond the scope of the paper.

When X_t is *iid*, Theorem 1 of Brown and Newey (2002) shows that the empirical distribution function of the EL-weighted X_t's is a more efficient estimator of the population distribution function of X_t than the ordinary empirical distribution function of the X_t's. Brown and Newey (2002) combine it with an Edgeworth expansion to show that the EL bootstrap improves test level accuracy over an *iid* bootstrap for some cases, for example in a one-sided test of the null hypothesis of $E[g(X_t, \theta_0)] = 0$.

In our case, we attach the EL weights to the blocks, instead of individual observations. By analogy to Brown and Newey (2002), using the EL weights would provide a more efficient estimate of the distribution function of the blocks. Therefore, the EL block bootstrap would estimate the distribution of the sample moments more efficiently than the standard block bootstrap. Our simulation results suggest that efficient estimation of the distribution of the blocks by the EL block bootstrap contributes to improvements in test level accuracy, at least in some cases.

On the other hand, it is not clear whether an Edgeworth-expansion based analysis can demonstrate a higher-order improvement of the EL block bootstrap over the first-order asymptotics. Inoue and Shintani (2006) demonstrate that a higher-order analysis of the block bootstrap is handicapped by the bias of the HAC covariance matrix estimator, unless one uses a kernel whose characteristic exponent is greater than two. This excludes standard kernels such as the Bartlett, Parzen, and quadratic spectrature kernel.

Another attractive feature of using the empirical likelihood bootstrap rather than the standard bootstrap is that re-centering is not required, as is the case in Hall and Horowitz (1996). The EL weights provide a probability measure under which the moment conditions hold exactly.

3.2.1 EMB

First consider the overlapping bootstrap. Let $N = n - \ell + 1$ be the total number of overlapping blocks. Define the ith overlapping block of the sample moment as (^{o} stands for "overlapping"):

$$T_i^o(\theta) = \ell^{-1} \sum_{t=1}^{\ell} g(X_{i+t-1}, \theta), \quad i = 1, \ldots, N,$$

and the Lagrangian as:

$$L = \sum_{i=1}^{N} \log(\pi_i^o) + \mu \left(1 - \sum_{i=1}^{N} \pi_i^o\right) - N\gamma' \sum_{i=1}^{N} \pi_i^o T_i^o(\theta).$$

It is known that the solution for the probability weights are given by:

$$\pi_i^o = \frac{1}{N}\left(\frac{1}{1+\gamma^o(\theta)' T_i^o(\theta)}\right),$$

where

$$\gamma^o(\theta) = \underset{\lambda \in \Lambda_n(\theta)}{\operatorname{argmax}} \sum_{i=1}^{N} \log(1 + \gamma' T_i^o(\theta)). \tag{9}$$

Solving out the Lagrange multipliers and the coefficients simultaneously requires solving a difficult saddle point problem outlined in Kitamura (1997). Instead, one can use the GMM estimate of θ to compute π_i^o and attach these weights to the bootstrapped (blocks of) samples. Given the GMM estimate $\hat{\theta}$, compute $\gamma^o(\hat{\theta})$, which is a much smaller dimensional problem. Then solve for the empirical probability weights:

$$\hat{\pi}_i^o = \frac{1}{N}\left(\frac{1}{1+\gamma^o(\hat{\theta})' T_i^o(\hat{\theta})}\right), \tag{10}$$

which satisfy the moment condition $\sum_{i=1}^{N} \hat{\pi}_i^o T_i^o(\hat{\theta}) = 0$. The EMB version of $\hat{\theta}$ is defined as:

$$\theta^*_{MBB} = \arg\min Q^*_{MBB,n}(\theta), \qquad Q^*_{MBB,n}(\theta) = \left(b^{-1} \sum_{i=1}^{b} T_i^{o*}(\theta)\right)' W^*_{MBB,n} \left(b^{-1} \sum_{i=1}^{b} T_i^{o*}(\theta)\right),$$

where $W^*_{MBB,n}$ is a weighting matrix and $\{T_i^{o*}(\theta)\}_{i=1}^{b}$ are b *iid* samples (with replacement) from the distribution with $\Pr(T_i^{o*}(\theta) = T_k^o(\theta)) = \hat{\pi}_k^o$ for $k = 1, \ldots, N$. Note that $E^* T_i^{o*}(\hat{\theta}) = \sum_{i=1}^{N} \hat{\pi}_i^o T_i^o(\hat{\theta}) = 0$.

The long-run autocovariance matrix estimator for EMB takes the form:

$$S^*_{MBB,n}(\theta) = \ell b^{-1} \sum_{i=1}^{b} T_i^{o*}(\theta) T_i^{o*}(\theta)', \tag{11}$$

and the second-stage (optimal) weighting matrix is given by $S^*_{MBB,n}(\tilde{\theta}^*_{MBB})^{-1}$, where $\tilde{\theta}^*_{MBB}$ is the first-stage EMB estimator. The overlapping block Wald tests are based on the long-run autocovariance matrix $S^*_{MBB,n}(\theta)$. The EMB version of the J-statistic, $\mathcal{J}^*_{MBB,n}$, is defined analogously to $\mathcal{J}_n$ but using $(S^*_{MBB,n}(\tilde{\theta}^*_{MBB}))^{-1/2}$ and $n^{1/2} b^{-1} \sum_{i=1}^{b} T_i^{o*}(\theta)$.

3.2.2 ENB

The ENB uses b non-overlapping blocks rather than overlapping blocks. The ith non-overlapping block is defined as:

$$T_i(\theta) = \ell^{-1} \sum_{t=1}^{\ell} g(X_{(i-1)\ell+t}, \theta), \quad i = 1, \ldots, b,$$

and the Lagrange multiplier and empirical probability weights are given by:

$$\gamma(\hat{\theta}) = \underset{\lambda \in \Lambda_n(\hat{\theta})}{\operatorname{argmax}} \sum_{i=1}^{b} \log(1 + \gamma' T_i(\hat{\theta})), \quad \hat{\pi}_i = \frac{1}{b} \left(\frac{1}{1 + \gamma(\hat{\theta})' T_i(\hat{\theta})} \right). \tag{12}$$

The ENB estimator is defined as:

$$\theta^*_{NBB} = \arg\min Q^*_{NBB,n}(\theta), \qquad Q^*_{NBB,n}(\theta) = \left(b^{-1} \sum_{i=1}^{b} T_i^*(\theta) \right)' W^*_{NBB,n} \left(b^{-1} \sum_{i=1}^{b} T_i^*(\theta) \right),$$

where $W^*_{NBB,n}$ is a weighting matrix and $\{T_i^*(\theta)\}_{i=1}^{b}$ are b *iid* samples (with replacement) from the distribution with $\Pr(T_i^*(\theta) = T_k(\theta)) = \hat{\pi}_k$ for $k = 1, \ldots, b$. The long-run autocovariance matrix estimator for ENB is:

$$S^*_{NBB,n}(\theta) = \ell b^{-1} \sum_{i=1}^{b} T_i^*(\theta) T_i^*(\theta)', \tag{13}$$

and the optimal weighting matrix is given by $S^*_{NBB,n}(\tilde{\theta}^*_{NBB})^{-1}$, where $\tilde{\theta}^*_{NBB}$ is the first-stage ENB estimator. The non-overlapping block Wald tests are based on the long-run autocovariance matrix, $S^*_{NBB,n}(\theta)$. The ENB version of the J-statistic, $\mathcal{J}^*_{NBB,n}$, is defined analogously to $\mathcal{J}^*_{MBB,n}$.

It may also be possible to attach EL weights to the blocks and draw *iid* bootstrap observations. For example, in EMB, draw b *iid* samples from $\{\hat{\pi}_j^o T_j^o(\theta) : j = 1, \ldots, N\}$. This variant of EL block bootstrap will have the same first-order asymptotic property, but it is not clear whether this variant will have the same higher-order property. While Theorems 2.1 and 2.2 of Hall and Mammen (1994) provide sufficient conditions for higher-order equivalence of weighted bootstraps in the *iid* case,[5] applying these theorems to the EL bootstrap requires more detailed bounds on the EL weights than those in this paper.

4 Consistency of the bootstrap-based inference

The following lemmas establish the consistency of the bootstrap-based inference. The proofs are based on the results in Gonçalves and White (2004), hereafter referred to as GW04, and Mason and Newton (1992). As in GW04, let P denote the probability measure that governs the behavior of the original time-series and let P^* be the probability measure induced by bootstrapping. For a bootstrap statistic T_n^* we write $T_n^* \to 0$ prob-P*, prob-P (or $T_n^* \to_{P^*,P} 0$) if for any $\varepsilon > 0$ and any $\delta > 0, \lim_{n\to\infty} \mathrm{P}[\mathrm{P}^*[|T_n^*| > \varepsilon] > \delta] = 0$. Also following GW04 we use the notation $x_n \to_{d^*} x$ prob-P when weak convergence under P^* occurs in a set with probability converging to one.

Theorem 1 *Let Assumptions A and B in the mathematical appendix hold. If $\ell \to \infty$, $\ell = o(n^{1/2-1/r})$, and $W_n^{**}, W_{MBB,n}^* \to_{P^*,P} W$, then for any $\varepsilon > 0$,* $\Pr\{\sup_{x\in\mathbb{R}^p} |P^*[\sqrt{n}(\theta^*_{MBB} - \hat{\theta}) \le x] - P[\sqrt{n}(\hat{\theta} - \theta_0) \le x]| > \varepsilon\} \to 0$ *and* $\Pr\{\sup_{x\in\mathbb{R}^p} |P^*[\sqrt{n}(\theta^{**}_{MBB} - \hat{\theta}) \le x] - P[\sqrt{n}(\hat{\theta} - \theta_0) \le x]| > \varepsilon\} \to 0$.

Theorem 2 *Let Assumptions A and B in the mathematical appendix hold. If $\ell \to \infty$, $\ell = o(n^{(r-2)/2(r-1)})$, and $W_n^{**}, W_{NBB,n}^* \to_{P^*,P} W$, then for any $\varepsilon > 0$,* $\Pr\{\sup_{x\in\mathbb{R}^p} |P^*[\sqrt{n}(\theta^*_{NBB} - \hat{\theta}) \le x] - P[\sqrt{n}(\hat{\theta} - \theta_0) \le x]| > \varepsilon\} \to 0$ *and* $\Pr\{\sup_{x\in\mathbb{R}^p} |P^*[\sqrt{n}(\theta^{**}_{NBB} - \hat{\theta}) \le x] - P[\sqrt{n}(\hat{\theta} - \theta_0) \le x]| > \varepsilon\} \to 0$.

Theorem 3 *Let Assumptions A and B in the mathematical appendix hold. Assume $S_n \to_P \Sigma$. If $\ell \to \infty$ and $\ell = o(n^{1/2-1/r})$, then the bootstrap-based inference using the Wald statistic is consistent. Further, $\mathcal{J}_n \to_d \chi^2_{m-p}$, and $\mathcal{J}^*_{MBB,n}$, $\mathcal{J}^*_{NBB,n}$, $\mathcal{J}^{**}_{MBB,n}$, $\mathcal{J}^{**}_{NBB,n} \to_{d^*} \chi^2_{m-p}$ prob-P.*

5 Monte Carlo Experiments

In this section, a comparison of the finite sample performance differences of the standard block bootstrapping approaches to the empirical likelihood block bootstrap approaches is undertaken in a number of Monte Carlo experiments. The Monte Carlo design includes both linear and nonlinear models. For each experiment we report actual and nominal size at the 1, 5, and 10 per cent level for

[5]Barbe and Bertail (1995) analyze the asymptotics of the generalized bootstrap of a large class of statistics including Fréchet differentiable functionals.

the t-test and J-test. Parameter settings are deliberately chosen to illustrate the most challenging size problems. There are sample sizes: 100, 250, and 1000. Each experiment has 2000 replications and 499 bootstrap samples. This number of bootstrap samples does not lead to appreciable distortions in size for any of the experiments.

5.1 Case I: Linear models

5.1.1 Symmetric Errors

Consider the same linear process as Inoue and Shintani (2006):

$$y_t = \theta_1 + \theta_2 x_t + u_t \quad \text{for } t = 1, \ldots T, \tag{14}$$

where $(\theta_1, \theta_2) = (0,0)$, $u_t = \rho u_{t-1} + \varepsilon_{1t}$ and $x_t = \rho x_{t-1} + \varepsilon_{2t}$. The error structure, $\varepsilon = (\varepsilon_1, \varepsilon_2)$ are uncorrelated *iid* normal processes with mean 0 and variance 1. The approach is instrumental variable estimation of θ_1 and θ_2 with instruments $z_t = (\iota \; x_t \; x_{t-1} \; x_{t-2})$. There are two overidentifying restrictions. The null hypothesis being tested is: $H_o : \theta_2 = 0$. The statistics based on the GMM estimator are Studentized using a Bartlett kernel applied to pre-whitened series (see Andrews and Monahan (1992)). The bootstrap sample is not smoothed since the b blocks are *iid*. Both the non-overlapping block bootstrap and the overlapping block bootstrap are considered in the experiment.

Results are reported in Table 1. The amount of dependence in the moment conditions is relatively high, $\rho = 0.9$. The block length is set equal to the lag window in the HAC estimator, which is chosen using a data-dependent method (Newey and West (1994)). One immediate observation is that the asymptotic test-statistics severely over-reject the true null hypothesis. For example, with 100 observations the actual level for a 10% t-test is 42.25%. The actual level of the J-test is closer to the nominal level, although there is still over-rejection. The block bootstrap, with block size averaging from 1.96 for 100 observations to 4.48 for 1,000 observations, reduces the amount of over-rejection of the t-test substantially. The greatest improvements for the t-test are with the standard bootstrap. For the J-test the empirical likelihood bootstrap produces actual size much closer to the nominal size than the alternatives. Interestingly, the overlapping bootstrap has worse size than the non-overlapping block bootstrap for the t-test.

5.1.2 Heteroscedastic Errors

The subsequent DGP is the same as in the previous section with the addition of conditional heteroscedasticity, modeled as a $GARCH(1,1)$. The DGP is:

$$y_t = \theta_1 + \theta_2 x_t + \sigma_t u_t \quad \text{for } t = 1, ...T, \tag{15}$$

where $(\theta_1, \theta_2) = (0,0)$, $x_t = 0.75x_{t-1} + \varepsilon_{1t}$, and $u_t \sim N(0, \sigma_t)$. $\sigma_t^2 = 0.0001 + 0.6\sigma_{t-1}^2 + 0.3\varepsilon_{2t-1}^2$ and $\varepsilon \sim N(0, I)$. The unconditional variance is 1. The instrument set is $z_t = [\iota \; x_t \; x_{t-1} \; x_{t-2}]$.

Results with 2,000 replications and 499 bootstrap samples are presented in Table 2. There are three sample sizes: 100, 250, and 1000. The actual size of the asymptotic tests are close to the nominal size for sample size 250 and greater. The moving block bootstrap tests have good size and the empirical likelihood bootstrap performs best out of the bootstrap procedures. Using the standard block bootstrap actually leads to more severe under-rejection of the true null hypothesis than the asymptotic tests.

5.2 Case II: Nonlinear Models

Two experiments are considered. First the chi-squared experiment from Imbens, Spady, and Johnson (1998). Second, the asset pricing DGP outlined in Hall and Horowitz (1996) and used by Gregory, Lamarche, and Smith (2002). Imbens, Spady, and Johnson (1998) also consider this DGP. In addition this section looks at the empirical likelihood bootstrap in a framework with dependent data. It is the case of nonlinear models where the asymptotic t-test and J-test tend to severely over-reject.

5.2.1 Asymmetric Errors

First consider a model with Chi-squared moments. Imbens, Spady, and Johnson (1998) provide evidence that average moment tests like the J-test can substantially over-reject a true null hypothesis under a DGP with Chi-squared moments. The authors find that tests based on the exponential tilting parameter perform substantially better.

The moment vector is:

$$g(X_t, \theta_1) = (X_t - \theta_1 \quad X_t^2 - \theta_1^2 - 2\theta_1)'.$$

The parameter θ_1 is estimated using the two moments.

Results for 2,000 replications and 499 bootstrap samples are presented in Table 3. There is severe over-rejection of the true null hypothesis when using the asymptotic distribution. The bootstrap procedures correct for this over-rejection; the empirical likelihood bootstrap performs very well for the t-tests. For small sample sizes the standard and empirical likelihood bootstrap both outperform the asymptotic approximation but there is still is an over-rejection.

5.2.2 Asset Pricing Example

Finally consider an asset pricing model with the following moment conditions.[6]:

$$E[\exp(\mu - \theta(x+z) + 3z) - 1] = 0, \quad Ez[\exp(\mu - \theta(x+z) + 3z) - 1] = 0$$

.

It is assumed that

$$\log x_t = \rho \log x_{t-1} + \sqrt{(1-\rho^2)}\varepsilon_{xt}, \quad z_t = \rho z_{t-1} + \sqrt{(1-\rho^2)}\varepsilon_{zt}$$

, where ε_{xt} and ε_{zt} are independent normal with mean 0 and variance 0.16. In the experiment $\rho = 0.6$.

Results for 2,000 replications and 499 bootstrap samples are presented in Table 4. Again, the asymptotic tests severely over-reject the true null hypothesis. The bootstrap procedures produce tests with reasonable size, especially for the t-tests. As was the case in the model with asymmetric errors, the empirical likelihood bootstrap performs best.

6 Conclusion

This paper extends the ideas put forth by Brown and Newey (2002) to bootstrap test-statistics based on empirical likelihood. Where Brown and Newey (2002) consider bootstrapping in an *iid* context, this paper provides a proof of the first-order asymptotic validity of empirical likelihood block bootstrapping in the context of dependent data. Given the test-statistics considered, the size distortions of those tests based on the asymptotic distribution are severe, especially in the case of nonlinear moment conditions and substantial serial correlation. The empirical likelihood bootstrap largely corrects for these size distortions and produces promising results. This is especially true when the regression errors are non-spherical. The significance of using the empirical likelihood estimator is that it satisfies the moment conditions identically while supplying a probability measure

[6]Derivation of the example can be found in Gregory, Lamarche, and Smith (2002).

under which these conditions hold. As highlighted by Brown and Newey (2002), the empirical likelihood bootstrap is the same as the conventional bootstrap, except that it is based on a more efficient distribution estimator. Two possible avenues for future research include combining subsampling methods with empirical likelihood probability weights and establishing higher order improvements for the ENB and EMB.

7 Implementing the Block Bootstrap

The procedure for implementing the GMM overlapping (MBB) and empirical likelihood (EMB) bootstrap procedures are outlined below. The procedure is similar for the non-overlapping bootstrap.

1. Given the random sample $(X_1,...,X_n)$, calculate $\hat{\theta}$ using 2-stage GMM

2. For EMB calculate $\hat{\pi}_i^o$ using equation (10)

3a. For EMB sample with replacement from $\{T_j^o(\hat{\theta}) : j = 1,\ldots,N\}$ with probability $\{\hat{\pi}_j^o : j = 1,\ldots,N\}$

3b. For MBB uniformly sample with replacement to get $\{X^*\}_{t=1}^n = (\tilde{X}_1,...,\tilde{X}_b)$

4a. For EMB calculate the J-statistic ($\mathcal{J}^*_{MBB,n}$) and t-statistic (T^*_{nr})

4b. For MBB calculate J-statistic ($\mathcal{J}^{**}_{MBB,n}$) and t-statistic (T^{**}_{nr})

5. Repeat steps 3-4 B times, where B is the number of bootstraps.

6. Let $\hat{q}^{\pi}_{\alpha}$ be a $(1-\alpha)$ percentile of the distribution of T^*_{nr} or T^{**}_{nr}

7. Let $\overline{q}^{\pi}_{\alpha}$ be a $(1-\alpha)$ percentile of the distribution of $\mathcal{J}^*_{MBB,n}$ or $\mathcal{J}^{**}_{MBB,n}$

8. The bootstrap confidence interval for θ_{0r} is $\hat{\theta}_{nr} \pm \hat{q}^{\pi}_{\alpha} n^{-1/2} \hat{\sigma}_{nr}$

9. For the bootstrap J-test, the test rejects if $\mathcal{J}_n \geq \overline{q}^{\pi}_{\alpha}$

8 Mathematical Appendix

Assumptions A and B are a simplified version of Assumptions A and B in Gonçalves and White (2004), tailored to our GMM estimation framework. $||x||_p$ denotes the L_p norm $(E|X_{nt}|^p)^{1/p}$. For a $(m \times k)$ matrix x, let $|x|$ denote the 1-norm of x, so $|x| = \sum_{i=1}^{m} \sum_{j=1}^{k} |x_{ij}|$.

Assumption A

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

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

A.3 (i) θ_0 is identifiably unique with respect to $Eg(X_t, \theta)' W E g(X_t, \theta)$ and (ii) θ_0 is interior to Θ.

A.4 (i) $\{g(X_t, \theta)\}$ is Lipschitz continuous on Θ, i.e. $|g(X_t, \theta) - g(X_t, \theta^o)| \leq L_t |\theta - \theta^o|$ *a.s.*-P, $\forall \theta, \theta^o \in \Theta$, where $\sup_t E(L_t) = O(1)$. (ii) $\{\nabla g(X_t, \theta)\}$ is Lipschitz continuous on Θ.

A.5 For some $r > 2$: (i) $\{g(X_t, \theta)\}$ is r-dominated on Θ uniformly in t, i.e. there exists $D_t : \mathbb{R}^{lt} \to \mathbb{R}$ such that $|g(X_t, \theta)| \leq D_t$ for all θ in Θ and D_t is measurable such that $||D_t||_r \leq \Delta < \infty$ for all t. (ii) $\{\nabla g(X_t, \theta)\}$ is r-dominated on Θ uniformly in t.

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

A.7 The elements of (i) $\{g(X_t, \theta)\}$ are NED on $\{V_t\}$ of size -1 uniformly on (Θ, ρ), where ρ is any convenient norm on $\mathbb{R}^p$, and (ii) $\{\nabla g(X_t, \theta)\}$ are NED on $\{V_t\}$ of size -1 uniformly on (Θ, ρ).

A.8 $\Sigma \equiv \lim_{n\to\infty} \text{var}(n^{-1/2} \sum_{t=1}^{n} g(X_t, \theta_0))$ is positive definite, and $G \equiv \lim_{n\to\infty} E(n^{-1} \sum_{t=1}^{n} \nabla g(X_t, \theta_0))$ is of full rank.

Assumption B

B.1 $\{g(X_t, \theta)\}$ is $3r$-dominated on Θ uniformly in t, $r > 2$.

B.2 For some small $\delta > 0$ and some $r > 2$, the elements of $\{g(X_t, \theta)\}$ are $L_{2+\delta}$–NED on $\{V_t\}$ of size $-(2(r-1))/(r-2)$ uniformly on (Θ, ρ); $\{V_t\}$ is an α -mixing sequence of size $-((2+\delta)r)/(r-2)$.

The following two lemmas are required to prove Theorems 1-3.

Lemma 1 *Suppose Assumption A in the mathematical appendix hold. Then* $\hat{\theta} - \theta_0 \to_P 0$. *If also* $\ell \to \infty$ *and* $\ell = o(n)$, *then* $\theta^{**}_{MBB} - \hat{\theta} \to_{P^*,P} 0$. *If also Assumption B in Appendix hold and* $\ell = o(n^{1/2-1/r})$, *then* $\theta^*_{MBB} - \hat{\theta} \to_{P^*,P} 0$.

Lemma 2 *Suppose Assumption A in the mathematical appendix hold,* $\ell \to \infty$, *and* $\ell = o(n)$. *Then* $\theta^{**}_{NBB} - \hat{\theta} \to_{P^*,P} 0$. *If also* $\ell = o(n^{(r-2)/2(r-1)})$, *then* $\theta^*_{NBB} - \hat{\theta} \to_{P^*,P} 0$. *Note that* ℓ *must satisfy* $\ell = o(n^{1/2})$ *because* $(r-2)/2(r-1) < 1/2$.

If we compare conditions on ℓ, the condition with the NBB is slightly weaker because $(r-2)/2(r-1) = 1/2 - 1/2(r-1)$ and $2(r-1) > r$.

8.1 Proof of Lemma 1

The proof follows the proof of Theorem 2.1 of GW04, with two differences: (i) the objective function is a GMM objective function, and (ii) in the case of EMB, the bootstrapped objective function depends on the probability weight $\hat{\pi}^o_i$. $\hat{\theta} - \theta_0 \to_P 0$ follows from applying Lemma A.2 of GW04 to the GMM objective function, because conditions (a1)-(a3) in Lemma A.2 of GW04 are satisfied by Assumption A. The consistency of θ^{**}_{MBB} is proved by applying Lemma A.2 of GW04. Their conditions (b1)-(b2) are satisfied by Assumptions A.2. Define $\tilde{Q}_n(\theta) = (n^{-1}\sum_{t=1}^n g(X^*_t,\theta))'W^*_n(n^{-1}\sum_{t=1}^n g(X^*_t,\theta))$, then their condition (b3) holds because $\sup_\theta |Q^{**}_{MBB,n}(\theta) - \tilde{Q}_n(\theta)| \to_{P^*,P} 0$ from a standard argument and $\sup_\theta |\tilde{Q}_n(\theta) - Q_n(\theta)| \to_{P^*,P} 0$ by Lemmas A.4 and A.5 of GW04.

We prove the consistency of θ^*_{MBB} by approximating the EMB sample moment condition with an uncentered MBB moment condition, namely, by showing $\sup_\theta |b^{-1}\sum_{i=1}^b T^{o*}_i(\theta) - n^{-1}\sum_{t=1}^n g(X^*_t,\theta)| \to_{P^*,P} 0$ for suitably chosen $T^{o*}_i(\theta)$'s and X^*_t's. Then the consistency of θ^*_{MBB} follows from the proof of the consistency of θ^{**}_{MBB}. We will use the following result, which we prove later:

$$N\hat{\pi}^o_i = 1 + \delta_{ni}, \quad \max_{1\le i\le N} |\delta_{ni}| = o_P(1). \tag{16}$$

Partition the interval $[0,1]$ into $A_1,\dots,A_N$, where $A_i = [\hat{\pi}^o_0 + \cdots + \hat{\pi}^o_{i-1}, \hat{\pi}^o_0 + \cdots + \hat{\pi}^o_i]$ with $\hat{\pi}^o_0 = 0$. Partition the interval $[0,1]$ into N sets, $B_1,\dots,B_N$, where the B_i's are chosen such that $\mu(B_i) = 1/N$ and $\max_{1\le i\le N}\mu(D_i) = o(N^{-1})$, where μ denotes the Lebesgue measure on $[0,1]$, and $D_i = (A_i - B_i) \cup (B_i - A_i)$, i.e., the symmetric difference between A_i and B_i. Such a construction of $B_1,\dots,B_N$ is possible by virtue of (16). One way to construct $\{T^{o*}_k(\theta)\}_{k=1}^b$ and $\{\tilde{X}^*_k\}_{k=1}^b$ is to draw *iid* uniform$[0,1]$ random variables $U_1,\dots,U_b$ and set $T_k^{o*}(\theta) = T_i^o(\theta)$ if $U_k \in A_i$ and set $\tilde{X}_k^* = \tilde{X}_i$ if $U_k \in B_i$. Then we may write $b^{-1}\sum_{i=1}^b T_i^{o*}(\theta) = b^{-1}\sum_{k=1}^b \sum_{i=1}^N 1\{U_k \in A_i\} T_i^o(\theta)$ and $|b^{-1}\sum_{i=1}^b T_i^{o*}(\theta) - n^{-1}\sum_{t=1}^n g(X_t^*,\theta)| = b^{-1}\sum_{k=1}^b \sum_{i=1}^N 1\{U_k \in D_i\}|T_i^o(\theta)|$. Taking the bootstrap expectation of its supremum over θ gives $E^* \sup_\theta b^{-1}\sum_{k=1}^b \sum_{i=1}^N 1\{U_k \in D_i\}|T_i^o(\theta)| \le E^* b^{-1}\sum_{k=1}^b \sum_{i=1}^N 1\{U_k \in D_i\}\sup_\theta |T_i^o(\theta)| = E^*\sum_{i=1}^N 1\{U_1 \in D_i\}\sup_\theta |T_i^o(\theta)| \le \max_{1\le i\le N}\mu(D_i)\sum_{i=1}^N \sup_\theta |T_i^o(\theta)| = o_P(1)$. Therefore, $\sup_\theta |b^{-1}\sum_{i=1}^b T_i^{o*}(\theta) - n^{-1}\sum_{t=1}^n g(X_t^*,\theta)| = o_{P^*,P}(1)$, and the consistency of θ^*_{MBB} follows.

It remains to show (16). First we show $\gamma^o(\hat\theta) = O_P(\ell n^{-1/2})$. In view of the argument in pp. 100-101 of Owen (1990) (see also Kitamura (1997)), $\gamma^o(\hat\theta) = O_P(\ell n^{-1/2})$ holds if (a) $\ell N^{-1}\sum_{i=1}^N T_i^o(\hat\theta)T_i^o(\hat\theta)' \to_P \Sigma$, (b) $\ell N^{-1}\sum_{i=1}^N T_i^o(\hat\theta) = O_P(\ell n^{-1/2})$, and (c) $\max_{1\le i\le N}|T_i^o(\hat\theta)| = o_P(n^{1/2}\ell^{-1})$. For (a), using a mean value expansion and Assumption A.5 gives

$$
\begin{aligned}
& \left| \ell N^{-1}\sum_{i=1}^N T_i^o(\hat\theta)T_i^o(\hat\theta)' - \ell N^{-1}\sum_{i=1}^N T_i^o(\theta_0)T_i^o(\theta_0)' \right| \\
\le\; & |\hat\theta - \theta_0| 2\ell N^{-1}\sum_{i=1}^N \sup_\theta |\nabla T_i^o(\theta)||T_i^o(\theta)| = O_P(n^{-1/2}\ell) = o_P(1).
\end{aligned}
$$

Define $\bar{G}_n^* = n^{-1}\sum_{t=1}^n g(X_t^*,\theta_0)$, then we have (cf. Lahiri (2003), p. 48) $\ell N^{-1}\sum_{i=1}^N T_i^o(\theta_0)T_i^o(\theta_0)' = \text{var}^*(\sqrt{n}\bar{G}_n^*) + \ell\bar{T}_n\bar{T}_n'$, where $\bar{T}_n = N^{-1}\sum_{i=1}^N T_i^o(\theta_0)$. $\text{var}^*(\sqrt{n}\bar{G}_n^*) - \Sigma \to_P 0$ from Corollary 2.1 of Gonçalves and White (2002) (hereafter GW02). $\bar{T}_n$ is equal to $\bar{X}_{\gamma,n}$ defined in p. 1371 of GW02 if we replace their X_t with $g(X_t,\theta_0)$. GW02 p.1381 shows $\bar{X}_{\gamma,n} = o_P(\ell^{-1})$, and hence $\ell\bar{T}_n^2 = o_P(1)$. Therefore, $\ell N^{-1}\sum_{i=1}^N T_i^o(\theta_0)T_i^o(\theta_0)' \to_P \Sigma$, and (a) follows. (b) follows from expanding $T_i^o(\hat\theta)$ around θ_0 and using $N^{-1}\sum_{i=1}^N T_i^o(\theta_0) = n^{-1}\sum_{t=1}^n g(X_t,\theta_0) + O_p(n^{-1}\ell)$ (cf. Lemma A.1 of Fitzenberger (1997)), and applying the central limit theorem. (c) holds because $\max_{1\le i\le N}|T_i^o(\hat\theta)| = O_{a.s.}(N^{1/r})$ from Lemma 3.2 of Künsch (1989) and $\ell = o(n^{1/2-1/r})$. Therefore, we have

$$
\gamma^o(\hat\theta) = O_P(\ell n^{-1/2}), \quad \max_{1\le i\le N}|\gamma^o(\hat\theta)'T_i^o(\hat\theta)| = o_P(1). \tag{17}
$$

(16) follows from expanding $N\pi_i^o = (1+\gamma^o(\hat\theta)'T_i^o(\hat\theta))^{-1}$ around $\gamma^o(\hat\theta)'T_i^o(\hat\theta) = 0$. □

8.2 Proof of Lemma 2

In view of the proof of Lemma 1, the consistency of θ^{**}_{NBB} holds because condition (b3) of Lemma A.2 of GW04 holds because $\sup_\theta |\tilde{Q}_n(\theta) - Q_n(\theta)| \to_{P^*,P} 0$ by Lemmas 3 and 4.

In view of the proof of the consistency of θ^*_{MBB} in Lemma 1, θ^*_{NBB} is consistent if

$$\gamma(\hat{\theta}) = O_P(\ell n^{-1/2}), \quad \max_{1\le i\le b} |\gamma(\hat{\theta})' T_i(\hat{\theta})| = o_P(1). \tag{18}$$

Equation (18) holds if (a) $\ell b^{-1}\sum_{i=1}^b T_i(\hat{\theta})T_i(\hat{\theta})' \to_P \Sigma$, (b) $\ell b^{-1}\sum_{i=1}^b T_i(\hat{\theta}) = O_P(\ell n^{-1/2})$, and (c) $\max_{1\le i\le b}|T_i(\hat{\theta})| = o_P(n^{1/2}\ell^{-1})$. (a) follows from expanding $T_i(\hat{\theta})$ around θ_0 and using Corollary 2. (b) follows from expanding $T_i(\hat{\theta})$ around θ_0 and applying the central limit theorem. (c) follows because $\max_{1\le i\le b}|T_i(\hat{\theta})| = O_{a.s.}(b^{1/r})$ and $\ell = o(n^{(r-2)/2(r-1)})$. □

8.3 Proof of Theorem 1

Define $H = (G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1}$, then the stated result follows from Polya's theorem if we show $\sqrt{n}(\hat{\theta}-\theta_0) \to_d N(0,H)$, $\sqrt{n}(\theta^*_{MBB}-\hat{\theta}) \to_{d^*} N(0,H)$ prob-P, and $\sqrt{n}(\theta^{**}_{MBB}-\hat{\theta}) \to_{d^*} N(0,H)$ prob-P. The limiting distribution of $\sqrt{n}(\hat{\theta}-\theta_0)$ follows from a standard argument.

The proof of the asymptotic normality of θ^*_{MBB} and θ^{**}_{MBB} uses Theorem 2.1 of Mason and Newton (1992), who prove the consistency of generalized bootstrap (weighted bootstrap) procedures. We first derive the asymptotics of the EMB estimator. The first order condition for the EMB estimator is $0 = b^{-1}\sum_{i=1}^b \nabla T_i^{o*}(\theta^*_{MBB})'W^*_{MBB,n} b^{-1}\sum_{i=1}^b T_i^{o*}(\theta^*_{MBB})$. Expanding $b^{-1}\sum_{i=1}^b T_i^{o*}(\theta^*_{MBB})$ around $\hat{\theta}$ and approximating $b^{-1}\sum_{i=1}^b \nabla T_i^{o*}(\theta)$ by $n^{-1}\sum_{i=1}^b \nabla g(X_t^*,\theta)$ as in the proof of Lemma 1 gives $n^{1/2}(\theta^*_{MBB}-\hat{\theta}) = -(\tilde{G}_n' W^*_{MBB,n}\tilde{G}_n)^{-1}\tilde{G}_n' W^*_{MBB,n} n^{1/2} b^{-1}\sum_{i=1}^b T_i^{o*}(\hat{\theta})$, where $\tilde{G}_n$ is a generic notation for $G + o_{P^*,P}(1)$. We proceed to rewrite $n^{1/2}b^{-1}\sum_{i=1}^b T_i^{o*}(\hat{\theta})$ so that we can apply the results in Mason and Newton (1992). For $i = 1,\dots,N$, let w_{Ni} be the number of times $T_i^o(\theta)$ appears in a bootstrap sample $\{T_k^{o*}(\theta)\}_{k=1}^b$. Conditional on $X_1,\dots,X_n$, an N-vector $w_N = (w_{N1},\dots,w_{NN})'$ follows a multinomial distribution such that $w_N \sim \text{Mult}(b;\hat{\pi}_1^o,\dots,\hat{\pi}_N^o)$. Using w_{Ni} in conjunction with $\sum_{i=1}^N \hat{\pi}_i^o T_i^o(\hat{\theta}) = 0$ and $b\ell = n$, we may rewrite $n^{1/2}b^{-1}\sum_{i=1}^b T_i^{o*}(\hat{\theta}) = N^{-1/2}\sum_{i=1}^N (N/b)^{1/2}(w_{Ni} - b\hat{\pi}_i^o)\ell^{1/2}T_i^o(\hat{\theta})$.

Therefore, the asymptotic normality of θ^*_{MBB} follows if we show

$$N^{-1/2}\sum_{i=1}^N (N/b)^{1/2}(w_{Ni} - b\hat{\pi}_i^o)\ell^{1/2}T_i^o(\hat{\theta}) \to_{d^*} N(0,\Sigma) \quad \text{prob-P}. \tag{19}$$

We apply Theorem 2.1 of Mason and Newton (1992) to the left hand side of (19) with two minor changes. First, the weights in Theorem 2.1 of Mason and Newton (1992) do not depend on the data, whereas our w_N depends on the data through $\hat{\pi}_i^o$. As Mason and Newton (1992) discuss on p. 1618, their Theorem 2.1 holds if the weights are exchangeable given the data. Second, in Mason and Newton (1992), condition (2.4) and result (2.7) hold P-almost surely. We can weaken both to hold in P-probability because $x_n \to x$ in probability if and only if every subsequence of $\{x_n\}$ has a further subsequence that converges almost surely to x (see, for example, Theorem 6.2 in p. 46 of Durrett (2005)).

For simplicity, we assume $T_i^o(\theta)$ to be a scalar without loss of generality. Note that our $\{N, \ell^{1/2}T_i^o(\hat{\theta}), (N/b)^{1/2}(w_{Ni} - b\hat{\pi}_i^o)\}$ correspond to $\{k_n, X_{n,k}, Y_{n,k}\}$ in Mason and Newton (1992). From Theorem 2.1 of Mason and Newton (1992), (19) follows if we show (recall $\sum_{i=1}^n (w_{Ni} - b\hat{\pi}_i^o) = 0$ by construction)

$$N^{-1}\sum_{i=1}^{N}(\ell^{1/2}T_i^o(\hat{\theta}) - \ell^{1/2}\bar{T}^o(\hat{\theta}))^2 \to_P \Sigma, \quad N^{-1}\sum_{i=1}^{N}((N/b)^{1/2}(w_{Ni} - b\hat{\pi}_i^o))^2 \to_{P^*,P} 1, \tag{20}$$

where $\bar{T}^o(\hat{\theta}) = N^{-1}\sum_{i=1}^N T_i^o(\hat{\theta})$, and, for all $\tau > 0$,

$$(a) \max_{1\le i\le N} U_{Ni}^2 \to_P 0, \quad (b) \max_{1\le i\le N} V_{Ni}^2 \to_{P^*,P} 0, \tag{21}$$

$$D_N(\tau) = \sum_{i=1}^{N}\sum_{j=1}^{N} U_{Ni}^2 V_{Nj}^2 1\{NU_{Ni}^2 V_{Nj}^2 > \tau\} \to_{P^*,P} 0, \tag{22}$$

where

$$U_{Ni} = \frac{\ell^{1/2}T_i^o(\hat{\theta}) - \ell^{1/2}\bar{T}^o(\hat{\theta})}{(\sum_{i=1}^N(\ell^{1/2}T_i^o(\hat{\theta}) - \ell^{1/2}\bar{T}^o(\hat{\theta}))^2)^{1/2}}, \quad V_{Ni} = \frac{(N/b)^{1/2}(w_{Ni} - b\hat{\pi}_i^o)}{(\sum_{i=1}^N((N/b)^{1/2}(w_{Ni} - b\hat{\pi}_i^o))^2)^{1/2}}.$$

We proceed to check (20)-(22). The first part of (20) holds because (a) and (b) in the proof of Lemma 1 show $\ell N^{-1}\sum_{i=1}^N T_i^o(\hat{\theta})^2 \to_P \Sigma$ and $\bar{T}^o(\hat{\theta}) = O_P(N^{-1/2})$. The second part of (20) follows from applying Lemma 5 to the left hand side with $r = 2$ because w_N satisfies the assumptions in Lemma 5. (a) of (21) follows from the first part of (20), $\bar{T}^o(\hat{\theta}) = O_P(N^{-1/2})$, and $\max_{1\le i\le N}|T_i^o(\hat{\theta})| = o_P(N^{1/2}\ell^{-1})$, which is shown in (c) in the proof of Lemma 1. (b) of (21) follows from Theorem 1 of Hoeffding (1951) in conjunction with the second part of (20) and Lemma 5 with $r = 4$. Finally, (22) can be be shown by a similar argument to Corollary 2.2 of Mason and Newton (1992). For any $\varepsilon \in (0,1)$, from (b) of (21) we have, for sufficiently large N, with prob-P^*, prob-P greater than $1-\varepsilon$, $D_N(\tau) \le \sum_{i=1}^N\sum_{j=1}^N U_{Ni}^2 V_{Nj}^2 1\{NU_{Ni}^2 > \tau/\varepsilon\} = \sum_{i=1}^N U_{Ni}^2 1\{NU_{Ni}^2 > \tau/\varepsilon\}$. From the first part of (20) and the order of $\bar{T}^o(\theta)$, this is bounded by $\Sigma^{-1}N^{-1}\sum_{i=1}^N \ell T_i^o(\theta_0)^2 1\{NU_{Ni}^2 > \tau/\varepsilon\} + o_P(1)$. Consequently, choosing ε sufficiently small gives $D_n(\tau) \to_{P^*,P} 0$ from $E\ell|T_i^o(\theta_0)|^2 = O(1)$ (see Lemmas A.1 and A.2 of GW02) and the dominated convergence theorem.

For the standard bootstrap estimator, expanding the first order condition and applying a routine argument gives $n^{1/2}(\theta^{**}_{MBB}-\hat{\theta}) = -(\tilde{G}_n' W_n^{**}\tilde{G}_n)^{-1}\tilde{G}_n' W_n^{**} n^{-1/2}\sum_{t=1}^n g^*(X_t^*,\hat{\theta})$. For $i=1,\dots,N$, let w^*_{Ni} be the number of times $\tilde{X}_i$ appears in a bootstrap sample $\{\tilde{X}^*_k\}_{k=1}^b$. Conditional on $X_1,\dots,X_n$, an N-vector $w_N^*=(w_{N1},\dots,w_{NN})'$ follows $\text{Mult}(b;1/N,\dots,1/N)$. Using $N^{-1}\sum_{i=1}^N T_i^o(\hat{\theta}) = n^{-1}\sum_{t=1}^n g(X_t,\hat{\theta}) + O_P(n^{-1}\ell)$ (cf. Lemma A.1 of Fitzenberger (1997)), we may write $n^{-1/2}\sum_{t=1}^n g^*(X_t^*,\hat{\theta})$ $= N^{-1/2}\sum_{i=1}^N (N/b)^{1/2}(w^*_{Ni}-b/N)\ell^{1/2}T_i^o(\hat{\theta}) + o_P(1)$. Since w_N^* satisfies the assumptions in Lemma 5, repeating the proof for the EMB estimator with replacing w_N by w_N^* gives $N^{-1/2}\sum_{i=1}^N (N/b)^{1/2}(w^*_{Ni}-b/N)\ell^{1/2}T_i^o(\hat{\theta}) \to_{d^*} N(0,\Sigma)$ prob-P, and the stated result follows. □

8.4 Proof of Theorem 2

The proof closely follows the proof of Theorem 1. Because we sample from b blocks, instead of N, we use Corollary 1 in place of Lemma 5.

We first derive the asymptotics of the ENB estimator. Expanding the first order condition for the ENB estimator gives $n^{1/2}(\theta^*_{NBB}-\hat{\theta}) = -(\tilde{G}_n' W^*_{NBB,n}\tilde{G}_n)^{-1}\tilde{G}_n' W^*_{NBB,n} n^{1/2} b^{-1}\sum_{i=1}^b T_i^*(\hat{\theta})$, where $\tilde{G}_n$ is a generic notation for $G + o_{P^*,P}(1)$. The required result follows if we show $n^{1/2}b^{-1}\sum_{i=1}^b T_i^*(\hat{\theta}) \to_{d^*} N(0,\Sigma)$ prob-P. For $i=1,\dots,b$, let w_{bi} be the number of times $T_i(\theta)$ appears in a bootstrap sample $\{T_k^*(\theta)\}_{k=1}^b$. Conditional on $X_1,\dots,X_n$, an b-vector $w_b=(w_{b1},\dots,w_{bb})'$ follows $\text{Mult}(b;\hat{\pi}_1,\dots,\hat{\pi}_b)$. Using $\sum_{i=1}^b \hat{\pi}_i T_i(\hat{\theta}) = 0$ and $b\ell = n$, we may rewrite $n^{1/2}b^{-1}\sum_{i=1}^b T_i^*(\hat{\theta}) = b^{-1/2}\sum_{i=1}^b (w_{bi}-b\hat{\pi}_i)\ell^{1/2}T_i(\hat{\theta})$. From Theorem 2.1 of Mason and Newton (1992), $b^{-1/2}\sum_{i=1}^b (w_{bi}-b\hat{\pi}_i)\ell^{1/2}T_i(\hat{\theta}) \to_{d^*} N(0,\Sigma)$ prob-P follows if we show

$$b^{-1}\sum_{i=1}^b (\ell^{1/2}T_i(\hat{\theta}) - \ell^{1/2}\bar{T}(\hat{\theta}))^2 \to_P \Sigma, \quad b^{-1}\sum_{i=1}^b (w_{bi}-b\hat{\pi}_i)^2 \to_{P^*,P} 1, \tag{23}$$

where $\bar{T}(\hat{\theta}) = b^{-1}\sum_{i=1}^b T_i(\hat{\theta})$, and, for all $\tau > 0$,

$$(a)\ \max_{1\le i\le b} U_{bi}^2 \to_P 0, \quad (b)\ \max_{1\le i\le b} V_{bi}^2 \to_{P^*,P} 0, \tag{24}$$

$$D_b(\tau) = \sum_{i=1}^b\sum_{j=1}^b U_{bi}^2 V_{bj}^2 1\{bU_{bi}^2 V_{bj}^2 > \tau\} \to_{P^*,P} 0, \tag{25}$$

where $U_{bi} = (\ell^{1/2}T_i(\hat{\theta}) - \ell^{1/2}\bar{T}(\hat{\theta}))(\sum_{i=1}^b(\ell^{1/2}T_i(\hat{\theta}) - \ell^{1/2}\bar{T}(\hat{\theta}))^2)^{-1/2}$ and $V_{bi} = (w_{bi}-b\hat{\pi}_i)(\sum_{i=1}^N (w_{bi}-b\hat{\pi}_i)^2)^{-1/2}$.

We proceed to check (23)-(25). The first part of (23) holds because (a) and (b) in the proof of Lemma 2 show $\ell b^{-1}\sum_{i=1}^b T_i(\hat{\theta})^2 \to_P \Sigma$ and $\bar{T}(\hat{\theta}) = O_P(n^{-1/2})$. The second part of (23) follows from applying Corollary 1 with $r=2$. (a) of (24) follows from the first part of (23), $\bar{T}(\hat{\theta}) = O_P(n^{-1/2})$, and $\max_{1\le i\le b}|T_i(\hat{\theta})| = o_P(n^{1/2}\ell^{-1})$, which is shown in (c) in the proof of Lemma 2. (b) of (24) follows from Theorem 1 of Hoeffding (1951) in conjunction with the second part of (23) and Corollary 1 with $r = 4$. Finally, (25) is shown by repeating the argument of the proof of (22) since $U_{bi}^2 = b^{-1}\ell T_i(\theta_0)^2(\Sigma^{-1} + o_{P^*,P}(1))$, and we derive the asymptotics of θ^*_{NBB}. The proof for the standard bootstrap estimator θ^{**}_{NBB} is very similar and omitted. □

8.5 Proof of Theorem 3

The validity of the bootstrap Wald test is proven if we show $S_n^{**}(\theta^*), S^*_{MBB,n}(\theta^*), S^*_{NBB,n}(\theta^*) \to_{P^*,P} \Sigma$ for any root-n consistent θ^*. Using a similar argument to the consistency proof of θ^*_{MBB}, we can show $S^*_{MBB,n}(\theta^*)$ is asymptotically equivalent in distribution to $S_n^{**}(\theta^*)$ that is constructed from a standard MBB sample. $S_n^{**}(\theta^*) \to_{P^*,P} \Sigma$ then follows from result (iii) in the proof of Theorem 3.1 of GW04. Similarly, $S^*_{NBB,n}(\theta^*)$ is asymptotically equivalent in distribution to $S_n^{**}(\theta^*)$ that is constructed from a standard NBB sample, which converges to Σ from Corollary 2.

$\mathcal{J}_n \to_d \chi^2_{m-p}$ if $W_n \to_P \Sigma^{-1}$ and $n^{-1/2}\sum_{t=1}^n g(X_t,\theta_0) \to_d N(0,\Sigma)$, which follows from Assumptions A and B and a standard argument. $\mathcal{J}^*_{MBB,n} \to_{d^*} \chi^2_{m-p}$ prob-P because $S^*_{MBB,n}(\tilde{\theta}^*_{MBB}) \to_{P^*,P} \Sigma$ and $n^{1/2}b^{-1}\sum_{i=1}^b T_i^{o*}(\hat{\theta}) \to_{d^*} N(0,\Sigma)$ prob-P. $\mathcal{J}^{**}_{MBB,n} \to_{d^*} \chi^2_{m-p}$ prob-P follows because $S_n^{**}(\theta^{**}_{MBB}) \to_{P^*,P} \Sigma$ and we have shown in the proof of Theorem 1 that $n^{-1/2}\sum_{t=1}^n g^*(X_t,\hat{\theta}) \to_{d^*} N(0,\Sigma)$ prob-P. The convergence of $\mathcal{J}^*_{NBB,n}$ and $\mathcal{J}^{**}_{NBB,n}$ are proven by a similar argument. □

9 Auxiliary results

Lemma 3 *(NBB uniform WLLN). Let $\{q^*_{nt}(\cdot,\omega,\theta)\}$ be an NBB resample of $\{q_{nt}(\omega,\theta)\}$ and assume: (a) For each $\theta \in \Theta \subset \mathbb{R}^p$, Θ a compact set, $n\sum_{t=1}^n(q^*_{nt}(\cdot,\omega,\theta) - q_{nt}(\omega,\theta)) \to 0$, prob-$P^*_{n,\omega}$, prob-P; and (b) $\forall\theta,\theta_0 \in \Theta$, $|q_{nt}(\cdot,\theta) - q_{nt}(\cdot,\theta_0)| \le L_{nt}|\theta-\theta_0|$ a.s.-P, where $\sup_n\{n^{-1}\sum_{t=1}^n E(L_{nt})\} = O(1)$. Then, if $\ell = o(n)$, for any $\delta > 0$ and $\xi > 0$,*

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

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

Lemma 4 *(NBB pointwise WLLN). For some $r > 2$, let $\{q_{nt} : \Omega \times \Theta \to \mathbb{R}^m : m \in \mathbb{N}\}$ 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 an NBB resample of $\{q_{nt}(\omega,\theta)\}$. If $\ell = o(n)$, then for any $\delta > 0$,*

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

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

Proof Fix $\theta \in \Theta$, and we suppress θ and ω henceforth. Since q^*_{nt} is a NBB resample, $E^*q^*_{nt} = n^{-1}\sum_{t=1}^{n} q_{nt} = \bar{q}_n$ and hence $\sum_{t=1}^{n}(q^*_{nt} - q_{nt}) = \sum_{t=1}^{n}(q^*_{nt} - E^*q_{nt})$. From the arguments in the proof of Lemma A.5 of GW04, the stated result follows if $||\text{var}^*(n^{-1/2}\sum_{t=1}^{n} q^*_{nt})||_{r/2} = O(\ell)$ for some $r > 2$. Define $U_{ni} = \ell^{-1}\sum_{t=1}^{\ell} q_{n,(i-1)\ell+t}$, the average of the ith block. Since the blocks are independently sampled, we have (cf. Lahiri (2003), p.48) $\text{var}^*(n^{-1/2}\sum_{t=1}^{n} q^*_{nt}) = b^{-1}\ell\sum_{i=1}^{b}(U_{ni} - \bar{q}_n)(U_{ni} - \bar{q}_n)' = b^{-1}\ell^{-1}\sum_{i=1}^{b}[\sum_{t=1}^{\ell}(q_{n,(i-1)\ell+t} - \bar{q}_n)\sum_{s=1}^{\ell}(q_{n,(i-1)\ell+s} - \bar{q}_n)'] = R_n(0) + b^{-1}\sum_{i=1}^{b}\sum_{\tau=1}^{\ell-1}(R_{ni}(\tau) + R'_{ni}(\tau))$, where $R_n(0) = n^{-1}\sum_{t=1}^{n}(q_{nt} - \bar{q}_n)(q_{nt} - \bar{q}_n)'$, and $R_{ni}(\tau) = \ell^{-1}\sum_{t=1}^{\ell-\tau}(q_{n,(i-1)\ell+t} - \bar{q}_n)(q_{n,(i-1)\ell+t+\tau} - \bar{q}_n)'$, $\tau = 1,\ldots,\ell-1$. Applying Minkowski and Cauchy-Schwartz inequalities gives $||R_n(\tau)||_{r/2} = O(1)$, $\tau = 0,\ldots,\ell-1$, and $||\text{var}^*(n^{-1/2}\sum_{t=1}^{n} q^*_{nt})||_{r/2} = O(\ell)$ follows. □

Lemma 5 *Suppose* $w_N = (w_{N1},\ldots,w_{NN})'$ *follows a multinomial distribution such that* $w_N \sim Mult(b; p_1,\ldots,p_N)$. *Assume further* $\max_{1\le i\le N}|Np_i - 1| \to 0$ *and* $N/b^2 \to 0$ *as* $N \to \infty$. *Then, for* $r = 2,4$, *as* $N \to \infty$,

$$N^{-1}\sum_{i=1}^{N}|(bp_i)^{-1/2}(w_{Ni} - bp_i)|^r \to_P \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E|(bp_i)^{-1/2}(Z(bp_i) - bp_i)|^r,$$

where $Z(c)$ *is a Poisson random variable with mean* c. *The limit on the right hand side exists because* $EZ(c) = c$, $E(Z(c)-c)^2 = c$, *and* $E(Z(c)-c)^4 = 3c^2 + c$.

Corollary 1 *Suppose* $w_b = (w_{b1},\ldots,w_{bb})'$ *follows* $w_b \sim Mult(b; p_1,\ldots,p_b)$. *Assume further* $\max_{1\le i\le b}|bp_i - 1| \to 0$ *as* $b \to \infty$. *Then, for* $r = 2,4$, *as* $b \to \infty$, $b^{-1}\sum_{i=1}^{b}|w_{bi} - bp_i|^r \to_P \lim_{b\to\infty} b^{-1}\sum_{i=1}^{b} E|(Z(bp_i) - bp_i)|^r$, *where* $Z(c)$ *is a Poisson random variable with mean* c.

Proof The proof closely follows that of Lemma 4.1 of Mason and Newton (1992). Their $\{n, j, M_{n,j}\}$ correspond to our $\{N, i, w_{Ni}\}$. We need to adjust the proof of Mason and Newton (1992) because we assume w_N follows a multinomial distribution $(b; p_1,\ldots,p_N)$ whereas Mason and Newton (1992) assume nM_n follows a multinomial distribution $(n; 1/n,\ldots,1/n)$.

Let $U_1, U_2, \ldots$ be a sequence of *iid* $U[0,1]$ random variables. Similar to Mason and Newton (1992), define $G_b(t) = \sum_{k=1}^{b} 1\{U_k \le t\}$ and $G^*_b(t) = \sum_{k=1}^{N(b)} 1\{U_k \le t\}$, where $N(t)$ is a Poisson process independent of U_k's. We can then write $w_{Ni} = \{G_b(p_1 + \cdots + p_i) - G_b(p_1 + \cdots + p_{i-1})\}$ with $p_0 = 0$ for $1 \le i \le N$. Further, analogously to $M^*_{n,j}$ in (4.3) of Mason and Newton (1992), define

$w^*_{Ni} = \{G^*_b(p_1 + \cdots + p_i) - G^*_b(p_1 + \cdots + p_{i-1})\}$, then the elements of $w^*_N = (w^*_{N1}, \ldots, w^*_{NN})'$ are independent Poisson(bp_i) random variables. Consequently, it follows from the weak law of large numbers, $\max_{1 \le i \le N} |Np_i - 1| \to 0$, and $N^{-1}\sum_{i=1}^N |(bp_i)^{-1/2}(bp_i - b/N)|^r \to 0$ that, as in (4.4) of Mason and Newton (1992),

$$N^{-1}\sum_{i=1}^N |(bp_i)^{-1/2}(w^*_{Ni} - \overline{w}^*_N)|^r \to_P \lim_{N\to\infty} N^{-1}\sum_{i=1}^N E|(bp_i)^{-1/2}(Z(bp_i) - bp_i)|^r,$$

where $\overline{w}^*_N = N^{-1}\sum_{i=1}^N w^*_{Ni}$. The stated result follows from replacing S_n and T_n in Mason and Newton (1992) with $S_N = N^{-1}\sum_{i=1}^N |(bp_i)^{-1/2}(w^*_{Ni} - \overline{w}^*_N - w_{Ni} + bp_i)|^r$ and $T_N = E(N^{-1}\sum_{i=1}^N |(bp_i)^{-1/2}(w^*_{Ni} - \overline{w}^*_N - w_{Ni} + bp_i)|^r | N(b))$ and repeating their argument in conjunction with $N^{-1}\sum_{i=1}^N |(bp_i)^{-1/2}(bp_i - b/N)|^r \to 0$ and $n/b^2 \to 0$.

The proof of Corollary 1 follows from repeating the above argument with replacing N with b. □

Lemma 6 *(Consistency of NBB conditional variance). Assume $\{X_t\}$ satisfies $EX_t = 0$ for all t, $||X_t||_{3r} \le \Delta < \infty$ for some $r > 2$ and all $t = 1, 2, \ldots$. Assume $\{X_t\}$ is L_2-NED on $\{V_t\}$ of size $-(2(r-1))/(r-2)$, and $\{V_t\}$ is an α-mixing sequence of size $-(2r/(r-2))$. Let $\{X^*_t\}$ be an NBB resample of $\{X_t\}$. Define $\bar{X}_n = n^{-1}\sum_{t=1}^n X_t$, $\bar{X}^*_n = n^{-1}\sum_{t=1}^n X^*_t$, $\Sigma_n = var(\sqrt{n}\bar{X}_n)$, and $\hat{\Sigma}_n = var^*(\sqrt{n}\bar{X}^*_n)$. Then, if $\ell \to \infty$ and $\ell = o(n^{1/2})$, $\Sigma_n - \hat{\Sigma}_n \to_P 0$.*

Corollary 2 *Assume X_t satisfies the assumptions of Lemma 6. Define $U_i = \ell^{-1}\sum_{t=1}^{\ell} X_{(i-1)\ell+t}$, the average of the ith non-overlapping block. Then, if $\ell \to \infty$ and $\ell = o(n^{1/2})$, $b^{-1}\ell\sum_{i=1}^b U_iU_i' - \Sigma_n \to_P 0$.*

Proof For simplicity, we assume X_t to be a scalar. The extension to the vector-valued X_t is straightforward, see GW02. Define $U_i = \ell^{-1}\sum_{t=1}^{\ell} X_{(i-1)\ell+t}$, the average of the ith block. Since the blocks are independently sampled, we have

$$\begin{aligned}
\hat{\Sigma}_n &= b^{-1}\ell\sum_{i=1}^b U_i^2 - \ell\bar{X}_n^2 \\
&= b^{-1}\ell^{-1}\sum_{i=1}^b \left[\sum_{t=1}^{\ell} X_{(i-1)\ell+t}\sum_{s=1}^{\ell} X_{(i-1)\ell+s}\right] - \ell\bar{X}_n^2 \qquad (26)\\
&= b^{-1}\sum_{i=1}^b \hat{R}_i(0) + 2b^{-1}\sum_{i=1}^b\sum_{\tau=1}^{\ell-1}\hat{R}_i(\tau) - \ell\bar{X}_n^2. \qquad (27)
\end{aligned}$$

where $\hat{R}_i(\tau) = \ell^{-1}\sum_{t=1}^{\ell-\tau} X_{(i-1)\ell+t}X_{(i-1)\ell+t+\tau}$, $\tau = 0, \ldots, \ell-1$. First we show $E(\hat{\Sigma}_n) - \Sigma_n = o(1)$. For the third term on the right of (27), $E|\ell\bar{X}_n^2| = o(1)$ holds because if follows from Lemmas A.1 and A.2

of GW02 that $E(\bar{X}_n^2) = n^{-2}E|\sum_{t=1}^n X_t|^2 \leq n^{-2}E(\max_{1\leq j\leq n}|\sum_{t=1}^j X_t|^2) \leq Cn^{-2}(\sum_{t=1}^n c_t^2) = O(n^{-1})$, where c_t are (uniformly bounded) mixingale constants of X_t. Define $R_i(\tau) = \ell^{-1}\sum_{t=1}^{\ell-\tau} E(X_{(i-1)\ell+t}X_{(i-1)\ell+t+\tau})$ and $R_{ij} = \ell^{-1}\sum_{t=1}^{\ell}\sum_{s=1}^{\ell} E(X_{(i-1)\ell+t}X_{(j-1)\ell+s})$ so that $E(\hat{R}_i(\tau)) = R_i(\tau)$, then $\Sigma_n = b^{-1}\sum_{i=1}^b R_i(0) + 2b^{-1}\sum_{i=1}^b\sum_{\tau=1}^{\ell-1} R_i(\tau) + b^{-1}\sum_{i=1}^b\sum_{j\neq i}^b R_{ij}$, and $E(\hat{\Sigma}_n) - \Sigma_n = b^{-1}\sum_{i=1}^b\sum_{j\neq i}^b R_{ij}$. From Gallant and White (1988) (pp.109-110), $E(X_tX_{t+\tau})$ is bounded by $|EX_tX_{t+\tau}| \leq \Delta(5\alpha_{[\tau/4]}^{1/2-1/r} + 2v_{[\tau/4]}) \leq C\tau^{-1-\xi}$ for some $\xi \in (0,1)$, where v_m is the NED coefficient. Therefore, for $|i-j| = k \geq 2$, we have $|R_{ij}| \leq C\ell^{-1}\sum_{t=1}^{\ell}\sum_{s=1}^{\ell}((k-1)\ell)^{-1-\xi} = O((k-1)^{-1-\xi}\ell^{-\xi})$, and $|R_{i,i+1}| \leq C\ell^{-1}\sum_{t=1}^{\ell}\sum_{s=1}^{\ell}|\ell+s-t|^{-1-\xi} \leq C\ell^{-1}\sum_{h=-\ell+1}^{\ell-1}(\ell-|h|)|\ell+h|^{-1-\xi} = O(\ell^{-\xi})$, where the last equality follows from evaluating the sums with $h > 0$ and $h < 0$ separately. It follows that $b^{-1}\sum_{i=1}^b\sum_{j\neq i}^b R_{ij} = O(\ell^{-\xi} + b^{-1}\sum_{k=2}^{b-1}(b-k)(k-1)^{-1-\xi}\ell^{-\xi}) = O(\ell^{-\xi})$, and we establish $E(\hat{\Sigma}_n) - \Sigma_n = o(1)$.

It remains to show $\mathrm{var}(\hat{\Sigma}_n) = o(1)$. It suffices to show that the variance of

$$b^{-1}\sum_{i=1}^{b}\left(\hat{R}_i(0) - R_i(0)\right) + 2b^{-1}\sum_{i=1}^{b}\sum_{\tau=1}^{\ell-1}\left(\hat{R}_i(\tau) - R_i(\tau)\right) \tag{28}$$

is $o(1)$. Following the derivation in GW02 leading to their equation (A.4), we obtain $\mathrm{var}(\hat{R}_i(\tau)) \leq \ell^{-2}\sum_{t=1}^{\ell-\tau}\mathrm{var}(X_{(i-1)\ell+t}X_{(i-1)\ell+t+\tau}) + 2\ell^{-2}\sum_{t=1}^{\ell-\tau}\sum_{s=t+1}^{\ell-\tau}|\,\mathrm{cov}(X_{(i-1)\ell+t}X_{(i-1)\ell+t+\tau}, X_{(i-1)\ell+s}X_{(i-1)\ell+s+\tau})| \leq C\ell^{-1}\{\Delta + \sum_{k=1}^{\infty}\alpha_{[k/4]}^{1/2-1/r} + \sum_{k=1}^{\infty}v_{[k/4]} + \sum_{k=1}^{\infty}v_{[k/4]}^{(r-2)/2(r-1)}\} + C\ell^{-1}(\tau\alpha_{[\tau/4]}^{1-2/r} + \tau v_{[\tau/4]}^2 + 2\tau\alpha_{[\tau/4]}^{1/2-1/r}v_{[\tau/4]}) = O(\ell^{-1})$. Observe that, when $|i-j| \geq 7$, from Lemma 6.7(a) of Gallant and White (1988) we have, for some $\xi \in (0,1)$, $\mathrm{cov}(\hat{R}_i(\tau), \hat{R}_j(\tau)) \leq \ell^{-2}\sum_{t=1}^{\ell-\tau}\sum_{s=1}^{\ell-\tau}|\mathrm{cov}(X_{(i-1)\ell+t}X_{(i-1)\ell+t+\tau}, X_{(j-1)\ell+s}X_{(j-1)\ell+s+\tau})| \leq \ell^{-2}\sum_{t=1}^{\ell-\tau}\sum_{s=1}^{\ell-\tau}(\alpha_{[(|i-j|-6)\ell/4]}^{1/2-1/r} + v_{[(|i-j|-6)\ell/4]}^{(r-2)/2(r-1)}) = O(\ell^{-2}\sum_{t=1}^{\ell-\tau}\sum_{s=1}^{\ell-\tau}[(|i-j|-6)\ell/4]^{-1-\xi}) \leq C(\ell|i-j|)^{-1-\xi}$. Define $B_r = \{1 \leq i \leq b : i = 7k + r, k \in \mathbb{N}\}$ for $r = 1, \ldots, 7$, so that all $i \in B_r$ are at least 7 apart from each other. Rewrite (28) as $\sum_{r=1}^7 b^{-1}\sum_{i\in B_r}(\hat{R}_i(0) - R_i(0)) + 2\sum_{r=1}^7\sum_{\tau=1}^{\ell-1} b^{-1}\sum_{i\in B_r}(\hat{R}_i(\tau) - R_i(\tau))$. Then, for $\tau = 0, \ldots, \ell-1$, we have $\mathrm{var}(b^{-1}\sum_{i\in B_r}(\hat{R}_i(\tau) - R_i(\tau))) = b^{-2}\sum_{i\in B_r}\sum_{j\in B_r}\mathrm{cov}(\hat{R}_i(\tau), \hat{R}_j(\tau)) = O(b^{-1}\ell^{-1} + \ell^{-1-\xi}b^{-2}\sum_{i=1}^b\sum_{j\neq i}^b|i-j|^{-1-\xi}) = O(b^{-1}\ell^{-1} + \ell^{-1-\xi}b^{-2}\sum_{h=1}^{b-1}(b-h)h^{-1-\xi}) = O(b^{-1}\ell^{-1})$. Therefore, the variance of (28) is $O(\ell b^{-1}) = O(\ell^2 n^{-1}) = o(1)$, giving the stated result. Corollary 2 follows because $b^{-1}\ell\sum_{i=1}^b U_iU_i' = \hat{\Sigma}_n + o_P(1)$ from (26). $\square$

Table 1: Linear Model - symmetric errors

Replications=2000; Bootstraps=499; auto-selection block length

$y_t = \theta_1 + \theta_2 x_t + u_t$; $u_t = 0.9u_{t-1} + \varepsilon_{1t}$;

$x_t = 0.9x_{t-1} + \varepsilon_{2t}$; $z_t = (\iota\ x_t\ x_{t-1}\ x_{t-2})$

$(\theta_1, \theta_2) = (0,0)$; $[\varepsilon_{1t}, \varepsilon_{2t}] \sim N(0, I_2)$

	T-Test			Sargan Test		
	10	05	01	10	05	01
100						
Asymptotic	0.4225	0.3420	0.2335	0.1360	0.0735	0.0245
SNB	0.2725	0.2070	0.1085	0.1505	0.0945	0.0320
SMB	0.3760	0.2885	0.1640	0.1330	0.0755	0.0255
ENB	0.2265	0.1830	0.1150	0.0675	0.0460	0.0220
EMB	0.2290	0.2260	0.1120	0.0775	0.0560	0.0250
250						
Asymptotic	0.3485	0.2755	0.1625	0.1225	0.0745	0.0235
SNB	0.2090	0.1460	0.0720	0.1320	0.0840	0.0310
SMB	0.3255	0.2390	0.1320	0.1315	0.0790	0.0260
ENB	0.1385	0.0990	0.0455	0.0815	0.054	0.0260
EMB	0.1500	0.1250	0.0500	0.1140	0.0830	0.0480
1000						
Asymptotic	0.2735	0.1945	0.0955	0.0925	0.0460	0.0075
SNB	0.1675	0.1140	0.0425	0.0930	0.0505	0.0090
SMB	0.2550	0.1815	0.0830	0.0970	0.0450	0.0070
ENB	0.0995	0.0605	0.0230	0.0875	0.0480	0.0145
EMB	0.0960	0.0590	0.0020	0.1045	0.0560	0.0180

Table 2: Linear Model - GARCH(1,1) errors

Replications=2000; Bootstraps=499; auto-selection block length

$y_t = \theta_1 + \theta_2 x_t + \sigma_t u_t$; $u_t \sim N(0, \sigma_t), \sigma_t^2 = 0.0001 + 0.6\sigma_{t-1}^2 + 0.3\varepsilon_{1t-1}$;

$x_t = 0.75x_{t-1} + \varepsilon_{2t}$, where $\varepsilon_{1t} \sim N(0,1)$; $z_t = (\iota\, x_t\, x_{t-1}\, x_{t-2})$

$(\theta_1, \theta_2) = (0,0)$; $\varepsilon_{1t} \sim N(0,1)$

	T-Test			Sargan Test		
	10	05	01	10	05	01
100						
Asymptotic	0.1420	0.0840	0.0280	0.070	0.0240	0.0040
SNB	0.0820	0.0340	0.0060	0.0530	0.0180	0.0050
SMB	0.0920	0.0480	0.0060	0.0590	0.0160	0.0050
ENB	0.0875	0.0405	0.0006	0.0730	0.0300	0.0040
EMB	0.1405	0.0870	0.0200	0.1100	0.0600	0.0100
250						
Asymptotic	0.1150	0.0580	0.0150	0.0840	0.0270	0.0040
SNB	0.0630	0.0300	0.0060	0.0820	0.0230	0.0030
SMB	0.0830	0.0370	0.0080	0.0760	0.0260	0.0040
ENB	0.0995	0.0410	0.0065	0.0845	0.0330	0.0035
EMB	0.1130	0.0570	0.0210	0.1510	0.0950	0.0140
1000						
Asymptotic	0.1050	0.0560	0.0150	0.0880	0.0390	0.0060
SNB	0.0700	0.0340	0.0070	0.0840	0.0420	0.0050
SMB	0.0910	0.0470	0.0110	0.0860	0.0410	0.0060
ENB	0.0995	0.0490	0.0090	0.0885	0.0370	0.0090
EMB	0.1000	0.0560	0.0090	0.0900	0.0490	0.0100

Note: The mean block length is 1.96 when $T = 100$, 2.84 when $T = 250$, and 4.48 when $T = 1000$.

Table 3: Nonlinear Model - Chi-Square Moment Conditions

Replications=2000; Bootstraps=499; auto-selection block length

$g(X_t, \theta_1) = (X_t - \theta_1 \quad X_t^2 - \theta_1^2 - 2\theta_1)'$.

	T-Test			Sargan Test		
	10	05	01	10	05	01
100						
Asymptotic	0.1845	0.1250	0.0625	0.2655	0.2065	0.1195
SNB	0.1535	0.1000	0.0380	0.1895	0.1505	0.0870
SMB	0.1800	0.0875	0.0070	0.1825	0.1465	0.0780
ENB	0.1175	0.0575	0.0100	0.2135	0.1470	0.0750
EMB	0.1100	0.0620	0.0090	0.2000	0.1550	0.0700
250						
Asymptotic	0.1245	0.0700	0.0250	0.1990	0.1560	0.0840
SNB	0.1095	0.0585	0.0200	0.1615	0.1290	0.0790
SMB	0.1240	0.0710	0.0175	0.1520	0.1225	0.0695
ENB	0.1050	0.0560	0.0120	0.1720	0.1200	0.0420
EMB	0.1040	0.0610	0.0110	0.1780	0.1280	0.0380
1000						
Asymptotic	0.0975	0.0515	0.0100	0.1325	0.0835	0.0400
SNB	0.0985	0.0620	0.0205	0.1335	0.0985	0.0580
SMB	0.0795	0.0395	0.0075	0.1180	0.0870	0.0430
ENB	0.0965	0.0480	0.0095	0.1120	0.0695	0.0240
EMB	0.0940	0.0400	0.0060	0.1340	0.0700	0.0400

Note: The mean block length is 1.29 when $T = 100$, 1.99 when $T = 250$, and 3.33 when $T = 1000$.

Table 4: Nonlinear Model - Asset Pricing Model

Replications=2000; Bootstraps=499; auto-selection block length
$g = (\exp(\mu - \theta(x+z) + 3z) - 1 \quad z[\exp(\mu - \theta(x+z) + 3z) - 1])$,
$\log x_t = \rho \log x_{t-1} + \sqrt{(1-\rho^2)}\varepsilon_{xt}, \quad z_t = \rho z_{t-1} + \sqrt{(1-\rho^2)}\varepsilon_{zt}$,
where ε_{xt} and ε_{zt} are independent normal with mean 0 and variance 0.16. In the experiment $\rho = 0.6$.

	T-Test			Sargan Test		
	10	05	01	10	05	01
100						
Asymptotic	0.4010	0.3235	0.2195	0.3080	0.2350	0.1460
SNB	0.1550	0.0985	0.0400	0.1880	0.1260	0.0385
SMB	0.1540	0.1015	0.0435	0.1930	0.1300	0.0420
ENB	0.1300	0.0780	0.0245	0.1250	0.0700	0.0150
EMB	0.1360	0.0825	0.0260	0.1880	0.0810	0.0200
250						
Asymptotic	0.3005	0.2275	0.1240	0.2470	0.1850	0.0995
SNB	0.1270	0.0755	0.0290	0.1435	0.1005	0.0510
SMB	0.1285	0.0780	0.0290	0.1430	0.0985	0.0535
ENB	0.1200	0.0620	0.0140	0.1210	0.0670	0.0180
EMB	0.1290	0.0600	0.0210	0.1245	0.0650	0.0270
1000						
Asymptotic	0.2205	0.1440	0.0545	0.1975	0.1335	0.0685
SNB	0.1440	0.0825	0.0280	0.1005	0.0715	0.0220
SMB	0.1420	0.0820	0.0250	0.1040	0.0660	0.0220
ENB	0.1180	0.0600	0.0220	0.1300	0.0695	0.0210
EMB	0.1160	0.0560	0.0160	0.1090	0.0700	0.0150

Note: The mean block length is 1.51 when $T = 100$, 2.62 when $T = 250$, and 4.96 when $T = 1000$.

References

AHN, S., AND P. SCHMIDT (1995): "Efficient Estimation of Models for Dynamic Panel Data," Journal of Econometrics, 68, 5–27.

ALTONJI, J., AND L. SEGAL (1996): "Small-sample Bias in GMM Estimation of Covariance Structures," Journal of Business & Economic Statistics, 14, 353–366.

ANATOLYEV, S. (2005): "GMM, GEL, Serial Correlation, and Asymptotic Bias," Econometrica, 73, 983–1002.

ANDREWS, D. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," Econometrica, 59.

——— (2002): "Higher-ordeer Improvements of a Computationally Attractive k-step Bootstrap for Extremum Estimators," Econometrica, 70, 119–262.

ANDREWS, D., AND J. MONAHAN (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," Econometrica, 60, 953–966.

ARELLANO, M., AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," Review of Economic Studies, 58, 277–297.

BARBE, P., AND P. BERTAIL (1995): The Weighted Bootstrap. Springer-Verlag, Berlin.

BERKOWITZ, J., AND L. KILIAN (2000): "Recent Developments in Bootstrapping Time Series," Econometric Reviews, 19, 1–48.

BRAVO, F. (2005): "Blockwise Empirical Entropy Tests for Time Series Regressions," Journal of Time Series Analysis, 26, 185–210.

BROWN, B., AND W. NEWEY (2002): "Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference," Journal of Business & Economic Statistics, 20, 507–517.

BÜHLMANN, P., AND H. KÜNSCH (1999): "Block Length Selection in the Bootstrap for Time Series," Computational Statistics & Data Analysis, 31, 295–310.

CARLSTEIN, E. (1986): “The Use of Subseries Methods for Estimating the Variance of a General Statistic from a Stationary Time Series,” The Annals of Statistics, 14, 1171–1179.

CHRISTIANO, L., AND W. HAAN (1996): “Small-sample Properties of GMM for Business-cycle Data,” Journal of Business & Economic Statistics, 14, 309–327.

CLARK, T. (1996): “Small-sample Properties of Estimators of Nonlinear Models of Covariance Structure,” Journal of Business and Economic Statistics, 14, 367–373.

DAVIDSON, R., AND J. G. MACKINNON (1999): “Bootstrap Testing in Nonlinear Models,” International Economic Review, 40, 487–508.

DE JONG, R., AND J. DAVIDSON (2000): “Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices,” Econometrica, 68, 407–423.

DURRETT, R. (2005): Probability: Theory and Examples. Duxbury Press, third edn.

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., AND H. WHITE (1988): A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Blackwell.

GONÇALVES, S., AND H. WHITE (2002): “The Bootstrap of the Mean for Dependent Heterogeneous Srrays,” Econometric Theory, 18, 1367–1384.

——— (2004): “Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models,” Journal of Econometrics, 119, 199–219.

GONZALEZ, A. (2007): “Empirical Likelihood Estimation in Dynamic Panel Models,” mimeo.

GREGORY, A., J. LAMARCHE, AND G. W. SMITH (2002): “Information-theoretic Estimation of preference Parameters: Macroecnomic Applications and Simulation Evidence,” Journal of Econometrics, 107, 213–233.

HAHN, J. (1996): “A Note on Bootstrapping Generalized Method of Moments Estimators,” Econometric Theory, 12, 187–197.

HALL, P., AND J. HOROWITZ (1996): "Bootstrap Critical Values for Tests Based on Generalized-method-of-moments Estimators," Econometrica, 64, 891–916.

HALL, P., AND E. MAMMEN (1994): "On General Resampling Algorithms and Their Performance in Distribution Estimation," Annals of Statistics, 22, 2011–2030.

HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, 50, 1029–1054.

HANSEN, L., AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," Econometrica, 50, 1296–1286.

HÄRDLE, W., J. HOROWITZ, AND J. KREISS (2003): "Bootstrapping Methods for Time Series," International Statistical Review, 71, 435–459.

HOEFFDING, W. (1951): "A Combinatorial Central Limit Theorem," Annals of Mathematical Statistics, 22, 558–566.

HONG, H., AND O. SCAILLET (2006): "A Fast Subsampling Method for Nonlinear Dynamic Models," Journal of Econometrics, 133.

IMBENS, G., R. SPADY, AND P. JOHNSON (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," Econometrica, 66, 333–357.

INOUE, A., AND M. SHINTANI (2006): "Bootstrapping GMM Estimators for Time Series," Journal of Econometrics, 133, 531–555.

KITAMURA, Y. (1997): "Empirical Likelihood Methods with Weakly Dependent Processes," The Annals of Statistics, 25, 2084–2102.

——— (2007): Empirical Likelihood Methods in Econometrics: Theory and Practice,vol. Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, 3, Volume III of Econometric Society Monograph ESM 4, pp. 174–237. Cambridge: Cambrdige University Press.

KITAMURA, Y., AND M. STUTZER (1997): "An Information-theoretic Alternative to Generalized Method of Moments Estimation," Econometrica, 65, 861–874.

KOCHERLAKOTA, N. (1990): "On Tests of Representative Consumer Asset Pricing Models," Journal of Monetary Economics, 25, 43–48.

KÜNSCH, H. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," The Annals of Statistics, 17, 1217–1261.

LAHIRI, S. (1999): "Theoretical Comparisons of Block Bootstrap Methods," Annals of Statistics, 27, 384–404.

——— (2003): Resampling Methods for Dependent Data. Springer.

MASON, D., AND M. NEWTON (1992): "A Rank Statistics Approach to the Consistency of a General Bootstrap," Annals of Statistics, 20, 1611–1624.

NEWEY, W., AND R. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," Econometrica, 72, 219–256.

NEWEY, W., AND K. WEST (1994): "Automatic Lag Selection in Covariance Matrix Estimation," Review of Economic Studies, 61, 631–654.

OWEN, A. (1990): "Empirical Likelihood Ratio Confidence Regions," Annals of Statistics, 18, 90–120.

POLITIS, D., AND J. ROMANO (1994): "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," The Annals of Statistics, 22.

——— (1995): "Bias-Corrected Nonparametric Spectral Estimation," Journal of Time Series Analysis, 16.

POLITIS, D., J. ROMANO, AND M. WOLF (1999): Subsampling. New York: Springer.

POLITIS, D., AND H. WHITE (2004): "Automatic Block-Length Selection for the Dependent Bootstrap," Econometric Reviews, 23, 53–70.

QIN, J., AND J. LAWLESS (1994): "Empirical Likelihood and General Estimating Equations," The Annals of Statistics, 22, 300–325.

RAMALHO, J. (2006): "Bootstrap Bias-Adjusted GMM Estimators," Economics Letters, 92, 149–155.

RUGE-MURCIA, F. (2007): “Methods to Estimate Dynamic Stochastic General Equilibrium Models,” Journal of Economic Dynamics and Control, 31, 2599–2636.

RUIZ, E., AND L. PASCUAL (2002): “Bootstrapping Financial Time Series,” Journal of Economic Surveys, 16, 271–300.

SMITH, R. (1997): “Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation,” The Economic Journal, 107, 503–519.

ZVINGELIS, J. (2003): On Bootstrap Coverage Probability with Dependent Data,vol. Computer-Aided Econometrics, pp. 69–90. New York: Marcel Dekker.