EMPIRICAL LIKELIHOOD BLOCK BOOTSTRAPPING

Jason Allen‡, Allan Gregory♭, and Katsumi Shimotsu♭∗

Department of Monetary and Financial Analysis, Bank of Canada‡
and Department of Economics, Queen’s University♭

January 19, 2007

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 and 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

PRELIMINARY: COMMENTS WELCOME

*Correspondence can be sent to J. Allen (JAllen@bankofcanada.ca). We thank Don Andrews, Geoffrey Dunbar, Atushi Inoue, Gregor Smith, and Tim Vogelsang for helpful comments and insightful discussion. We also thank seminar participants at the Bank of Canada, the Econometric Society World Congress in London (2005), and the Canadian Econometric Study Group in Vancouver (2005). We acknowledge the Social Sciences and Humanities Research Council of Canada for support of this research.
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 (2003)).

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. The rise of calibration methods may in part be a response of theorists to existing GMM methods.

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. 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 Zwingelis (2002) 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. Zwingelis (2002) 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 and Stutzer (1997), and Imbens, Spady, and Johnson (1998) have proposed and/or tested new statistics, ones based on GEL-estimators. 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 authors 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 dis-similar to the Hall and Horowitz (1996) bootstrap.

In this chapter, the approach of Brown and Newey (2002) is extended to the case of dependent data, using the empirical likelihood (Owen (1990)). The 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. The first-order asymptotic validity of the proposed empirical likelihood block bootstrap is proven using the results in Gonçalves and White (2004). This chapter reports on the finite-sample properties of $t$-ratios, and overidentification test-statistics, including a Lagrange Multiplier test first proposed by Imbens, Spady, and Johnson (1998) that uses empirical likelihood. 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. Furthermore, the empirical likelihood block bootstrap does not require solving the difficult empirical likelihood saddle point problem since estimation of the probability weights can be conducted using the first-round GMM estimates. This is a common argument amongst applied researchers for not switching from GMM to EL even though the latter is higher-order efficient.

The chapter is organized as follows. Chapter 2.2 presents an overview of GMM and EL. Chapter 2.3 presents a discussion of how resampling methods might improve inference in GMM. Chapter 2.4 presents the Monte Carlo design for both linear and nonlinear models. The technical assumption and proofs are collected at the end of the chapter in the mathematical appendix.
2 Overview of GMM and GEL

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 \( \theta_0 (p \times 1) \) the following moment conditions (m equations) hold and \( p \leq m < n \):

\[
E[g(X_t, \theta_0)] = 0, \quad (1)
\]

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

\[
\hat{\theta} = \arg \min Q_n(\theta), \quad Q_n(\theta) = \left( n^{-1} \sum_{i=1}^{n} g(X_t, \theta) \right)' W_n \left( n^{-1} \sum_{i=1}^{n} g(X_t, \theta) \right), \quad (2)
\]

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 asymptotic covariance matrix is \( G' \Sigma G \) where

\[
G = \lim_{n \to \infty} E\left( n^{-1} \sum_{t=1}^{n} \nabla g(X_t, \theta_0) \right), \quad \text{with } \nabla g(x, \theta) = \partial g(x, \theta) / \partial \theta. \]

The elements of \( \{g(X_t, \theta)\} \) and \( \{\nabla g(x, \theta)\} \) are assumed to be near epoch dependent (NED) on the \( \alpha \)-mixing sequence \( \{V_t\} \) of size \( -1 \) uniformly on \( (\Theta, \rho) \) where \( \rho \) is any convenient norm on \( \mathbb{R}^p \). \( ||x||_p \) denotes the \( L_p \) norm \( (E|X_n|^p)^{1/p} \). For a \((m \times k)\) matrix \( x \), let \( |x| \) denote the 1-norm of \( x \), so \( |x| = \sum_{i=1}^{m} \sum_{j=1}^{k} |x_{ij}| \).

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

\[
J_n = K_n(\hat{\theta}_n)' K_n(\hat{\theta}_n) \to_d \chi_{m-r}, \quad (3)
\]

where

\[
K_n(\theta) = S_n^{1/2} n^{-1/2} \sum_{i=1}^{n} g(x_t, \theta_n),
\]

and \( S_n \) is a consistent estimate of \( \Sigma \). The t-statistic for testing the null hypothesis \( H_0 : \hat{\theta}_{nr} = \theta_{nr} \) is:

\[
T_{nr} = \frac{\sqrt{m}(\hat{\theta}_{nr} - \theta_{nr})}{(S_{nr})^{1/2}} \to_d N(0,1), \quad (4)
\]

where \( \hat{\theta}_{nr} \) and \( \theta_{nr} \) are the \( r \)th element of \( \hat{\theta}_n \) and \( \theta_0 \). The standard long-run kernel autocovariance estimate for the GMM estimator is

\[
S_n(\theta) = \sum_{h=-n}^{n} k \left( \frac{h}{m} \right) \hat{\Gamma}(h, \theta), \quad (5)
\]

where \( k(\cdot) \) is a kernel and \( \hat{\Gamma}(h, \theta) = n^{-1} \sum_{t=h+1}^{n} g(X_t, \hat{\theta})g(X_{t+h}, \hat{\theta})' \) for \( h \geq 0 \) and \( n^{-1} \sum_{t=1}^{n-h} g(X_t, \hat{\theta})g(X_{t+h}, \hat{\theta})' \) for \( h < 0 \). It is known that \( S_n(\hat{\theta}) \) converges to \( \Sigma \) in probability under weak conditions on kernel and bandwidth; see de Jong and Davidson (2000). The optimal weighting matrix is given by \( S_n(\hat{\theta})^{-1} \).
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.

Smith (2000) generalizes EL to include several specifications. The Generalize 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 \(\pi_t\) gives

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

In the case of EL, \(h(\cdot) = \log(\pi_t)\), and exponential tilting (ET) corresponds to \(h(\cdot) = \pi_t \log(\pi_t)\). Kitamura and Stutzer (1997) show for ET the Lagrangian amounts to a saddle point problem where instead of solving the Lagrangian one can solve the dual. Kitamura and Stutzer (1997) also 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) addresses serial correlation in the moment conditions by using averages across blocks of data.

### 3 Improving Inference: Resampling Methods

Under the assumption of finite autocorrelation of the moment conditions, Hall and Horowitz (1996) show that block bootstrapping provides asymptotic refinements to the critical values of t-tests and Hansen’s J-test. A small Monte Carlo experiment, consisting of two nonlinear moment conditions and
one parameter, is used to show that the block bootstrap usually reduces the errors in level from the critical values based on first-order asymptotic theory.\footnote{This chapter follows this design in the Monte Carlo experiments and also includes cases with persistence, heteroscedasticity, and asymmetry in the moment conditions.}

### 3.1 The Block Bootstrap

The bootstrap amounts to treating the estimation data as if they were the population and carrying a Monte Carlo in which bootstrap data is generated 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. 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 (2002) 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.

Two forms of the block bootstrap are implemented in this chapter. The first approach implements the overlapping bootstrap (MBB, Künsch (1989)). Let $b$ be the number of blocks and $\ell$ the block length, such that $n = b\ell$. The $i$th overlapping block is $\hat{X}_i = \{X_i, \ldots, X_i + \ell - 1\}$, $i = 1, \ldots, n - \ell + 1$. The MBB resample is $\{X^*_i\}_{i=1}^n = \{\hat{X}_1^*, \ldots, \hat{X}_b^*\}$, where $\hat{X}_i^* \sim iid(\hat{X}_1, \ldots, \hat{X}_{n-\ell+1})$. The GMM estimator is therefore:

$$\theta_{MBB}^* = \arg\min \mathcal{Q}_{MBB,n}^*(\theta),$$

$$\mathcal{Q}_{MBB,n}^*(\theta) = \left(n^{-1} \sum_{t=1}^{n} g^*(X^*_t, \theta) \right)' W_n^* \left(n^{-1} \sum_{t=1}^{n} g^*(X^*_t, \theta) \right),$$

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 $\theta_{MBB}^*$.

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 $\ell$ the block length, and assume $b\ell = n$. We resample $b$ blocks with replacement from $\{\hat{X}_i : i = 1, \ldots, b\}$ where $\hat{X}_i = (X_{i-1}\ell+1, \ldots, X_{i-1}\ell+\ell)$. The NBB resample is $\{X^*_i\}_{i=1}^n$. The NBB version of the GMM problem is identical to the MBB version, except for the way one resamples the data.
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 (c.f. equation (3.1) of GW04):

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

where $\theta^{**}$ denotes either $\theta_{NBB}^{**}$ or $\theta_{MBB}^{**}$. The optimal weighting matrix is given by $(S_n^{**}(\tilde{\theta}^{**}))^{-1}$, where $\tilde{\theta}^{**}$ is the first-stage MBB/NBB estimator.

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 $\theta$ 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 $\hat{\theta}_n$ (c.f. Hahn (1996)), but is necessary for the first-order validity of the bootstrap J-test.

Both bootstrap approaches are considered because there is little known about the finite sample properties of either method. It is, however, known that the bias and variance of a block bootstrap estimator depends on the block length (Hall, Horowitz, and Jing (1992)), and that the MBB is more efficient than the NBB in estimating the variance (Lahiri (1999)).

3.2 Empirical Likelihood Bootstrap

In this section we develop the empirical likelihood approach to estimating time-series models. Two cases are considered: (i) the overlapping empirical likelihood block bootstrap (EL-MBB), and (ii) the non-overlapping empirical likelihood block bootstrap (EL-NBB). The procedure for implementing the empirical block bootstrap is straightforward and outlined at the end of this chapter in section 2.5.

3.2.1 EL-MBB

First consider the overlapping bootstrap. Let $N = n - \ell + 1$ be the total number of overlapping blocks. Define the $i$th 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
\[ L = \sum_{i=1}^{N} \log(\pi_i^\omega) + \mu \left( 1 - \sum_{i=1}^{N} \pi_i^\omega \right) - N\gamma \sum_{i=1}^{N} \pi_i^\omega T_i^\omega(\theta). \]

It is known that the solution for the probability weights are given by:
\[ \pi_i^\omega = \frac{1}{N} \left( \frac{1}{1 + \gamma(\theta) T_i^\omega(\theta)} \right), \]
where
\[ \gamma(\theta) = \max_{\lambda^o \in \Lambda_0(\theta)} \sum_{i=1}^{N} \log(1 + \gamma T_i^\omega(\theta)). \]  

(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 \( \theta \) to compute \( \hat{\pi}_i^\omega \) and attach these weights to the bootstrapped (blocks of) samples. Given the GMM estimate \( \hat{\theta} \), compute \( \gamma(\hat{\theta}) \), which is a much smaller dimensional problem. Then solve for the empirical probability weights:
\[ \hat{\pi}_i^\omega = \frac{1}{N} \left( \frac{1}{1 + \gamma(\hat{\theta}) T_i^\omega(\hat{\theta})} \right), \]
which satisfy the moment condition \( \sum_{i=1}^{N} \hat{\pi}_i^\omega T_i^\omega(\hat{\theta}) = 0 \). The EL-MBB version of \( \hat{\theta} \) is defined as:
\[ \theta_{MBB}^* = \arg\min Q_{MBB,n}(\theta), \]
\[ Q_{MBB,n}(\theta) = \left( \frac{1}{b} \sum_{i=1}^{b} \sum_{j=1}^{N} \hat{\pi}_i^\omega T_i^\omega(\theta) \right)' W_{MBB,n} \left( \frac{1}{b} \sum_{i=1}^{b} \sum_{j=1}^{N} \hat{\pi}_i^\omega T_i^\omega(\theta) \right), \]
where \( W_{MBB,n} \) is a weighting matrix and \( \{ \hat{\pi}_i^\omega T_i^\omega(\theta) \} \) are \( b \) iid samples (with replacement) from \( \{ \hat{\pi}_i^\omega T_i^\omega(\theta) : j = 1, \ldots, N \} \). The multiplicative numbers \( b^{-1} \) and \( N \) are included so that the order of \( Q_{MBB,n}(\theta) \) mimics that of \( Q_n(\theta) \). Note that \( E^* \hat{\pi}_i^\omega T_i^\omega(\hat{\theta}) = N^{-1} \sum_{i=1}^{N} \hat{\pi}_i^\omega T_i^\omega(\hat{\theta}) = 0 \).

The long-run autocovariance matrix estimator for EL-MBB takes the form:
\[ S_{MBB,n}(\theta) = \ell b^{-1} \sum_{i=1}^{b} (N \hat{\pi}_i^\omega T_i^\omega(\theta) \hat{\pi}_i^\omega T_i^\omega(\theta))' = \ell b^{-1} N^2 \sum_{i=1}^{b} \hat{\pi}_i^\omega T_i^\omega(\theta) \hat{\pi}_i^\omega T_i^\omega(\theta)', \]
and the second-stage (optimal) weighting matrix is given by \( S_{MBB,n}(\hat{\theta}_{MBB}^*)^{-1} \), where \( \hat{\theta}_{MBB}^* \) is the first-stage EL-MBB estimator. The validity of the overlapping block Wald tests are based on the long-run autocovariance matrix \( S_{MBB,n}(\theta) \).
3.2.2 EL-NBB

The EL-NBB uses $b$ non-overlapping blocks rather than overlapping blocks. The $i$th 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:

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

(12)

The EL-NBB estimator is defined as:

$$\theta^*_{NBB} = \arg \min_{\theta} Q^*_{NBB,n}(\theta), \quad Q^*_{NBB,n}(\theta) = \left( \sum_{i=1}^{b} \hat{\pi}_i^* T_i^* (\theta) \right)' W^*_{NBB,n} \left( \sum_{i=1}^{b} \hat{\pi}_i^* T_i^* (\theta) \right),$$

where $W^*_{NBB,n}$ is a weighting matrix. The long-run autocovariance matrix estimator for EL-NBB is:

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

(13)

and the optimal weighting matrix is given by $S^*_{NBB,n}(\tilde{\theta}^*_{NBB})^{-1}$, where $\tilde{\theta}^*_{NBB}$ is the first-stage EL-NBB estimator. The validity of the non-overlapping block Wald tests are based on the long-run autocovariance matrix, $S^*_{NBB,n}(\theta)$.

3.3 Test-statistic based on the Lagrange Multiplier

This section explores a block-bootstrap version of the Lagrange multiplier overidentifying restrictions test (based on GMM) introduced by Imbens, Spady, and Johnson (1998).

Given the evidence concerning the poor finite sample properties of Hansen’s $J$-test, Imbens, Spady, and Johnson (1998) suggest an alternative overidentifying restrictions tests based on exponential tilting. Monte Carlo evidence presented in their paper reveal that the Lagrange multiplier (LM)-test has actual size closer to nominal size in finite samples that Hansen’s $J$-test. Following the statistic $T^{LM}_{gmm(r)}$ in Imbens, Spady, and Johnson (1998) we consider a similar test of overidentifying restrictions based on the Lagrange multiplier associated with the GMM estimator $\hat{\theta}$.\(^2\) The form of the LM-test is dependent on the bootstrap approach.

\(^2\)Note that Imbens, Spady, and Johnson (1998) derive the tests for exponential tilting and not empirical likelihood.
When overlapping blocks are employed, The LM-test has the form:

\[ T^o = \ell^2 N \gamma'(\hat{\theta}) S_n(\hat{\theta}) \gamma'(\hat{\theta}) \]

where \( \gamma'(\theta) \) is the Lagrange multiplier defined in equation 9, and \( S_n(\theta) \) is the long-run autocovariance estimator defined in equation 5.

In view of \( N \hat{\pi}_i \sim 1 \), we define the corresponding Lagrange multiplier from the bootstrapped sample as:

\[ \gamma^o(\theta^*_{MBB}) = \arg \max_{\gamma \in \Lambda_n(\theta^*_{MBB})} \left\{ \frac{1}{N} \sum_{i=1}^N \log \left( 1 + \gamma \hat{\pi}_i T_i^* (\theta^*_{MBB}) \right) \right\}, \]

and the test-statistic takes the form:

\[ T^{o*} = \ell^2 N \gamma^{o*}(\theta^*_{MBB}) S^{*}_{MBB,n}(\theta^*_{MBB}) \gamma^{o*}(\theta^*_{MBB}), \]

where \( S^{*}_{MBB,n}(\theta) \) is the long-run autocovariance matrix of the EL-MBB estimator and is defined in equation 11.

When non-overlapping blocks are employed the test-statistics are defined slightly differently. The Lagrange multiplier test based on the original time-series is defined as:

\[ T = b^2 n^{-1} \gamma(\hat{\theta}) S_n(\hat{\theta}) \gamma(\hat{\theta}), \]

where \( \gamma(\theta) \) is defined in equation 12. Since \( \hat{\pi}_i \sim b^{-1} \), its bootstrap counterpart is:

\[ \gamma^*(\theta^*_{NBB}) = \arg \max_{\gamma \in \Lambda_n(\theta^*_{NBB})} \left\{ \frac{1}{b} \sum_{i=1}^b \log \left( 1 + \gamma b \hat{\pi}_i T_i^* (\theta^*_{NBB}) \right) \right\}, \]

and the test-statistic takes the form:

\[ T^* = b^2 n^{-1} \gamma^*(\theta^*_{NBB}) S^{*}_{NBB,n}(\theta^*_{NBB}) \gamma^*(\theta^*_{NBB}) \]

where \( S^{*}_{NBB,n}(\theta) \) is the long-run autocovariance matrix of the EL-NBB estimator and is defined in equation 13.

The following lemmas establish the consistency of the bootstrap-based inference. The proofs are based on the results in Gonçalves and White (2004). As in GW04, let \( P \) denote the probability measure that governs the behaviour 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 \( \delta > 0 \), \( \lim_{n \to \infty} P[P^*|\|T^*_n\| > \varepsilon] > \delta = 0 \). 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.
Lemma 1 Suppose Assumption A in Appendix hold. Then \( \hat{\theta} - \theta_0 \rightarrow^{p} 0 \). If also \( \ell \rightarrow \infty \) and \( \ell = o(n) \), then \( \theta^{**}_{MBB} - \hat{\theta} \rightarrow^{p} 0 \). If also Assumption B in Appendix hold and \( \ell = o(n^{1/2 - 1/r}) \), then \( \theta^{**}_{MBB} - \hat{\theta} \rightarrow^{p} 0 \).

Lemma 2 Suppose Assumption A in Appendix hold, \( \ell \rightarrow \infty \), and \( \ell = o(n) \). Then \( \theta^{**}_{NBB} - \hat{\theta} \rightarrow^{p} 0 \). If also \( \ell = o(n^{(r-2)/(2(r-1))}) \), then \( \theta^{**}_{NBB} - \hat{\theta} \rightarrow^{p} 0 \). Note that \( \ell \) must satisfy \( \ell = o(n^{1/2}) \) because \( (r-2)/2(r-1) < 1/2 \).

Lemma 3 Let Assumptions A and B in Appendix hold. If \( \ell \rightarrow \infty \), \( \ell = o(n^{1/2 - 1/r}) \), and \( W^{**}, W^{*}_{MBB,n} \rightarrow^{p} P \), then for any \( \varepsilon > 0 \), \( \Pr\{ \sup_{x \in \mathbb{R}^p} |P^x[\sqrt{n}(\theta^{**}_{MBB} - \hat{\theta})] - P[\sqrt{n}(\hat{\theta} - \theta_0)]| > \varepsilon \} \rightarrow 0 \) and \( \Pr\{ \sup_{x \in \mathbb{R}^p} |P^x[\sqrt{n}(\theta^{**}_{NBB} - \hat{\theta})] - P[\sqrt{n}(\hat{\theta} - \theta_0)]| > \varepsilon \} \rightarrow 0 \).

Lemma 4 Let Assumptions A and B in Appendix hold. If \( \ell \rightarrow \infty \), \( \ell = o(n^{(r-2)/(2(r-1))}) \), and \( W^{**}, W^{*}_{NBB,n} \rightarrow^{p} P \), then for any \( \varepsilon > 0 \), \( \Pr\{ \sup_{x \in \mathbb{R}^p} |P^x[\sqrt{n}(\theta^{**}_{NBB} - \hat{\theta})] - P[\sqrt{n}(\hat{\theta} - \theta_0)]| > \varepsilon \} \rightarrow 0 \) and \( \Pr\{ \sup_{x \in \mathbb{R}^p} |P^x[\sqrt{n}(\theta^{**}_{NBB} - \hat{\theta})] - P[\sqrt{n}(\hat{\theta} - \theta_0)]| > \varepsilon \} \rightarrow 0 \).

Lemma 5 Let Assumptions A and B in Appendix hold. Assume \( S_n \rightarrow^{p} \Sigma \). If \( \ell \rightarrow \infty \) and \( \ell = o(n^{1/2 - 1/r}) \), then the Wald statistic converges to \( \chi^2_{d} \) in distribution, \( f_n \rightarrow^{d} \chi^2_{m-p} \), and \( J^*_{MBB,n}, J^*_{NBB,n}, J^{**}_{MBB,n}, J^{**}_{NBB,n} \rightarrow^{d^*} \chi^2_{m-p} \) prob-P. Therefore, the bootstrap inference is consistent.

Lemma 6 Let Assumptions A and B in Appendix hold. Assume \( S_n \rightarrow^{p} \Sigma \). If \( \ell \rightarrow \infty \) and \( \ell = o(n^{1/2 - 1/r}) \), then \( T^o, T \rightarrow^{d} \chi^2_{m-p} \) and \( T^{o*}, T^* \rightarrow^{d^*} \chi^2_{m-p} \) prob-P.

4 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 size at the 1, 5, and 10 per cent level.

4.1 Case I: Linear models

4.1.1 Symmetric Errors

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

\[ y_t = \theta_1 + \theta_2 x_t + u_t \quad \text{for } t = 1, \ldots, T, \]  

(14)
where \((\theta_1, \theta_2) = (0, 0)\), \(u_t = \rho u_{t-1} + \varepsilon_1\), and \(x_t = \rho x_{t-1} + \varepsilon_2\). 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 \(\theta_1\) and \(\theta_2\) with instruments \(z_t = (1, x_t, x_{t-1}, x_{t-2})\). There are two overidentifying restrictions. The null hypothesis being tested is: \(H_0 : \theta_2 = 0\). 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.

DISCUSS RESULTS

4.1.2 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:

\[
y_t = \theta_1 + \theta_2 x_t + \sigma_t u_t \quad \text{for} \quad t = 1, …, T,
\]

where \((\theta_1, \theta_2) = (0, 0)\), \(x_t = 0.75 x_{t-1} + \varepsilon_1\), and \(u_t \sim N(0, \sigma_t)\). \(\sigma_t^2 = 0.0001 + 0.6 \sigma_{t-1}^2 + 0.3 \varepsilon_{t-1}^2\) and \(\varepsilon \sim N(0, I)\). The unconditional variance is 1. The instrument set is \(z_t = [1, x_t, x_{t-1}, x_{t-2}]\).

DISCUSS RESULTS

4.2 Case II: Nonlinear Models

Two experiments are consider. 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.

4.2.1 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. This section examines the finite sample properties of bootstrapped \(J\)-tests as well as a bootstrap test based on the Lagrange multiplier (\(\gamma(\theta)\)).
The moment vector is:
\[ g(X_t, \theta_1) = (X_t - \theta_1, X_t^2 - \theta_1^2 - 2\theta_1)' \]

The parameter \( \theta_1 \) is estimated using the two moments.

DISCUSS RESULTS

4.2.2 Asset Pricing Model: Environment

Finally consider an asset pricing model with the following moment conditions.\(^3\):
\[
E[\exp(\mu - \theta(x + z) + 3z) - 1] = 0, \quad Ez[\exp(\mu - \theta(x + z) + 3z) - 1] = 0.
\]

It is assumed that
\[
\log x_t = \rho \log x_{t-1} + \sqrt{(1 - \rho^2)} \varepsilon_{x_t}, \quad z_t = \rho z_{t-1} + \sqrt{(1 - \rho^2)} \varepsilon_{z_t},
\]
where \( \varepsilon_{x_t} \) and \( \varepsilon_{z_t} \) are independent normal with mean 0 and variance 0.16. In the experiment \( \rho = 0.6 \).

DISCUSS RESULTS

5 Conclusion

---

\(^3\)Derivation of the example can be found in Gregory, Lamarche, and Smith (2002).
6 Implementing the Block Bootstrap

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

1. Given the random sample \((x_1, \ldots, x_n)\), calculate \(\hat{\theta}\) using 2-stage GMM
2. For EL-MBB calculate \(\hat{\pi}_o\) using equation (10)
3a. For EL-MBB sample with replacement from \(\{\hat{\pi}_o T_j^o(\theta) : j = 1, \ldots, N\}\)
3b. For MBB uniformly sample with replacement to get \(\{X^*_1, \ldots, X^*_b\} = (\hat{x}_1, \ldots, \hat{x}_b)\)
4a. For EL-MBB calculate the J-statistic \(J_{MBB,n}^{**}\) and t-statistic \(T_{nr}^{**}\)
4b. For MBB calculate J-statistic \(J_n\) and t-statistic \(T_{nr}\)
5. Repeat steps 3-4 B times
6. Let \(\overline{q}_\alpha\) be a \((1 - \alpha)\) percentile of the distribution of \(T_{nr}^{**}\) or \(T_{nr}\)
7. Let \(\overline{q}_\alpha\) be a \((1 - \alpha)\) percentile of the distribution of \(J_{MBB,n}^{**}\) or \(J_n\)
8. The bootstrap confidence interval is \(\hat{\theta}_{nj} \pm \overline{q}_\alpha \sqrt{(\hat{\Sigma}_{jj}/N)}\)
9. For the bootstrap J-test, the test rejects if \(J \geq \overline{q}_\alpha\)
7 Mathematical Appendix

**Assumption A** Assumptions A and B are a simplified version of Assumptions A and B in Gonçalves and White (2004), tailored to our GMM estimation framework.

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
\[
X_t(\omega) = W_t(\ldots, V_{t-1}(\omega), V_t(\omega), V_{t+1}(\omega), \ldots, V_t : \Omega \to \mathbb{R}^v, v \in \mathbb{N}, \text{and } W_t : \prod_{t=\tau}^{\infty} \mathbb{R}^v \to \mathbb{R}^l \text{ is such that } X_t \text{ is measurable for all } t.
\]

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 \(\Theta\) a.s. for \(t = 1, 2, \ldots\).

A.3 (i) \(\theta_0\) is identifiably unique with respect to \(Eg(X_t, \theta)\) i.e. \(\lim_{\theta \to \theta_0} |g(X_t, \theta) - g(X_t, \theta_0)| = L_t |\theta - \theta_0|\) a.s. for all \(\theta, \theta_0 \in \Theta\), where \(\sup_{\theta} E(L_t) = O(1)\). (ii) \(\nabla g(X_t, \theta)\) is Lipschitz continuous on \(\Theta\).

A.4 For some \(r > 2\) : (i) \(\{g(X_t, \theta)\}\) is \(r\)-dominated on \(\Theta\) uniformly in \(t\), i.e. there exists \(D_t : \mathbb{R}^R \to \mathbb{R}\) such that \(|g(X_t, \theta)| \leq D_t\) for all \(\theta \in \Theta\) and \(D_t\) is measurable such that \(\|D_t\|_{\infty} \leq \Delta < \infty\) for all \(t\).

(ii) \(\nabla g(X_t, \theta)\) is \(r\)-dominated on \(\Theta\) uniformly in \(t\).

A.6 \(\{V_t\}\) is an \(\alpha\)-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 \((\Theta, \rho)\), where \(\rho\) 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 \((\Theta, \rho)\).

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 \(\Theta\) 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 \((\Theta, \rho); \{V_t\}\) is an \(\alpha\)-mixing sequence of size \(-(2+\delta)r/(r-2)\).
7.1 Proof of Lemma 1

The proof closely 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 EL–MBB, the bootstrapped objective function contains the probability weight \( \hat{\pi}_i \). \( \hat{\theta} - \theta_0 \to_p 0 \) 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 \( \theta_{MBB}^{**} \) is proved by applying Lemma A.2 of GW04. Their conditions (b1)-(b2) are satisfied by Assumptions A.2. Define \( \hat{Q}_n(\theta) = (n^{-1} \sum_{i=1}^n g(X_i^*, \theta)) W_n(n^{-1} \sum_{i=1}^n g(X_i^*, \theta)) \), then their condition (b3) holds because sup_\theta |\( \hat{Q}_{MBB,n}(\theta) - Q_n(\theta) \)| \( \to_p \) 0 from a standard argument and sup_\theta |\( \hat{Q}_n(\theta) - Q_n(\theta) \)| \( \to_p \) 0 by Lemmas A.4 and A.5 of GW04.

Deriving the asymptotics of \( \theta_{MBB} \) requires the bound of the difference between \( \pi_i \) and \( 1/N \). First we show \( \gamma'(\hat{\theta}) = O_P(\ell n^{-1/2}) \). In view of the argument in pp. 100-101 of Owen (1990) (see also Kitamura (1997)), \( \gamma'(\hat{\theta}) = O_P(\ell n^{-1/2}) \) 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) \( \max_{1 \leq i \leq N} |T_i^o(\hat{\theta})| = o_p(n^{1/2} \ell^{-1}) \). For (a), a mean value expansion gives, with \( \hat{\theta} \in [\theta_0, \hat{\theta}] \),

\[
\left| \ell N^{-1} \sum_{i=1}^N T_i^o(\hat{\theta}) T_i^o(\hat{\theta})' - \ell N^{-1} \sum_{i=1}^N T_i^o(\theta_0) T_i^o(\theta_0)' \right| \\
\leq (\hat{\theta} - \theta_0) |2 \ell N^{-1} \sum_{i=1}^N \|T_i^o(\hat{\theta})\|T_i^o(\hat{\theta})| = O_P(n^{-1/2} \ell) = o_p(1),
\]

where the second equality follows because \( |T_i^o(\theta)| \) and \( |\|T_i^o(\theta)\|| \) are \( r \)-dominated on \( \Theta \) with \( r > 2 \). Define \( G_n^o = -n^{-1} \sum_{i=1}^n g(X_i, \theta) \), then we have (c.f. Lahiri (2003, p. 48)) \( \ell N^{-1} \sum_{i=1}^N T_i^o(\theta_0) T_i^o(\theta_0)' = \text{var}^o(\sqrt{n}G_n^o) + \ell T_n T_n' \), where \( T_n = N^{-1} \sum_{i=1}^N T_i^o(\theta_0) \). \( \text{var}^o(\sqrt{n}G_n^o) - \Sigma \to_p 0 \) from Corollary 2.1 of Gonçalves and White (2002) (hereafter GW02). \( \tilde{T}_n \) is equal to \( \tilde{X}_{\ell,n} \) defined in p. 1371 of GW02 if we replace their \( X_i \) with \( g(X_i, \theta) \). GW02 p.1381 shows \( \tilde{X}_{\ell,n} = O_p(\ell^{-1}) \), and hence \( \ell \tilde{T}_n^2 = o_p(1) \). 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 \( \theta_0 \), using \( N^{-1} \sum_{i=1}^N T_i^o(\theta_0) = n^{-1} \sum_{i=1}^n g(X_i, \theta_0) + O_p(n^{-1} \ell) \) (c.f. Lemma A.1 of Fitzenberger (1997)), and applying the central limit theorem. (c) holds because \( \max_{1 \leq i \leq N} |T_i^o(\hat{\theta})| = O_{a.s.}(N^{1/r}) \) from Lemma 3.2 of Künsch (1989) and \( \ell = o(n^{1/2-1/r}) \). Therefore, we have

\[
\gamma'(\hat{\theta}) = O_P(\ell n^{-1/2}), \quad \max_{1 \leq i \leq N} |\gamma'(\hat{\theta})' T_i^o(\hat{\theta})| = o_p(1).
\]
Since \((1 + \alpha)^{-1} = 1 - (1 + \alpha)^{-2}\alpha, \alpha \in [0, \alpha]\), it follows that
\[
\hat{\pi}_i^\alpha = N^{-1}(1 + \delta_i), \quad \max_{1 \leq i \leq N} |\delta_i| = o_P(1).
\]  (18)

Consequently, \(\sup_\theta |Q_{\theta}^n(\theta) - \tilde{Q}_n(\theta)| \to_P P 0\), and the stated result follows since the conditions (b1)-(b2) of Lemma A.2 of GW04 are satisfied by Assumptions A.2. □

### 7.2 Proof of Lemma 2

In view of the proof of Lemma 1, the consistency of \(\theta_{MBB}^n\) 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 7 and 8.

Similarly, \(\theta_{MBB}^n\) is consistent if
\[
\gamma(\hat{\theta}) = O_P(\ell n^{1/2}), \quad \max_{1 \leq i \leq b} |\gamma(\hat{\theta})^i T_i(\hat{\theta})| = o_P(1). \]  (19)

(19) 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) \(\max_{1 \leq i \leq b} |T_i(\hat{\theta})| = o_P(n^{1/2} \ell^{-1})\). (a) follows from expanding \(T_i(\hat{\theta})\) around \(\theta_0\) and using Corollary 1. (b) follows from expanding \(T_i(\hat{\theta})\) around \(\theta_0\) and applying the central limit theorem. (c) follows because \(\max_{1 \leq i \leq b} |T_i(\hat{\theta})| = O_{as.}(b^{1/r})\) and \(\ell = o(n^{(r-2)/(r-1)})\). □

### 7.3 Proof of Lemma 3

The proof follows the argument in the proof of Theorem 2.2 of GW04. Define
\[
H = (G'WG)^{-1} G' \Sigma W G (G'WG)^{-1},
\]
then the stated result follows from Polya’s theorem if we show \(\sqrt{n}(\hat{\theta} - \theta_0) \to_P N(0, H)\), \(\sqrt{n}(\theta_{MBB}^n - \hat{\theta}) \to_P N(0, H)\) prob-P, and \(\sqrt{n}(\theta_{MBB}^n - \hat{\theta}) \to_{d'} N(0, H)\) prob-P.

The limiting distribution of \(\sqrt{n}(\hat{\theta} - \theta_0)\) follows from a standard argument. First, we derive the limiting distribution of \(\theta_{MBB}^n\). We need to strengthen the bound on \(\hat{\pi}_i^\alpha - 1/N\). Since \((1 + \alpha)^{-1} = 1 - \alpha + 2(1 + \alpha)^{-3} \alpha^2, \alpha \in [0, \alpha]\), it follows that
\[
\hat{\pi}_i^\alpha = N^{-1} \left(1 - \gamma(\hat{\theta})^i T_i(\hat{\theta}) + A_{ni} \gamma(\hat{\theta})^i T_i(\hat{\theta})^2\right), \quad \max_{1 \leq i \leq N} |A_{ni}| \leq 1 \text{ with prob-P approaching one.} \]  (20)  (21)

The first order condition gives:
\[
0 = \left(\sum_{i=1}^b \hat{\pi}_i^\alpha \nabla T_i^\alpha(\theta_{MBB}^n)\right)' W_{MBB,n} \left(\sum_{i=1}^b \hat{\pi}_i^\alpha T_i^\alpha(\theta_{MBB}^n)\right).
\]
Expanding $\sum_{i=1}^{b} \hat{\mathbf{r}}_{i}\mathbf{T}_{i}^{\alpha}(\theta_{MBB}^{*})$ around $\hat{\theta}$ gives, with $\hat{\theta} \in [\hat{\theta}, \theta_{MBB}^{*}]$,

$$0 = \left( \sum_{i=1}^{b} \hat{\mathbf{r}}_{i}^{\alpha}\nabla T_{i}^{\alpha}(\theta_{MBB}^{*}) \right)^{'} W_{MBB,n}^{\alpha} \left( \sum_{i=1}^{b} \hat{\mathbf{r}}_{i}^{\alpha}\nabla T_{i}^{\alpha}(\hat{\theta}) \right)$$

$$+ \left( \sum_{i=1}^{b} \hat{\mathbf{r}}_{i}\mathbf{T}_{i}^{\alpha}(\theta_{MBB}^{*}) \right)^{'} W_{MBB,n}^{\alpha} \left( \sum_{i=1}^{b} \hat{\mathbf{r}}_{i}\nabla T_{i}^{\alpha}(\hat{\theta}) \right) (\theta_{MBB}^{*} - \hat{\theta}).$$

Note that

$$b^{-1} \sum_{i=1}^{b} N\hat{\mathbf{r}}_{i}\mathbf{T}_{i}^{\alpha}(\theta_{MBB}^{*}) - G = b^{-1} \sum_{i=1}^{b} (N\hat{\mathbf{r}}_{i}^{\alpha} - 1) \mathbf{T}_{i}^{\alpha}(\theta_{MBB}^{*})$$

$$+ b^{-1} \sum_{i=1}^{b} \nabla T_{i}^{\alpha}(\theta_{MBB}^{*}) - G.$$

In view of (18) and $E(E^{*}b^{-1} \sum_{i=1}^{b} \sup_{\theta}|\mathbf{T}_{i}^{\alpha}(\theta)|) = O(1)$, the first term on the right is $o_{P_{n}}(1)$. Define $G_{n}(\theta) = n^{-1} \sum_{i=1}^{n} \nabla g(X_{i}, \theta)$. The second term on the right is $o_{P_{n}}(1)$ because $b^{-1} \sum_{i=1}^{b} \mathbf{T}_{i}^{\alpha}(\theta) - G_{n}(\theta)$ converges to 0 uniformly in prob-$P_{n}$, prob-$P$ from Lemmas A.4 and A.5 of GW04, $G_{n}(\theta)$ converges to $G(\theta) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \nabla g(X_{i}, \theta)$ uniformly, $G(\theta)$ is continuous, and $\theta_{MBB}^{*}$ is consistent. Therefore, $b^{-1} \sum_{i=1}^{b} N\hat{\mathbf{r}}_{i}\mathbf{T}_{i}^{\alpha}(\theta_{MBB}^{*})$ converges to $G$ in prob-$P_{n}$, prob-$P$. $b^{-1} \sum_{i=1}^{b} N\hat{\mathbf{r}}_{i}\nabla T_{i}^{\alpha}(\theta_{MBB}^{*})$ converges to $G$ from the same argument.

We proceed to derive the limiting distribution of $\sqrt{nb^{-1}} \sum_{i=1}^{b} N\hat{\mathbf{r}}_{i}\mathbf{T}_{i}^{\alpha}(\hat{\theta})$. Since $\sum_{i=1}^{N} \hat{\mathbf{r}}_{i}\mathbf{T}_{i}^{\alpha}(\hat{\theta}) = 0$ by the construction of $\hat{\mathbf{r}}_{i}^{\alpha}$, we can write $\sqrt{nb^{-1}} \sum_{i=1}^{b} N\hat{\mathbf{r}}_{i}\mathbf{T}_{i}^{\alpha}(\hat{\theta}) = I_{n} + II_{n}$, where

$$I_{n} = \sqrt{nb^{-1}} \sum_{i=1}^{b} T_{i}^{\alpha}(\hat{\theta}) - \sqrt{n}N^{-1} \sum_{i=1}^{N} T_{i}^{\alpha}(\hat{\theta}),$$

$$II_{n} = \sqrt{nb^{-1}} \sum_{i=1}^{b} (N\hat{\mathbf{r}}_{i}^{\alpha} - 1) T_{i}^{\alpha}(\hat{\theta}) - \sqrt{n}N^{-1} \sum_{i=1}^{N} (N\hat{\mathbf{r}}_{i}^{\alpha} - 1) T_{i}^{\alpha}(\hat{\theta}).$$

Since $N^{-1} \sum_{i=1}^{N} T_{i}^{\alpha}(\hat{\theta}) = n^{-1} \sum_{i=1}^{n} g(X_{i}, \hat{\theta}) + O_{P}(n^{-1/2} \ell)$ from Lemma A.1 of Fitzenberger (1997),

$$I_{n} = n^{-1/2} \sum_{i=1}^{n} g(X_{i}, \hat{\theta}) - n^{-1/2} \sum_{i=1}^{n} g(X_{i}, \hat{\theta}) + O_{P}(n^{-1/2} \ell) \to^{d} N(0, \Sigma) \text{ prob-P},$$

where the convergence of $n^{-1/2} \sum_{i=1}^{n} g(X_{i}, \hat{\theta})$ follows from the proof of Theorem 2.2 of GW04.
The limiting distribution of $\theta^*_\text{MBB}$ is obtained if we show $H_n = o_{P^*}(1)$. It follows from (20) that

$$H_n = H_n^1 + \sqrt{n}b^{-1} \sum_{i=1}^{b} A_{ni} \bigg| \gamma'(\hat{\theta})' T_i^{\alpha^*}(\hat{\theta}) \bigg|^2 T_i^{\alpha^*}(\hat{\theta}) - \sqrt{n}N^{-1} \sum_{i=1}^{N} A_{ni} \bigg| \gamma'(\hat{\theta})' T_i^{\alpha^*}(\hat{\theta}) \bigg|^2 T_i^{\alpha^*}(\hat{\theta}),$$

(22)

where

$$H_n^1 = -\sqrt{n}b^{-1} \sum_{i=1}^{b} T_i^{\alpha^*}(\theta_0) T_i^{\alpha^*}(\theta_0)' \gamma'(\hat{\theta}) + \sqrt{n}N^{-1} \sum_{i=1}^{N} T_i^{\alpha^*}(\theta_0)' \gamma'(\hat{\theta}).$$

Expanding $T_i^{\alpha^*}(\hat{\theta})$ and $T_i^{\alpha^*}(\hat{\theta})$ around $\theta_0$ and using (17), we obtain

$$H_n^1 = -\sqrt{n}b^{-1} \sum_{i=1}^{b} T_i^{\alpha^*}(\theta_0) T_i^{\alpha^*}(\theta_0)' \gamma'(\hat{\theta}) + \sqrt{n}N^{-1} \sum_{i=1}^{N} T_i^{\alpha^*}(\theta_0)' \gamma'(\hat{\theta}) + o_P(1)$$

$$= -b^{-1} \sum_{i=1}^{b} \{ \ell T_i^{\alpha^*}(\theta_0) T_i^{\alpha^*}(\theta_0)' - E^* [\ell T_i^{\alpha^*}(\theta_0) T_i^{\alpha^*}(\theta_0)'] \} \sqrt{n}\ell^{-1} \gamma'(\hat{\theta}) + o_P(1).$$

We assume $T_i^{\alpha^*}(\theta_0)$ is a scaler and derive the bound on $H_n^1$, because the bound for the vector-valued case follows from the elementwise bounds and the Cauchy-Schwarz inequality. Let $p = 1 + \delta/2$ with $0 < \delta \leq 2$. Then, proceeding in a similar manner as the proof of Lemma B.1 of GW04 (p.217), we obtain

$$E^* \left| b^{-1} \sum_{i=1}^{b} \left( \ell^{1/2} T_i^{\alpha^*}(\theta_0) \right)^2 - E^* (\ell^{1/2} T_i^{\alpha^*}(\theta_0))^2 \right| = b^{-p} CE^* \left| b^{-1} \sum_{i=1}^{b} \left( \ell^{1/2} T_i^{\alpha^*}(\theta_0) \right)^2 - E^* (\ell^{1/2} T_i^{\alpha^*}(\theta_0))^2 \right|^{p/2}$$

$$\leq b^{-p-1} CE^* \left| \ell^{1/2} T_i^{\alpha^*}(\theta_0) \right|^2 - E^* (\ell^{1/2} T_i^{\alpha^*}(\theta_0))^2 \right|^{p}$$

$$\leq b^{-(p-1)2} CE^* |\ell^{1/2} T_i^{\alpha^*}(\theta_0)|^{2p}.$$ 

From Lemmas A.1 and A.2 of GW02, we have, for $i = 1, \ldots, N$,

$$E|T_i^{\alpha^*}(\theta_0)|^{2p} \leq E \left( \max_{1 \leq i \leq \ell} \left| \sum_{i=1}^{i+i-1} g(X_i, \theta_0) \right|^{2p} \right) \leq C \left( \sum_{i=1}^{\ell} c_i^2 \right)^{2p/2} = O(1),$$

(24)

where $c_i$ are (uniformly bounded) mixingale constants of $\{g(X_i, \theta_0)\}$. Therefore, $E(E^*|\ell^{1/2} T_i^{\alpha^*}(\theta_0)|^{2p}) = N^{-1} \sum_{i=1}^{N} \ell^{-p} E|T_i^{\alpha^*}(\theta_0)|^{2p} = O(1)$, and (23) $= O_P(b^{-(p-1)})$ and $H_n^1 = o_{P^*}(1)$ follow.
For the other terms in (22), note that the Liapunov inequality implies \( E|T_i^\gamma(\theta_0)|^2 \leq (E|T_i^\gamma(\theta_0)|^{2p})^{1/p} = O(\ell^{-1}) \). Therefore, the third term on the right of (22) is bounded by

\[
\sqrt{n}N^{-1} \left( \max_{1 \leq j \leq N} |A_{nj}| \right) \left( \max_{1 \leq j \leq N} |\gamma'(\hat{\theta})^\gamma T_i^\gamma(\hat{\theta})| \right) |\gamma'(\hat{\theta})| \sum_{i=1}^N |T_i^\gamma(\hat{\theta})|^2 = o_P(1),
\]

and the second term on the right of (22) is also \( o_P(1) \) by a similar argument. Therefore, \( H_n = o_P(1) \) and \( \sqrt{n}(\theta_{MBB} - \hat{\theta}) \to_d N(0, H) \) prob-P follows.

For the standard bootstrap estimator \( \theta_{MBB}^* \), expanding the first order condition gives:

\[
0 = (G + o_P(1))' W_n^* n^{-1} \sum_{i=1}^n (g(X_i^*, \hat{\theta}) - g(X_i, \hat{\theta})) + (G + o_P(1))' W_n^* (G + o_P(1)) (\theta_{MBB}^* - \hat{\theta}),
\]

and the limiting distribution of \( \theta_{MBB}^* \) follows immediately. \( \square \)

### 7.4 Proof of Lemma 4

Expanding the first order condition gives, with \( \hat{\theta} \in [\hat{\theta}, \theta_{MBB}^*] \),

\[
0 = \left( \sum_{i=1}^b \hat{\pi}_i^* \nabla T_i^*(\theta_{MBB}^*) \right)' W_{n,MBB}^* \left( \sum_{i=1}^b \hat{\pi}_i^* \nabla T_i^*(\hat{\theta}) \right) + \left( \sum_{i=1}^b \hat{\pi}_i^* \nabla T_i^*(\theta_{MBB}^*) \right)' W_{n,MBB}^* \left( \sum_{i=1}^b \hat{\pi}_i^* \nabla T_i^*(\hat{\theta}) \right) (\theta_{MBB}^* - \hat{\theta}).
\]

In view of (19), the weights \( \hat{\pi}_i \) satisfy the bound (18), (20), and (21) with \( (b, \gamma(\hat{\theta}), T_i(\hat{\theta})) \) replacing \( (N, \gamma'(\hat{\theta}), T_i^\gamma(\hat{\theta})) \). Therefore, \( \sum_{i=1}^b \hat{\pi}_i^* \nabla T_i^*(\theta_{MBB}^*), \sum_{i=1}^b \hat{\pi}_i^* \nabla T_i^*(\hat{\theta}) \to_{p} G \) follows from repeating the argument of the proof of Lemma 3 using Lemmas 7 and 8 in place of Lemmas A.4 and A.5 of GW04.

We proceed to derive the limiting distribution of \( \sqrt{n} \sum_{i=1}^b \hat{\pi}_i^* T_i^*(\hat{\theta}) \). Since \( \sum_{i=1}^b \hat{\pi}_i \hat{T}_i(\hat{\theta}) = 0 \) by the construction of \( \hat{\pi}_i \), we can rewrite \( \sqrt{n} \sum_{i=1}^b \hat{\pi}_i^* T_i^*(\hat{\theta}) = \sqrt{n} \sum_{i=1}^b [\hat{\pi}_i^* T_i^*(\hat{\theta}) - \hat{\pi}_i \hat{T}_i(\hat{\theta})] \). The argument leading to (24) can be used to show \( E|T_i(\theta_0)|^{2p} = O(\ell^p) \) for \( i = 1, \ldots, b \). Then, using this bound and the bounds of \( \hat{\pi}_i - 1/b \) and proceeding as in the proof of Lemma 3, we obtain

\[
\sqrt{n} \sum_{i=1}^b \hat{\pi}_i^* T_i^*(\hat{\theta}) = \sqrt{nb}^{-1} \sum_{i=1}^b [T_i^*(\hat{\theta}) - T_i(\hat{\theta})] + o_P(1).
\]

Rewrite

\[
\sqrt{nb}^{-1} \sum_{i=1}^b [T_i^*(\hat{\theta}) - T_i(\hat{\theta})] = \zeta_{1n} + \zeta_{2n} + \zeta_{3n}, \quad \text{where} \quad \zeta_{1n} = \sqrt{n}b^{-1} \sum_{i=1}^b [T_i^*(\theta_0) - T_i(\theta_0)], \quad \zeta_{2n} = \sqrt{n}b^{-1} \sum_{i=1}^b [T_i^*(\hat{\theta}) - T_i^*(\theta_0)], \quad \text{and} \quad \zeta_{3n} = \sqrt{n}b^{-1} \sum_{i=1}^b [T_i(\theta_0) - T_i^*(\theta_0)] - E(T_i(\theta_0)),
\]

and proceed with the analysis of each term.
Observe that $\zeta_{2n} + \zeta_{3n} = b^{-1} \sum_{i=1}^{b} \nabla T_{i}^*(\hat{\theta}) - \nabla T_{i}(\bar{\theta})$ where $\zeta_{2n}, \zeta_{3n} \in [\theta_{0}, \hat{\theta}]$. Then $\zeta_{2n} + \zeta_{3n} = o_{P_{n}}(1)$ because both $\hat{\theta}$ and $\bar{\theta}$ converge to $\theta_{0}$, $b^{-1} \sum_{i=1}^{b} \nabla T_{i}^*(\bar{\theta}) - \nabla T_{i}(\theta_{0})$ and $b^{-1} \sum_{i=1}^{b} \nabla T_{i}(\theta_{0}) - G(\theta)$ converges to 0 uniformly, and $G(\theta)$ is continuous.

In view of the proof of Theorem 2.2 of GW02, $\zeta_{1n} \rightarrow_{d} N(0, \Sigma) \text{ prob-P}$ follows if, for some small $\delta > 0$,

(a) $\text{var}^{*}(\zeta_{1n}) - \Sigma \rightarrow_{p} 0, \quad \Sigma$ is positive definite,

(b) $bE^{*}|Z_{ml}|^{2+\delta} \rightarrow_{p} 0$.

where $Z_{ml} = \Sigma^{-1/2} n^{-1/2} \ell[T_{i}^*(\theta_{0}) - E^{*} T_{i}^*(\theta_{0})]$. Lemma 9 implies (a), because $\zeta_{1n} = n^{-1/2} \sum_{i=1}^{n} [g(X_{i}, \theta_{0}) - E^{*} g(X_{i}, \theta_{0})]$. For (b), first observe that $E(\ell^{1/2} T_{i}^*(\theta_{0})^{2p}) = O(1)$ because $E|\ell T_{i}(\theta_{0})|^{2p} = O(\ell^{p})$. Therefore, by setting $p = 1 + \delta/2$,

$$E(bE^{*}|Z_{ml}|^{2+\delta}) \leq C(\ell E^{*}|n^{-1/2} \ell T_{i}^*(\theta_{0})|^{2+\delta}) = O(b n^{-1-\delta/2} \ell^{1+\delta/2}) = O(b^{-\delta/2}) = o(1),$$

and $\zeta_{1n} \rightarrow_{d} N(0, \Sigma) \text{ prob-P}$ and the limiting distribution of $\theta_{NBB}^{*}$ follows.

For the standard bootstrap estimator $\theta_{NBB}^{*}$, expanding the first order condition gives:

$$0 = (G + o_{P_{n}}(1))^{\prime} W_{n}^{*} \sum_{i=1}^{b} b^{-1} (T_{i}^*(\hat{\theta}) - T_{i}(\bar{\theta})) + (G + o_{P_{n}}(1))^{\prime} W_{n}^{*} (G + o_{P_{n}}(1)) (\theta_{NBB}^{*} - \hat{\theta}),$$

and the limiting distribution of $\sqrt{n}(\theta_{NBB}^{*} - \hat{\theta})$ follows immediately. $\square$

### 7.5 Proof of Lemma 5

The validity of the bootstrap Wald test with the EL bootstrap is proven if we show $S_{MBB,n}^{*}(\theta^{*}) \rightarrow_{P_{n}} \Sigma$ and $S_{NBB,n}^{*}(\theta^{*}) \rightarrow_{P_{n}} \Sigma$ for any root-$n$ consistent $\theta^{*}$. First,

$$S_{MBB,n}^{*}(\theta^{*}) = \ell b^{-1} \sum_{i=1}^{b} (N^{T_{i}^{*}T_{i}^{*}(\theta_{0})} T_{i}^{*}(\theta_{0})) + o_{P_{n}}(1)$$

and

$$S_{NBB,n}^{*}(\theta^{*}) = \ell b^{-1} \sum_{i=1}^{b} T_{i}^{*}(\theta_{0}) T_{i}^{*}(\theta_{0}) + o_{P_{n}}(1) = \Sigma + o_{P_{n}}(1).$$
where the first equality follows from expanding $T_{i}^{*}\left(\theta^{*}\right)$ around $\theta_{0}$, the second equality follows from (18) and $E|T_{i}^{*}(\theta_{0})|^{2} = O(\ell^{-1})$, and the third equality follows from the proof of Theorem 3.1 of GW04. Similarly, we obtain

$$S_{NBB,n}^{*}(\theta) = \ell b^{-1} \sum_{i=1}^{b} T_{i}^{*}(\theta) T_{i}^{*}(\theta)^{'} + o_{p\cdot}p(1)$$

$$= \ell b^{-1} \sum_{i=1}^{b} T_{i}(\theta) T_{i}(\theta)^{'} + o_{p\cdot}p(1) = \Sigma + o_{p\cdot}p(1),$$

where the second equality follows because the argument following (23) is valid even if we replace $T_{i}^{*}(\theta_{0})$ in (23) with $T_{i}^{*}(\theta_{0})$, and the third equality follows Corollary 1. The proof for the standard MBB and NBB bootstrap is very similar and omitted.

$$J_{n} \rightarrow_{d} \chi_{m-p}^{2} \text{ if } W_{n} \rightarrow_{d} \Sigma^{-1} \text{ and } n^{-1/2} \sum_{i=1}^{n} g(X_{i}, \theta) \rightarrow_{d} N(0, \Sigma), \text{ which follows from Assumptions A and B and a standard argument. } J_{MBB,n}^{*} \rightarrow_{d} \chi_{m-p}^{2} \text{ prob-P because } S_{MBB,n}^{*}(\theta)_{MBB} \rightarrow_{p\cdot}p \Sigma \text{ and } \sqrt{n} b^{-1} \sum_{i=1}^{b} \Lambda_{i}^{*} \rightarrow_{d} \chi_{m-p}^{2} \text{ prob-P. } J_{MBB,n}^{**} \rightarrow_{d} \chi_{m-p}^{2} \text{ prob-P follows because } S_{n}^{*}(\theta)_{MBB}^{*} \rightarrow_{p\cdot}p \Sigma \text{ and we have shown in the proof of Lemma 3 that } n^{-1/2} \sum_{i=1}^{n} g(X_{i}, \hat{\theta}) = n^{-1/2} \sum_{i=1}^{n} g(X_{i}, \hat{\theta}) - n^{-1/2} \sum_{i=1}^{n} g(X_{i}, \hat{\theta}) \rightarrow_{d} N(0, \Sigma) \text{ prob-P. The convergence of } J_{NBB,n}^{*} \text{ and } J_{NBB,n}^{**} \text{ are proven by a similar argument.}$$

**Proof of Lemma 6**

We prove the limiting distribution of $T^{*}$ and $T^{**}$ first. The first order condition from equation 9 is:

$$\sum_{i=1}^{N} T_{i}^{*}(\hat{\theta}) (1 + \gamma(\hat{\theta})^{'} T_{i}^{*}(\hat{\theta}))^{-1} = 0.$$  

Using $(1 + \alpha)^{-1} = 1 - \alpha + 2(1 + \alpha)^{-3} \alpha^{2}$, $\bar{\alpha} \in [0, \alpha]$ and (17), we obtain

$$0 = \sum_{i=1}^{N} T_{i}^{*}(\hat{\theta})' - \gamma'(\hat{\theta})^{'} \sum_{i=1}^{N} T_{i}^{*}(\hat{\theta}) T_{i}^{*}(\hat{\theta})' + \sum_{i=1}^{N} A_{i} [\gamma'(\hat{\theta})^{'} T_{i}^{*}(\hat{\theta})]^{2} T_{i}^{*}(\hat{\theta})',$$

where $\max_{1 \leq i \leq N} |A_{i}| \leq 1$ with prob-P approaching one. The third term on the right of (26) is equal to $N n^{-1/2}$ times the third term on the right of (22) in the proof of Lemma 3, hence it follows from (25) that it is $o_{p}(N n^{-1/2}) = o_{p}(N^{1/2})$. Therefore,

$$\ell^{-1} N^{1/2} \gamma'(\hat{\theta}) = \left[ \ell N \sum_{i=1}^{N} T_{i}^{*}(\hat{\theta}) T_{i}^{*}(\hat{\theta})' \right]^{-1} \left[ \frac{1}{N^{1/2}} \sum_{i=1}^{N} T_{i}^{*}(\hat{\theta}) + o_{p}(1) \right].$$
Define \( V_i = \Sigma^{-1} (I - G(G'\Sigma^{-1}G)^{-1}G') \Sigma^{-1} \) = \( \Sigma^{-1/2}(I - \Sigma^{-1/2}G(G'\Sigma^{-1}G)^{-1}G'\Sigma^{-1/2})\Sigma^{-1/2} \). Then, \( \ell^{-1}N^{1/2}\gamma(\hat{\theta}) \to d \ N(0, V_1) \) if (i) \( \ell N^{-1/2} \sum_{i=1}^N T_i'(\hat{\theta})T_i(\hat{\theta})' \to d. \Sigma \), and (ii) \( N^{-1/2} \sum_{i=1}^N T_i'(\hat{\theta}) \to d N(0, \Sigma V_1 \Sigma) \). (i) is shown in the proof of Lemma 1. For (ii), \( N^{-1} \sum_{i=1}^n g(\hat{X}_i, \hat{\theta}) + O_p(n^{-1} \ell) \) from Lemma A.1 of Fitzenberger (1997), and a standard derivation (e.g., Newey and McFadden (1994) p. 2231) gives

\[
n^{-1/2} \sum_{i=1}^n g(\hat{X}_i, \hat{\theta}) = (I - G(G'\Sigma^{-1}G)^{-1}G')\Sigma^{-1/2} \sum_{i=1}^n g(\hat{X}_i, \theta_0) + o_p(1)
= \Sigma V_1 n^{-1/2} \sum_{i=1}^n g(\hat{X}_i, \theta_0) + o_p(1) \to d N(0, \Sigma V_1 \Sigma).\]

Since \( \Sigma \) is a generalized inverse of \( V_1 \), i.e., \( V_1 \Sigma V_1 = V_1 \) from \( \Sigma = \Sigma^{-1/2} \text{proj}((\Sigma^{-1/2}G)\Sigma^{-1/2}) \), the limiting variance is \( \Sigma V_1 \Sigma \), and (ii) follows. Finally, \( T^o \to d \chi^2_{m-p} \) because \( \Sigma_n(\hat{\theta}) \to d \Sigma, \text{rank}(V_1) = m - p \), and \( \Sigma \) is a generalized inverse of \( V_1 \) (Theorem 9.2.2, Rao and Mitra (1971)).

For \( \gamma^o(\theta_{MBB}) \), the first order condition gives an equation that corresponds to (26):

\[
\sum_{i=1}^b \frac{N\kappa_i^o T_i'(\theta_{MBB})}{1 + \gamma^o(\theta_{MBB}) N\kappa_i^o T_i'(\theta_{MBB})} = 0.
\]

Repeating the above argument with replacing \( T_i'(\hat{\theta}) \) with \( N\kappa_i^o T_i'(\theta_{MBB}) \) gives

\[
\ell^{-1}n^{1/2}\gamma^o(\theta_{MBB}) = b \left[ \ell b^{-1} \sum_{i=1}^b N\kappa_i^o T_i'(\theta_{MBB}) \kappa_i^o T_i'(\theta_{MBB}) \right]^{-1} \left[ \sqrt{b} \left( \ell b^{-1} \sum_{i=1}^b N\kappa_i^o T_i'(\theta_{MBB}) + o_p(1) \right) \right].
\]

The first term is \( S_{MBB,\theta}^o(\theta_{MBB}) = \Sigma + o_{p, \ell}(1) \). Expanding the second term around \( \hat{\theta} \) and applying the arguments in the proof of Lemma 3 gives

\[
\sqrt{b} \left( \ell b^{-1} \sum_{i=1}^b N\kappa_i^o T_i'(\theta_{MBB}) \right) = (I - G(G'\Sigma^{-1}G)^{-1}G') \sqrt{b} \left( \ell b^{-1} \sum_{i=1}^b N\kappa_i^o T_i'(\hat{\theta}) + o_p(1) \right).
\]

The required follows from repeating the argument for \( T^o \).

For the nonoverlapping block version, we only provide the sketch of the proof. The first order condition is \( \sum_{i=1}^b T_i'(\hat{\theta})[(1 + \gamma(\hat{\theta})^o T_i(\hat{\theta}))^{-1} = 0. \) Expanding the denominator around 1 and rearranging, we obtain

\[
bn^{-1/2} \gamma(\hat{\theta}) = \left[ \ell b^{-1} \sum_{i=1}^b T_i'(\hat{\theta}) T_i(\hat{\theta})' \right]^{-1} \left[ n^{-1/2} \ell b^{-1} \sum_{i=1}^b T_i(\hat{\theta}) + o_p(1) \right].
\]

23
\( \ell b^{-1} \sum_{i=1}^{b} T_i(\hat{\theta}) T_i(\hat{\theta})' \rightarrow_p \Sigma \) is shown in the proof of Lemma 2. From the definition of \( T_i(\hat{\theta}) = n^{-1/2} \ell \sum_{i=1}^{b} T_i(\hat{\theta}) = n^{-1/2} \sum_{i=1}^{n} g(X_i, \hat{\theta}) \), which converges to \( N(0, \Sigma V \Sigma) \) in distribution. Therefore, \( b^2 n^{-1} \gamma(\hat{\theta})' S_n(\hat{\theta}) \gamma(\hat{\theta}) \rightarrow_d \chi_m^2 \).

For \( \gamma'(\theta_{NBB}) \), the first order condition is

\[
\sum_{i=1}^{b} \frac{b \hat{\gamma}_i^* T_i(\theta_{NBB})}{1 + \gamma'(\theta_{NBB}) b \hat{\gamma}_i^* T_i(\theta_{NBB})} = 0,
\]

and proceeding as above gives (note that \( \ell b = n \))

\[
b n^{-1/2} \gamma'(\theta_{NBB}) = \left[ \ell b^{-1} \sum_{i=1}^{b} b \hat{\gamma}_i^* T_i(\theta_{NBB}) b \hat{\gamma}_i^* T_i(\theta_{NBB}) \right]^{-1} \left[ \sqrt{n} b^{-1} \sum_{i=1}^{b} b \hat{\gamma}_i^* T_i(\theta_{NBB}) + op_p(1) \right]
\]

The first term is \( S_{NBB,n}(\theta_{NBB})^{-1} = \Sigma^{-1} + op_p(1) \). Expanding the second term around \( \hat{\theta} \) and applying the arguments in the proof of Lemma 4 gives

\[
\sqrt{n} \sum_{i=1}^{b} \hat{\gamma}_i^* T_i(\hat{\theta}) = (I - G(G' \Sigma^{-1} G')^{-1} G' \Sigma^{-1}) \sqrt{n} \sum_{i=1}^{b} \hat{\gamma}_i^* T_i(\hat{\theta}) + op_p(1).
\]

\( \sqrt{n} \sum_{i=1}^{b} \hat{\gamma}_i^* T_i(\hat{\theta}) \rightarrow_d N(0, \Sigma) \) prob-P is shown in the proof of Lemma 4, and the stated result follows. \( \square \)

\section{Auxiliary results}

\textbf{Lemma 7 (NBB uniform WLLN).} Let \( \{q_m^*(\cdot, \omega, \theta)\} \) be an NBB resample of \( \{q_m(\omega, \theta)\} \) and assume:
(a) For each \( \theta \in \Theta \subset \mathbb{R}^p \), \( \Theta \) a compact set, \( n \sum_{n=1}^{n} (q_m^*(\cdot, \omega, \theta) - q_m(\omega, \theta)) \rightarrow 0 \), prob-\( P_{n,\omega} \), prob-\( P \); and
(b) \( \forall \theta, \theta_0 \in \Theta, |q_m^*(\cdot, \theta) - q_m^*(\cdot, \theta_0)| \leq L_{m|\theta - \theta_0| a.s.-P \text{; and}} \), where \( \sup_n \{n^{-1} \sum_{i=1}^{n} E(L_{m|i})\} = O(1) \). Then, if \( \ell = o(n) \), for any \( \delta > 0 \) and \( \xi > 0 \),

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

\textbf{Proof} \hspace{1cm} The proof closely follows that of Lemma 8 of Hall and Horowitz (1996). \( \square \)

\textbf{Lemma 8 (NBB pointwise WLLN).} For some \( r > 2 \), let \( \{q_{m,t} : \Omega \times \Theta \rightarrow \mathbb{R}^m : m \in \mathbb{N} \} \) be such that for all \( n, t \), there exists \( D_{m,t} : \Omega \rightarrow \mathbb{R} \) with \( |q_{m,t}(\cdot, \theta)| \leq D_{m,t} \) for all \( \theta \in \Theta \) and \( \|D_{m,t}\|_r \leq \Delta < \infty \). For each \( \theta \in \Theta \)
let \( \{ q_m(\cdot, \omega, \theta) \} \) be an NBB resample of \( \{ q_m(\omega, \theta) \} \). If \( \ell = o(n) \), then for any \( \delta > 0, \xi > 0 \) and for each \( \theta \in \Theta \),

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

**Proof** Fix \( \theta \in \Theta \), and we suppress \( \theta \) and \( \omega \) henceforth. Since \( q_m^* \) is a NBB resample, \( E^* q_m^* = n^{-1} \sum_{i=1}^{n} q_m = \bar{q}_n \) and hence \( \sum_{i=1}^{n} (q_m^* - q_m) = \sum_{i=1}^{n} (q_m^* - E^* q_m) \). From the arguments in the proof of Lemma A.5 of GW04, the stated result follows if \( \| \text{var}^* (n^{-1/2} \sum_{i=1}^{n} q_m^*) \|_r/2 = O(\ell) \) for some \( r > 2 \).

Define \( U_m = \ell^{-1} \sum_{i=1}^{\ell} q_{ni,(i-1)\ell+t} \), the average of the \( i \)th non-overlapping block. Since the blocks are independently sampled, we have (c.f. Lahiri (2003), p.48)

\[
\text{var}^* \left( n^{-1/2} \sum_{i=1}^{n} q_m^* \right) = b^{-1} \ell \sum_{i=1}^{b} (U_m - \bar{q}_n) (U_m - \bar{q}_n)',
\]

\[
= b^{-1} \ell \sum_{i=1}^{b} \left[ \ell \sum_{t=1}^{\ell} (q_{ni,(i-1)\ell+t} - \bar{q}_n) (q_{ni,(i-1)\ell+t} - \bar{q}_n)' \right]
\]

\[
= R_n(0) + b^{-1} \sum_{i=1}^{b} \sum_{\tau=1}^{\ell-1} \left( R_m(\tau) + R_m'^{\tau} \right),
\]

where

\[
R_n(0) = n^{-1} \sum_{i=1}^{n} (q_m - \bar{q}_n) (q_m - \bar{q}_n)',
\]

\[
R_m(\tau) = \ell^{-1} \sum_{i=1}^{\ell-\tau} (q_{ni,(i-1)\ell+t} - \bar{q}_n) (q_{ni,(i-1)\ell+t+\tau} - \bar{q}_n)', \quad \tau = 1, \ldots, \ell - 1.
\]

Applying Minkowski and Cauchy-Schwartz inequalities gives \( \| R_n(\tau) \|_r/2 = O(1), \tau = 0, \ldots, \ell - 1, \) and \( \| \text{var}^* (n^{-1/2} \sum_{i=1}^{n} q_m^*) \|_r/2 = O(\ell) \) follows. □

**Lemma 9** (Consistency of NBB conditional variance). Assume \( \{ X_t \} \) satisfies \( E X_t = 0 \) for all \( t \), \( \| X_t \|_r \leq \Delta < \infty \) 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 \( \alpha \)-mixing sequence of size \( -(2r/(r-2)) \). Let \( \{ X_t^r \} \) be an NBB resample of \( \{ X_t \} \). Define \( \tilde{X}_n = n^{-1} \sum_{t=1}^{n} X_t, \tilde{X}_n^r = n^{-1} \sum_{t=1}^{n} X_t^r, \Sigma_n = \text{var}(\sqrt{n}\tilde{X}_n), \) and \( \Sigma_n^r = \text{var}^*(\sqrt{n}\tilde{X}_n^r) \). Then, if \( \ell \to \infty \) and \( \ell = o(n^{1/2}) \), \( \Sigma_n - \Sigma_n^r \to_p 0 \).

**Corollary 1** Assume \( X_t \) satisfies the assumptions of Lemma 9. Define \( U_i = \ell^{-1} \sum_{t=1}^{\ell} X_{ni,(i-1)\ell+t} \), the average of the \( i \)th non-overlapping block. Then, if \( \ell \to \infty \) and \( \ell = o(n^{1/2}) \),

\[
b^{-1} \ell \sum_{i=1}^{b} U_i^r - \Sigma_n \to_p 0.
\]
Proof. For simplicity, we assume $X_t$ to be a scalar. The extension to the vector-valued $X_t$ is straightforward, see GW02. Define $U_t = \ell^{-1} \sum_{t=1}^{\ell} X_{(i-1) \ell + t}$, the average of the $i$th block. Since the blocks are independently sampled, we have

$$
\hat{\Sigma}_n = b^{-1} \ell \sum_{i=1}^{b} U_i^2 - \ell \bar{X}_n^2
$$

$$
= b^{-1} \ell^{-1} \sum_{i=1}^{b} \left[ \sum_{t=1}^{\ell} X_{(i-1) \ell + t} \sum_{s=1}^{\ell} X_{(i-1) \ell + s} \right] - \ell \bar{X}_n^2
$$

$$
= b^{-1} \sum_{i=1}^{b} \hat{R}_i(0) + 2b^{-1} \sum_{i=1}^{b} \sum_{\tau=1}^{\ell - 1} \hat{R}_i(\tau) - \ell \bar{X}_n^2.
$$

where $\hat{R}_i(\tau) = \ell^{-1} \sum_{t=1}^{\ell} X_{(i-1) \ell + t} X_{(i-1) \ell + \tau}$, $\tau = 0, \ldots, \ell - 1$. First we show $E(\hat{\Sigma}_n) - \Sigma_n = o(1)$. From Lemmas A.1 and A.2 of GW02, we have, for $i = 1, \ldots, b$,

$$
E(\hat{X}_n^2) = n^{-2} E \left| \sum_{j=1}^{n} X_j \right|^2 \leq n^{-2} E \left( \max_{1 \leq j \leq n} \left| \sum_{i=1}^{j} X_i \right|^2 \right) \leq C n^{-2} \left( \sum_{i=1}^{b} \tau_i^2 \right) = O(n^{-1}),
$$

where $\tau_i$ are (uniformly bounded) mixingale constants of $X_t$, and $E|\ell \bar{X}_n^2| = o(1)$ follows. Define $R_i(\tau) = \ell^{-1} \sum_{t=1}^{\ell} E(X_{(i-1) \ell + t} X_{(i-1) \ell + \tau})$ and $R_{ij} = \ell^{-1} \sum_{t=1}^{\ell} E(X_{(i-1) \ell + t} X_{(j-1) \ell + t})$ so that $E(\hat{R}_i(\tau)) = R_i(\tau)$, then

$$
\Sigma_n = b^{-1} \sum_{i=1}^{b} R_i(0) + 2b^{-1} \sum_{i=1}^{b} \sum_{\tau=1}^{\ell - 1} R_i(\tau) + b^{-1} \sum_{i=1}^{b} \sum_{j \neq i} R_{ij},
$$

and $E(\hat{\Sigma}_n) - \Sigma_n = b^{-1} \sum_{i=1}^{b} \sum_{j \neq i} R_{ij}$. From Gallant and White (1988) (pp.109-110), $E(X_t X_{t+\tau})$ is bounded by

$$
|EX_t X_{t+\tau}| \leq \Delta (5 \alpha_{\sqrt{\tau}/4}^{-1/2} + 2 v_{\sqrt{\tau}/4}) \leq C \tau^{-1 - \xi},
$$

for some $\xi \in (0, 1)$, where $v_m$ is the NED coefficient. Therefore, for $|i - j| = k \geq 2$, we have $|R_{ij}| \leq C \ell^{-1} \sum_{i=1}^{\ell} \sum_{\tau=1}^{\ell} ((k-1) \ell)^{-1 - \xi} = O((k-1)^{-1} \ell^{-1 - \xi})$, and

$$
|R_{i,i+1}| \leq C \ell^{-1} \sum_{\ell=1}^{\ell} \sum_{s=1}^{\ell} |s - i|^{-1 - \xi}
$$

$$
\leq C \ell^{-1} \sum_{h=0}^{\ell+1} (\ell - |h|) |h + h|^{-1 - \xi} = O(\ell^{-\xi}),
$$

where the last equality follows from evaluating the sums with $h > 0$ and $h < 0$ separately. It follows that

$$
b^{-1} \sum_{i=1}^{b} \sum_{j \neq i} R_{ij} = O \left( \ell^{-\xi} + b^{-1} \sum_{k=2}^{b-1} (b - k) (k-1)^{-1} \ell^{-\xi} \right) = O \left( \ell^{-\xi} \right),
$$

26
and we establish $E(\hat{\Sigma}_n) - \Sigma_n = o(1)$. It remains to show \( \text{var}(\text{var}(n^{1/2}X^*_n)) = o(1) \). It suffices to show that the variance of

\[
\begin{align*}
 b^{-1} \sum_{i=1}^{b} (\hat{R}_i(0) - R_i(0)) + 2b^{-1} \sum_{\tau=1}^{b-1} \sum_{i=1}^{\ell-1} (\hat{R}_i(\tau) - R_i(\tau))
\end{align*}
\]

is $o(1)$. Following the derivation in GW02 leading to their equation (A.4), we obtain

\[
\begin{align*}
\text{var}(\hat{R}_i(\tau)) &\leq \ell^{-2} \sum_{\tau=1}^{\ell-1} \text{var}(X_{(\ell-1)\ell+i}X_{(i-1)\ell+i+\tau}) \\
&\quad + 2\ell^{-2} \sum_{\tau=1}^{\ell-1} \sum_{i=1}^{\ell-1} \text{cov}(X_{(\ell-1)\ell+i}X_{(i-1)\ell+i+\tau}, X_{(\ell-1)\ell+s}X_{(i-1)\ell+s+\tau}) \\
&\leq C\ell^{-1} \left\{ \Delta + \sum_{k=1}^{\infty} \alpha_{[k/4]}^{1/2-1/r} + \sum_{k=1}^{\infty} v_{[k/4]} + \sum_{k=1}^{\infty} v_{[k/4]}^{(r-2)/2(r-1)} \right\} \\\n&\quad + C\ell^{-1} \left( \tau \alpha_{[\tau/4]}^{1/2-1/r} + \tau v_{[\tau/4]}^{2} + 2\tau \alpha_{[\tau/4]}^{1/2-1/r} v_{[\tau/4]} \right) = O(\ell^{-1}).
\end{align*}
\]

Observe that, when $|i-j| \geq 7$, from Lemma 6.7(a) of Gallant and White (1988) we have, for some $\xi \in (0, 1)$,

\[
\begin{align*}
\text{cov}(\hat{R}_i(\tau), \hat{R}_j(\tau)) &\leq \ell^{-2} \sum_{\tau=1}^{\ell-1} \sum_{i=1}^{\ell-1} \text{cov}(X_{(\ell-1)\ell+i}X_{(i-1)\ell+i+\tau}X_{(\ell-1)\ell+s}X_{(i-1)\ell+s+\tau}) \\
&\leq \ell^{-2} \sum_{\tau=1}^{\ell-1} \sum_{i=1}^{\ell-1} \left( \alpha_{[\ell(i-j)-6]\ell/4]}^{1/2-1/r} + v_{[\ell(i-j)-6]\ell/4]}^{(r-2)/2(r-1)} \right) \\
&= O \left( \ell^{-2} \sum_{\tau=1}^{\ell-1} \sum_{i=1}^{\ell-1} \left[ (\ell-6)\ell/4]^{-1-\xi} \right) \leq C(\ell|i-j|)^{-1-\xi}
\end{align*}
\]

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 (29) 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 $\tau = 0, \ldots, \ell - 1$,

\[
\begin{align*}
\text{var} \left( b^{-1} \sum_{i \in B_r} (\hat{R}_i(\tau) - R_i(\tau)) \right) &= b^{-2} \sum_{i \in B_r} \sum_{j \in B_r} \text{cov}(\hat{R}_i(\tau), \hat{R}_j(\tau)) \\
&= O \left( b^{-1} \ell^{-1} + \ell^{-1-\xi} \sum_{i=1}^{b} \sum_{j \neq i} |i-j|^{-1-\xi} \right) \\
&= O \left( b^{-1} \ell^{-1} + \ell^{-1-\xi} b^{-2} \sum_{h=1}^{b-1} (b-h)h^{-1-\xi} \right) \\
&= O \left( b^{-1} \ell^{-1} \right).
\end{align*}
\]
Therefore, the variance of (29) is $O(\ell b^{-1}) = O(\ell^2 n^{-1}) = o(1)$, giving the stated result. Corollary 1 follows because $b^{-1} \sum_{i=1}^{b} U_i U_i' = \hat{\Sigma}_n + o_P(1)$ from (27). \[\Box\]
References


