Bootstrapping M-Estimators with Multi-Dimensional Heterogeneity

Xueyuan Liu

February 23, 2022

Abstract

This paper proves the validity of pigeonhole bootstrap for M-estimators with multi-dimensional heterogeneity. And I conjecture that there exists a uniform bootstrap (block bootstrap) method that works not only for the nondegenerate case, i.e., multi-dimensional heterogeneity exists, but also for the degenerate case when the data is i.i.d.

1 Introduction

Empirical data may exhibit cluster dependence in multiple dimensions because observations are heterogeneous in multiple dimensions. In empirical work in economics, it is common to account for the clustering of units. This is because unobserved components in outcomes for units within clusters are correlated. However, the correlation may occur across more than one dimension, such as geographic, age cohorts, gender, etc. Moreover, in an e-commerce dataset, consumers are different in many dimensions. Suppose that we have a large dataset containing variables like cookies, customer IDs, query strings, IP addresses, product IDs (e.g., SKUs), URLs, etc. Another example is dyadic data, which means outcomes yin reflect pairwise interaction among sampled units, which frequently arise in social science research. Such data play central roles in contemporary empirical trade and international relations analysis. They also feature in work on international financial flows development economics, among other fields.

This paper proves the validity of a bootstrap procedure for M-estimators with multi-dimensional heterogeneity for the nondegenerate case is proved. For example, in the bipartite network, two-dimensional heterogeneity exists. In the literature, Menzel (2021) proposes wild bootstrap procedures for linear parameters, e.g., sample mean. MacKinnon et al.(2021) propose wild bootstrap procedures for regression models with clustering in two dimensions. But bootstrapping nonlinear estimators with multi-dimensional heterogeneity has not been studied in the literature.

There are some well-known advantages of bootstrapping. The bootstrap is convenient from a programming point of view because it relies on the same estimation procedure that delivers the point estimates. For instance, sometimes, we need to make choices regarding tuning parameters such as bandwidths or the number of nearest neighbors. The bootstrap avoids this. Moreover, it is useful when the analytical expression does not exist for complicated models. Likewise, estimation of the asymptotic variance of two-step estimators requires the calculation of the derivative of the estimating equation in the second step with respect to the first-step parameters. The bootstrap can also avoid this calculation. Bootstrap would be more helpful in two-step estimation or more complicated structural models, such as the game-theoretical network formation model.

I conjecture that a uniform bootstrap procedure called block bootstrap may exist (see section 4).

2 An example

Consider the following two-way interaction model, one leading example is a gravity equation for bilateral trade flows between countries; they feature both importer and exporter effects:

$$Y_{ij} = \exp\left(R_{ij}'\theta_0\right) U_i U_j V_{ij} \tag{1}$$

where Y_{ij} is outcome variable; $R_{ij} \overset{\text{def}}{\equiv} w\left(X_i, X_j\right)$ is a vector of constructed dyad-specific attributes; U_i, U_j, and V_{ij} are mean one random variables. Following Graham (2020), l_{ij}

denotes the corresponding log-likelihood.

$$l_{ij}(\theta) = Y_{ij} R_{ij}' \theta - \exp\left(R_{ij}'\theta\right) \tag{2}$$

which equals the log likelihood of a Poisson random variable Y_{ij} with mean $\exp\left(R_{ij}'\theta\right)$, and choose $\hat{\theta}$ to maximize

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} l_{ij}(\theta) \tag{3}$$

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) = \left[-H_N(\bar{\theta})\right]^{+} \sqrt{N} S_N\left(\theta_0\right) \tag{4}$$

where $\bar{\theta}$ is a mean value between $\hat{\theta}$ and θ_0 which may vary from row to row, the + superscript denoting a Moore-Penrose inverse, and a "score" vector of

$$S_N(\theta) = \frac{1}{N} \frac{1}{N-1} \sum_{i} \sum_{j \neq i} s_{ij}\left(Z_{ij}, \theta\right) \tag{5}$$

with $s\left(Z_{ij}, \theta\right) = \partial l_{ij}(\theta) / \partial \theta$ for $Z_{ij} = \left(Y_{ij}, R_{ij}'\right)'$ and $H_N(\theta) = \frac{1}{N} \frac{1}{N-1} \sum_i \sum_{j \neq i} \frac{\partial^2 l_{ij}(\theta)}{\partial \theta \partial \theta'}$, $H_N(\bar{\theta}) \xrightarrow{p} \Gamma_0$, then we have

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) = -\Gamma_0^{-1} \sqrt{N} S_N\left(\theta_0\right) + o_p(1) \tag{6}$$

we rewrite $S_N\left(\theta_0\right)$ as

$$S_N = \begin{pmatrix} N \\ 2 \end{pmatrix}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{s_{ij} + s_{ji}}{2} \tag{7}$$

where $s_{ij} \overset{\text{def}}{\equiv} s\left(Z_{ij}, \theta_0\right)$ and $S_N \overset{\text{def}}{\equiv} S_N\left(\theta_0\right)$. Let $\mathbf{U} = \left[U_i\right]_{1 \leq i \leq N}$, $\mathbf{X} \overset{\text{def}}{\equiv} \left[X_i\right]_{1 < i < N}$, $\mathbf{Y} \overset{\text{def}}{\equiv} \left[Y_{ij}\right]_{1 \leq i, j \leq N, i \neq j}$,

$$V_N \overset{\text{def}}{\equiv} \mathbb{E}\left[S_N \mid \mathbf{X}, \mathbf{U}\right] = \begin{pmatrix} N \\ 2 \end{pmatrix}^{-1} \sum_{i<j} \frac{\bar{s}_{ij} + \bar{s}_{ji}}{2} \tag{8}$$

where $\bar{s}_{ij} \overset{\text{def}}{\equiv} \bar{s}\left(X_i, U_i, X_j, U_j\right)$ and $\bar{s}\left(X_i, U_i, X_j, U_j\right) \overset{\text{def}}{\equiv} \mathbb{E}\left[s\left(Z_{ij}, \theta_0\right) \mid X_i, U_i, X_j, U_j\right]$, the projection error is

$$T_N = S_N - V_N \tag{9}$$

which consists of a summation of $\begin{pmatrix} N \\ 2 \end{pmatrix}$ conditionally uncorrelated summands; hence

$$\mathbb{V}\left(T_N\right) = \begin{pmatrix} N \\ 2 \end{pmatrix}^{-1} \mathbb{E}\left(\mathbb{V}\left(\frac{s_{12} + s_{21}}{2} \mid X_1, U_1, X_2, U_2\right)\right) = O\left(N^{-2}\right) \tag{10}$$

We define $\bar{s}^e(x,u) = \mathbb{E}\left[\bar{s}\left(x,u,X_1,U_1\right)\right]$ and $\bar{s}^a(x,u) = \mathbb{E}\left[\bar{s}\left(X_1,U_1,x,u\right)\right]$, then decompose V_N into two terms:

$$V_N = V_{1N} + V_{2N} \tag{11}$$

where

$$V_{1N} = \frac{2}{N}\sum_{i=1}^{N}\left\{\frac{\bar{s}_1^e\left(X_i,U_i\right)+\bar{s}_1^a\left(X_i,U_i\right)}{2}\right\} \tag{12}$$

and that

$$V_{2N} = \begin{pmatrix} N \\ 2 \end{pmatrix}^{-1}\sum_{i<j}\left\{\frac{\bar{s}_{ij}+\bar{s}_{ji}}{2} - \frac{\bar{s}_1^e\left(X_i,U_i\right)+\bar{s}_1^a\left(X_i,U_i\right)}{2} - \frac{\bar{s}_1^e\left(X_j,U_j\right)+\bar{s}_1^a\left(X_j,U_j\right)}{2}\right\} \tag{13}$$

We think about the variance of S_N in terms of the ANOVA decomposition :

$$\begin{aligned}
\mathbb{V}\left(S_N\right) &= \mathbb{V}\left(\mathbb{E}\left[S_N \mid \mathbf{X},\mathbf{U}\right]\right) + \mathbb{E}\left[\mathbb{V}\left(S_N \mid \mathbf{X},\mathbf{U}\right)\right] \\
&= \mathbb{V}\left(V_N\right) + \mathbb{V}\left(T_N\right) \\
&= \mathbb{V}\left(V_{1N}\right) + \mathbb{V}\left(V_{2N}\right) + \mathbb{V}\left(T_N\right)
\end{aligned} \tag{14}$$

Let

$$\Sigma_1 = \mathbb{C}\left(\mathbb{E}\left[\frac{s_{12}+s_{21}}{2} \mid X_1,U_1,X_2,U_2\right], \mathbb{E}\left[\frac{s_{13}+s_{31}}{2} \mid X_1,U_1,X_3,U_3\right]'\right) \tag{15}$$

and that

$$\begin{aligned}
\Sigma_2 &= \mathbb{C}\left(\mathbb{E}\left[\frac{s_{12}+s_{21}}{2} \mid X_1,U_1,X_2,U_2\right], \mathbb{E}\left[\frac{s_{12}+s_{21}}{2} \mid X_1,U_1,X_2,U_2\right]'\right) \\
&= \mathbb{V}\left(\mathbb{E}\left[\frac{s_{12}+s_{21}}{2} \mid X_1,U_1,X_2,U_2\right]\right)
\end{aligned} \tag{16}$$

$$\Sigma_3 \overset{\text{def}}{\equiv} \mathbb{E}\left[\mathbb{V}\left(\frac{s_{12}+s_{21}}{2} \mid X_1,U_1,X_2,U_2\right)\right] \tag{17}$$

Calculations analogous to those use in variance analyses for U-statistics variance analyses for U-statistics yield

$$\begin{aligned}
\mathbb{V}\left(V_{1N}\right) &= \frac{4\Sigma_1}{N} \\
\mathbb{V}\left(V_{2N}\right) &= \frac{2}{N(N-1)}\left(\Sigma_2 - 2\Sigma_1\right) \\
\mathbb{V}\left(T_N\right) &= \frac{2}{N(N-1)}\Sigma_3
\end{aligned} \tag{18}$$

such that, defining the notation $\Omega \overset{\text{def}}{\equiv} \mathbb{V}\left(\sqrt{N} S_N\right)$,

$$\Omega = 4\Sigma_1 + \frac{2}{N-1}\left(\Sigma_2 + \Sigma_3 - 2\Sigma_1\right) \tag{19}$$

the variances of V_{2N} and T_N are of smaller order. Therefore,

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) = \left[-H_N(\bar{\theta})\right]^{+} \sqrt{N} S_N\left(\theta_0\right) \xrightarrow{D} \mathcal{N}\left(0, 4\left(\Gamma_0' \Sigma_1^{-1} \Gamma_0\right)^{-1}\right) \tag{20}$$

2.1 Degeneracy

Consider the case when U_i and U_j are constants, e.g. $U_i = U_j = 1$, then the model becomes

$$Y_{ij} = \exp\left(R_{ij}\theta_0\right) V_{ij} \tag{21}$$

and that

$$l_{ij}(\theta) = Y_{ij} R_{ij}\theta - \exp\left(R_{ij}\theta\right) \tag{22}$$

$$s_{ij} = s\left(Z_{ij}, \theta_0\right) = \partial l_{ij}(\theta_0)/\partial\theta = Y_{ij} R_{ij} - \exp\left(R_{ij}\theta_0\right) R_{ij} \tag{23}$$

Since $\bar{s}_{ij} = \bar{s}\left(X_i, U_i, X_j, U_j\right) = \mathbb{E}\left[s\left(Z_{ij}, \theta_0\right) \mid X_i, U_i, X_j, U_j\right] = 0$,

$$V_N = \begin{pmatrix} N \\ 2 \end{pmatrix}^{-1} \sum_{i<j} \frac{\bar{s}_{ij} + \bar{s}_{ji}}{2} = 0 \tag{24}$$

$$S_N = \begin{pmatrix} N \\ 2 \end{pmatrix}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{s_{ij} + s_{ji}}{2} = V_N + T_N = T_N \tag{25}$$

which consists of a summation of $\begin{pmatrix} N \\ 2 \end{pmatrix}$ conditionally uncorrelated summands; hence

$$\mathbb{V}\left(S_N\right) = \mathbb{V}\left(T_N\right) = \frac{2}{N(N-1)}\Sigma_3 = O\left(N^{-2}\right) \tag{26}$$

If only U_i is a constant, e.g. $U_i = 1$, U_j is still a mean one random variable, then the model becomes

$$Y_{ij} = \exp\left(R_{ij}\theta_0\right) U_j V_{ij} \tag{27}$$

and that

$$l_{ij}(\theta) = Y_{ij} R_{ij}\theta - \exp\left(R_{ij}\theta\right) \tag{28}$$

$$s_{ij} = s\left(Z_{ij}, \theta_0\right) = \partial l_{ij}(\theta_0)/\partial\theta = Y_{ij}R_{ij} - \exp\left(R_{ij}\theta_0\right)R_{ij} \tag{29}$$

Since $\bar{s}_{ij} = \bar{s}\left(X_i, U_i, X_j, U_j\right) = \mathbb{E}\left[s\left(Z_{ij}, \theta_0\right) \mid X_i, U_i, X_j, U_j\right] = \exp\left(R_{ij}\theta_0\right)U_jR_{ij} - \exp\left(R_{ij}\theta_0\right)R_{ij}$, $\bar{s}^e(x,u) = \mathbb{E}\left[\bar{s}\left(x, u, X_1, U_1\right)\right] = \bar{s}^e(x)$ and $\bar{s}^a(x,u) = \mathbb{E}\left[\bar{s}\left(X_1, U_1, x, u\right)\right]$,

$$V_{1N} = \frac{2}{N}\sum_{i=1}^{N}\left\{\frac{\bar{s}_1^e\left(X_i, U_i\right) + \bar{s}_1^a\left(X_i, U_i\right)}{2}\right\} = \frac{2}{N}\sum_{i=1}^{N}\left\{\frac{\bar{s}_1^e\left(X_i\right) + \bar{s}_1^a\left(X_i, U_i\right)}{2}\right\} \tag{30}$$

and that

$$V_{2N} = \begin{pmatrix} N \\ 2 \end{pmatrix}^{-1}\sum_{i<j}\left\{\frac{\bar{s}_{ij} + \bar{s}_{ji}}{2} - \frac{\bar{s}_1^e\left(X_i\right) + \bar{s}_1^a\left(X_i, U_i\right)}{2} - \frac{\bar{s}_1^e\left(X_j\right) + \bar{s}_1^a\left(X_j, U_j\right)}{2}\right\} \tag{31}$$

We think about the variance of S_N in terms of the ANOVA decomposition :

$$\begin{aligned} \mathbb{V}\left(S_N\right) &= \mathbb{V}\left(\mathbb{E}\left[S_N \mid \mathbf{X}, \mathbf{U}\right]\right) + \mathbb{E}\left[\mathbb{V}\left(S_N \mid \mathbf{X}, \mathbf{U}\right)\right] \\ &= \mathbb{V}\left(V_N\right) + \mathbb{V}\left(T_N\right) \\ &= \mathbb{V}\left(V_{1N}\right) + \mathbb{V}\left(V_{2N}\right) + \mathbb{V}\left(T_N\right) \end{aligned} \tag{32}$$

Let

$$\Sigma_1 = \mathbb{C}\left(\mathbb{E}\left[\frac{s_{12} + s_{21}}{2} \mid X_1, U_1, X_2, U_2\right], \mathbb{E}\left[\frac{s_{13} + s_{31}}{2} \mid X_1, U_1, X_3, U_3\right]'\right) \tag{33}$$

and that

$$\begin{aligned} \Sigma_2 &= \mathbb{C}\left(\mathbb{E}\left[\frac{s_{12} + s_{21}}{2} \mid X_1, U_1, X_2, U_2\right], \mathbb{E}\left[\frac{s_{12} + s_{21}}{2} \mid X_1, U_1, X_2, U_2\right]'\right) \\ &= \mathbb{V}\left(\mathbb{E}\left[\frac{s_{12} + s_{21}}{2} \mid X_1, U_1, X_2, U_2\right]\right) \end{aligned} \tag{34}$$

$$\Sigma_3 \overset{\text{def}}{\equiv} \mathbb{E}\left[\mathbb{V}\left(\frac{s_{12} + s_{21}}{2} \mid X_1, U_1, X_2, U_2\right)\right] \tag{35}$$

Calculations analogous to those use in variance analyses for U-statistics variance analyses for U-statistics yield

$$\begin{aligned} \mathbb{V}\left(V_{1N}\right) &= \frac{4\Sigma_1}{N} \\ \mathbb{V}\left(V_{2N}\right) &= \frac{2}{N(N-1)}\left(\Sigma_2 - 2\Sigma_1\right) \\ \mathbb{V}\left(T_N\right) &= \frac{2}{N(N-1)}\Sigma_3 \end{aligned} \tag{36}$$

such that, defining the notation $\Omega \overset{\text{def}}{\equiv} \mathbb{V}\left(\sqrt{N}S_N\right)$,

$$\Omega = 4\Sigma_1 + \frac{2}{N-1}\left(\Sigma_2 + \Sigma_3 - 2\Sigma_1\right) \tag{37}$$

the variances of V_{2N} and T_N are of smaller order. Therefore,

$$\sqrt{N}\left(\hat{\theta}-\theta_0\right)=\left[-H_N(\bar{\theta})\right]^{+}\sqrt{N}S_N\left(\theta_0\right)\xrightarrow{D}\mathcal{N}\left(0,4\left(\Gamma_0'\Sigma_1^{-1}\Gamma_0\right)^{-1}\right) \tag{38}$$

Similarly, If only U_j is a constant, e.g. $U_j=1$, U_i is still a mean one random variable, we also have that

$$\Omega=4\Sigma_1+\frac{2}{N-1}\left(\Sigma_2+\Sigma_3-2\Sigma_1\right) \tag{39}$$

To sum up, in the nondegenerate case, V_{1N} drives the asymptotic distribution, while in degenerate case, T_N drives the asymptotic distribution. This means that the non-Gaussian term V_{2N} does not drive the asymptotic distribution in both cases.

3 Pigeonhole bootstrap validity

Definition 1 *(Pigeonhole Bootstrap)*

1. n units are sampled independently in $\{1,\ldots,n\}$ with replacement and equal probability. W_i denotes the number of times unit i is sampled.

2. the pair (i,j) is then selected $W_{ij}=W_iW_j$ times in the bootstrap sample.

The goal is to prove the bootstrap counterpart:

$$\sqrt{N}\left(\theta^*-\hat{\theta}\right)=\left[-H_N^*(\bar{\theta}^*)\right]^{+}\sqrt{N}S_N^*\left(\hat{\theta}\right)+o_p(1) \tag{40}$$

where

$$S_N^*(\theta)=\frac{1}{N}\frac{1}{N-1}\sum_i\sum_{j\neq i}s(Z_{ij}^*,\theta) \tag{41}$$

and that

$$H_N^*(\theta)=\frac{1}{N}\frac{1}{N-1}\sum_i\sum_{j\neq i}\frac{\partial s(Z_{ij}^*,\theta)}{\partial\theta} \tag{42}$$

There exist two sources of randomness for the bootstrapped quantity, for example, θ^*: one comes from the observed data; another comes from the resampling done by the bootstrap, that is, randomness in W_{ij} 's. Therefore, in order to rigorously state our theoretical results for the bootstrap, we need to specify relevant probability spaces and define the related stochastic orders.

We view Z_i as the i th coordinate projection from the canonical probability space $(\mathcal{Z}^{\infty}, \mathcal{A}^{\infty}, P_Z^{\infty})$ onto the i th copy of $\mathcal{X}$. For the joint randomness involved, the product probability space is defined as

$$(\mathcal{Z}^{\infty}, \mathcal{A}^{\infty}, P_Z^{\infty}) \times (\mathcal{W}, \Omega, P_W) = (\mathcal{Z}^{\infty} \times \mathcal{W}, \mathcal{A}^{\infty} \times \Omega, P_{ZW}) \tag{43}$$

we assume that W_{ij} is independent of the data Z_i 's, thus $P_{ZW} = P_Z^{\infty} \times P_W$. We write P_Z^{∞} as P_Z for simplicity thereafter. We define E_{ZW}^o as the outer expectation w.r.t. P_{ZW}. The notation $E_{W|Z}^o, E_Z^o$ and E_W are defined similarly. Given a real-valued function Δ_n defined on the above product probability space, for example, θ^*, we say that Δ_n is of an order $o_{P_W}^o(1)$ in P_Z^o-probability if for any $\varepsilon, \delta > 0$,

$$P_Z^o \left\{ P_{W|Z}^o \left(|\Delta_n| > \varepsilon \right) > \delta \right\} \longrightarrow 0 \quad \text{ as } n \to \infty \tag{44}$$

and that Δ_n is of an order $O_{P_W}^o(1)$ in P_Z^o-probability if for any $\delta > 0$, there exists a $0 < M < \infty$ such that

$$P_Z^o \left\{ P_{W|Z}^o \left(|\Delta_n| \geq M \right) > \delta \right\} \longrightarrow 0 \quad \text{ as } n \to \infty \tag{45}$$

Given a function Γ_n defined only on $(\mathcal{Z}^{\infty}, \mathcal{A}^{\infty}, P_Z^{\infty})$, if it is of an order $o_{P_Z}^o(1)\left[O_{P_Z}^o(1)\right]$, then it is also of an order $o_{P_{ZW}}^o(1)\left[O_{P_{ZW}}^o(1)\right]$ based on the following argument:

$$\begin{aligned} P_{ZW}^o \left(|\Gamma_n| > \varepsilon \right) &= E_{ZW}^o 1 \left\{ |\Gamma_n| > \varepsilon \right\} = E_Z E_{W|Z} 1 \left\{ |\Gamma_n| > \varepsilon \right\}^o \\ &= E_Z 1 \left\{ |\Gamma_n| > \varepsilon \right\}^o = P_Z^o \left\{ |\Gamma_n| > \varepsilon \right\} \end{aligned} \tag{46}$$

where the third equation holds since Γ_n does not depend on the bootstrap weight.

Lemma 1

$$\sqrt{N} \mathbb{P}_n^* \widehat{m} = \mathbb{G}_n^* m_0 + \left(\mathbb{G}_n^* - \mathbb{G}_n \right) \left(\widehat{m} - \widehat{m}^* \right) + o_{P_{ZW}}^o(1) \tag{47}$$

Proof.

$$\sqrt{N} \mathbb{P}_n^* \widehat{m} - \mathbb{G}_n^* m_0 = -I_1 + I_2 + I_3 + \left(\mathbb{G}_n^* - \mathbb{G}_n \right) \left(\widehat{m} - \widehat{m}^* \right)$$

where

$$\begin{aligned} I_1 &= -\mathbb{G}_n^* \left(\widehat{m}^* - m_0 \right) = \sqrt{N} \left(\mathbb{P}_n^* - \mathbb{P}_n \right) \left(m_0 - \widehat{m}^* \right) = o_{P_W}^o (1) \\ I_2 &= \mathbb{G}_n \left(\widehat{m} - m_0 \right) = \sqrt{N} \left(\mathbb{P}_n - P_Z \right) \left(\widehat{m} - m_0 \right) = o_{P_Z}^o (1) \\ I_3 &= -\mathbb{G}_n \left(\widehat{m}^* - m_0 \right) = \sqrt{N} \left(\mathbb{P}_n - P_Z \right) \left(m_0 - \widehat{m}^* \right) = o_{P_W}^o (1) \\ I_4 &= \sqrt{N} \mathbb{P}_n^* \widehat{m}^* - \sqrt{N} \mathbb{P}_n \widehat{m} = o_{P_Z}^o(1) + o_{P_{ZW}}^o(1) \end{aligned}$$

■

Lemma 2

$$\sqrt{N}\mathbb{P}_n^*\widehat{m}^* = \mathbb{G}_n^* m_0 + \sqrt{N} P_Z(\widehat{m}^* - \widehat{m}) + o^o_{P_{ZW}}(1) \tag{48}$$

Proof.

$$\begin{aligned}
\sqrt{N}\mathbb{P}_n^*\widehat{m} + \sqrt{N}\mathbb{P}_n^*(\widehat{m}^* - \widehat{m}) &= \mathbb{G}_n^* m_0 + (\mathbb{G}_n^* - \mathbb{G}_n)(\widehat{m} - \widehat{m}^*) + \sqrt{N}\mathbb{P}_n^*(\widehat{m}^* - \widehat{m}) + o^o_{P_{ZW}}(1) \\
&= \mathbb{G}_n^* m_0 + (\mathbb{G}_n^* - \mathbb{G}_n)(\widehat{m} - \widehat{m}^*) + \left(\mathbb{G}_n^* + \sqrt{N}\mathbb{P}_n\right)(\widehat{m}^* - \widehat{m}) + o^o_{P_{ZW}}(1) \\
&= \mathbb{G}_n^* m_0 - \mathbb{G}_n(\widehat{m} - \widehat{m}^*) + \sqrt{N}\mathbb{P}_n(\widehat{m}^* - \widehat{m}) + o^o_{P_{ZW}}(1) \\
&= \mathbb{G}_n^* m_0 + 2\mathbb{G}_n(\widehat{m}^* - \widehat{m}) + \sqrt{N} P_Z(\widehat{m}^* - \widehat{m}) + o^o_{P_{ZW}}(1) \\
&= \mathbb{G}_n^* m_0 + \sqrt{N} P_Z(\widehat{m}^* - \widehat{m}) - I_2 - I_3 + o^o_{P_{ZW}}(1) \\
&= \mathbb{G}_n^* m_0 + \sqrt{N} P_Z(\widehat{m}^* - \widehat{m}) + o^o_{P_{ZW}}(1)
\end{aligned}$$

Combing Lemma 1 and Lemma 2,

$$\begin{aligned}
\sqrt{N} P_Z(\widehat{m}^* - \widehat{m}) &= -\mathbb{G}_n^* m_0 + o^o_{P_{ZW}}(1) \\
\sqrt{N} P_Z m^{\bar{\theta}^*}\left(\theta^* - \hat{\theta}\right) &= -\mathbb{G}_n^* m_0 + o^o_{P_{ZW}}(1) \\
\sqrt{N}\left(\theta^* - \hat{\theta}\right) &= \left(-P_Z m^{\bar{\theta}^*}\right)^{-1}(\mathbb{G}_n^* m_0) + o^o_{P_{ZW}}(1)
\end{aligned}$$

■

Lemma 3 *Let M_n be random functions and let M be a fixed function of θ such that for every $\varepsilon > 0$,*

$$\begin{aligned}
&\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \xrightarrow{\mathrm{P}} 0 \\
&\sup_{\theta: d(\theta,\theta_0)\geq\varepsilon} M(\theta) < M(\theta_0)
\end{aligned} \tag{49}$$

Then any sequence of estimators $\hat{\theta}_n$ with $M_n\left(\hat{\theta}_n\right) \geq M_n(\theta_0) - o_P(1)$ converges in probability to θ_0.

Lemma 4 *Suppose that*

1.("well-separated" condition)

$$P_Z l(\theta_0) > \sup_{\theta\notin G} P_Z l(\theta) \tag{50}$$

for any open set $G \subset \Theta$ containing θ_0.

2.

$$\sup_{\theta \in \Theta} |\mathbb{P}_n^* m(\theta) - P_Z m(\theta)| \xrightarrow{P_{ZW}^o} 0 \tag{51}$$

We have that $\theta^* \xrightarrow{P_W^o} \theta_0$ in P_Z^o-probability.

Proof. See Lemma 3. ■

Lemma 5 *Suppose that*

$$\begin{aligned} Q_n &= o_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability}, \\ R_n &= O_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability} \\ Q_n' &= o_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability}, \\ R_n' &= O_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability} \end{aligned} \tag{52}$$

We have

$$A_n = o_{P_{ZW}}^o(1) \Longleftrightarrow A_n = o_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability}, \tag{53}$$

$$B_n = O_{P_{ZW}}^o(1) \Longleftrightarrow B_n = O_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability}, \tag{54}$$

$$C_n = Q_n \times O_{P_Z}^o(1) \Longrightarrow C_n = o_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability}, \tag{55}$$

$$D_n = R_n \times O_{P_Z}^o(1) \Longrightarrow D_n = O_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability}, \tag{56}$$

$$E_n = Q_n \times R_n \Longrightarrow E_n = o_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability}. \tag{57}$$

$$F_n = Q_n + Q_n' \Longrightarrow F_n = o_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability}. \tag{58}$$

$$G_n = Q_n + R_n \Longrightarrow G_n = O_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability}. \tag{59}$$

$$H_n = R_n + R_n' \Longrightarrow H_n = O_{P_W}^o(1) \quad \text{in } P_Z^o\text{-probability}. \tag{60}$$

Proof. For equation (53), we have for every $\varepsilon, v > 0$,

$$\begin{aligned} P_Z^o \left\{ P_{W|Z}^o \left(|A_n| \geq \varepsilon \right) \geq v \right\} &\leq \frac{1}{v} E_Z^o P_{W|Z}^o \left(|A_n| \geq \varepsilon \right) \\ &\leq \frac{1}{v} E_Z^o E_{W|Z}^o 1 \left\{ |A_n| \geq \varepsilon \right\} \end{aligned} \tag{61}$$

by Markov's inequality.

Lemma 6 *(Fubini's theorem)*

Let T be an arbitrary real valued map on a product probability space. Then $\mathrm{E}_1^*\mathrm{E}_2^*T \leq \mathrm{E}^*T$, where $\mathrm{E}_1^*\mathrm{E}_2^*T$ is the outer integral of E_2^*T.

According to Lemma 6,we have $E_Z^o E_{W|Z}^o 1\left\{|A_n| \geq \varepsilon\right\} \leq E_{ZW}^o 1\left\{|A_n| \geq \varepsilon\right\} = P_{ZW}^o\left(|A_n| \geq \varepsilon\right)$, and thus

$$P_Z^o\left\{P_{W|Z}^o\left(|A_n| \geq \varepsilon\right) \geq \nu\right\} \leq \frac{1}{v} P_{ZW}^o\left(|A_n| \geq \varepsilon\right) \tag{62}$$

which gives that $A_n = o_{P_{ZW}}^o(1) \Longrightarrow A_n = o_{P_W}^o(1) \quad$ in P_Z^o-probability, $\;A_n = o_{P_{ZW}}^o(1) \Longleftarrow A_n = o_{P_W}^o(1) \quad$ in P_Z^o-probability follows from the following inequalities, for any $\varepsilon, \eta > 0$:

$$\begin{aligned}
P_{ZW}^o\left(|A_n| \geq \varepsilon\right) =& E_Z^o\left\{P_{W|Z}^o\left(|A_n| \geq \varepsilon\right)\right\} \\
=& E_Z^o\left\{P_{W|Z}^o\left(|A_n| \geq \varepsilon\right) 1\left\{P_{W|Z}^o\left(|A_n| \geq \varepsilon\right) \geq \eta\right\}\right\} \\
&+ E_Z^o\left\{P_{W|Z}^o\left(|A_n| \geq \varepsilon\right) 1\left\{P_{W|Z}^o\left(|A_n| \geq \varepsilon\right) < \eta\right\}\right\} \\
\leq& E_Z^o\left\{1\left\{P_{W|Z}^o\left(|A_n| \geq \varepsilon\right) \geq \eta\right\}\right\} + \eta \\
\leq& P_Z^o\left\{P_W^o\left(|A_n| \geq \varepsilon\right) \geq \eta\right\} + \eta
\end{aligned} \tag{63}$$

Note that the first term in equation (63) can be made arbitrarily small by the assumption that $A_n = o_{P_W}^o(1)$ in P_Z^o-probability. Since η can be chosen arbitrarily small, we can show $\lim_{n\to\infty} P_{ZW}^o\left(|A_n| \geq \varepsilon\right) = 0$ for any $\varepsilon > 0$. For equation (54), we have for every $v > 0$, there exists a $0 < M < \infty$ such that

$$\begin{aligned}
P_Z^o\left\{P_{W|Z}^o\left(|A_n| \geq M\right) \geq v\right\} &\leq \frac{1}{v} E_Z^o P_{W|Z}^o\left(|A_n| \geq M\right) \\
&\leq \frac{1}{v} E_Z^o E_{W|Z}^o 1\left\{|A_n| \geq M\right\}
\end{aligned} \tag{64}$$

by Markov's inequality. According to Lemma 6, we have

$$E_Z^o E_{W|Z}^o 1\left\{|A_n| \geq M\right\} \leq E_{ZW}^o 1\left\{|A_n| \geq M\right\} = P_{ZW}^o\left(|A_n| \geq M\right)$$

and thus

$$P_Z^o\left\{P_{W|Z}^o\left(|A_n| \geq M\right) \geq \nu\right\} \leq \frac{1}{v} P_{ZW}^o\left(|A_n| \geq M\right) \tag{65}$$

which gives that $A_n = O_{P_{ZW}}^o(1) \Longrightarrow A_n = O_{P_W}^o(1) \quad$ in P_Z^o-probability, $\;A_n = O_{P_{ZW}}^o(1) \Longleftarrow$

$A_n = O^o_{P_W}(1)$ in P^o_Z-probability follows from the following inequalities, for any $\eta > 0$:

$$
\begin{aligned}
P^o_{ZW}\left(|A_n| \geq M\right) =& E^o_Z\left\{P^o_{W|Z}\left(|A_n| \geq M\right)\right\} \\
=& E^o_Z\left\{P^o_{W|Z}\left(|A_n| \geq M\right) 1\left\{P^o_{W|Z}\left(|A_n| \geq M\right) \geq \eta\right\}\right\} \\
& + E^o_Z\left\{P^o_{W|Z}\left(|A_n| \geq M\right) 1\left\{P^o_{W|Z}\left(|A_n| \geq M\right) < \eta\right\}\right\} \\
\leq& E^o_Z\left\{1\left\{P^o_{W|Z}\left(|A_n| \geq M\right) \geq \eta\right\}\right\} + \eta \\
\leq& P^o_Z\left\{P^o_W\left(|A_n| \geq M\right) \geq \eta\right\} + \eta
\end{aligned} \tag{66}
$$

Note that the first term in equation (66) can be made arbitrarily small by the assumption that $A_n = O^o_{P_W}(1)$ in P^o_Z-probability. Since η can be chosen arbitrarily small, we can show $\lim_{n\to\infty} P^o_{ZW}\left(|A_n| \geq M\right) = 0$. As for equation (55),

$$
\begin{aligned}
& P^o_Z\{\, P^o_{W|Z}\left(\left|Q_n \times O^o_{P_Z}(1)\right| \geq \varepsilon\right) \geq v\} \\
& \quad \leq P^o_Z\left\{P^o_{W|Z}\left(|Q_n| \geq \varepsilon / \left|O^o_{P_Z}(1)\right|\right) \geq v\right\} \\
& \quad \leq P^o_Z\left\{P^o_{W|Z}\left(|Q_n| \geq \varepsilon/M\right) + P^o_{W|Z}\left(\left|O^o_{P_Z}(1)\right| \geq M\right) \geq v\right\} \\
& \quad \leq P^o_Z\left\{P^o_{W|Z}\left(|Q_n| \geq \varepsilon/M\right) \geq v/2\right\} + P^o_Z\left\{P^o_{W|Z}\left(\left|O^o_{P_Z}(1)\right| \geq M\right) \geq v/2\right\} \\
& \quad \leq P^o_Z\left\{P^o_{W|Z}\left(|Q_n| \geq \varepsilon/M\right) \geq v/2\right\} + \frac{2}{v} P^o_Z\left(\left|O^o_{P_Z}(1)\right| \geq M\right)
\end{aligned} \tag{67}
$$

for any $\varepsilon, \nu, M > 0$. Since M can be chosen arbitrarily large, we can show equation (55) by considering the definition of $O^o_{P_Z}(1)$. As for equation (56),

$$
\begin{aligned}
& P^o_Z\{\, P^o_{W|Z}\left(\left|Q_n \times O^o_{P_Z}(1)\right| \geq M\right) \geq v\} \\
& \quad \leq P^o_Z\left\{P^o_{W|Z}\left(|Q_n| \geq M / \left|O^o_{P_Z}(1)\right|\right) \geq v\right\} \\
& \quad \leq P^o_Z\left\{P^o_{W|Z}\left(|Q_n| \geq M/M_1\right) + P^o_{W|Z}\left(\left|O^o_{P_Z}(1)\right| \geq M\right) \geq v\right\} \\
& \quad \leq P^o_Z\left\{P^o_{W|Z}\left(|Q_n| \geq M/M_1\right) \geq v/2\right\} + P^o_Z\left\{P^o_{W|Z}\left(\left|O^o_{P_Z}(1)\right| \geq M\right) \geq v/2\right\} \\
& \quad \leq P^o_Z\left\{P^o_{W|Z}\left(|Q_n| \geq M/M_1\right) \geq v/2\right\} + \frac{2}{v} P^o_Z\left(\left|O^o_{P_Z}(1)\right| \geq M\right)
\end{aligned} \tag{68}
$$

for any $\nu, M_1 > 0$. Since M can be chosen arbitrarily large, we can show equation (56) by considering the definition of $O^o_{P_Z}(1)$. To prove equation (57), we have that

$$
\begin{aligned}
& P^o_Z\left\{P^o_{W|Z}\left(|Q_n \times R_n| \geq \varepsilon\right) \geq \eta\right\} \\
& \quad \leq P^o_Z\left\{P^o_{W|Z}\left(|Q_n| \geq \varepsilon/M\right) \geq \eta/2\right\} + P^o_Z\left\{P^o_{W|Z}\left(|R_n| \geq M\right) \geq \eta/2\right\}
\end{aligned} \tag{69}
$$

for any $\varepsilon, \eta, M > 0$. Then by selecting sufficiently large M, we can show that

$$P_Z^o \left\{ P_{W|Z}^o \left(|Q_n \times R_n| \geq \varepsilon \right) \geq \eta \right\} \to 0$$

as $n \to \infty$ for any $\varepsilon, \eta > 0$. As for equation (58),

$$\begin{aligned} &P_Z^o \{ \, P_{W|Z}^o \left(|Q_n + Q_n'| \geq \varepsilon \right) \geq v \} \\ &\quad \leq P_Z^o \left\{ P_{W|Z}^o \left(|Q_n| \geq \varepsilon/2 \right) \geq v/2 \right\} + P_Z^o \left\{ P_{W|Z}^o \left(|Q_n'| \geq \varepsilon/2 \right) \geq v/2 \right\} \end{aligned} \tag{70}$$

for any $\varepsilon, \nu > 0$. We can show equation (58) by considering the definition of $o_{P_W}^o$ (1). As for equation (59),

$$\begin{aligned} &P_Z^o \{ \, P_{W|Z}^o \left(|Q_n + R_n| \geq M \right) \geq v \} \\ &\quad \leq P_Z^o \left\{ P_{W|Z}^o \left(|Q_n| \geq \varepsilon \right) \geq v/2 \right\} + P_Z^o \left\{ P_{W|Z}^o \left(|R_n| \geq M - \varepsilon \right) \geq v/2 \right\} \end{aligned} \tag{71}$$

for any $\varepsilon, \nu > 0$. We can show equation (59) by considering the definition of $o_{P_W}^o$ (1), $O_{P_W}^o$ (1) and selecting sufficiently large M. As for equation (63),

$$\begin{aligned} &P_Z^o \{ \, P_{W|Z}^o \left(|R_n + R_n'| \geq M \right) \geq v \} \\ &\quad \leq P_Z^o \left\{ P_{W|Z}^o \left(|R_n| \geq M/2 \right) \geq v/2 \right\} + P_Z^o \left\{ P_{W|Z}^o \left(|R_n'| \geq M/2 \right) \geq v/2 \right\} \end{aligned} \tag{72}$$

for any $\nu > 0$. We can show equation (59) by considering the definition of $o_{P_W}^o$ (1), $O_{P_W}^o$ (1) and selecting sufficiently large M. ■

Lemma 7

$$\|\theta^* - \theta_0\| = O_{P_W}^o \left(N^{-1/2} \right) \tag{73}$$

in P_Z^o-probability.

Proof. Since $P_Z m_0 = 0$, we have that

$$\mathbb{G}_n^* m_0 + \mathbb{G}_n m_0 + \sqrt{N} P_Z \left(\widehat{m}^* - m_0 \right) = \mathbb{G}_n^* \left(m_0 - \widehat{m}^* \right) + \mathbb{G}_n \left(m_0 - \widehat{m}^* \right) + \sqrt{N} \mathbb{P}_n^* \widehat{m}^* \tag{74}$$

Thus, we have the following inequality:

$$\begin{aligned} \left\| \sqrt{n} P_Z \left(\widehat{m}^* - m_0 \right) \right\| \leq & \|\mathbb{G}_n^* m_0\| + \|\mathbb{G}_n m_0\| + \|\mathbb{G}_n^* \left(m_0 - \widehat{m}^* \right)\| + \|\mathbb{G}_n \left(m_0 - \widehat{m}^* \right)\| + \left\| \sqrt{N} \mathbb{P}_n^* \widehat{m}^* \right\| \\ \equiv & L_1 + L_2 + L_3 + L_4 + L_5 \end{aligned} \tag{75}$$

$L_1 = O^o_{P_W}(1)$ (I will add the proof later), $L_2 = O^o_{P_Z}(1)$ (By CLT, lemma needs to be cited or added), Furthermore, I will prove $L_3 = O^o_{P_W}(1)$ and $L_4 = O^o_{P_W}(1)$, finally, $L_5 = o^o_{P_{ZW}}(1)$. To sum up,

$$\left\|\sqrt{N} P_Z \left(\widehat{m}^* - m_0\right)\right\| \leq O^o_{P_W}(1) + O^o_{P_Z}(1) \tag{76}$$

■

4 Block Bootstrap

Definition 2 *(Block Bootstrap)*

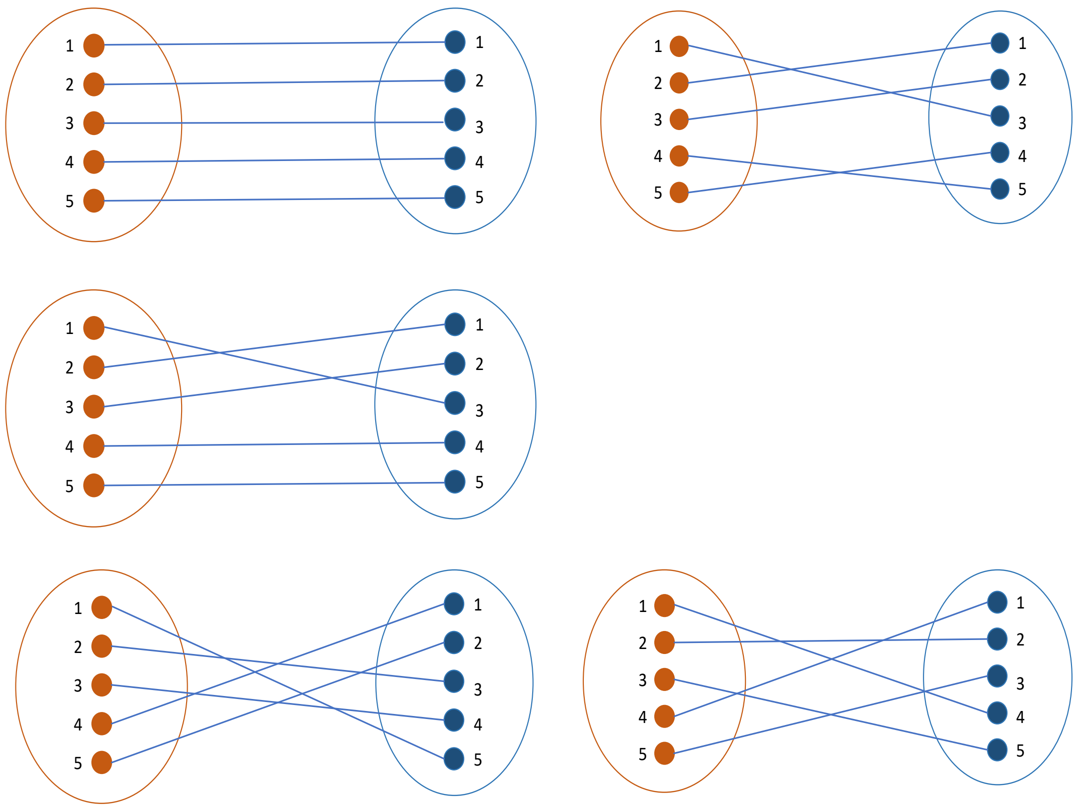
1. We want to sample b blocks with length l randomly with replacement from the original dataset, in which $bl = NT$.

2. For every random draw with block size l: select l integers from $\{1, \ldots, N\}$ without replacement and l integers from $\{1, \ldots, T\}$ without replacement, draw one permutation, W_{ij} is constructed as: i is from the selected l integers in $\{1, \ldots, N\}$ and j is from the selected l integers in $\{1, \ldots, T\}$.

For example $N = T = 5$, the original dataset is as follows:

11	12	13	14	15
21	22	23	24	25
31	32	33	34	35
41	42	43	44	45
51	52	53	54	55

Suppose that we choose block length $l = 5$, and number of blocks $b = 5$. In total, we have 25 different blocks to choose from, for every bootstrap sample, we randomly choose 5 blocks with replacements. For one bootstrap sample, we have for example:



Consider the additive factor model in Menzel (2021)'s example,

$$Y_{it} = \mu + \alpha_i + \gamma_t + \varepsilon_{it} \tag{77}$$

where μ is fixed, and $\alpha_i, \gamma_t, \varepsilon_{it}$ are zero zero-mean, i.i.d. random variables for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ with bounded second moments, and $N = T$. The parameter of interest is:

$$\bar{Y}_{NT} := \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} Y_{it} \tag{78}$$

The variance of the sample analogue is

$$Var\left(\bar{Y}_{NT}\right) = \frac{1}{N^2 T^2} \left(T^2 \sum_{i=1}^{N} \mathrm{Var}\left(\alpha_i\right) + N^2 \sum_{t=1}^{T} \mathrm{Var}\left(\gamma_t\right) + \sum_{i=1}^{N} \sum_{t=1}^{T} \mathrm{Var}\left(\varepsilon_{it}\right) \right) \tag{79}$$

Under naive bootstrap,

$$Var\left(\bar{Y}_{NT}\right) = \frac{1}{N^2T^2}\left[T\left(1-\frac{1}{N}\right)\sum_i \mathrm{Var}\left(\alpha_i\right) + N\left(1-\frac{1}{T}\right)\sum_t \mathrm{Var}\left(\gamma_t\right) + \sum_i\sum_t \mathrm{Var}\left(\varepsilon_{it}\right)\right] \tag{80}$$

Under pigeonhole bootstrap,

$$\begin{aligned} Var\left(\bar{Y}_{NT}\right) = \frac{1}{N^2T^2}\Bigg[T^2\left(1-\frac{1}{T}\right)\sum_i \left(\bar{Y}_{i\bullet}-\bar{Y}_{NT}\right)^2 + N^2\left(1-\frac{1}{N}\right)\sum_j \left(\bar{Y}_{\bullet j}-\bar{Y}_{NT}\right)^2 \\ + \sum_i\sum_j \left(Y_{ij}-\bar{Y}_{NT}\right)^2\Bigg] \end{aligned} \tag{81}$$

where $\bar{Y}_{i\bullet} = \frac{1}{T}\sum_{t=1}^{T} Y_{ij}$ and $\bar{Y}_{\bullet j} = \frac{1}{N}\sum_{i=1}^{N} Y_{ij}$. But under block bootstrap, we have correct variance.