Reliable Wild Bootstrap Inference with Multiway Clustering

Ulrich Hounyo[1] Jiahao Lin[2]

[1]Department of Economics
University at Albany - State University of New York

[2]Department of Economics
University at Albany - State University of New York

December 27, 2024

Multiway Clustering

- In the realm of statistical inference, the consideration of dependence structures within data has been a subject of enduring significance.
- Previous research has shown that failure to address the dependence in data analysis leads to significantly misleading inferences. (Bertrand, Duflo, and Mullainathan, 2004; Petersen, 2008)
- This paper considers a specific dependence structure where clustering of observations or disturbances (error terms) may occur in two or more dimensions.

Why Relevant?

Here are a few examples where multiway cluster-robust inference techniques can be applied in economics and finance:

- Panel Data Analysis: Economic studies often involve panel data where observations are collected across different entities (countries, firms) and over time (years, months).
- Asset Pricing Models: In finance, asset pricing models often incorporate multi-dimensional data, considering assets from different sectors, industries, or geographical regions.
- International Trade: Trade data involve various dimensions, such as countries, industries, and time periods.
- $\cdots$

Two-way Clustering Regression Model

We focus mainly on the two-way clustering:

$$\boldsymbol{y}_{gh} = \boldsymbol{X}_{gh}\boldsymbol{\beta} + \boldsymbol{u}_{gh}, \quad g = 1, \ldots, G, \quad h = 1, \ldots, H,$$

where $\boldsymbol{y}_{gh}$ is an N_{gh} by 1 vector consisting of observations in the intersection (g, h) of y that correspond to both the g^{th} cluster in the first dimension and the h^{th} cluster in the second dimension. Similarly, $\boldsymbol{u}_{gh}$ and $\boldsymbol{X}_{gh}$ adhere to this structure, though $\boldsymbol{X}_{gh}$ differs in that it is an N_{gh} by K matrix. As usual, the OLS estimator

$$\hat{\boldsymbol{\beta}} = \left(\sum_{g}\sum_{h}\boldsymbol{X}_{gh}^{\top}\boldsymbol{X}_{gh}\right)^{-1}\sum_{g}\sum_{h}\boldsymbol{X}_{gh}^{\top}\boldsymbol{y}_{gh}.$$

The null hypothesis $\mathcal{H}_0 : \boldsymbol{a}^{\top}\boldsymbol{\beta} = \boldsymbol{a}^{\top}\boldsymbol{\beta}_0$, where $\boldsymbol{a}$ is a known unit vector.

Variance of $\widehat{\beta}$

We begin with the standard two-way clustering, which assumes that two intersections are considered independent if no shared clusters:

$$E\left(\boldsymbol{X}_{gh}^{\top}\boldsymbol{u}_{gh}\boldsymbol{u}_{g'h'}^{\top}\boldsymbol{X}_{g'h'}\right)=0,\ \text{for } g'\neq g \text{ and } h'\neq h.$$

Cameron et al. (2011) (CGM hereafter) introduce a cluster-robust variance estimator denoted as $\widehat{\boldsymbol{V}}_{CGM}=\widehat{\boldsymbol{Q}}^{-1}\widehat{\boldsymbol{\Gamma}}_{CGM}\widehat{\boldsymbol{Q}}^{-1}$, with

$$\widehat{\boldsymbol{\Sigma}}_{CGM}=\sum_{g}\boldsymbol{X}_{g}^{\top}\widehat{\boldsymbol{u}}_{g}\widehat{\boldsymbol{u}}_{g}^{\top}\boldsymbol{X}_{g}+\sum_{h}\boldsymbol{X}_{h}^{\top}\widehat{\boldsymbol{u}}_{h}\widehat{\boldsymbol{u}}_{h}^{\top}\boldsymbol{X}_{h} - \sum_{g}\sum_{h}\boldsymbol{X}_{gh}^{\top}\widehat{\boldsymbol{u}}_{gh}\widehat{\boldsymbol{u}}_{gh}^{\top}\boldsymbol{X}_{gh}.$$

Under some regularity assumptions, MacKinnon et al. (2021) show that

$$\widehat{t}_{CGM}\equiv\frac{\boldsymbol{a}^{\top}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_{0}\right)}{\sqrt{\boldsymbol{a}^{\top}\widehat{\boldsymbol{V}}_{CGM}\boldsymbol{a}}}\rightarrow^{d}\mathcal{N}(0,1).$$

Bootstrap Algorithm

The bootstrap method is known to have the potential to improve the finite sample results. We now provide the algorithm procedure for a standard residual bootstrap method:

- Regress $\boldsymbol{y}$ on $\boldsymbol{X}$ to obtain the regression estimate $\widehat{\boldsymbol{\beta}}$, the residuals $\widehat{u}_{gh,i}$, and the cluster-robust variance estimate $\widehat{\boldsymbol{V}}$.
- Construct the original cluster-robust t-statistic $\widehat{t} = \frac{\boldsymbol{a}^\top(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)}{\sqrt{\boldsymbol{a}^\top\widehat{\boldsymbol{V}}\boldsymbol{a}}}$.
- Perturb the estimated residual $\widehat{u}_{gh,i}$ to generate the bootstrap residuals $u^b_{gh,i}$, b is an index for the bootstrap number. Different bootstrap methods amount to different ways of perturbing the estimated residual, which is the key difference. The general wild bootstrap generates bootstrap disturbances $\boldsymbol{u}^b$ as follows:

$$u^b_{gh,i} = \widehat{u}_{gh,i}\nu^b_{gh,i}.$$

Bootstrap Algorithm (Cont')

- Generate the bootstrap dependent variables $\boldsymbol{y}^b = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{u}^b$.
- Regress $\boldsymbol{y}^b$ on $\boldsymbol{X}$ to obtain the bootstrap regression estimate $\widehat{\boldsymbol{\beta}}^b$ and the bootstrap variance estimate $\widehat{\boldsymbol{V}}^b$. Obtain the bootstrap cluster-robust t-statistic.

$$\widehat{t}^b = \frac{\boldsymbol{a}^\top \left(\widehat{\boldsymbol{\beta}}^b - \widehat{\boldsymbol{\beta}}\right)}{\sqrt{\boldsymbol{a}^\top \widehat{\boldsymbol{V}}^b \boldsymbol{a}}}.$$

Bootstrap Algorithm (Cont')

- Repeat Step 3 to Step 5 for B times. For the two-sided alternative hypothesis $\mathrm{H}_2 : \boldsymbol{a}^\top \boldsymbol{\beta} \neq \boldsymbol{a}^\top \boldsymbol{\beta}_0$, compute the symmetric or equal-tail bootstrap P values:

$$P_{\mathsf{S}}^* = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I} \left(\left| \widehat{t}^b \right| > \left| \widehat{t} \right| \right),$$

$$P_{\mathsf{E}}^* = 2 \min \left(\frac{1}{B} \sum_{b=1}^{B} \mathbb{I} \left(\widehat{t}^b < \widehat{t} \right), \frac{1}{B} \sum_{b=1}^{B} \mathbb{I} \left(\widehat{t}^b > \widehat{t} \right) \right)$$
$$= 2 \min \left(P_{\mathsf{L}}^*, P_{\mathsf{R}}^* \right),$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. Reject the null hypothesis if the bootstrap P value is smaller than the predetermined significance level.

Existing Bootstrap Methods

- MacKinnon et al. (2021): WCB_G, WCB_H, WCB_I, WB.
  $\Rightarrow$ Can preserve the original data structure, but cannot faithfully mimic the DGP. e.g. $u^b_{gh,i} = \widehat{u}_{gh,i}\nu^b_{gh}$, where $\nu^b_{gh} \overset{i.i.d.}{\sim} (0,1)$.
- Davezies et al. (2021): The pigeonhole method (draw g and h with replacement independently).
- Menzel (2021)'s method.
  $\Rightarrow$ Both bootstrap methods rely on the Efron (1979) bootstrap, which is known to be inadequate in preserving the original data structure, especially in scenarios involving missing observations or unbalanced intersection sizes, common in practical settings.

These challenges motivate our multiway wild cluster bootstrap (MWCB) method, designed to preserve multi-dimensional dependence simultaneously, minimize the introduction of excessive "artificial" dependence, and remain robust to unbalanced data structures.

Multiway Wild Cluster Bootstrap (MWCB$_I$)

The bootstrap errors (in Step 3 of Bootstrap Algorithm) are obtained as follows:

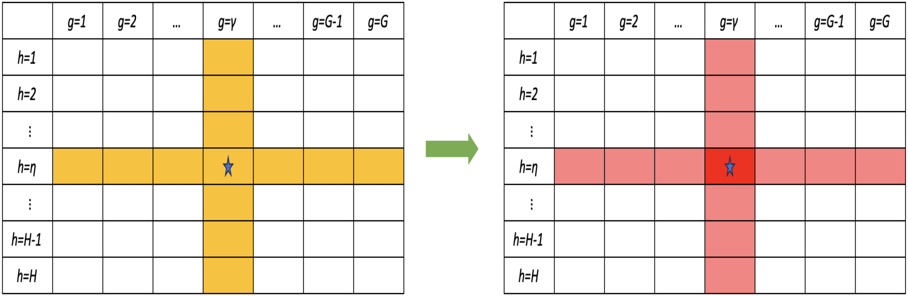
$$u^*_{gh,i} = \ddot{u}_{gh,i}\nu^*_{gh} \equiv \ddot{u}_{gh,i}\left(\frac{1}{\sqrt{G+H-1}}\sum_{\gamma=1}^{G}\sum_{\eta=1}^{H}\nu^{*\gamma\eta}_{gh}\right),$$

where given γ, η,

$$\nu^{*\gamma\eta}_{gh} = \begin{cases} 0, & \text{if } g \neq \gamma \text{ and } h \neq \eta, \\ \nu^{*\gamma\eta}, & \text{if } g = \gamma \text{ or } h = \eta, \end{cases}$$

with $\nu^{*\gamma\eta} \sim (0,1)$ i.i.d. over γ, η.

How the MWCB Method Mimic the Dependence Structure



The "fundamental" random variables $v_{gh}^{*\gamma\eta}$ are perfectly correlated when either $g=\gamma$ or $h=\eta$, given γ,η. That is, the intersections that share the same cluster along at least one dimension with the intersection (γ,η) will be assigned perfectly correlated random variables.

The "aggregated" random variables v_{gh}^{*} are strongly correlated with $v_{\gamma\eta}^{*}$ when either $g=\gamma$ or $h=\eta$. A darker shade suggests a higher degree of correlation with intersection (γ,η).

Figure: **An illustration of how the random variable is generated under the standard multiway clustering.**

A Simple Example When $G = H = 4$

ν^*_{gh} is the weighted sum of all elements in each row (g, h).

$\nu^{*\gamma\eta}$ is independent over each column (γ, η).

	ν^{*11}	ν^{*12}	ν^{*13}	ν^{*14}	ν^{*21}	ν^{*22}	ν^{*23}	ν^{*24}	ν^{*31}	ν^{*32}	ν^{*33}	ν^{*34}	ν^{*41}	ν^{*42}	ν^{*43}	ν^{*44}
ν^*_{11}	ν^{*11}	ν^{*12}	ν^{*13}	ν^{*14}	ν^{*21}				ν^{*31}				ν^{*41}			
ν^*_{12}	ν^{*11}	ν^{*12}	ν^{*13}	ν^{*14}		ν^{*22}				ν^{*32}				ν^{*42}		
ν^*_{13}	ν^{*11}	ν^{*12}	ν^{*13}	ν^{*14}			ν^{*23}				ν^{*33}				ν^{*43}	
ν^*_{14}	ν^{*11}	ν^{*12}	ν^{*13}	ν^{*14}				ν^{*24}				ν^{*34}				ν^{*44}
ν^*_{21}	ν^{*11}				ν^{*21}	ν^{*22}	ν^{*23}	ν^{*24}	ν^{*31}				ν^{*41}			
ν^*_{22}		ν^{*12}			ν^{*21}	ν^{*22}	ν^{*23}	ν^{*24}		ν^{*32}				ν^{*42}		
ν^*_{23}			ν^{*13}		ν^{*21}	ν^{*22}	ν^{*23}	ν^{*24}			ν^{*33}				ν^{*43}	
ν^*_{24}				ν^{*14}	ν^{*21}	ν^{*22}	ν^{*23}	ν^{*24}				ν^{*34}				ν^{*44}
ν^*_{31}	ν^{*11}				ν^{*21}				ν^{*31}	ν^{*32}	ν^{*33}	ν^{*34}	ν^{*41}			
ν^*_{32}		ν^{*12}				ν^{*22}			ν^{*31}	ν^{*32}	ν^{*33}	ν^{*34}		ν^{*42}		
ν^*_{33}			ν^{*13}				ν^{*23}		ν^{*31}	ν^{*32}	ν^{*33}	ν^{*34}			ν^{*43}	
ν^*_{34}				ν^{*14}				ν^{*24}	ν^{*31}	ν^{*32}	ν^{*33}	ν^{*34}				ν^{*44}
ν^*_{41}	ν^{*11}				ν^{*21}				ν^{*31}				ν^{*41}	ν^{*42}	ν^{*43}	ν^{*44}
ν^*_{42}		ν^{*12}				ν^{*22}				ν^{*32}			ν^{*41}	ν^{*42}	ν^{*43}	ν^{*44}
ν^*_{43}			ν^{*13}				ν^{*23}				ν^{*33}		ν^{*41}	ν^{*42}	ν^{*43}	ν^{*44}
ν^*_{44}				ν^{*14}				ν^{*24}				ν^{*34}	ν^{*41}	ν^{*42}	ν^{*43}	ν^{*44}

Figure: **A diagrammatic display for the connection between $\nu^{*\gamma\eta}_{gh}$ and ν^*_{gh} when $G = H = 4$.**

Kurtosis Correction

- As noted by Djogbenou et al. (2019), a wild bootstrap with random weights that incorporate kurtosis correction (i.e., having a fourth moment equal to 1) is important.
- However, the first method generates ν^*_{gh} by summing some independent fundamental random variables, which tends to be normally distributed with a fourth moment of 3.
- The second bootstrap DGP approach is proposed to address this issue while preserving the correlation structure within clusters, similar to the first method.

Second Multiway Wild Cluster Bootstrap ($MWCB_{II}$)

The second bootstrap random weights are obtained as follows:

$$\nu_{gh}^* = \begin{cases} \nu_g^*, & \text{with probability } p, \\ \nu_h^*, & \text{with probability } 1-p, \end{cases}$$

with ν_g^* and ν_h^* mutually independent over g and h, respectively. We let ν_g^* and ν_h^* to follow a Rademacher distribution, so that we have ν_{gh} also follows a Rademacher distribution.

When $p = 1$, our method becomes WCB_G; and when $p = 0$, our method becomes WCB_H.

Correlation of Different Methods

Notably, one can deduce that for $g \neq g'$ and $h \neq h'$,

	$corr^*(\nu^*_{gh}, \nu^*_{gh'})$	$corr^*(\nu^*_{gh}, \nu^*_{g'h})$	$corr^*(\nu^*_{gh}, \nu^*_{g'h'})$	fourth moment
Ideal	1	1	0	1
WCB_I	0	0	0	1
WCB_G	1	0	0	1
WCB_H	0	1	0	1
MWCB_I	$\frac{H}{G+H-1}$	$\frac{G}{G+H-1}$	$\frac{2}{G+H-1}$	(1,3)
MWCB_{II}	p^2	$(1-p)^2$	0	1

Table: Comparison of different wild bootstrap methods

Two MWCB methods account for partial dependence in both dimensions.

Discussion

- A wild bootstrap following the standard bootstrap Algorithm with the following correlations seems impossible:

$$corr^*(\nu^*_{gh}, \nu^*_{gh'}) = 1, corr^*(\nu^*_{gh}, \nu^*_{g'h}) = 1, \text{ and } corr^*(\nu^*_{gh}, \nu^*_{g'h'}) = 0.$$

- To see this, assume the first two equalities hold. In that case, we would expect $corr^*(\nu^*_{gh}, \nu^*_{g'h'}) = 1$, which contradicts the third equality.

- WCB$_G$ and WCB$_H$ are special cases of the two MWCB methods.

- One advantage of our approach is its robustness when the true cluster dependence is unknown.

Bootstrap Theory

We further establish the first-order asymptotic validity of two MWCB methods based on the CGM variance:

$$\widehat{t}^{*}_{CGM} = \frac{\boldsymbol{a}^{\top}\left(\widehat{\boldsymbol{\beta}}^{*} - \ddot{\boldsymbol{\beta}}\right)}{\left(\boldsymbol{a}^{\top}\widehat{\boldsymbol{V}}^{*}_{CGM}\boldsymbol{a}\right)^{1/2}} \rightarrow^{d^*} \mathcal{N}(0,1), \text{ in probability,}$$

Moreover, for any $\varepsilon > 0$,

$$P\left(\sup_{x\in\mathbb{R}}\left|P^{*}(\widehat{t}^{*}_{CGM} \leq x) - P(\widehat{t}_{CGM} \leq x)\right| > \varepsilon\right) \rightarrow 0.$$

The result holds regardless of whether the DGP involves clustering along two dimensions, one dimension, intersections, or no clustering at all. In other words, the MWCB approach demonstrates robustness against different forms of two-way cluster dependence.

Simulation Experiment

We generate data based on the linear model

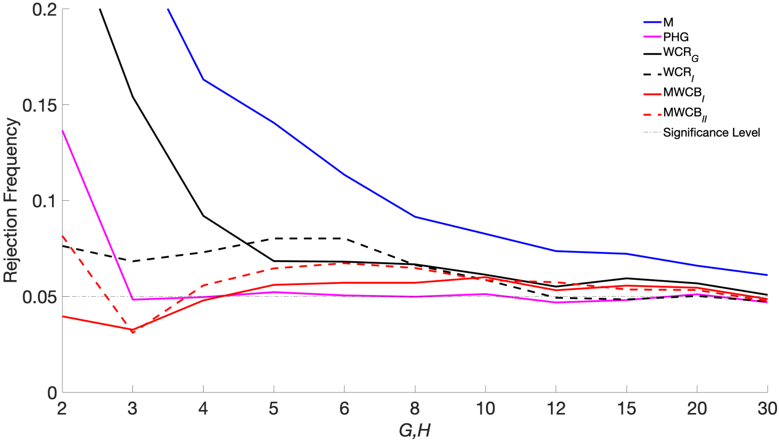
$$y_{gh} = \beta_0 + \beta_1 X_{gh} + u_{gh},$$

where

$$u_{gh} = \omega_\alpha \alpha_g^u + \omega_\xi \xi_h^u + \omega_\varepsilon \varepsilon_{gh}^u, \quad \text{and}$$
$$X_{gh} = \omega_\alpha \alpha_g^x + \omega_\xi \xi_h^x + \omega_\varepsilon \varepsilon_{gh}^x.$$

We set $(\beta_0, \beta_1) = (1, 1)$ and $(\alpha_g^x, \alpha_g^u, \xi_h^x, \xi_h^u, \varepsilon_{gh}^x, \varepsilon_{gh}^u)$ are all mutually independent standard normal random variables. Note that we let each intersection have exact one observation, so that the pigeonhole and Menzel method can be applied directly.

Simulation Experiment



Two-way Clustering with Time Dimension

The MWCB method can be extended to a scenario where clusters are not necessarily independent of each other, extending the concept from the first scenario. In this situation, we consider a prevalent case where the second (H) dimension represents time series data.

As elaborated by Chiang et al. (2024), the intersections (g, h) and (g', h') can exhibit dependence for arbitrary indices under such circumstances. In other words,

$E\left(\boldsymbol{X}_{gh}^{\top}\boldsymbol{u}_{gh}\boldsymbol{u}_{g'h'}^{\top}\boldsymbol{X}_{g'h'}\right) = 0$ may not hold for any combination of g, h, g', h'.

Variance for Two-way Clustering with Time Dimension

Chiang et al. (2024) propose a variance estimator by effectively handling serial dependence in common time effects and providing rigorous proof of asymptotic validity:

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{CHS} = & \sum_{g} \boldsymbol{X}_g^\top \widehat{\boldsymbol{u}}_g \widehat{\boldsymbol{u}}_g^\top \boldsymbol{X}_g + \sum_{h} \boldsymbol{X}_h^\top \widehat{\boldsymbol{u}}_h \widehat{\boldsymbol{u}}_h^\top \boldsymbol{X}_h - \sum_{g} \sum_{h} \boldsymbol{X}_{gh}^\top \widehat{\boldsymbol{u}}_{gh} \widehat{\boldsymbol{u}}_{gh}^\top \boldsymbol{X}_{gh} \\
& + \sum_{\iota=1}^{\ell-1} w(\iota, \ell) \Bigg(\sum_{h=1}^{H-\iota} \boldsymbol{X}_h^\top \widehat{\boldsymbol{u}}_h \widehat{\boldsymbol{u}}_{h+\iota}^\top \boldsymbol{X}_{h+\iota} + \sum_{h=1}^{H-\iota} \boldsymbol{X}_{h+\iota}^\top \widehat{\boldsymbol{u}}_{h+\iota} \widehat{\boldsymbol{u}}_h^\top \boldsymbol{X}_h \\
& - \sum_{h=1}^{H-\iota} \sum_{g=1}^{G} \boldsymbol{X}_{gh}^\top \widehat{\boldsymbol{u}}_{gh} \widehat{\boldsymbol{u}}_{gh+\iota}^\top \boldsymbol{X}_{gh+\iota} - \sum_{h=1}^{H-\iota} \sum_{g=1}^{G} \boldsymbol{X}_{gh+\iota}^\top \widehat{\boldsymbol{u}}_{gh+\iota} \widehat{\boldsymbol{u}}_{gh}^\top \boldsymbol{X}_{gh} \Bigg).
\end{aligned}
$$

However, this scenario is excluded by existing multiway clustered bootstrap inference methods (including MacKinnon et al. (2021), Menzel (2021), and Davezies et al. (2021)).

MWCB$_I$ with Time Dimension

The bootstrap errors (in Step 3 of Bootstrap Algorithm) are obtained as follows:

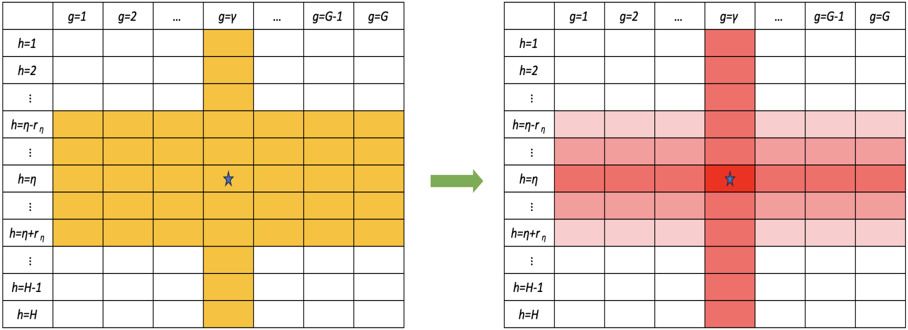
$$\nu^*_{gh} = \frac{1}{\sqrt{G\ell + H - 1}} \sum_{\gamma=1}^{G} \sum_{\eta=2-\ell}^{H} \nu^{*\gamma\eta}_{gh},$$

where given γ, η,

$$\nu^{*\gamma\eta}_{gh} = \begin{cases} 0, & \text{if } g \neq \gamma \text{ and } h \notin [\eta, \eta + \ell - 1], \\ \nu^{*\gamma\eta}, & \text{if } g = \gamma \text{ or } h \in [\eta, \eta + \ell - 1], \end{cases}$$

with $\nu^{*\gamma\eta} \sim (0, 1)$ i.i.d. over γ, η.

MWCB$_I$ with Time Dimension



The "fundamental" random variables $v_{gh}^{*\gamma\eta}$ are perfectly correlated when either $g = \gamma$ or $d(h,\eta) \leq r_\eta$, given γ, η. That is, the intersections in the cluster γ in the first dimension or in the neighborhood of cluster η in the second dimension will be assigned perfectly correlated random variables.

The "aggregated" random variables v_{gh}^* are strongly correlated with $v_{\gamma\eta}^*$ when either $g = \gamma$ or $d(h,\eta) \leq r_\eta$. A darker shade suggests a higher degree of correlation with intersection (γ,η).

Figure: **An illustration of how the random variable is generated under the multiway clustering with time dimension.**

MWCB$_{II}$ with Time Dimension

The second bootstrap random weights are obtained as follows:

$$\nu_{gh}^* = \begin{cases} \nu_g^*, & \text{with probability } p, \\ \nu_h^*, & \text{with probability } 1-p, \end{cases}$$

with ν_g^* independent over g. The key difference is that instead of letting ν_h^* to be independent over h, we now generate ν_h^* to exhibit dependence:

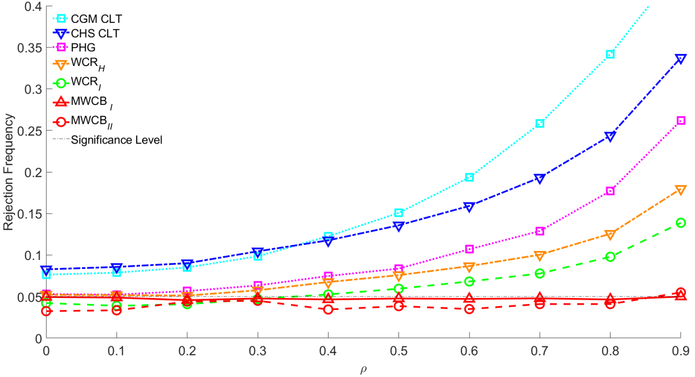
$$\nu_{h+1}^* = \begin{cases} \nu_h^*, & \text{with probability } \frac{1+q}{2}, \\ -\nu_h^*, & \text{with probability } \frac{1-q}{2}, \end{cases}$$

where $q \in (0,1)$, and we let the initial term ν_0^* to follow a Rademacher distribution. Hence, one can deduce that $Cov^*\left(\nu_h^*, \nu_{h+\iota}^*\right) = q^\iota$ for each h. The correlation decays exponentially, reflecting the mixing property, with q representing the mixing coefficients. Therefore, this approach is expected to capture the mixing structure effectively, particularly in the presence of serial dependence.

Simulation Experiment with Time Dimension

Instead of letting ξ_h to be independent over h, we allow it to be dependent by following an AR(1) procedure:

$$\xi_h = \rho\xi_{h-1} + \tilde{\xi}_h, \text{ where } \tilde{\xi}_h \text{ are independent draws from } \mathcal{N}(0, 1-\rho^2).$$



We consider the slave trade example studied by Nunn and Wantchekon (2011) and further studied by MacKinnon et al. (2021). They study the causal impact of the historical slave export on the current levels of trust toward neighbors in Africa:

$$\text{trust}_{iedc} = \boldsymbol{\alpha}_c + \beta \text{ slave exports}_e + \boldsymbol{X}_i^\top \boldsymbol{\phi}_1 + \boldsymbol{X}_d^\top \boldsymbol{\phi}_2 + \boldsymbol{X}_e^\top \boldsymbol{\phi}_3 + \varepsilon_{iedc},$$

where i, e, d, and c denote individual, ethnicity, district, and country, respectively.

Panel A. Cluster by district and ethnicity										
	CLT P value		Bootstrap P value					No. of clusters		
$\widehat{\boldsymbol{\beta}}$	EHW	CGM	WCR_G	WCR_H	WCR_I	MWCB_I	MWCB_{II}	G	H	I
-0.6971	0.0000	0.0000	0.0004	0.0015	0.0009	0.0007	0.0010	1257	185	3225
Panel B. Cluster by country and ethnicity										
	CLT P value		Bootstrap P value					No. of clusters		
$\widehat{\boldsymbol{\beta}}$	EHW	CGM	WCR_G	WCR_H	WCR_I	MWCB_I	MWCB_{II}	G	H	I
-0.6971	0.0000	0.0016	0.0712	0.0496	0.0451	0.0480	0.0460	16	185	223

Table: **Inference result for the effect of slave exports on trust in neighbors.** The first (G) and second (H) dimensions correspond to geography and ethnicity, respectively. I denotes the intersection-level cluster.

Empirical Studies (Cont')

- Panel A shows the result when clustering is by district in the geography dimension and ethnicity, which is suggested by Null and Wantchekon (2011).

- The overall result is consistent with previous findings: all methods indicate that the effect of slave export on trust level is significantly negative at the 1% level.

Empirical Studies (Cont')

- Panel B shows the results when clustering is by country and ethnicity following the recommendation of MacKinnon et al. (2021).
- The CLT result based on $\widehat{\boldsymbol{V}}_{CGM}$ suggests that we can reject the null hypothesis at the 1% level.
- However, all bootstrap methods suggest a different result – the coefficient estimate is no longer significant at the 1% level, but still significant at the 10% level.
- Considering the robustness of MWCB observed in the simulation across various dependence structures and data configurations, we claim that the effect of slave export on the trust level to neighbors is significantly negative at the 5% level, no matter whether the DGP is clustered by district or country.

Conclusions

- We propose and demonstrate the robustness of a new bootstrap procedure for inference in linear regression models with multiway clustering.
- The novel method demonstrates near-optimal performance across various scenarios, even when dealing with a small number of clusters.
- In the scenario of multiway clustering with a time dimension, our method emerges as the most effective approach, adeptly addressing serial dependence.
- The method also shows applicability and ease of implementation in contexts involving spatial dependence or dependence along more than two dimensions.

References

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics 119*(1), 249–275.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Chiang, H. D., B. E. Hansen, and Y. Sasaki (2024). Standard errors for two-way clustering with serially correlated time effects. *Review of Economics and Statistics*.

Davezies, L., X. D'Haultfœuille, and Y. Guyonvarch (2021). Empirical process results for exchangeable arrays. *Annals of Statistics 49*, 845–862.

Djogbenou, A. A., J. G. MacKinnon, and M. Ø. Nielsen (2019). Asymptotic theory and wild bootstrap inference with clustered errors. *Journal of Econometrics 212*(2), 393–412.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2021). Wild bootstrap and asymptotic inference with multiway clustering. *Journal of Business & Economic Statistics 39*(2), 505–519.

Menzel, K. (2021). Bootstrap with cluster-dependence in two or more dimensions. *Econometrica 89*(5), 2143–2188.

Nunn, N. and L. Wantchekon (2011). The slave trade and the origins of mistrust in africa. *American Economic Review 101*(7), 3221–3252.

Petersen, M. A. (2008). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies 22*(1), 435–480.