Wild Bootstrap Inference with Multiway Clustering and Serially Correlated Time Effects*

Ulrich Hounyo† **Jiahao Lin**‡

December 28, 2024

Abstract

This paper studies wild bootstrap-based inference for regression models with multiway clustering. Our proposed methods are multiway counterparts to the (one-way) wild cluster bootstrap approach introduced by Cameron et al. (2008). We establish the validity of our methods for studentized statistics. Theoretical results are provided, accommodating arbitrary serial dependence in the common time effects – an aspect excluded by existing two-way bootstrap-based approaches. Simulation experiments document the potential for enhanced inference with our novel approaches. We illustrate the effectiveness of the methods by revisiting an empirical study involving multiway clustered and serially correlated data.

JEL Classification: C15, C23, C31, C80

Keywords: Bootstrap, clustered standard errors, clustered data, two-way clustering, robust inference, wild cluster bootstrap.

*We are grateful for the helpful comments provided by James MacKinnon, Morten Nielsen, Harold Chiang, Bruce Hansen, and Kaicheng Chen. All the remaining errors are ours.

†Ulrich Hounyo: khounyo@albany.edu. Department of Economics, University at Albany – State University of New York, Albany, NY 12222, United States.

‡Jiahao Lin: jlin28@albany.edu. Department of Economics, University at Albany – State University of New York, Albany, NY 12222, United States.

1 Introduction

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 the assumption of independence, a cornerstone of classical statistical methods, is often untenable in practical empirical work, regardless of whether one is dealing with time series data or cross-sectional data. The failure to account for the intricacies of dependence within the data can have far-reaching consequences, often culminating in the emergence of misleading results. The seminal studies by Bertrand, Duflo, and Mullainathan (2004) and Petersen (2008), drew great attention to this issue, underlining the critical importance of addressing the dependence in data analysis. The presence of temporal or spatial correlations in data can render this assumption invalid. This violation of independence can manifest in the underestimation of standard errors, leading to inflated Type I error rates and, consequently, erroneous inferences regarding the statistical significance of observed effects.

This paper considers the application of wild bootstrap-based inference methods in a linear regression model, where clustering of observations or disturbances (error terms) may occur in two or more dimensions. In various fields including economics, finance, and social sciences, datasets frequently possess intricate patterns where observations are correlated not only within clusters but also across different dimensions. Two-way and more generally multiway clustering techniques recognize these multi-sources of dependence (and heterogeneity), enabling researchers to model the inherent complexities more accurately and to adjust for correlation among observations that arise from multiway dependence. Here are a few examples where multiway cluster-robust inference techniques can be applied in economics and finance to account for the complicated dependence present in the data:

Example 1 – Panel Data Analysis: Economic studies often involve panel data where observations are collected across different entities (countries, firms) and over time (years, months). There may be correlation both within entities across time periods and within time periods across jurisdictions. Two-way clustering allows researchers to control for both time and entity-specific effects simultaneously, providing more accurate estimations in panel data models. See, e.g., Cameron, Gelbach, and Miller (2011), Menzel (2021), Chiang, Hansen, and Sasaki (2024), and Chen and

Vogelsang (2024).

Example 2 – International Trade: Trade data involve various dimensions, such as countries, industries, and time periods. Multiway clustering methods can be used to analyze trade patterns, identify comparative advantages, and assess the impact of trade policies on different sectors and regions. See, e.g., Fieler (2011) and Korovkin and Makarin (2023).

Example 3 – Labor Economics: In labor data, individuals are often nested within various grouping variables, such as firms, households, industry sectors, regions, or job types, which can create various clusters and correlations within these groups. Multiway cluster-robust inference allows for potentially diverse populations with varying characteristics and leads to accurate modeling of the complex data dependence. See, e.g., Bloom et al. (2015) and Card, Cardoso, and Kline (2016).

There are many other situation-specific cases, where the application of multiway cluster-robust inference is useful in economics, see among others, Cameron et al. (2011). Two-way clustering has been introduced in economics and finance independently by Cameron et al. (2011) and Thompson (2011). Davezies, D'Haultfœuille, and Guyonvarch (2021) and Chiang et al. (2024), to name a few, study the asymptotic theory of multiway clustering, whereas MacKinnon, Nielsen, and Webb (2021) and Menzel (2021) derive asymptotic validity of bootstrap methods for multiway clustered errors. Recently, Chiang et al. (2024) propose improved standard errors and an asymptotic distribution theory for two-way clustered panels, under general conditions allowing for arbitrary serial dependence in the common time effects. For instance, in a panel data setting, serial correlation in the common time effects can create an extra layer of serial dependence over the two-way clustering dependence. This induces dependence among observations that do not share a common entity or time index. This form of dependence is excluded by existing two-way methods, including the popular two-way cluster standard errors of Cameron et al. (2011) and the cluster bootstrap procedures of MacKinnon et al. (2021), Menzel (2021), and Davezies et al. (2021).

In practice, when clustering occurs in just one dimension, it is now standard to combine cluster-robust variance estimator, or CRVE with the so-called wild cluster bootstrap method introduced by Cameron, Gelbach, and Miller (2008), see, e.g., Djogbenou, MacKinnon, and Nielsen

(2019), MacKinnon (2023), and references therein. This is because CRVE-based t-statistics can fail (sometimes disastrously) in finite samples. Similarly, multiway cluster-robust asymptotic-based inferences may not be at all reliable in finite samples. Chen and Vogelsang (2024) investigate the fixed-b asymptotic properties of the Chiang et al. (2024) variance estimator and tests, and propose bias-corrected variance estimators along with fixed-b critical values. In this paper, to improve the performance tests based on the Chiang et al. (2024) asymptotic-based approach, we propose two multiway wild cluster bootstrap-based methods (MWCBs). We shed new light on the literature of bootstrapping multiway clustering by proposing the novel MWCB approaches and formally showing their theoretical validity under conditions like Chiang et al. (2024), allowing time effects with arbitrary serial correlation, which is excluded by existing multiway clustered bootstrap inference methods (including MacKinnon et al. (2021), Menzel (2021), and Davezies et al. (2021)).

While our new methods, the MWCBs, introduce innovative perspectives, they are firmly grounded in prior research, drawing inspiration from works such as Cameron et al. (2008), Menzel (2021), MacKinnon et al. (2021), Hansen (1992), Shao (2010), Hounyo (2023), Chiang, Kato, and Sasaki (2023), and various others. Menzel (2021) proposes a hybrid bootstrap scheme, combining the wild bootstrap and nonparametric i.i.d. bootstrap. Similarly, MacKinnon et al. (2021) suggest various versions of the wild cluster bootstrap, clustering bootstrap sample along a specific dimension or by intersections. Our MWCB methods, similar to those developed by Cameron et al. (2008) and MacKinnon et al. (2021), are based on wild cluster bootstrapping. Moreover, the MWCB_{II} method can be considered as a linear combination of two variants (WCR_G and WCR_H) from MacKinnon et al. (2021) that perform optimally in scenarios where cluster dependence is present in only one dimension.

Our approaches relate to a specific variant (WCR_I) in MacKinnon et al. (2021), which generates bootstrap sample by intersections. They perturb the data within the same intersection by multiplying them with a random weight, where the weight is independent across intersections. This preserves dependence within each intersection while destroying potential dependence between any two intersections.

We propose to bootstrap sample by intersections, rather than a specific dimension. The rationale is as follows: when the bootstrap procedure accurately mimics the dependence properties of sample at the intersection level, it enables us to effortlessly recover the corresponding dependence properties, whether the DGP is clustered along any dimension, by intersections, or not clustered at all. This choice mirrors the relationship between joint and marginal distributions, where we can deduce marginal distributions from the joint distribution, but the reverse is not always true.

While both our methods and MacKinnon et al. (2021)'s use a wild cluster bootstrap procedure and obtain bootstrap sample by intersections, a fundamental difference lies in the construction of the random weight. For example, in the (simple and conventional) multiway clustering setting, where two intersections are considered asymptotically dependent only if they share any cluster in any dimension, in our MWCBs, the random weight associated with a given intersection is carefully generated to exhibit a correlation with that of another intersection, whenever both intersections share any cluster in any dimension. This design allows us to preserve potential dependence between any two intersections.

In the broadest scenario, where intersections are assumed to be asymptotically dependent even when they lack shared clusters in any dimension, the random weights of our method are generated to maintain potential dependence whenever they are close enough to each other. This ensures that the method accounts for possible dependence even in the absence of shared clusters.

Following MacKinnon et al. (2021) and Chiang et al. (2024), we state conditions that are sufficient for asymptotic normality and our bootstrap methods are proposed for this setting. We first provide general high-level conditions that imply the validity of the proposed bootstrap approach. We then verify these high-level conditions in two main types of clustering settings.

While our simulation results show clear finite sample superiority of the proposed wild bootstrap method over asymptotic theory-based inference and existing multiway clustered bootstrap methods (particularly when dependence exists along more than one dimension), we do not claim theoretical advantages in the conventional multiway clustering context. Instead, these advantages become more relevant in a multiway clustering with a time dimension, where existing methods lack validity. We further conduct an empirical study on the impact of working from home on employees' performance. Our findings reveal a significant improvement in the total number of phones answered per week. This increase in performance is primarily attributed to a rise in the total time employees spend on the phone each week, rather than an improvement in the number of calls answered per minute.

The rest of the paper is organized as follows. Section 2 introduces the two-way clustering model and two key kinds of two-way clustering. It also presents the existing variance estimators in both scenarios. Section 3 explains the novel multiway clustering bootstrap methods and provides proof of bootstrap validity. Section 4 presents the outcomes of various simulation experiments, comparing results obtained through the new method and other approaches. In Section 5, we elucidate our findings further with an empirical example. Section 6 concludes. All proofs are given in the appendix.

2 The Model and Variance Estimators

2.1 Two-way Clustering Regression Model

We focus on the linear regression model with two-way clustering written as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \tag{2.1}$$

where $\boldsymbol{y}$ and $\boldsymbol{u}$ are $N \times 1$ vectors containing observations and disturbances, respectively. $\boldsymbol{X}$ is an $N \times k$ matrix comprising covariates, and $\boldsymbol{\beta}$ is a k-vector of coefficients. As usual, the ordinary least squares (OLS) estimator $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}$. The model assumes two dimensions of clustering, with G clusters in the first dimension and H clusters in the second dimension. We can rewrite equation (2.1) as

$$\boldsymbol{y}_{gh} = \boldsymbol{X}_{gh}\boldsymbol{\beta} + \boldsymbol{u}_{gh}, \quad g = 1, \ldots, G, \quad h = 1, \ldots, H, \tag{2.2}$$

where $\boldsymbol{y}_{gh}$ and $\boldsymbol{u}_{gh}$, as well as the matrix $\boldsymbol{X}_{gh}$, consist of the rows from $\boldsymbol{y}$, $\boldsymbol{u}$, and $\boldsymbol{X}$ that correspond to both the g^{th} cluster in the first dimension and the h^{th} cluster in the second dimension. The GH clusters mentioned in (2.2) represent the intersections of the two clustering dimensions.

To denote the number of observations in each cluster for each dimension, we use N_g to represent the number of observations in cluster g in the first dimension and N_h for cluster h in the second dimension, along with N_{gh} to represent the number of observations in the intersection of cluster g in the first dimension and cluster h in the second dimension. The vector $\boldsymbol{y}_{gh} \equiv \left(y_{gh,1}, y_{gh,2}, ..., y_{gh,N_{gh}}\right)^\top$ can be empty if the g^{th} cluster in the first dimension has no intersection with the h^{th} cluster in the second dimension. Moreover, we let $\boldsymbol{X}_{gh,i}$ and $u_{gh,i}$ denote the i^{th} rows of $\boldsymbol{X}_{gh}$ and $\boldsymbol{u}_{gh}$, respectively.

Similarly, we use vectors $\boldsymbol{y}_g$ and $\boldsymbol{u}_g$ and the matrix $\boldsymbol{X}_g$ to represent vectors containing the rows of $\boldsymbol{y}$, $\boldsymbol{u}$, and $\boldsymbol{X}$ for the g^{th} cluster in the first dimension, and $\boldsymbol{y}_h$, $\boldsymbol{u}_h$, and $\boldsymbol{X}_h$ for rows of the h^{th} cluster in the second dimension. In terms of the notation used in (2.2), $\boldsymbol{y}_g$ contains the subvectors $\boldsymbol{y}_{g1}$ through $\boldsymbol{y}_{gH}$.

For motivational purposes only, it is noteworthy that, under certain regularity conditions, the variance matrix of $\widehat{\boldsymbol{\beta}}$ can be expressed as: $\boldsymbol{V} \equiv Var(\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{Q}}^{-1}\boldsymbol{\Gamma}\widehat{\boldsymbol{Q}}^{-1}$, where $\widehat{\boldsymbol{Q}} \equiv \frac{1}{GH}\sum_g \sum_h \boldsymbol{X}_{gh}^\top \boldsymbol{X}_{gh}$, $\boldsymbol{\Gamma} \equiv \frac{1}{(GH)^2}\boldsymbol{\Sigma}$, and $\boldsymbol{\Sigma}$ is the variance matrix of the scores

$$\boldsymbol{\Sigma} \equiv E\left(\boldsymbol{X}^\top \boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{X}\right) = \sum_g \sum_{g'} \sum_h \sum_{h'} E\left(\boldsymbol{X}_{gh}^\top \boldsymbol{u}_{gh}\boldsymbol{u}_{g'h'}^\top \boldsymbol{X}_{g'h'}\right).$$

The existing literature examines various dependence scenarios, incorporating different assumptions about the variance matrix $\boldsymbol{\Sigma}$. These scenarios encompass cases such as the absence of clustering (see, e.g., White (1980)). Additionally, one-way clustering dependence is explored in works like Liang and Zeger (1986) and Arellano (1987), while one-way cluster dependence across individuals with weak serial dependence across time is investigated by Driscoll and Kraay (1998). More recently, Cameron et al. (2011) and Thompson (2011) have expanded on these assumptions by permitting more than one-way cluster dependence, leading to the development of two main forms of two-way clustering.

2.2 Two Main Types of Two-way Clustering

We categorize the two-way cluster setting into primarily two types. The first is the standard two-way clustering where two intersections are (asymptotically) independent if they do not share any cluster in any dimension. As discussed in Cameron et al. (2011) (CGM hereafter), MacKinnon et al. (2021), Menzel (2021), and Davezies et al. (2021), we have that

$$E\left(\boldsymbol{X}_{gh}^{\top}\boldsymbol{u}_{gh}\boldsymbol{u}_{g'h'}^{\top}\boldsymbol{X}_{g'h'}\right) = 0, \text{ for } g' \neq g \text{ and } h' \neq h. \tag{2.3}$$

Therefore, $\boldsymbol{\Sigma}$ can be rewrite as

$$\boldsymbol{\Sigma} = \sum_{g} E\left(\boldsymbol{X}_{g}^{\top}\boldsymbol{u}_{g}\boldsymbol{u}_{g}^{\top}\boldsymbol{X}_{g}\right) + \sum_{h} E\left(\boldsymbol{X}_{h}^{\top}\boldsymbol{u}_{h}\boldsymbol{u}_{h}^{\top}\boldsymbol{X}_{h}\right) - \sum_{g}\sum_{h} E\left(\boldsymbol{X}_{gh}^{\top}\boldsymbol{u}_{gh}\boldsymbol{u}_{gh}^{\top}\boldsymbol{X}_{gh}\right). \tag{2.4}$$

The second scenario introduces a condition 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. In instances where the time series dependence gradually diminishes, a cluster may manifest a substantial shared influence with other clusters within a specific time period. This phenomenon is often referred to as a natural order, contingent on a bandwidth parameter, denoted as ℓ.

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

$$E\left(\boldsymbol{X}_{gh}^{\top}\boldsymbol{u}_{gh}\boldsymbol{u}_{g'h'}^{\top}\boldsymbol{X}_{g'h'}\right) = 0 \text{ may not hold for any combination of } g, h, g', h'. \tag{2.5}$$

Although each term of $\boldsymbol{\Sigma}$ is not necessarily zero, we are still able to decompose them in a more intuitive form:

$$\boldsymbol{\Sigma} = \sum_{g} E\left(\boldsymbol{X}_{g}^{\top}\boldsymbol{u}_{g}\boldsymbol{u}_{g}^{\top}\boldsymbol{X}_{g}\right) + \sum_{h} E\left(\boldsymbol{X}_{h}^{\top}\boldsymbol{u}_{h}\boldsymbol{u}_{h}^{\top}\boldsymbol{X}_{h}\right) - \sum_{g}\sum_{h} E\left(\boldsymbol{X}_{gh}^{\top}\boldsymbol{u}_{gh}\boldsymbol{u}_{gh}^{\top}\boldsymbol{X}_{gh}\right) \tag{2.6}$$

$$+ \sum_{\iota=1}^{H-1}\left[\sum_{h=1}^{H-\iota} E\left(\boldsymbol{X}_{h}^{\top}\boldsymbol{u}_{h}\boldsymbol{u}_{h+\iota}^{\top}\boldsymbol{X}_{h+\iota}\right) + \sum_{h=1}^{H-\iota} E\left(\boldsymbol{X}_{h+\iota}^{\top}\boldsymbol{u}_{h+\iota}\boldsymbol{u}_{h}^{\top}\boldsymbol{X}_{h}\right)\right. \tag{2.7}$$

$$\left. - \sum_{h=1}^{H-\iota}\sum_{g=1}^{G} E\left(\boldsymbol{X}_{gh}^{\top}\boldsymbol{u}_{gh}\boldsymbol{u}_{gh+\iota}^{\top}\boldsymbol{X}_{gh+\iota}\right) - \sum_{h=1}^{H-\iota}\sum_{g=1}^{G} E\left(\boldsymbol{X}_{gh+\iota}^{\top}\boldsymbol{u}_{gh+\iota}\boldsymbol{u}_{gh}^{\top}\boldsymbol{X}_{gh}\right)\right]. \tag{2.8}$$

The decomposition represents two-way clustering with arbitrary stationary serial dependence. The term (2.6) equals (2.4), the term (2.7) represents the autocovariances of the time sums, and the term (2.8) corrects the double-counting.

2.3 Cluster Robust Variance Estimators

Several variance estimators exist for the matrix $\boldsymbol{V}$, all sharing a common structure denoted as $\widehat{\boldsymbol{V}}_m = \widehat{\boldsymbol{Q}}^{-1}\widehat{\boldsymbol{\Gamma}}_m\widehat{\boldsymbol{Q}}^{-1}$, where $\widehat{\boldsymbol{\Gamma}}_m \equiv \frac{1}{(GH)^2}\widehat{\boldsymbol{\Sigma}}_m$. In this context, the pivotal element is $\widehat{\boldsymbol{\Sigma}}_m$, serving as the estimation for the variance of scores. For the standard multiway clustering scenario, where two intersections are considered independent if no shared clusters, i.e., (2.3) holds, CGM 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

$$\widehat{\boldsymbol{\Sigma}}_{CGM} \;=\; \sum_g \widehat{\boldsymbol{s}}_g\widehat{\boldsymbol{s}}_g^\top + \sum_h \widehat{\boldsymbol{s}}_h\widehat{\boldsymbol{s}}_h^\top - \sum_g\sum_h \widehat{\boldsymbol{s}}_{gh}\widehat{\boldsymbol{s}}_{gh}^\top, \tag{2.9}$$

where the empirical score $\widehat{\boldsymbol{s}}_\bullet = \boldsymbol{X}_\bullet^\top\widehat{\boldsymbol{u}}_\bullet$. The three terms in (2.9) have similar forms as (2.4). In practice, let I denote the number of nonempty intersections, and the three terms are usually multiplied by the corresponding scalar factors

$$\frac{G(N-1)}{(G-1)(N-k)},\ \frac{H(N-1)}{(H-1)(N-k)},\ \text{and}\ \frac{I(N-1)}{(I-1)(N-k)}, \tag{2.10}$$

respectively, as suggested by Cameron et al. (2011) and MacKinnon et al. (2021).

However, $\widehat{\boldsymbol{V}}_{CGM}$ is not guaranteed to be positive definite in practice, which means one may not be able to make inferences relying on $\widehat{\boldsymbol{V}}_{CGM}$.[1] That is because $\widehat{\boldsymbol{\Sigma}}_{CGM}$ deduct the double-counting term $\sum_g\sum_h \widehat{\boldsymbol{s}}_{gh}\widehat{\boldsymbol{s}}_{gh}^\top$, which ensures consistency when there is no cluster dependence. Recently, Davezies et al. (2021) propose an asymptotically equivalent variant under specific conditions which guarantees positive definiteness by adding back the double-counting term:

$$\widehat{\boldsymbol{\Sigma}}_{DHG} \;=\; \sum_g \widehat{\boldsymbol{s}}_g\widehat{\boldsymbol{s}}_g^\top + \sum_h \widehat{\boldsymbol{s}}_h\widehat{\boldsymbol{s}}_h^\top. \tag{2.11}$$

MacKinnon et al. (2021) further formally prove the asymptotic validity of these two cluster-robust variance estimators when the cluster dependence exists.

[1]If the matrix lacks positive semidefiniteness, its eigenvalues will encompass negative values. In their work, CGM suggest substituting these negative values with zeros. We adopt this approach in our simulation, constructing the new variance estimator using the updated non-negative eigenvalues and their corresponding eigenvectors.

In the setting of the second multiway clustering with serially correlated time effects where (2.5) holds, Thompson (2011) introduces a variance estimator robust to two-way clustering with potential temporal dependence. CHS further enhance Thompson (2011)'s variance estimator by effectively handling serial dependence in common time effects and providing rigorous proof of asymptotic validity:

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{CHS} = & \sum_{g} \widehat{\boldsymbol{s}}_g \widehat{\boldsymbol{s}}_g^{\top} + \sum_{h} \widehat{\boldsymbol{s}}_h \widehat{\boldsymbol{s}}_h^{\top} - \sum_{g} \sum_{h} \widehat{\boldsymbol{s}}_{gh} \widehat{\boldsymbol{s}}_{gh}^{\top} & (2.12) \\
& + \sum_{\iota=1}^{\ell-1} w(\iota, \ell) \left(\sum_{h=1}^{H-\iota} \widehat{\boldsymbol{s}}_h \widehat{\boldsymbol{s}}_{h+\iota}^{\top} + \sum_{h=1}^{H-\iota} \widehat{\boldsymbol{s}}_{h+\iota} \widehat{\boldsymbol{s}}_h^{\top} \right. & (2.13) \\
& \left. - \sum_{h=1}^{H-\iota} \sum_{g=1}^{G} \widehat{\boldsymbol{s}}_{gh} \widehat{\boldsymbol{s}}_{gh+\iota}^{\top} - \sum_{h=1}^{H-\iota} \sum_{g=1}^{G} \widehat{\boldsymbol{s}}_{gh+\iota} \widehat{\boldsymbol{s}}_{gh}^{\top} \right). & (2.14)
\end{aligned}
$$

In the work of CHS, the index h is represented as t, which serves as the time series index. $w(\iota, \ell)$ is the weight function, and we choose the triangle weight $w(\iota, \ell) = 1 - \iota$ suggested by Newey and West (1986) and Chiang et al. (2024). The form of $\widehat{\boldsymbol{\Sigma}}_{CHS}$ is similar as $\boldsymbol{\Sigma}$ as shown in (2.6)-(2.8). The term (2.12) is the same as $\widehat{\boldsymbol{\Sigma}}_{CGM}$, which captures the dependence in the cluster g in the first dimension and cluster h in the second temporal dimension. While the term (2.13) estimates the autocovariances of the time sums. Likewise, $\widehat{\boldsymbol{\Sigma}}_{CHS}$ subtracts the double-counting terms $\sum_g \sum_h \widehat{\boldsymbol{s}}_{gh} \widehat{\boldsymbol{s}}_{gh}^{\top}$ and (2.14), potentially introducing a practical concern regarding the lack of a guarantee of positive definiteness.[2] In their recent work, Chen and Vogelsang (2024) suggest adding back the double-counting terms to address this issue:

$$
\widehat{\boldsymbol{\Sigma}}_{CV} = \sum_{g} \widehat{\boldsymbol{s}}_g \widehat{\boldsymbol{s}}_g^{\top} + \sum_{h} \widehat{\boldsymbol{s}}_h \widehat{\boldsymbol{s}}_h^{\top} + \sum_{\iota=1}^{\ell-1} w(\iota, \ell) \sum_{h=1}^{H-\iota} \left(\widehat{\boldsymbol{s}}_h \widehat{\boldsymbol{s}}_{h+\iota}^{\top} + \widehat{\boldsymbol{s}}_{h+\iota} \widehat{\boldsymbol{s}}_h^{\top} \right). \tag{2.15}
$$

Our null hypothesis is $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, and we define the t-statistic $\widehat{t}_m = \frac{\boldsymbol{a}^{\top}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)}{\sqrt{\boldsymbol{a}^{\top} \widehat{\boldsymbol{V}}_m \boldsymbol{a}}}$, $m \in \{CGM, DHG, CHS, CV\}$. Here $\boldsymbol{a}$ is a known unit vector such that $\boldsymbol{a}^{\top}\boldsymbol{a} = 1$. If (2.3) holds, under some regularity assump-

[2]Similarly, CHS propose substituting any negative eigenvalue by zero to ensure positive semidefiniteness.

tions, MacKinnon et al. (2021) show that

$$\widehat{t}_m \to^d \mathcal{N}(0, \sigma_m^2), \text{ for } m \in \{CGM, DHG\}, \tag{2.16}$$

where $\sigma^2_{CGM} = 1$ and $\sigma^2_{DHG} \in \{1, \frac{1}{2}\}$.

If (2.5) holds with one dimension being time series, under some regularity assumptions, following CHS, we have

$$\widehat{t}_m \to^d \mathcal{N}(0, \sigma_m^2), \text{ for } m \in \{CHS, CV\}, \tag{2.17}$$

where $\sigma^2_{CHS} = 1$ and $\sigma^2_{CV} \in \{1, \frac{1}{2}\}$. Specifically, if the DGP is clustered along at least one dimension, then σ^2_{DHG} and σ^2_{CV} equal 1; if the DGP is clustered only by intersections or no clustering at all, then σ^2_{DHG} and σ^2_{CV} equal $\frac{1}{2}$. Note that under the multiway clustering setting, these two scenarios are mutually exclusive but not complementary. Menzel (2021) discusses the degenerate case where $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$ may converge to a non-normal limit, making the usual t-test invalid, and proposes a general bootstrap method that adapts to this situation. Since our bootstrap procedure relies on existing variance estimators, such cases fall beyond the scope of this paper.

3 Bootstrap Inference

The bootstrap method is known to have the potential to improve the asymptotic results. MacKinnon et al. (2021), Menzel (2021), and Davezies et al. (2021) provide several bootstrap methods that have promising simulation performance under the multiway clustering setting. We now provide the algorithm procedure for a standard bootstrap method.

3.1 Bootstrap Algorithm

Algorithm 1. Standard Bootstrap Algorithm

Step 1: Regress $\boldsymbol{y}$ on $\boldsymbol{X}$ to obtain the regression estimate $\widehat{\boldsymbol{\beta}}$, the residual $\widehat{u}_{gh,i}$, the empirical score $\widehat{\boldsymbol{s}}_{gh,i} = \boldsymbol{X}^\top_{gh,i}\widehat{u}_{gh,i}$, and the cluster-robust variance estimate $\widehat{\boldsymbol{V}}_m$. For restricted bootstrap, we perform regression once more to obtain the restricted estimate $\widetilde{\boldsymbol{\beta}}$, the restricted residual

$\widetilde{u}_{gh,i}$, and the restricted empirical score $\widetilde{\boldsymbol{s}}_{gh,i} = \boldsymbol{X}_{gh,i}^\top \widetilde{u}_{gh,i}$. Let $\ddot{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}$ and $\ddot{\boldsymbol{s}}_{gh,i} = \widehat{\boldsymbol{s}}_{gh,i}$ for non-restricted bootstrap, while $\ddot{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}$ and $\ddot{\boldsymbol{s}}_{gh,i} = \widetilde{\boldsymbol{s}}_{gh,i}$ for restricted bootstrap.

Step 2: Construct the original cluster-robust t-statistic $\widehat{t}_m = \frac{\boldsymbol{a}^\top (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)}{\sqrt{\boldsymbol{a}^\top \widehat{\boldsymbol{V}}_m \boldsymbol{a}}}$ under the null hypothesis $\mathcal{H}_0 : \boldsymbol{a}^\top \boldsymbol{\beta} = \boldsymbol{a}^\top \boldsymbol{\beta}_0$.

Step 3: Perturb the empirical score $\ddot{\boldsymbol{s}}_{gh,i}$ to generate the bootstrap score $\boldsymbol{s}_{gh,i}^{*b}$, where b is an index for the bootstrap number. Different bootstrap methods amount to different ways of perturbing the empirical score, which is the key difference. The general wild bootstrap generates bootstrap scores as: $\boldsymbol{s}_{gh,i}^{*b} = \ddot{\boldsymbol{s}}_{gh,i} \nu_{gh,i}^{*b}$.

Step 4: Compute the bootstrap cluster-robust t-statistic: $\widehat{t}_m^{*b} = \frac{\boldsymbol{a}^\top (\widehat{\boldsymbol{\beta}}^{*b} - \ddot{\boldsymbol{\beta}})}{\sqrt{\boldsymbol{a}^\top \widehat{\boldsymbol{V}}_m^{*b} \boldsymbol{a}}}$, where in the numerator $\widehat{\boldsymbol{\beta}}^{*b} - \ddot{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \sum_g \sum_h \boldsymbol{s}_{gh}^{*b}$, with $\boldsymbol{s}_{gh}^{*b} = \sum_{i=1}^{N_{gh}} \boldsymbol{s}_{gh,i}^{*b}$. Unless otherwise specified, $\widehat{\boldsymbol{V}}_m^{*b}$ is computed in the same way as $\widehat{\boldsymbol{V}}_m$, except that $\widehat{\boldsymbol{s}}_{\bullet}^{*b}$ is used in place of $\widehat{\boldsymbol{s}}_{\bullet}$ and $\widehat{\boldsymbol{s}}_g^{*b} = \sum_h \widehat{\boldsymbol{s}}_{gh}^{*b}$, $\widehat{\boldsymbol{s}}_h^{*b} = \sum_g \widehat{\boldsymbol{s}}_{gh}^{*b}$, with $\widehat{\boldsymbol{s}}_{gh}^{*b} = \boldsymbol{s}_{gh}^{*b} - \boldsymbol{X}_{gh}^\top \boldsymbol{X}_{gh} \left(\widehat{\boldsymbol{\beta}}^{*b} - \ddot{\boldsymbol{\beta}}\right)$.[3]

Step 5: Repeat Step 3 to Step 4 for B times. When focusing on the one-sided alternative hypotheses $\mathrm{H_L} : \boldsymbol{a}^\top \boldsymbol{\beta} < \boldsymbol{a}^\top \boldsymbol{\beta}_0$ and $\mathrm{H_R} : \boldsymbol{a}^\top \boldsymbol{\beta} > \boldsymbol{a}^\top \boldsymbol{\beta}_0$, calculate the bootstrap P values for the left and right tails as follows:

$$P_{\mathrm{L}}^* = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\widehat{t}_m^{*b} < \widehat{t}_m\right) \text{ and } P_{\mathrm{R}}^* = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\widehat{t}_m^{*b} > \widehat{t}_m\right).$$

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:

$$P_{\mathrm{S}}^* = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(|\widehat{t}_m^{*b}| > |\widehat{t}_m|\right) \text{ and } P_{\mathrm{E}}^* = 2 \min\left(P_{\mathrm{L}}^*, P_{\mathrm{R}}^*\right),$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. Reject the null hypothesis if the bootstrap P value is smaller than the predetermined significance level.[4]

[3] The derivation of the formula for $\widehat{\boldsymbol{s}}_{gh}^{*b}$ follows the residual bootstrap approach.

[4] Throughout the paper, we present our findings using P_E^* without explicit mention, following the recommendation of MacKinnon (2012).

In a recent work, within a one-way clustering setting, MacKinnon, Nielsen, and Webb (2023b) introduced an innovative bootstrap algorithm that involves bootstrapping the transformed score vectors, referred to as WCR-S in their paper. This approach demonstrates improved performance compared to traditional wild cluster restricted bootstrap in specific experiments.[5]

MacKinnon et al. (2021) provide three different wild cluster bootstrap approaches, all of which follow Algorithm 1 and are asymptotically valid based on $\widehat{\boldsymbol{V}}_{CGM}$. We follow MacKinnon et al. (2021) and call these methods WCR$_G$, WCR$_H$, and WCR$_I$.[6] The key difference among these methods is how the random variable $\nu^*_{gh,i}$ in Step 3 of Algorithm 1 is correlated across observations within the same cluster or in different clusters. WCR$_G$ bootstraps along the first (G) dimension, so $\nu^*_{gh,i}$ will be the same for all observations in the same cluster in the first (G) dimension: $\nu^*_{gh,i} = \nu^*_g$, where ν^*_g is i.i.d. over each cluster g; WCR$_H$, on the contrary, lets $\nu^*_{gh,i} = \nu^*_h$, where ν^*_h is i.i.d. over each cluster h. WCR$_I$ bootstraps by intersections, that is, $\nu^*_{gh,i}$ will be the same for all observations in the same intersection and independent across each intersection (g, h): $\nu^*_{gh,i} = \nu^*_{gh}$.

Nonetheless, as mentioned in MacKinnon et al. (2021), when the DGP is clustered along both dimensions, all three methods are not able to mimic the DGP. Although keeping the dependence in the same cluster along one dimension, WCR$_G$ and WCR$_H$ will destroy the cluster dependence along another dimension. For example, for two intersections (g, h) and (g', h) in the same h^{th} cluster in the second dimension, but in different clusters in the first dimension, WCR$_G$ will destroy the

[5]Although, in our multiway clustering setting, there is the potential to enhance existing restricted bootstrap methods by transforming residuals or scores before generating bootstrap statistics, introducing such extensions or transformations would significantly complicate the exposition and divert attention from the main ideas. Therefore, we present the discovery relying on Algorithm 1 throughout the paper.

[6]All three methods are based on the restricted scores. While the non-restricted versions are also valid, we focus exclusively on reporting results for the restricted versions in this paper due to their similarity and superior finite-sample performance. For a thorough discussion of both methods, please refer to Djogbenou et al. (2019), MacKinnon et al. (2023), and MacKinnon, Nielsen, and Webb (2023a).

potential dependence in the second dimension between these two intersections. Furthermore, when considering two intersections (g, h) and (g, h'), WCR_G upholds dependence in both dimensions by assigning the same external variable to both, even if independence exists along the second dimension. This results in an "artificial" dependence in the second dimension. In contrast, WCR_I assigns independent external variables to each intersection, and consequently, the dependence within each intersection (g, h) will persist, while the dependence between any two intersections will be eliminated, regardless of whether they belong to the same cluster or not.

Davezies et al. (2021) establish the asymptotic properties of the pigeonhole bootstrap, originally introduced by McCullagh (2000) and subsequently examined by Owen (2007) and Menzel (2021), within the framework of multiway clustering. The pigeonhole bootstrap resamples cluster index i.i.d. for each dimension separately, allowing it to preserve the dependence on both dimensions simultaneously. Their Monte Carlo simulations indicate that the pigeonhole bootstrap performs well even with a minimal number of clusters. Menzel (2021) shows that the pigeonhole bootstrap is conservative in the absence of clustering, and further introduces a novel bootstrap method that involves decomposing the estimated score into three parts, projection along two dimensions. Menzel (2021) further argues that this method can offer refinement under certain conditions.

Despite these advancements, 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. Notably, all the mentioned methods are designed exclusively for standard multiway clustering settings where (2.3) holds. When two intersections are no longer (asymptotically) independent, even if they do not share any cluster in any dimension, current bootstrap methods, to the best of our knowledge, fail to faithfully reproduce the underlying DGP, consequently limiting their ability to deliver promising outcomes.

These challenges motivate our two MWCB methods, designed to preserve multi-dimensional dependence simultaneously, minimize the introduction of excessive "artificial" dependence, and remain robust to unbalanced data structures. Additionally, our methods can be extended to scenarios of multiway clustering where dependence among intersections that do not share any cluster is allowed.

3.2 Two Multiway Wild Cluster Bootstraps

3.2.1 Standard Multiway Clustering

In this subsection, we introduce and discuss two MWCB methods under the standard multiway clustering (i.e., (2.3) holds).[7] The bootstrap methods follow the standard procedure introduced in Algorithm 1, with the main difference being the way we perturb the empirical scores to generate the bootstrapped scores in Step 3 of Algorithm 1. In this paper, and as usual in the bootstrap literature, P^* (E^*, Var^*, and $corr^*$) denotes the probability measure (expected value, variance, and correlation) induced by the bootstrap resampling, conditional on the original sample. We define $\boldsymbol{V}^* \equiv Var^*\left(\widehat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}}\right)$.

Our bootstraps perturb the scores by intersection, and hence the bootstrap scores (in Step 3 of Algorithm 1) are obtained as follows:

$$\boldsymbol{s}^*_{gh,i} = \ddot{\boldsymbol{s}}_{gh,i}\nu^*_{gh,i} = \ddot{\boldsymbol{s}}_{gh,i}\nu^*_{gh}.$$

MWCB$_I$:

The first bootstrap generates the random weight ν^*_{gh} as:

$$\nu^*_{gh} \equiv \frac{1}{\sqrt{G+H-1}}\sum_{\gamma=1}^{G}\sum_{\eta=1}^{H}\nu^{*\gamma\eta}_{gh}, \tag{3.1}$$

[7]We present our method and prove the bootstrap validity in the context of a two-way cluster setting, a choice supported by MacKinnon et al. (2021). They assert that in empirical research, the utilization of multiway clustering seldom extends beyond two dimensions. However, we provide the generalized bootstrap version which can be applied to more than two dimensions in (D.2) in Appendix D.

where given γ, η,

$$\nu_{gh}^{*\gamma\eta} = \begin{cases} 0, & \text{if } g \neq \gamma \text{ and } h \neq \eta, \\ \chi_1 \nu^{*\gamma\eta}, & \text{if } g = \gamma, \\ \chi_2 \nu^{*\gamma\eta}, & \text{if } g \neq \gamma \text{ and } h = \eta, \end{cases}$$

with $\nu^{*\gamma\eta} \sim (0, \sigma_\nu^2)$ i.i.d. over γ, η.

The scaling terms χ_1 and χ_2 are designed to balance the "level" of (co)variation of scores in both dimensions simultaneously. For self-normalized and studentized statistics, both χ_1 and χ_2 can be set to any nonzero finite constant, e.g., $\chi_1 = \chi_2 = 1$. However, it will be shown that when the following conditions are met:

$$\chi_1 = \sqrt{1 + \frac{G}{H}} \quad \text{and} \quad \chi_2 = \sqrt{1 + \frac{H}{G}}, \tag{3.2}$$

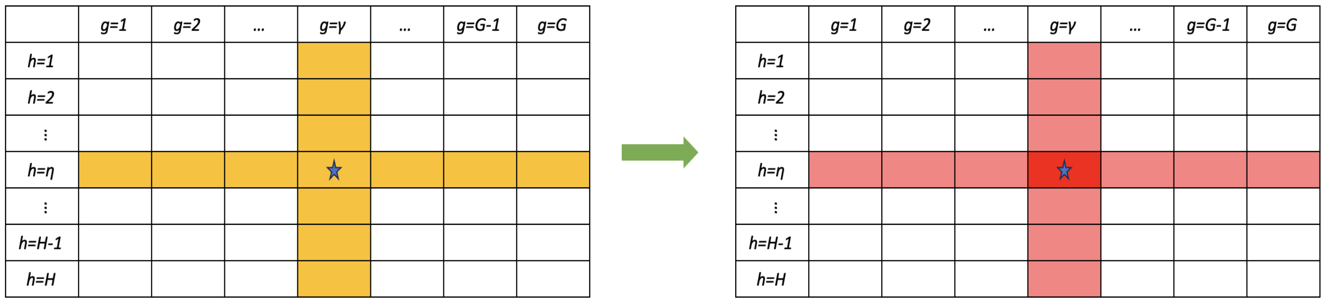
under suitable additional regularity conditions (as stated in subsection 3.3.1), we obtain

$$\boldsymbol{V}^* = \sigma_\nu^2 \widehat{\boldsymbol{V}}_{DHG} + \widehat{\boldsymbol{e}}_{GH}, \tag{3.3}$$

where $\widehat{\boldsymbol{e}}_{GH}$ is an asymptotically negligible term, i.e., $\left(\sigma_\nu^2 \widehat{\boldsymbol{V}}_{DHG}\right)^{-1} \widehat{\boldsymbol{e}}_{GH} = o_P(1)$. It is noteworthy that (3.3) holds irrespective of whether the DGP involves clustering along at least one dimension, clustering at the intersection level, or complete independence among observations.

The result in (3.3) aims to underscore that the novel MWCB approach produces a bootstrap covariance matrix $\boldsymbol{V}^*$ that is (asymptotically) equivalent to $\sigma_\nu^2 \widehat{\boldsymbol{V}}_{DHG}$, where $\widehat{\boldsymbol{V}}_{DHG}$ is a variance estimator robust to the two-way clustering recently investigated by Davezies et al. (2021). Given the connection between $\boldsymbol{V}^*$ and the robust covariance matrix estimator $\widehat{\boldsymbol{V}}_{DHG}$, we can expect this new bootstrap method to be valid under conditions similar to $\widehat{\boldsymbol{V}}_{DHG}$. Moreover, it will be shown that our bootstrap DGP can consistently estimate the asymptotic variance of $\widehat{\boldsymbol{\beta}}$, provided that we choose the variance of the external random variable σ_ν^2 appropriately (as shown in Theorem 3.7).

We then illustrate our proposed bootstrap DGP by providing further intuition. As depicted in Figure 1, the principal concept behind the novel approach is to ensure perfect correlation in $\nu_{gh}^{*\gamma\eta}$ whenever intersections share a common cluster along at least one dimension, i.e., when $g = \gamma$



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

Figure 1: **An illustration of how the random variable is generated under the standard multiway clustering.**

or $h=\eta$, we have $corr^*(\nu_{\gamma\eta}^{*\gamma\eta},\nu_{gh}^{*\gamma\eta})=1$. Thus, the "aggregated" external random variable ν_{gh}^* associated with the intersection (g,h), formed by taking a weighted average of $\nu_{gh}^{*\gamma\eta}$, exhibits correlation with that of another intersection (g',h') when both intersections share any cluster in any dimension. This design preserves the interdependence across both dimensions simultaneously. The simultaneous preservation of dependence enables our bootstrap method to mimic the main features of the DGP and ensure its theoretical validity. The generation concept of ν_{gh}^* is also displayed in detail in Figure D.1 for a simple case where $G=H=4$.

Notably, as displayed in Table 1, one can deduce that when $\chi_1=\chi_2=1$, for $g\neq g'$ and $h\neq h'$,

$$corr^*(\nu_{gh}^*,\nu_{gh'}^*)=\frac{H}{G+H-1},\ corr^*(\nu_{gh}^*,\nu_{g'h}^*)=\frac{G}{G+H-1},\ \text{and } corr^*(\nu_{gh}^*,\nu_{g'h'}^*)=\frac{2}{G+H-1}.$$

Hence, as $(G,H)\to\infty$, it becomes evident that the correlation $corr^*(\nu_{gh}^*,\nu_{g'h'}^*)$ tends to zero. This behavior mirrors the pattern wherein the intersections of (g,h) and (g',h') are asymptotically independent. The formulas for $corr^*(\nu_{gh}^*,\nu_{gh'}^*)$ and $corr^*(\nu_{gh}^*,\nu_{g'h}^*)$ exhibit a discernible pattern for adjusting dependence automatically. Specifically, when $G/H\to 0$, there is a proportionally greater number of small, independent clusters along the second dimension. Concurrently, in the first dimension, there are fewer clusters with large sizes, resulting in a relatively higher level of dependence (among all observations in the same cluster) that necessitates control compared to the second dimension. This concept aligns with the approach advocated by MacKinnon et al. (2021) to preserve dependence along the dimension with fewer clusters. In such scenarios, $corr^*(\nu_{gh}^*,\nu_{gh'}^*)\to 1$ and $corr^*(\nu_{gh}^*,\nu_{g'h}^*)\to 0$, resulting in a bootstrap clustering primarily along the first dimension, similar to WCR_G.

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. This may result in an unsatisfactory finite sample performance because the bootstrap fails to replicate the correct kurtosis. The second bootstrap DGP approach is proposed to address this issue while preserving the correlation structure within clusters, similar to the first method.

MWCB$_{II}$:

The second bootstrap random weights are obtained as follows:

$$\nu^*_{gh} = \begin{cases} \nu^*_g, & \text{with probability } p, \\ \nu^*_h, & \text{with probability } 1-p, \end{cases} \tag{3.4}$$

with ν^*_g and ν^*_h mutually independent with $(0, \sigma^2_\nu)$ over g and h, respectively. In practice, we let ν^*_g and ν^*_h follow Rademacher distributions, so that we have ν_{gh} also follows a Rademacher distribution. Note that we can write $\nu^*_{gh} = c_{gh}\nu^*_g + (1 - c_{gh})\,\nu^*_h$, where c_{gh} is a binomial random variable equaling 1 with probability p and 0 with probability $1-p$, independent over g and h.

Moreover, for $g \neq g'$ and $h \neq h'$, we can derive the correlations:

$$corr^*(\nu^*_{gh}, \nu^*_{gh'}) = p^2,\ corr^*(\nu^*_{gh}, \nu^*_{g'h}) = (1-p)^2,\ \text{and } corr^*(\nu^*_{gh}, \nu^*_{g'h'}) = 0.$$

Observe that the MWCB_{II} method is a generalization of the WCR_G and WCR_H approaches. When $p = 1$, the MWCB_{II} method simplifies to the WCR_G approach, and when $p = 0$, it corresponds to the WCR_H approach. In practice, when the true DGP is unknown, we recommend applying an adaptive $p = \frac{H}{G+H}$. This allows the method to naturally adjust: when G is small, it shifts towards WCB_G, preserving dependence along the dimension with fewer clusters, and when H is small, it shifts towards WCB_H, yielding improved results.

The advantage of this MWCB_{II} method over the MWCB_I method is that it avoids generating artificial correlations between intersections (g, h) and (g', h'), and the fourth moment of the random

	$corr^*(\nu^*_{gh}, \nu^*_{gh'})$	$corr^*(\nu^*_{gh}, \nu^*_{g'h})$	$corr^*(\nu^*_{gh}, \nu^*_{g'h'})$	fourth moment
Ideal Wild Bootstrap	1	1	0	1
WCR_I	0	0	0	1
WCR_G	1	0	0	1
WCR_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 1: Comparison of different wild bootstrap methods

variable ν^*_{gh} is 1. However, the trade-off is that it does not preserve as strong a correlation as the first method: when $G = H$ and $p = \frac{1}{2}$, we have $corr^*(\nu^*_{gh}, \nu^*_{gh'}) = \frac{1}{4}$ for MWCB_{II}, whereas MWCB_I yields $corr^*(\nu^*_{gh}, \nu^*_{gh'}) \approx \frac{1}{2}$.

Note that the two novel MWCB methods should demonstrate similar finite sample performance to WCR_G or WCR_H, as they account for partial dependence in both dimensions. A wild bootstrap following Algorithm 1 achieves the ideal 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. The key advantage of our approaches is its robustness when the true cluster dependence is unknown. Regardless of whether the DGP is clustered along both dimensions, along the G dimension only, along the H dimension only, by intersections, or not clustered at all, our methods remain robust.

Our two bootstrap DGPs build upon the wild cluster bootstrap method introduced by Cameron et al. (2008). Originally designed for one-way clustering, Cameron et al. (2008)'s method assigns an independent weight to each cluster, assuming independence among clusters. In the context of two-way clustering, MacKinnon et al. (2021) present a similar approach assigning an *independent* external random variable to each intersection (g, h), denoted as WCR_I. However, as highlighted by MacKinnon et al. (2021), this method faces challenges in faithfully replicating multiway clustering, leading to incomplete preservation of intra-cluster correlations and consequently, diminished performance. The limitation arises from considering intersections as independent even if they share a cluster in one dimension.

Similar to WCR_I, our bootstrap DGPs perturb the score at the intersection level. However, to address the dependence among intersections, we allow external variables to exhibit dependence, provided they share any cluster in any dimension. Hence, our bootstrap procedures generalize the standard wild bootstrap and the wild cluster bootstrap, resulting in bootstrap DGPs that exhibit correlations similar to the dependent wild bootstrap introduced by Shao (2010) and Hounyo (2023) for the time series case, and the spatial dependent wild bootstrap of Conley et al. (2023) for spatial (cross-sectional) dependent data. Furthermore, it is noteworthy to underscore that MWCB_I is closely connected to the groundbreaking concept initially introduced by Hansen (1992).[8]

The standard wild bootstrap was introduced for linear regression in the non-clustered context by Wu (1986). Its asymptotic validity and refinement were proven by Liu (1988) and Mammen (1993). It was later popularized by Cameron et al. (2008) for a one-way cluster setting. Related work includes Carter, Schnepel, and Steigerwald (2017), Djogbenou et al. (2019), Hansen and Lee (2019), and Canay, Santos, and Shaikh (2021), among others.

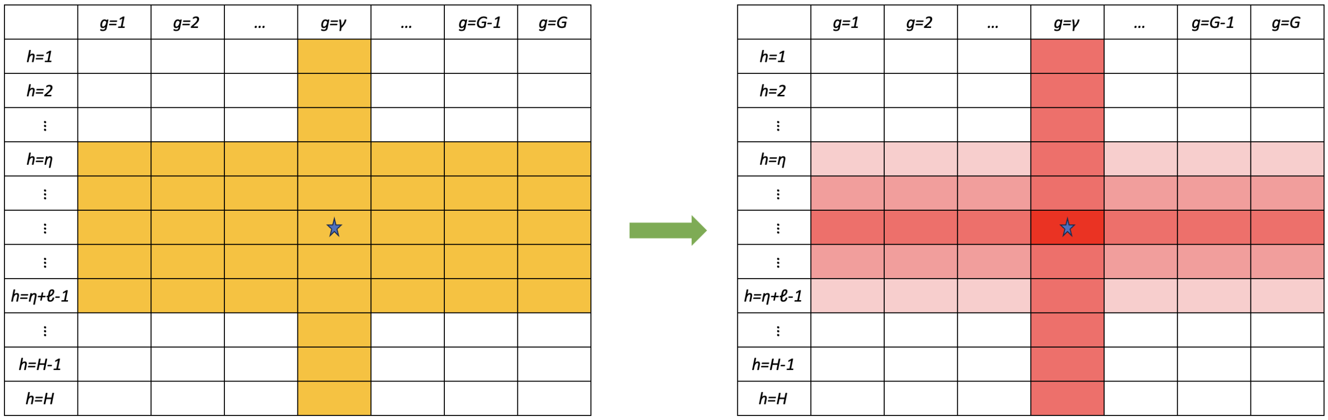
3.2.2 Multiway Clustering with Time Dimension

In the more general context of multiway clustering, when condition (2.5) is satisfied, our two MWCB methods can be extended to preserve the temporal dependence. This preservation is achieved by allowing for dependence in the random variable ν^*_{gh} even when intersections do not share a common cluster in both dimensions, provided they are sufficiently close in the second (H) dimension.

MWCB_I:

In the context of two-way clustering with time dimension, where the bandwidth parameter is

[8]In the Errata, Hansen (1992) extends the concept of the standard wild bootstrap method to address serial correlation by multiplying the sample by a moving average of i.i.d. shocks. This results in a serial correlation pattern in the generated wild bootstrap observations, mirroring an identical pattern found in a Bartlett-weighted HAC estimator.



The "fundamental" random variables $\nu_{gh}^{*\gamma\eta}$ are perfectly correlated when either $g=\gamma$ or $h \in [\eta, \eta+\ell-1]$, 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 ν_{gh}^{*} are strongly correlated with $\nu_{\gamma\eta}^{*}$ when either $g=\gamma$ or $h \in [\eta, \eta+\ell-1]$. A darker shade suggests a higher degree of correlation with intersection (γ,η).

Figure 2: **An illustration of how the random variable is generated under the multiway clustering with time dimension.**

ℓ, the external random variable ν_{gh}^{*} in (3.1) can be generated as follows:[9]

$$\nu_{gh}^{*} = \frac{1}{\sqrt{G\ell + H - 1}} \sum_{\gamma=1}^{G} \sum_{\eta=2-\ell}^{H} \nu_{gh}^{*\gamma\eta}, \tag{3.5}$$

where given γ, η,

$$\nu_{gh}^{*\gamma\eta} = \begin{cases} 0, & \text{if } g \neq \gamma \text{ and } h \notin [\eta, \eta+\ell-1], \\ \chi_1 \nu^{*\gamma\eta}, & \text{if } g = \gamma, \\ \chi_2 \nu^{*\gamma\eta}, & \text{if } g \neq \gamma \text{ and } h \in [\eta, \eta+\ell-1], \end{cases}$$

with $\nu^{*\gamma\eta} \sim (0, \sigma_\nu^2)$ i.i.d. over γ, η. Figure 2 illustrates how the novel method preserves cluster dependence, encompassing temporal dependence. Two intersections are considered "close" if they are situated in the neighborhood of some elements. Moreover, the extent of dependence diminishes with a reduction in the number of elements sharing a neighborhood with both intersections. In the context of the time series dimension, when a "natural" distance is present, the count of such elements declines as the distance of two intersections in the second dimension increases, mirroring the behavior of the DGP where temporal dependence diminishes after a certain period. Similarly,

[9] A more general bootstrap DGP, which can adapt to potential spatial dependence, is provided in (D.1) in Appendix D.

when the parameter of interest is studentized, we can simply choose $\chi_1 = \chi_2 = 1$. While when

$$\chi_1 = \sqrt{1 + \frac{G\ell}{H}} \quad \text{and} \quad \chi_2 = \sqrt{1 + \frac{H}{G\ell}}, \tag{3.6}$$

under suitable additional regularity conditions (as stated in subsection 3.3.2), we obtain

$$\boldsymbol{V}^* = \sigma_\nu^2 \widehat{\boldsymbol{V}}_{CV} + \widehat{\boldsymbol{e}}_{GH}, \tag{3.7}$$

where $\widehat{\boldsymbol{V}}_{CV}$ is a variance estimator robust to the two-way clustering with serial dependence proposed by Chen and Vogelsang (2024).

To compute the bootstrap t-statistic, we propose novel bootstrap variances $\widehat{\boldsymbol{V}}_m^*$. This involves employing a consistent estimator of $\boldsymbol{V}^*$, which is computed in the same manner as $\widehat{\boldsymbol{V}}_m$ but substituting the regression scores $\widehat{\boldsymbol{s}}_{gh}$ with bootstrap-estimated scores $\widehat{\boldsymbol{s}}_{gh}^*$. Another crucial difference is the assignment of the weight functions to be 1 (refer to (7.1) and (7.2) in the appendix). This adjustment enables the MWCB method to effectively address serial dependence.

MWCB$_{II}$:

Moreover, we introduce the variant of the second bootstrap that is robust to arbitrary serial dependence along the second dimension. With time series dependence, the random variable ν_{gh}^* is still obtained as in (3.4), with the key difference being that instead of letting ν_h^* to be independent over h, we now generate ν_h^* to exhibit dependence: for each $h \geq 1$, we let

$$\nu_{h+1}^* = \begin{cases} \nu_h^*, & \text{with probability } \frac{1+q}{2}, \\ -\nu_h^*, & \text{with probability } \frac{1-q}{2}, \end{cases} \tag{3.8}$$

where $q \in (0,1)$, and we let the initial term ν_0^* to follow a Rademacher distribution. One can deduce that the correlation $corr^*\left(\nu_h^*, \nu_{h+\iota}^*\right) = q^\iota$ for each h and ι, which decays exponentially. Therefore, this approach is expected to capture the mixing structure effectively, particularly in the presence of serial dependence. The rule of thumb for choosing the parameter q is provided in the appendix. If $q = 0$, the method reduces to the simple MWCB$_{II}$ in the standard multiway clustering setting where there is no time dependence. Note that we can also rewrite $\nu_{gh}^* = c_{gh}\nu_g^* + (1 - c_{gh}) \prod_{\iota=1}^{h} \kappa_\iota \nu_0^*$, where c_{gh} is independent over g and h, and κ_ι is independent over ι, with

$$c_{gh} = \begin{cases} 1, & \text{with probability } p, \\ 0, & \text{with probability } 1-p, \end{cases} \quad \text{and} \quad \kappa_\iota = \begin{cases} 1, & \text{with probability } \frac{1+q}{2}, \\ -1, & \text{with probability } \frac{1-q}{2}. \end{cases}$$

Moreover, we modify the original variance estimator as well to match the bootstrap correlation. The variants of the CHS and CV estimator of the score variance are

$$\widehat{\boldsymbol{V}}_{CHS,V} = \widehat{\boldsymbol{Q}}^{-1}\widehat{\boldsymbol{\Gamma}}_{CHS,V}\widehat{\boldsymbol{Q}}^{-1} \quad \text{and} \quad \widehat{\boldsymbol{V}}_{CV,V} = \widehat{\boldsymbol{Q}}^{-1}\widehat{\boldsymbol{\Gamma}}_{CV,V}\widehat{\boldsymbol{Q}}^{-1}, \tag{3.9}$$

where

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}_{CHS,V} =& \sum_g \widehat{\boldsymbol{s}}_g \widehat{\boldsymbol{s}}_g^\top + \sum_h \widehat{\boldsymbol{s}}_h \widehat{\boldsymbol{s}}_h^\top - \sum_g \sum_h \widehat{\boldsymbol{s}}_{gh} \widehat{\boldsymbol{s}}_{gh}^\top \\ &+ \sum_{\iota=1}^{H-1} k(\iota, q) \sum_{h=1}^{H-\iota} \left(\widehat{\boldsymbol{s}}_h \widehat{\boldsymbol{s}}_{h+\iota}^\top + \widehat{\boldsymbol{s}}_{h+\iota} \widehat{\boldsymbol{s}}_h^\top - \sum_{g=1}^{G} \widehat{\boldsymbol{s}}_{gh} \widehat{\boldsymbol{s}}_{gh+\iota}^\top - \sum_{g=1}^{G} \widehat{\boldsymbol{s}}_{gh+\iota} \widehat{\boldsymbol{s}}_{gh}^\top \right), \quad \text{and} \\ \widehat{\boldsymbol{\Sigma}}_{CV,V} =& \sum_g \widehat{\boldsymbol{s}}_g \widehat{\boldsymbol{s}}_g^\top + \sum_h \widehat{\boldsymbol{s}}_h \widehat{\boldsymbol{s}}_h^\top + \sum_{\iota=1}^{H-1} k(\iota, q) \sum_{h=1}^{H-\iota} \left(\widehat{\boldsymbol{s}}_h \widehat{\boldsymbol{s}}_{h+\iota}^\top + \widehat{\boldsymbol{s}}_{h+\iota} \widehat{\boldsymbol{s}}_h^\top \right), \end{aligned}$$

and the weight function $k(\iota, q) = q^\iota$. Notice we have also adjusted the total number of terms in the sum for autocorrelation (from $\ell - 1$ to $H - 1$). This is because, as $\iota \to \infty$, the weight function converges to zero automatically. There is no need to truncate the autocorrelation sum, otherwise it may result in the absence of positive semi-definiteness. In Proposition 3.1, we further show that with these two adjustments, we continue to have valid variance estimators and $\widehat{\boldsymbol{V}}_{CV,V}$ is guaranteed to be positive semi-definite.

Hence, our two proposed methods can be viewed as the multiway counterpart to the (one-way) wild cluster bootstrap approach of Cameron et al. (2008). We generate bootstrap scores at the intersection level by multiplying each empirical score by a variable that accounts for the multiway dimension appropriately and is dependent within a serial distance, provided that dependence between observations decays with the given distance.

3.3 Bootstrap Theory

Our goal in this subsection is to provide a set of bootstrap high-level conditions such that the MWCBs are valid for inference on $\boldsymbol{\beta}$. Next, we provide a set of primitive assumptions under which the bootstrap high-level conditions are satisfied. Simultaneously, under these assumptions, we undertake a comprehensive review and presentation of the asymptotic theory of the OLS estimator.

This exploration into the asymptotic theory of the OLS estimator serves a crucial purpose – it aids in understanding the essential properties required for our proposed bootstrap method to be asymptotically valid. Following the approaches in Section 2.3, we examine two scenarios. Firstly, when clusters are asymptotically independent, as outlined in equation (2.3). Secondly, we delve into the more general case, allowing for interdependence among distinct intersections that do not share any cluster in any dimension; see (2.5).

As usual in the bootstrap literature, we write $T^*_{GH} \to^{d^*} D$, in probability, if conditional on a sample with probability that converges to one, T^*_{GH} weakly converges to the distribution D under P^*, i.e., $E^*\left(f\left(T^*_{GH}\right)\right) \to^{P} E\left(f\left(D\right)\right)$ for all bounded and uniformly continuous function f. Then the first-order asymptotic validity of the multiway wild cluster bootstrap can be established under the conditions below.

Condition 3.1. *There exists a scalar M (function of G and/or H) and a matrix $\boldsymbol{V}_0 > 0$ such that i) $M\boldsymbol{V}^* \to^{P} \boldsymbol{V}_0$, as $M \to \infty$. ii) $E^*\left(M\left|\boldsymbol{a}^\top\left(\sigma_m^2\widehat{\boldsymbol{V}}_m^* - \boldsymbol{V}^*\right)\boldsymbol{a}\right|\right) = o_P(1)$, as $M \to \infty$.*

Condition 3.2. $\frac{\boldsymbol{a}^\top\left(\widehat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}}\right)}{\left(\boldsymbol{a}^\top\boldsymbol{V}^*\boldsymbol{a}\right)^{1/2}} \to^{d^*} \mathcal{N}(0,1)$, *in probability.*

Condition 3.3. *For all $x \in \mathbb{R}$, $P(\widehat{t}_m \le x) \to \Phi(\frac{x}{\sigma_m})$, where $\Phi(\cdot)$ is the cumulative density function (CDF) of a standard Gaussian function.*

Theorem 3.4. *For two MWCB methods, under Condition 3.1, we have*

$$\frac{\boldsymbol{a}^\top\boldsymbol{V}^*\boldsymbol{a}}{\boldsymbol{a}^\top\widehat{\boldsymbol{V}}_m^*\boldsymbol{a}} \to^{P^*} \sigma_m^2, \text{ in probability,} \tag{3.10}$$

If in addition Condition 3.2 holds, then

$$\widehat{t}_m^* = \frac{\boldsymbol{a}^\top \left(\widehat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}}\right)}{\left(\boldsymbol{a}^\top \widehat{\boldsymbol{V}}_m^* \boldsymbol{a}\right)^{1/2}} \rightarrow^{d^*} \mathcal{N}(0, \sigma_m^2), \ \textit{in probability.} \tag{3.11}$$

Moreover, if Condition 3.3 is also satisfied, then for any $\varepsilon > 0$*,*

$$P\left(sup_{x \in \mathbb{R}} \left| P^*(\widehat{t}_m^* \leq x) - P(\widehat{t}_m \leq x) \right| > \varepsilon\right) \rightarrow 0. \tag{3.12}$$

3.3.1 Bootstrap Consistency Under Standard Multiway Clustering

In the context of standard two-way clustering where (2.3) holds, we present the conditions for the asymptotic theory, formulating them at the intersection level. Let

$$\boldsymbol{S}_{gh} = (\boldsymbol{S}_{gh,i})_{i \geq 1} = (\boldsymbol{X}_{gh,i}^\top, u_{gh,i})_{i \geq 1}^\top, \qquad g \geq 1, h \geq 1,$$

Throughout, for a matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\| = \left(trace\left(\boldsymbol{A}^\top \boldsymbol{A}\right)\right)^{1/2}$ denotes the Euclidean norm. C represents a generic finite constant. We adopt the assumptions as follows:

Assumption 1. $E(\boldsymbol{X}_{gh}^\top \boldsymbol{u}_{gh}) = \boldsymbol{0}$ *and* $E\left(\left(\sum_{i=1}^{N_{gh}} \|\boldsymbol{S}_{gh,i}\|^2\right)^2\right) < \infty$.

Assumption 2. *The regressors are such that* $\boldsymbol{Q}_0 = E(\boldsymbol{X}_{gh}^\top \boldsymbol{X}_{gh})$ *is non-singular.*

Assumption 3. $E(N_{gh}) > 0$.

Assumption 4. $(N_{gh}, \boldsymbol{S}_{gh})$ *is independent of* $(N_{g'h'}, \boldsymbol{S}_{g'h'})$ *if* $g \neq g'$ *and* $h \neq h'$.

Assumption 5. $R = min(G, H) \rightarrow \infty$.

Assumption 6. $(N_{gh}, \boldsymbol{S}_{gh}) \overset{d}{=} (N_{\pi_1(g)\pi_2(h)}, \boldsymbol{S}_{\pi_1(g)\pi_2(h)})$*, where* $\pi_1(\cdot)$ *and* $\pi_2(\cdot)$ *are permutations of* $\mathbb{N}$.

Assumptions 1-6 are standard in the asymptotic theory for linear regressions under standard multiway clustering. These assumptions correspond precisely to those delineated by MacKinnon et al. (2021). As pointed out by MacKinnon et al. (2021), the condition specified in Assumption 4 assumes paramount significance for standard multiway clustering. This condition will be relaxed in Section 3.3.2.

Before presenting the next results, we find it convenient to introduce additional notations that will be required later. Specifically, we define the asymptotic variance matrices:

$$\mathbf{\Gamma}_G \equiv \lim \frac{1}{GH^2} \sum_{g=1}^{G} E\left(\boldsymbol{X}_g^\top \boldsymbol{u}_g \boldsymbol{u}_g^\top \boldsymbol{X}_g\right), \tag{3.13}$$

$$\mathbf{\Gamma}_H \equiv \lim \frac{1}{G^2 H} \sum_{h=1}^{H} E\left(\boldsymbol{X}_h^\top \boldsymbol{u}_h \boldsymbol{u}_h^\top \boldsymbol{X}_h\right), \tag{3.14}$$

$$\mathbf{\Gamma}_I \equiv \lim \frac{1}{GH} \sum_{g=1}^{G} \sum_{h=1}^{H} E\left(\boldsymbol{X}_{gh}^\top \boldsymbol{u}_{gh} \boldsymbol{u}_{gh}^\top \boldsymbol{X}_{gh}\right), \tag{3.15}$$

$$\mathbf{\Gamma}_{NC} \equiv \lim \frac{1}{GH} \sum_{g=1}^{G} \sum_{h=1}^{H} \sum_{i=1}^{N_{gh}} E\left(\boldsymbol{X}_{gh,i}^\top u_{gh,i} u_{gh,i} \boldsymbol{X}_{gh,i}\right). \tag{3.16}$$

Let $\boldsymbol{Q} \equiv \lim \frac{1}{GH} \sum_{g=1}^{G} \sum_{h=1}^{H} E\left(\boldsymbol{X}_{gh}^\top \boldsymbol{X}_{gh}\right)$, $\boldsymbol{V}_m = \boldsymbol{Q}^{-1} \mathbf{\Gamma}_m \boldsymbol{Q}^{-1}$ for $m \in \{G, H, I, NC\}$, and $\lambda_m = \lim(m^{-1}R)$ for $m \in \{G, H\}$.

Theorem 3.5. *Suppose Assumptions 1-6 are satisfied and the true value of $\boldsymbol{\beta}$ is given by $\boldsymbol{\beta}_0$.*

a) If the DGP is clustered along at least one dimension ($\lambda_G \boldsymbol{V}_G + \lambda_H \boldsymbol{V}_H > 0$), then

$$\begin{aligned}
\sqrt{R}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &\to^d \mathcal{N}(\mathbf{0}, \lambda_G \boldsymbol{V}_G + \lambda_H \boldsymbol{V}_H), \\
R \widehat{\boldsymbol{V}}_{CGM} &\to^P \lambda_G \boldsymbol{V}_G + \lambda_H \boldsymbol{V}_H, \text{ and} \\
R \widehat{\boldsymbol{V}}_{DHG} &\to^P \lambda_G \boldsymbol{V}_G + \lambda_H \boldsymbol{V}_H.
\end{aligned}$$

b) If the DGP is only clustered by intersections, then

$$\begin{aligned}
\sqrt{GH}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &\to^d \mathcal{N}(\mathbf{0}, \boldsymbol{V}_I), \\
GH \widehat{\boldsymbol{V}}_{CGM} &\to^P \boldsymbol{V}_I, \text{ and} \\
GH \widehat{\boldsymbol{V}}_{DHG} &\to^P 2\boldsymbol{V}_I.
\end{aligned}$$

c) If there is no clustering at all in the DGP, i.e., all observations are independent, then

$$\begin{aligned}
\sqrt{GH}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &\to^d \mathcal{N}(\mathbf{0}, \boldsymbol{V}_{NC}), \\
GH \widehat{\boldsymbol{V}}_{CGM} &\to^P \boldsymbol{V}_{NC}, \text{ and} \\
GH \widehat{\boldsymbol{V}}_{DHG} &\to^P 2\boldsymbol{V}_{NC}.
\end{aligned}$$

Results presented in Theorem 3.5 have been widely acknowledged in the literature (see MacKinnon et al. (2021)), and are only reiterated here for completeness. The rate of convergence of $\widehat{\boldsymbol{\beta}}$ is readily seen from Theorem 3.5 and is a function of the premise that clustering is present or not in the DGP. Specifically, when the DGP is clustered along at least one dimension, the rate of convergence of $\widehat{\boldsymbol{\beta}}$ is $1/\sqrt{R}$; otherwise, it is $1/\sqrt{GH}$. In the following, we adhere to the literature (see, e.g., Hansen and Lee (2019), MacKinnon et al. (2021), and Chiang et al. (2024)), and primarily focus on the bootstrap and asymptotic limit theory for studentized (or self-normalized) statistics.

Theorem 3.6. *Assume that the bootstrap scores $\boldsymbol{s}^*_{gh,i}$ are generated following the process in (3.1) or (3.4), and $0 < E^*(|\nu^*_{gh}|^4) < \infty$. Moreover, suppose Assumptions 1-6 are satisfied and the null hypothesis $\mathcal{H}_0 : \boldsymbol{a}^\top \boldsymbol{\beta} = \boldsymbol{a}^\top \boldsymbol{\beta}_0$ holds true. If $m = CGM, DHG$, then no matter whether the DGP is clustered along at least one dimension, by intersections, or not clustered at all, results in Theorem 3.4 holds.*

The result presented in Theorem 3.6 provides a theoretical foundation for the utilization of P values computed in Step 5 of Algorithm 1. This computation is based on our proposed MWCB procedures, employing standard error estimates derived from the cluster-robust variance estimators by Cameron et al. (2011) and Davezies et al. (2021), denoted as $m = CGM, DHG$. Specifically, this justification holds under standard multiway clustering conditions, as specified when condition (2.3) is satisfied.

This outcome demonstrates the validity of our MWCB bootstrap tests, irrespective of whether we employ $\widehat{\boldsymbol{V}}_{DHG}$ (i.e., the two-term CRVE) or $\widehat{\boldsymbol{V}}_{CGM}$ (i.e., the three-term CRVE) to compute the t-statistics. Furthermore, this validity holds regardless of whether the DGP involves clustering along at least one dimension, by intersections, or exhibits no clustering.

In contrast, the results presented by MacKinnon et al. (2021, cf. Theorem 3) show that when the DGP involves clustering along at least one dimension, the distribution of their bootstrap t-statistics (clustering by intersections, i.e., WCR_I) does not coincide with that of the original t-statistic computed using $\widehat{\boldsymbol{V}}_{DHG}$. Specifically, MacKinnon et al. (2021) demonstrate that inference based on their WCR_I distribution of $\widehat{t}^*_{DHG}$ leads to over-rejection in cases of clustering along at least one dimension. This discrepancy does not arise when relying on our proposed MWCB-based test, which utilizes the two-term CRVE.

Next, we provide conditions under which our MWCB approaches can be used to consistently estimate the asymptotic variance of $\widehat{\boldsymbol{\beta}}$. This requires strengthening Assumption 5 as follows.

Assumption 5'. $G/H \to c \in (0, \infty)$ as $(G, H) \to \infty$.

Assumption 5' imposes constraints on the ratio G/H and establishes the asymptotic framework wherein the number of clusters tends to infinity along both dimensions of clustering. This assumption aims to prevent the divergence of the scaling terms χ_1 and χ_2 as defined in (3.2). It is worth noting that, in the context of our study, this restriction on the ratio G/H is exclusively employed to demonstrate the consistency of bootstrap variance. Interestingly, it is not a prerequisite for establishing the bootstrap validity for studentized statistics when $m = CGM, DHG$. In such case (as stated in Theorem 3.6), the scaling terms can be set to any nonzero finite constant, eliminating the need for the aforementioned ratio restriction.

Theorem 3.7. *Assume that the bootstrap scores* $\boldsymbol{s}^*_{gh,i}$ *are generated following the process in (3.1), the scaling terms* χ_1, χ_2 *are defined as in (3.2), and* $0 < \sigma^2_\nu < \infty$*. Moreover, suppose Assumptions 1-6 hold, strengthened by Assumption 5'.*

a) If the DGP is clustered along at least one dimension, then

$$R\boldsymbol{V}^* \to^P \sigma^2_\nu \cdot (\lambda_G \boldsymbol{V}_G + \lambda_H \boldsymbol{V}_H) .$$

b) If the DGP is clustered only by intersections, then

$$GH\boldsymbol{V}^* \to^P 2\sigma^2_\nu \cdot \boldsymbol{V}_I .$$

c) If there is no clustering in the GDP, then

$$GH\boldsymbol{V}^* \to^P 2\sigma^2_\nu \cdot \boldsymbol{V}_{NC} .$$

Result a) also holds for a general $MWCB_{II}$ *as provided in (7.5).*

Remark 1. *Given results in Theorems 3.5 and 3.7, it is evident that, by setting* $\sigma_\nu^2 = 1$ *(as is customary),* $\boldsymbol{V}^*$ *exhibits identical asymptotic properties as* $\widehat{\boldsymbol{V}}_{DHG}$. *However, the consistency of* $\boldsymbol{V}^*$ *toward the asymptotic variance of* $\widehat{\boldsymbol{\beta}}$ *is impeded by the absence of a universally applicable choice for* σ_ν^2. *This observation aligns with McCullagh (2000) and Menzel (2021, cf. Footnote 4). Consequently, due to the uncertainty in DGP clustering, we focus on and recommend the use of studentized statistics for consistent scaling and independence from the choice of* σ_ν^2 *in the bootstrap DGP.*

3.3.2 Bootstrap Consistency Under Multiway Clustering with Time Dimension

In this subsection, we consider the multiway clustering with serially correlated time effects where (2.5) holds. For simplicity, we focus on the setting where there is exactly one observation in each intersection.[10] Set $\boldsymbol{a}_g = E(\boldsymbol{X}_{gh}^\top \boldsymbol{u}_{gh}|\boldsymbol{\alpha}_g)$ and $\boldsymbol{b}_h = E(\boldsymbol{X}_{gh}^\top \boldsymbol{u}_{gh}|\boldsymbol{\xi}_h)$. Let $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_b$ be the variance/long-run variance matrices of $\boldsymbol{a}_g$ and $\boldsymbol{b}_h$, respectively. We adhere to assumptions as follows: For some $\delta > 0$ and $\zeta > 1$,

Assumption 7. $(\boldsymbol{Y}_{gh}, \boldsymbol{X}_{gh}, \boldsymbol{u}_{gh}) = f(\boldsymbol{\alpha}_g, \boldsymbol{\xi}_h, \boldsymbol{\varepsilon}_{gh})$ *for* $1 \leq g \leq G, 1 \leq h \leq H$, *where* $f(\cdot)$ *is a Borel-measurable function,* $\{\boldsymbol{\alpha}_g\}$, $\{\boldsymbol{\xi}_h\}$, *and* $\{\boldsymbol{\varepsilon}_{gh}\}$ *are mutually independent sequences,* $\boldsymbol{\alpha}_g$ *is i.i.d. across* g, $\boldsymbol{\varepsilon}_{gh}$ *is i.i.d. across* g, h, *and the sequence* $\boldsymbol{\xi}_h$ *is a strictly stationary serially correlated process.*

Assumption 8. $E(\boldsymbol{X}_{gh}^\top \boldsymbol{u}_{gh}) = 0$, $E(\boldsymbol{X}_{gh}^\top \boldsymbol{X}_{gh}) > 0$, $E(\|\boldsymbol{X}_{gh}\|^{8(\zeta+\delta)}) < \infty$, *and* $E(\|\boldsymbol{u}_{gh}\|^{8(\zeta+\delta)}) < \infty$.

Assumption 9. $\boldsymbol{\xi}_h$ *is a* β*-mixing sequence with size* $2\zeta/(\zeta-1)$, *that is,* $\beta(\ell) = O(\ell^{-\lambda})$ *for a* $\lambda > 2\zeta/(\zeta-1)$.

---

[10]The main findings outlined in this section can be easily extrapolated to a broader scenario, allowing for the presence of zero or multiple observations in each intersection; see Chiang et al. (2024, cf. Section G). Furthermore, our MWCB procedures are directly applicable to these more general cases.

Assumption 10. *One of the following two conditions hold: (i) Either* $\boldsymbol{\Sigma}_a > 0$ *or* $\boldsymbol{\Sigma}_b > 0$*, and* $G/H \to c \in (0, \infty)$ *as* $(G, H) \to \infty$*; or (ii)* $(\boldsymbol{X}_{gh}, \boldsymbol{u}_{gh})$ *are i.i.d. across* g, h*, and* $Var(\boldsymbol{X}_{gh}^\top \boldsymbol{u}_{gh}) > 0$.

Assumption 11. *The weight functions* $w(\iota, \ell) = (1 - \iota/\ell)\mathbb{I}(1 - \iota/\ell > 0)$ *and* $k(\iota, q) = q^\iota$.

Assumption 12. $\ell \to \infty$*, as* $R \to \infty$*, and* $\ell = o(R^{1/2})$*;* $q \to 1$ *as* $R \to \infty$*, and for some constant* $0 < b < 1$*,* $\log_b q = o(R^{-1/3})$.

Assumptions 7-12 are standard in the asymptotic theory for linear regressions as proposed by CHS. The i.i.d. condition of $\boldsymbol{\alpha}_g$ in Assumption 7 can be relaxed in some ways. For instance, one could consider spatial dependence over individuals associated with panel data, but we defer this exploration to future research. As the current literature on two-way clustering typically assumes the i.i.d. condition, we maintain this assumption. Nevertheless, the MWCB random weights can be generated to accommodate this situation when $\boldsymbol{\alpha}_g$ does not adhere to i.i.d. condition in two-way clustering (refer to (D.2) in Appendix D).

Assumption 12 establishes the guidelines for selecting parameters in the weight functions. The primary objective is to ensure that both weight functions tend to 1 as $R \to \infty$ and tend to 0 as $\iota \to \infty$, keeping all other factors constant. For the weight function $w(\iota, \ell)$, Andrews (1991) offers a useful approach for determining the optimal value, where $\ell = O_P(R^{1/3})$. As $k(\iota, q)$ is a new function, we illustrate its behavior with an example: the parameter value $q = b^{R^{-1/4}}$ satisfies the required condition. Notably, as $R \to \infty$, $R^{-1/4}$ converges to 0, so for fixed ι, $b^{\iota R^{-1/4}}$ asymptotically approaches 1; while when $\iota \geq R^{1/3}$, we have $q^\iota \leq b^{R^{1/12}}$ and hence it decays to 0.

Define the new asymptotic variance matrices: $\boldsymbol{V}_{H_2} = \boldsymbol{Q}^{-1}\boldsymbol{\Gamma}_{H_2}\boldsymbol{Q}^{-1}$, where

$$\boldsymbol{\Gamma}_{H_2} \equiv \lim \frac{1}{G^2 H} \left[\sum_{h=1}^{H} E\left(\boldsymbol{X}_h^\top \boldsymbol{u}_h \boldsymbol{u}_h^\top \boldsymbol{X}_h\right) + \sum_{\iota=1}^{H-1} \sum_{h=1}^{H-\iota} E\left(\boldsymbol{X}_h^\top \boldsymbol{u}_h \boldsymbol{u}_{h+\iota}^\top \boldsymbol{X}_{h+\iota} + \boldsymbol{X}_{h+\iota}^\top \boldsymbol{u}_{h+\iota} \boldsymbol{u}_h^\top \boldsymbol{X}_h\right) \right]. \tag{3.17}$$

Theorem 3.8. *Suppose Assumptions 7-12 are satisfied and the true value of* $\boldsymbol{\beta}$ *is given by* $\boldsymbol{\beta}_0$.

a) If the DGP is clustered along at least one dimension ($\lambda_G \boldsymbol{V}_G + \lambda_H \boldsymbol{V}_{H_2} > 0$)*, then*

$$\sqrt{R}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to^d \mathcal{N}(\boldsymbol{0}, \lambda_G \boldsymbol{V}_G + \lambda_H \boldsymbol{V}_{H_2}),$$

$$R\widehat{\boldsymbol{V}}_{CHS} \to^{P} \lambda_G \boldsymbol{V}_G + \lambda_H \boldsymbol{V}_{H_2}, \text{ and}$$
$$R\widehat{\boldsymbol{V}}_{CV} \to^{P} \lambda_G \boldsymbol{V}_G + \lambda_H \boldsymbol{V}_{H_2}.$$

b) If the DGP is only clustered by intersections, then

$$\sqrt{GH}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to^{d} \mathcal{N}(\mathbf{0}, \boldsymbol{V}_I),$$
$$GH\widehat{\boldsymbol{V}}_{CHS} \to^{P} \boldsymbol{V}_I, \text{ and}$$
$$GH\widehat{\boldsymbol{V}}_{CV} \to^{P} 2\boldsymbol{V}_I.$$

c) If there is no clustering in the DGP, then

$$\sqrt{GH}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to^{d} \mathcal{N}(\mathbf{0}, \boldsymbol{V}_{NC}),$$
$$GH\widehat{\boldsymbol{V}}_{CHS} \to^{P} \boldsymbol{V}_{NC}, \text{ and}$$
$$GH\widehat{\boldsymbol{V}}_{CV} \to^{P} 2\boldsymbol{V}_{NC}.$$

Proposition 3.1. *Suppose Assumptions 7-12 are satisfied and the true value of $\boldsymbol{\beta}$ is given by $\boldsymbol{\beta}_0$. Then results in theorem 3.8 still hold for $\widehat{\boldsymbol{V}}_{CHS,V}$ and $\widehat{\boldsymbol{V}}_{CV,V}$, respectively, as defined in (3.9). Moreover, $\widehat{\boldsymbol{V}}_{CV,V}$ is positive semi-definite.*

Results in Theorem 3.8 can be deduced directly from Chiang et al. (2024), and are only stated here for completeness. Next, we show the consistency of our bootstrap method within the framework of multiway clustering with a temporal dimension.

Theorem 3.9. *Assume that the bootstrap random variables ν^*_{gh} are generated following the process in (3.5) or (3.4) with ν^*_h generated as in (3.8), and $0 < E^*(|\nu^*_{gh}|^4) < \infty$. Moreover, suppose Assumptions 7-12 are satisfied and the null hypothesis $\mathcal{H}_0 : \boldsymbol{a}^\top \boldsymbol{\beta} = \boldsymbol{a}^\top \boldsymbol{\beta}_0$ holds true. If $m = CHS, CV$, then no matter whether the DGP is clustered along at least one dimension, by intersections, or not clustered at all, results in Theorem 3.4 holds.*

This result provides a theoretical justification (and the conditions) for using the P values computed in Step 5 of Algorithm 1, based on our proposed MWCB procedure, with standard errors calculated using the Chiang et al. (2024) and Chen and Volgesang (2024) cluster-robust variance estimators (i.e., $m = CHS, CV$). Specifically, this justification holds under multiway clustering with time dimension conditions (where (2.5) holds).

Likewise, the results showcase the robustness of our MWCB bootstrap tests, regardless of whether we utilize $\widehat{\boldsymbol{V}}_{CHS}$ or $\widehat{\boldsymbol{V}}_{CV}$ for computing the t-statistics. Moreover, this validity persists irrespective of whether the DGP involves clustering along at least one dimension, by intersections, or exhibits no clustering at all.

Furthermore, we provide conditions under which our MWCB method can be used to consistently estimate the asymptotic variance of $\widehat{\boldsymbol{\beta}}$, which also requires strengthening Assumption 10 to prevent the scaling terms χ_1 and χ_2 from diverging.

Assumption 10'. One of the following two conditions hold: (i) Either $\boldsymbol{\Sigma}_a > 0$ or $\boldsymbol{\Sigma}_b > 0$, and $G/H \to c \in (0, \infty)$ as $(G, H) \to \infty$; or (ii) $(\boldsymbol{X}_{gh}, \boldsymbol{u}_{gh})$ are i.i.d. across g, h, $Var(\boldsymbol{X}_{gh}^{\top}\boldsymbol{u}_{gh}) > 0$, and $G/H \to c \in (0, \infty)$ as $(G, H) \to \infty$.

Theorem 3.10. *Assume that the external random variables are generated following the process in (3.5), the scaling terms χ_1 and χ_2 are defined as in (3.6), and $0 < \sigma_\nu^2 < \infty$. Additionally, suppose Assumptions 7-12 are satisfied, strengthened by Assumption 10'.*

a) If the DGP is clustered along at least one dimension, then

$$R\boldsymbol{V}^* \to^P \sigma_\nu^2 \cdot (\lambda_G \boldsymbol{V}_G + \lambda_H \boldsymbol{V}_{H_2}).$$

b) If the DGP is clustered only by intersections, then

$$GH\boldsymbol{V}^* \to^P 2\sigma_\nu^2 \cdot \boldsymbol{V}_I.$$

c) If there is no clustering in the DGP, then

$$GH\boldsymbol{V}^* \to^P 2\sigma_\nu^2 \cdot \boldsymbol{V}_{NC}.$$

Result a) also holds for a general $MWCB_{II}$ as provided in (7.5), with ν_h^* generated as in (3.8).

Similarly, based on Theorems 3.8 and 3.10, our bootstrap variance $\boldsymbol{V}^*$ exhibits asymptotic properties comparable to those of $\widehat{\boldsymbol{V}}_{CV}$ when opting for $\sigma_\nu^2 = 1$.

4 Simulation Experiments

We conduct a comprehensive series of simulation experiments and present the most noteworthy findings. Our simulations aim to support the validity of our approach, as outlined in Section 3.3. We primarily present the finite sample performance of various wild cluster bootstrap methods across diverse scenarios to enhance the clarity of the graphs. Wild cluster bootstrap methods are particularly useful due to their adaptability to cases with heterogeneous cluster sizes, a common feature in practice. Additional simulation results and discussions for other methods are provided in Appendix C.

We generate data based on the linear model

$$y_{gh} = \beta_1 + \sum_{k=2}^{10} \beta_k X_{gh,k} + u_{gh}, \tag{4.1}$$

where

$$u_{gh} = \omega_\alpha \alpha_g^u + \omega_\xi \xi_h^u + \omega_\varepsilon \varepsilon_{gh}^u, \quad \text{and} \tag{4.2}$$

$$X_{gh,k} = \omega_\alpha \alpha_{g,k}^x + \omega_\xi \xi_{h,k}^x + \omega_\varepsilon \varepsilon_{gh,k}^x. \tag{4.3}$$

We set $\beta_k = 1$ for each k and $(\alpha_{g,k}^x, \alpha_g^u, \varepsilon_{gh,k}^x, \varepsilon_{gh}^u)$ are standard normal random variables mutually independent over g, h, and k. The latent components $\left(\xi_{h,k}^x, \xi_h^u\right)$ are generated according to two different multiway clustering settings. We choose the number of regressors to be a large number 10 intentionally because the recent work of MacKinnon (2023) suggests that the performances of many methods deteriorate when the number of regressors is large. We examine the true value of β_{10} by different methods.

4.1 Standard Multiway Clustering

In this subsection, we consider the standard multiway clustering where (2.3) holds. Hence, $\left(\xi_{h,k}^x, \xi_h^u\right)$ are set to be mutually independent over h. For MWCB$_{II}$, recall that this method becomes WCR$_G$ when $p = 1$, and WCR$_H$ when $p = 0$. We report results for parameter values $p = 0.2, 0.4, 0.6, 0.8$ to observe variations in performance and include an adaptive value of $p = \frac{H}{G+H}$.

In Panel (a) of Figure 3, we set $(\omega_\xi, \omega_\varepsilon) = (0, 1)$ and vary ω_α from 0 to 2 in increments of 0.2. In this setup, the DGP is clustered only along the first G dimension, and the level of dependence increases as ω_α increases. We observe that WCR_G performs best, which is expected since WCR_G maintains dependence along the first dimension while leaving dependence along the second dimension unaffected. Conversely, WCR_H performs the worst, as it maintains dependence along the incorrect dimension. Notably, MWCB_{II} ($p = 0.8$) performs second best, slightly behind WCR_G, as it assigns more weight to the first dimension than the other variants, though not as much as WCR_G. The performances of MWCB_I and MWCB_{II} (adaptive p) lie between MWCB_{II} ($p = 0.4$) and MWCB_{II} ($p = 0.6$), as these methods address dependence along both dimensions equally when the number of clusters is balanced. Notice that when $\omega_\alpha = 0$, meaning the DGP is clustered by intersections, MWCB_I shows a slight tendency to under-reject the null hypothesis. This aligns with the expectation that wild bootstrap methods may underperform when the fourth moment is not 1.

Panel (b) is a symmetric case of Panel (a), as we set $(\omega_\alpha, \omega_\varepsilon) = (0, 1)$ and vary ω_ξ from 0 to 2 in increments of 0.2. Hence, we find that WCR_H performs the best, with MWCB_{II} ($p = 0.2$) the second best. The performances of MWCB_I and MWCB_{II} (adaptive p) are again moderate.

Panel (c) displays the results when we set $\omega_\varepsilon = 1$ and vary ω_α and ω_ξ from 0 to 2. In this setup, the DGP is clustered across both dimensions. Note that to improve clarity, we zoom in on the y-axis. Since both dimensions are balanced, similar performance is observed among MWCB_{II} ($p = 0.6$) and MWCB_{II} ($p = 0.4$), MWCB_{II} ($p = 0.8$) and MWCB_{II} ($p = 0.2$), as well as WCR_G and WCR_H. We find that MWCB_{II} (adaptive p) performs the best in most cases. MWCB_I tends to under-reject when the level of dependence is low but over-rejects more than MWCB_{II} (adaptive p) when the level of dependence is high. In this scenario, WCR_G and WCR_H perform the worst, as they are highly unbalanced in controlling dependence across both dimensions.

In Panel (d), we explore the effect of the unbalanced number of clusters. Setting $(\omega_\alpha, \omega_\xi, \omega_\varepsilon) = (1, 1, 1)$ ensures an equivalent level of dependence. We keep the number of clusters in the first dimension constant at $G = 15$, while varying the number of clusters in the second dimension, H, from 5 to 40. When $H = 5$, which is much smaller than G, maintaining dependence in the

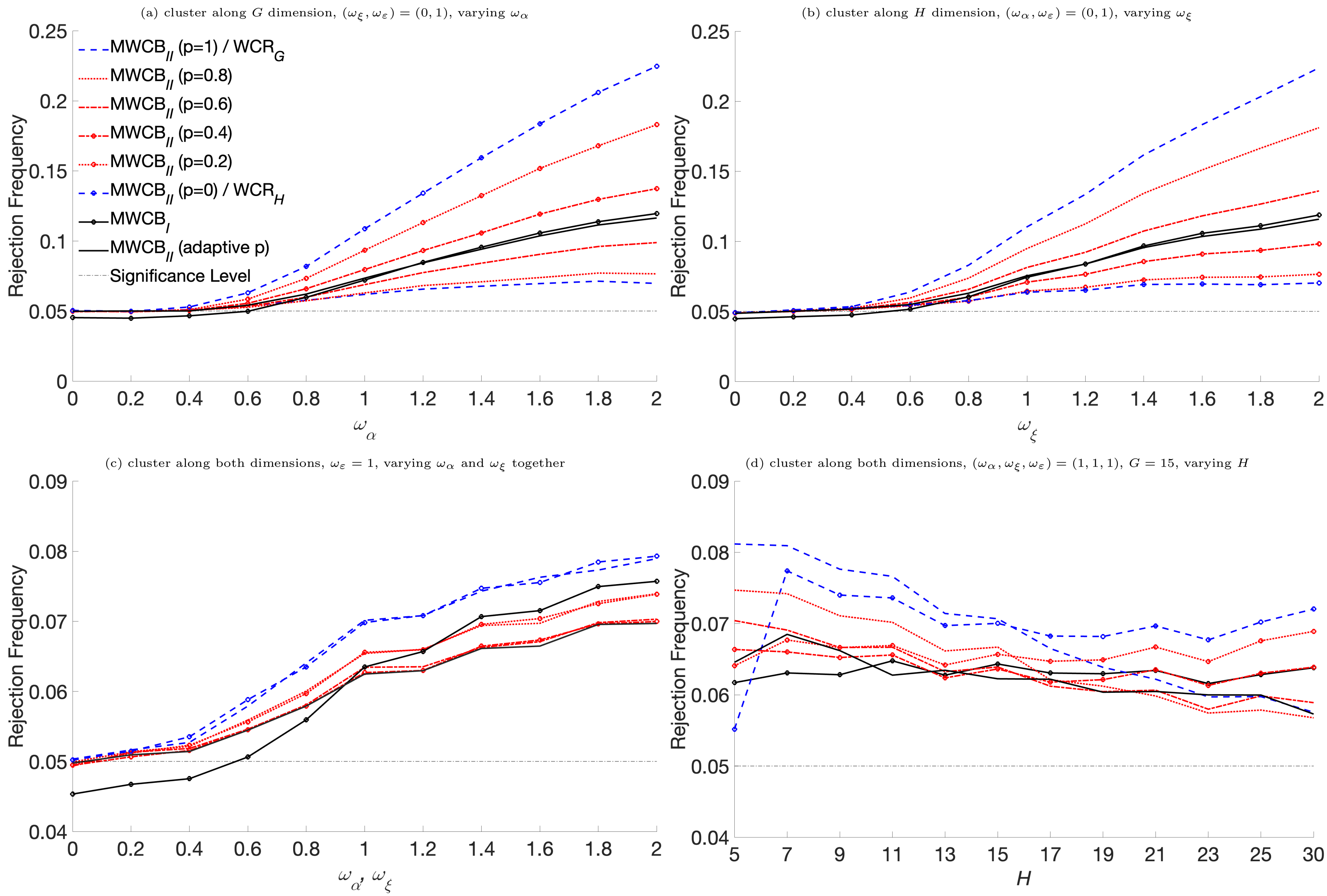


Figure 3: **Simulations with different data structures for the standard multiway clustering.** The simulation is based on the standard multiway clustering setting. We perform 100,000 replications for each simulation in (4.1). We set the number of bootstrap replications to $B = 1,999$, the default number of clusters $G = H = 15$, and rejection criteria based on the equal-tail bootstrap P value ($P_E^* < 0.05$). All bootstrap methods apply the CGM variance estimator. The y-axis in Panels (c) and (d) has been zoomed in for greater detail.

second dimension becomes crucial. Consequently, WCR_H performs well under these conditions. MWCB_{II} ($p = 0.2$) underperforms relatively compared to WCR_H, as it assigns relatively less weight to the second dimension. At the other extreme, when $H = 30$, dependence in the first dimension takes precedence, making WCR_G the better-performing choice. Interestingly, MWCB_{II} (with adaptive p) performs as well as WCR_G when adapting to large values of G. MWCB_{II} (adaptive p) adjusts effectively to varying conditions, performing well across different scenarios by automatically calibrating to the cluster configuration. While MWCB_I also adapts to the number of clusters, it performs better when H is small but is outperformed by MWCB_{II} (adaptive p) when H is large.

4.2 Multiway Clustering with Time Series

Next, we set up the DGP allowing the time serial dependence. The model still follows the setting in (4.1). We choose the balanced level of dependence with parameter values $(\omega_\alpha, \omega_\xi, \omega_\varepsilon) = (1, 1, 1)$. The key difference is that instead of letting ξ_h to be independent over h, we allow it to be dependent by following an AR(1) procedure:

$$\xi_h = \rho \xi_{h-1} + \widetilde{\xi}_h, \text{ where } \widetilde{\xi}_h \text{ are independent draws from } \mathcal{N}(0, 1 - \rho^2). \tag{4.4}$$

We compare various methods in this setting, including the CLT outcomes of the CGM variance, the CHS variance, and our new variant of the CHS variance, $\widehat{\boldsymbol{V}}_{CHS,V}$, as defined in (3.9). Additionally, we consider the CV variance and two bias-corrected variances proposed by Chen and Vogelsang (2024): the bias-corrected CHS variance, $\widetilde{\boldsymbol{V}}_{CHS}$, given in (7.3), and the bias-corrected CV variance, $\widetilde{\boldsymbol{V}}_{CV}$, given in (7.4). Finally, we test various wild bootstrap methods based on the CGM, CHS, and CV variances, respectively.

The most relevant results outcome depicted in Figure 4, suggests that the variance variant $\widehat{\boldsymbol{V}}_{CHS,V}$ performs similarly to, but slightly better than, $\widehat{\boldsymbol{V}}_{CHS}$. The finding from comparing $\widehat{\boldsymbol{V}}_{CHS}$ and $\widetilde{\boldsymbol{V}}_{CHS}$ aligns with the findings of Chen and Vogelsang (2024), indicating that the incorporation of the bias-corrected term results in improved finite sample performance.

Moreover, the result suggests that the utilization of $\widehat{\boldsymbol{V}}_{CHS}$ to generate original and bootstrap

t-statistics enhances the results based on the CLT: WCR_G and WCR_H based on $\widehat{\boldsymbol{V}}_{CHS}$ outperform the CLT result of $\widehat{\boldsymbol{V}}_{CHS}$ for all levels of dependence assumed. Even though the bootstrap DGP does not mimic the DGP, simply employing the bootstrap method with the identical variance estimator to obtain the original and bootstrap statistics can still enhance the finite sample performance. This pattern is discussed formally in Goncalves and Vogelsang (2011) and Davidson and MacKinnon et al. (2010).

In cases where ρ is small, WCR_G and two MWCB methods exhibit similar and reasonably robust rejection frequencies. WCR_H performs slightly worse since the number of clusters H is greater than G. However, with an increase in ρ, WCR_G and WCR_H generate rejection rates much higher than the significance level. Consequently, it is evident that the two MWCB approaches consistently exhibit minimal deviation across various levels of serial dependence, showcasing their effective management of temporal dependence.

Additional results presented in Table C.1 in Appendix C illustrate the performance of other methods. Overall, the findings suggest that, compared to using a bootstrap method and a variance estimator that are both not robust to time dependence, simply employing either a robust bootstrap method with a non-robust variance estimator or a non-robust bootstrap method with a variance estimator that accounts for time dependence can improve finite-sample performance. However, the most effective approach is to use a robust bootstrap method paired with a robust variance estimator. Consequently, based on the performance of various methods, the MWCB_I method with the CHS variance is recommended.

4.3 Fixed Effect Setting

In this subsection, we further consider the effect of fixed effect. There is a common yet incorrect belief that a two-way cluster robust estimator is not necessary for a two-way fixed effect model. This misconception arises because the cluster dependence can often be absorbed by the two-way fixed effects, e.g. the simulation settings in Sections 4.1 and 4.2. To address this, Chiang et al. (2024) and MacKinnon, Nielsen, and Webb (2024) provide innovative examples demonstrating situations in which two-way fixed effects fail to fully account for cluster dependence. Following

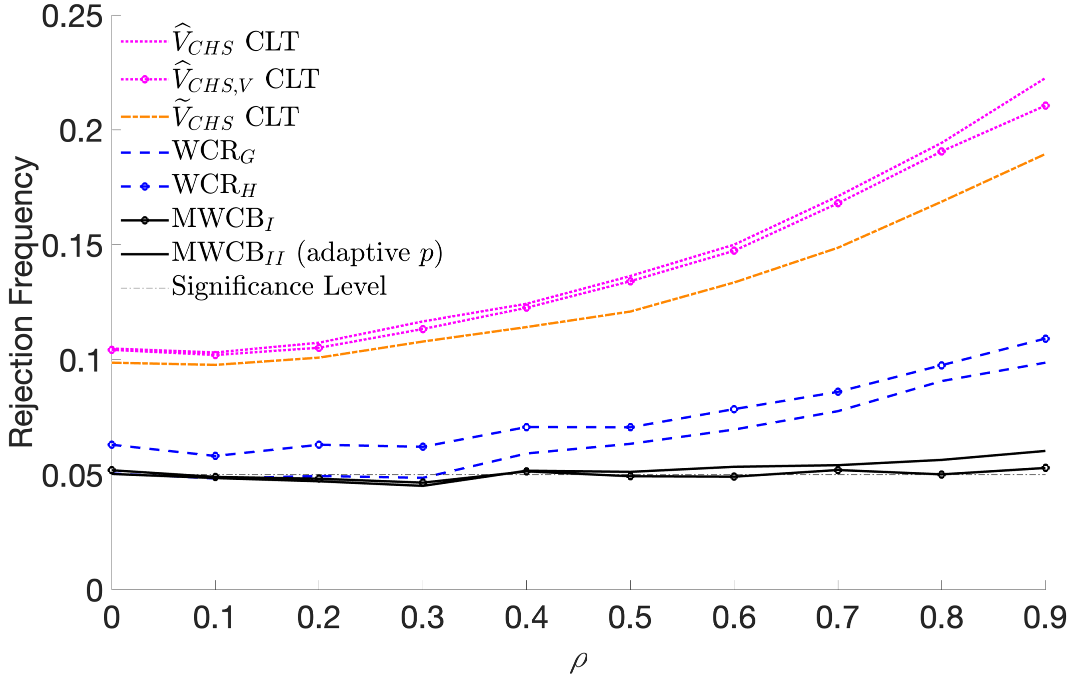


Figure 4: **Multiway clustering simulation with varying levels of serial dependence.** This figure shows the rejection frequency under the multiway clustering setting with varying levels of serial dependence. ρ represents the level of serial dependence. We perform 10,000 replications for each simulation. We set the number of bootstrap replications to $B = 1,999$, ensuring $G = 25$ and $H = 50$, an equal level of dependence in both dimensions $(\omega_\alpha, \omega_\xi, \omega_\varepsilon) = (1, 1, 1)$, and rejection criteria based on the equal-tail bootstrap P value $(P_E^* < 0.05)$. All bootstrap methods apply the CHS variance estimator.

their ideas, we consider a scenario where the cluster-robust methods are necessary:

$$y_{gh} = \beta_1 + \sum_{k=2}^{10} \beta_k X_{gh,k} + \gamma_g + \gamma_h + u_{gh}, \tag{4.5}$$

where γ_g and γ_h are fixed effects in the two dimensions; u_{gh} and $X_{gh,k}$ are generated as follows:

$$u_{gh} = \omega_\alpha \alpha_{gh}^u + \omega_\xi \xi_{gh}^u + \omega_\varepsilon \varepsilon_{gh}^u, \quad \text{and}$$
$$X_{gh,k} = \omega_\alpha \alpha_{gh,k}^x + \omega_\xi \xi_{gh,k}^x + \omega_\varepsilon \varepsilon_{gh,k}^x,$$

where

$$\alpha_{gh}^u = \delta_h^\alpha \alpha_g^{u,1} + (1 - \delta_h^\alpha) \alpha_g^{u,2}, \qquad \xi_{gh}^u = \delta_g^\xi \xi_h^{u,1} + (1 - \delta_g^\xi) \xi_h^{u,2},$$
$$\alpha_{gh,k}^x = \delta_h^\alpha \alpha_{g,k}^{x,1} + (1 - \delta_h^\alpha) \alpha_{g,k}^{x,2}, \qquad \xi_{gh,k}^x = \delta_g^\xi \xi_{h,k}^{x,1} + (1 - \delta_g^\xi) \xi_{h,k}^{x,2},$$

and δ_h^α and δ_g^ξ are binary random variables that can be either 0 or 1, independent over h and g, respectively. Other random variables are generated similarly to those in model (4.1). For instance, the variables $\alpha_g^{u,1}$ and $\alpha_g^{u,2}$ are produced as α_g^u in model (4.1). Here, the superscripts

1 and 2 denote two groups within the same cluster. Random variables from different groups are mutually independent.

As discussed in MacKinnon, Nielsen, and Webb (2024), the main concept is that within each cluster, there exist different types/sub-clusters of observations. For example, suppose the first dimension, G, represents industries, and the second dimension, H, represents time. In a given year h, some seasons may see strong sales (favorable conditions), while others may experience weak sales (unfavorable conditions). The key observation here is that while researchers can observe annual data, they lack access to finer seasonal or monthly details (favorable or unfavorable). Consequently, the time-fixed effect (γ_h) captures the average impact across all seasons within a given year h but cannot distinguish the specific temporal effects encountered by individual sub-groups.

A similar concept applies to the industry dimension. For example, within the car industry, some electric vehicle firms may be in a growth phase (sunrise period), while traditional gas car firms may be experiencing a decline (sunset period). While researchers can observe industry-level data, they lack visibility into specific firm types or categories (sunrise or sunset). As a result, the industry-fixed effect (γ_g) captures the average impact across all firms within a given industry g but cannot capture the specific effects experienced by distinct sub-groups. The probability of δ_h^α and δ_g^ξ being 1 determines the sub-cluster size, and we set this probability to 0.5.

Note that when the two-way fixed effect model is considered, we follow all the same discussions to generate the variance and bootstrap, with the main difference being that we first demean the regressors and regressands:

$$\dot{\boldsymbol{X}}_{gh} = \boldsymbol{X}_{gh} - \bar{\boldsymbol{X}}_g - \bar{\boldsymbol{X}}_h + \bar{\boldsymbol{X}},$$

$$\dot{\boldsymbol{y}}_{gh} = \boldsymbol{y}_{gh} - \bar{\boldsymbol{y}}_g - \bar{\boldsymbol{y}}_h + \bar{\boldsymbol{Y}},$$

where $\bar{\boldsymbol{X}}_g$ ($\bar{\boldsymbol{y}}_g$), $\bar{\boldsymbol{X}}_h$ ($\bar{\boldsymbol{y}}_h$), and $\bar{\boldsymbol{X}}$ ($\bar{\boldsymbol{y}}$) are the mean values of $\boldsymbol{X}_{gh,i}$ ($y_{gh,i}$) for all N_g observations in the g^{th} cluster in the first dimension, all N_h observations in the h^{th} cluster in the second dimension, and all N observations, respectively. We then substitute $\boldsymbol{X}_{gh}$ and $\boldsymbol{y}_{gh}$ with $\dot{\boldsymbol{X}}_{gh}$ and $\dot{\boldsymbol{y}}_{gh}$, respectively, applying this change throughout all discussions above to estimate original and bootstrap t-statistics.[11] The simulation results are very similar to the results in Sections 4.1 and 4.2, so we display the results in Appendix C to save space.

5 Revisiting the Working From Home (WFH) Case

We revisit the influential study by Bloom, Liang, Roberts, and Ying (hereafter BLRY) (2015) on the impact of working from home (WFH) on employee performance. Their study was based on a WFH experiment conducted at CTrip, a large Chinese travel agency with 16,000 employees listed on NASDAQ. The core model they estimated is specified as follows:

$$\text{performance}_{it} = \alpha \ \text{treat}_i \times \text{experiment}_t + \beta_t + \gamma_i + \varepsilon_{it}, \tag{5.1}$$

where i and t denote individual and time, respectively. Here, performance_{it} represents the performance level of employee i at time t. performance_{it} can take various forms, such as the log of weekly phone calls answered (log phones per week), log of phone calls answered per minute (log phones per minute), log of the weekly sum of minutes on the phone (log sum minute by week), or an overall performance z-score measure (overall performance). treat_i is a dummy variable indicating if the individual belongs to the treatment group, and experiment_t is a dummy variable that equals 1 during the experimental period. β_t and γ_i reflect time fixed effect and individual fixed effect, respectively. The coefficient of interest, α, captures the effect of WFH on performance.

BLRY employed two-way clustered standard errors as proposed by CGM and found that working from home led to a statistically significant 13%(=exp(0.1198)-1) increase in performance. This improvement was driven by a 9%(=exp(0.0881)-1) increase in total minutes worked per week and a 4%(=exp(0.0317)-1) increase in the number of calls handled per minute.

However, it is plausible that the error terms ε_{it} are correlated over time due to skill accumulation, especially in a high-turnover environment like a call center. This setting often includes many new employees with limited prior experience, leading to a learning curve that affects performance

[11] The β_1 term, serving as the intercept in the simulation, is omitted because it is eliminated during the demeaning process.

Performance Measures	$\widehat{\alpha}$	CLT P value				Bootstrap P value			
		CGM	CHS	BCCHS	CHS-V	WCR_G	WCR_H	MWCB_I	MWCB_{II}
Overall performance	0.2324	0.0007	0.0006	0.0015	0.0006	0.0000	0.0020	0.0065	0.0290
Log phones per week	0.1198	0.0000	0.0001	0.0001	0.0000	0.0002	0.0004	0.0014	0.0050
Log phones per minute	0.0317	0.0365	0.0207	0.0338	0.0197	0.0142	0.0380	0.0530	0.0764
Log sum minutes by week	0.0881	0.0034	0.0034	0.0056	0.0034	0.0040	0.0058	0.0210	0.0348

Table 2: **Inference result for the effects of working from home.** This table shows the coefficient estimates, CLT P values, and bootstrap P values for the working-from-home (WFH) model studies by Bloom et al. (2015). The first (G) and second (H) dimensions correspond to individual and time, respectively. CGM and CHS denote the variance estimators proposed by Cameron et al. (2011) and Chiang et al. (2024), respectively. BCCHS is the bias-corrected variant of CHS by Chen and Volgesang (2024) as defined in (7.3), and CHS-V is the new variant of CHS as defined in (3.9). MWCB_{II} applies the adaptive $p = \frac{H}{G+H}$. All bootstraps utilize the CHS variance estimator. All bootstrap P values are equal-tail and based on the null hypothesis that the true coefficient value is zero. The number of bootstrap replications $B = 9,999$.

over time. A Ljung-Box Q-test on the overall performance data reveals that 58.23% of individuals exhibit significant autocorrelation at the 5% significance level. The effect is even more pronounced within the treatment group, where 61.83% of individuals show autocorrelation, suggesting stronger serial dependence among those working from home. Developing a consistent remote work routine takes time. The longer someone works remotely, the better they might become at managing distractions or using digital collaboration tools.

Table 2 reproduces and extends the results from BLRY's Table 2 using the publicly available dataset at `www.stanford.edu/~nbloom/WFH.zip`. In the first row, we analyze the model by utilizing overall performance as the measure of performance. Applying the CHS variance, the CLT results show a similar effect of WFH compared to the CGM variance, with a P-value of 0.0006. Both BCCHS and CHS-V estimators confirm significance at the 1% level. Most bootstraps align with this finding, except for MWCB_{II}, which yields a weaker significance level with a P-value of 0.0290. However, given the stronger performance of MWCB_I in simulations under time dependence, we maintain that the effect remains significant at the 1% level.

The second row focuses on the effect of WFH on weekly phone calls answered. Here, all methods consistently show a significant improvement at the 1% level, confirming that WFH leads to more calls handled per week.

The third and fourth rows explore the components of this improvement—specifically, phone calls answered per minute and total weekly call time. While all other methods suggest that WFH significantly boosts calls per minute at the 5% significance level, both MWCB methods indicate only marginal significance. We observe that bootstrap results using MWCB_I and MWCB_{II} tend to produce higher P-values, likely due to their robustness against serial dependence, reducing the likelihood of over-rejecting the null hypothesis, as shown in the simulations. Similarly, while all other methods suggest a significant improvement in total minutes by week at the 1% level, the MWCB approaches suggest a weaker significance, only at the 5% level.

In summary, at the 5% significance level, WFH appears to enhance performance, primarily by increasing the total time spent on calls, rather than by a rise in the number of phone calls handled per minute.

6 Conclusion

The main contribution of this paper is the proposal of two new bootstrap procedures for inference in linear regression models with multiway clustering. We demonstrated the robustness of our proposed multiway wild cluster bootstraps (MWCBs) in two main types of multiway clustering settings: (1) standard multiway clustering, assuming independence between intersections lacking shared clusters, and (2) multiway clustering with a time dimension, accounting for serial dependence across clusters in time.

Our methodologies, compared to the standard wild cluster bootstrap by Cameron et al. (2008), offer two significant advantages. Firstly, this approach allows for the customization of dependence levels, offering flexibility through the chosen weights assigned to each dimension. Additionally, our method enables simultaneous control of dependence along multiple dimensions, a feature absent in the standard wild cluster bootstrap.

We provide a set of regularity conditions under which MWCBs are asymptotically valid, along with theoretical evidence supporting its validity in two scenarios. In the standard multiway clustering context, our method remains valid irrespective of whether $\widehat{\boldsymbol{V}}_{CGM}$ or $\widehat{\boldsymbol{V}}_{DHG}$ (with the double-counting term added back) is employed, regardless of whether the DGP involves clustering along at least one dimension, by intersections, or not clustered at all. The broad applicability of our method is not replicated in existing wild cluster bootstrap methods. Furthermore, in the context of multiway clustering with a time dimension, our method remains valid regardless of whether

$\widehat{\boldsymbol{V}}_{CHS}$ or $\widehat{\boldsymbol{V}}_{CV}$ (with the double-counting term added back) is used, irrespective of whether the DGP involves clustering along at least one dimension, by intersections, or not clustered at all.

Our simulation results indicate that MWCB delivers consistent and reliable performance across a wide range of scenarios, including cases with dependence along a single dimension, dependence across both dimensions, and different cluster structures. Furthermore, in the scenario of multiway clustering with a time dimension, our method MWCB_I, particularly with $\widehat{\boldsymbol{V}}_{CHS}$, emerges as the most effective approach, adeptly addressing serial dependence – a finding consistent with our theoretical demonstration.

The empirical findings on the working-from-home case reveal a significant impact of home working on the performance of employees. Notably, the improvement primarily stems from an increase in total time spent on the phone each week, rather than a higher rate of calls handled per minute.

MWCB_I demonstrates applicability and ease of implementation in contexts involving spatial information or network dependence. Nevertheless, it is crucial to acknowledge that our examination of the theoretical asymptotic validity did not encompass scenarios where the individuals' or firms' effects exhibit spatial dependence. This extension is reserved for future research.

7 Appendix

This appendix collects some extra findings and details omitted from the main text. We present the rule of thumb for choosing the parameter q in MWCB_{II}, along with the bootstrap variance estimators employed for the MWCB methods when the serial dependence is presented. We also provide the explicit form of the two bias-corrected CRVE $\widetilde{\boldsymbol{V}}_{CHS}$ and $\widetilde{\boldsymbol{V}}_{CV}$, proposed by Chen and Vogelsang (2024). Then, we introduce a generalized MWCB_{II} that works when the statistic of interest is not studentized. Finally, we state the Lemma regarding the bootstrap variance $\boldsymbol{V}^*$.

In practice, a key objective when selecting the parameter is to minimize the mean squared error. We choose $q = 0.25^{\omega^{-1/5} R^{-1/5}}$, where ω represents a measure of autocorrelation, as defined in Andrews (1991): for $k = 1, \ldots, K$, let $\widehat{s}_{h,k}$ be the kth element of $\widehat{\boldsymbol{s}}_h$, and $\widehat{\rho}_k$ is obtained by regressing $\widehat{s}_{h,k}$ on $\widehat{s}_{h-1,k}$, then $\omega = \sum_{k=1}^{K} \frac{\widehat{\rho}_k^2}{(1-\widehat{\rho}_k)^4} / \sum_{k=1}^{K} \frac{(1-\widehat{\rho}_k^2)^2}{(1-\widehat{\rho}_k)^4}$.

We define the bootstrap CHS and CV variance estimators for both MWCB methods as follows:

$$\begin{aligned}\widehat{\boldsymbol{V}}_{CHS}^{*} =& \frac{1}{(GH)^2}\widehat{\boldsymbol{Q}}^{-1}\Bigg[\sum_{g}\widehat{\boldsymbol{s}}_{g}^{*}\widehat{\boldsymbol{s}}_{g}^{*\top} + \sum_{h}\widehat{\boldsymbol{s}}_{h}^{*}\widehat{\boldsymbol{s}}_{h}^{*\top} - \sum_{g}\sum_{h}\widehat{\boldsymbol{s}}_{gh}^{*}\widehat{\boldsymbol{s}}_{gh}^{*\top} \\ &+ \sum_{\iota=1}^{\ell-1}\sum_{h=1}^{H-\iota}\left(\widehat{\boldsymbol{s}}_{h}^{*}\widehat{\boldsymbol{s}}_{h+\iota}^{*\top} + \widehat{\boldsymbol{s}}_{h+\iota}^{*}\widehat{\boldsymbol{s}}_{h}^{*\top} - \sum_{g=1}^{G}\left(\widehat{\boldsymbol{s}}_{gh}^{*}\widehat{\boldsymbol{s}}_{gh+\iota}^{*\top} + \widehat{\boldsymbol{s}}_{gh+\iota}^{*}\widehat{\boldsymbol{s}}_{gh}^{*\top}\right)\right)\Bigg]\widehat{\boldsymbol{Q}}^{-1},\end{aligned} \tag{7.1}$$

and

$$\widehat{\boldsymbol{V}}_{CV}^{*} = \frac{1}{(GH)^2}\widehat{\boldsymbol{Q}}^{-1}\left[\sum_{g}\widehat{\boldsymbol{s}}_{g}^{*}\widehat{\boldsymbol{s}}_{g}^{*\top} + \sum_{h}\widehat{\boldsymbol{s}}_{h}^{*}\widehat{\boldsymbol{s}}_{h}^{*\top} + \sum_{\iota=1}^{\ell-1}\sum_{h=1}^{H-\iota}\left(\widehat{\boldsymbol{s}}_{h}^{*}\widehat{\boldsymbol{s}}_{h+\iota}^{*\top} + \widehat{\boldsymbol{s}}_{h+\iota}^{*}\widehat{\boldsymbol{s}}_{h}^{*\top}\right)\right]\widehat{\boldsymbol{Q}}^{-1}. \tag{7.2}$$

In Chen and Volgesang (2024), the bias-corrected CHS and CV variance estimators are defined as follows:

$$\begin{aligned}\widetilde{\boldsymbol{V}}_{CHS} =& \left(1 - \frac{\ell}{H} + \frac{1}{3}\left(\frac{\ell}{H}\right)^2\right)^{-1}\frac{1}{(GH)^2}\widehat{\boldsymbol{Q}}^{-1}\Bigg[\sum_{g}\widehat{\boldsymbol{s}}_{g}\widehat{\boldsymbol{s}}_{g}^{\top} + \sum_{h}\widehat{\boldsymbol{s}}_{h}\widehat{\boldsymbol{s}}_{h}^{\top} - \sum_{g}\sum_{h}\widehat{\boldsymbol{s}}_{gh}\widehat{\boldsymbol{s}}_{gh}^{\top} \\ &+ \sum_{\iota=1}^{\ell-1}w(\iota,\ell)\sum_{h=1}^{H-\iota}\left(\widehat{\boldsymbol{s}}_{h}\widehat{\boldsymbol{s}}_{h+\iota}^{\top} + \widehat{\boldsymbol{s}}_{h+\iota}\widehat{\boldsymbol{s}}_{h}^{\top} - \sum_{g=1}^{G}\left(\widehat{\boldsymbol{s}}_{gh}\widehat{\boldsymbol{s}}_{gh+\iota}^{\top} + \widehat{\boldsymbol{s}}_{gh+\iota}\widehat{\boldsymbol{s}}_{gh}^{\top}\right)\right)\Bigg]\widehat{\boldsymbol{Q}}^{-1},\end{aligned} \tag{7.3}$$

and

$$\begin{aligned}\widetilde{\boldsymbol{V}}_{CV} =& \frac{1}{(GH)^2}\widehat{\boldsymbol{Q}}^{-1}\Bigg\{\sum_{g}\widehat{\boldsymbol{s}}_{g}\widehat{\boldsymbol{s}}_{g}^{\top} \\ &+ \left(1 - \frac{\ell}{H} + \frac{1}{3}\left(\frac{\ell}{H}\right)^2\right)^{-1}\left[\sum_{h}\widehat{\boldsymbol{s}}_{h}\widehat{\boldsymbol{s}}_{h}^{\top} + \sum_{\iota=1}^{\ell-1}w(\iota,\ell)\sum_{h=1}^{H-\iota}\left(\widehat{\boldsymbol{s}}_{h}\widehat{\boldsymbol{s}}_{h+\iota}^{\top} + \widehat{\boldsymbol{s}}_{h+\iota}\widehat{\boldsymbol{s}}_{h}^{\top}\right)\right]\Bigg\}\widehat{\boldsymbol{Q}}^{-1}.\end{aligned} \tag{7.4}$$

A general MWCB$_{II}$ bootstrap is provided for the non-standardized statistic $\boldsymbol{a}^{\top}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$. The random weights are obtained as follows:

$$\nu_{gh}^{*} = \begin{cases} \varsigma_1\nu_g^{*}, & \text{with probability } p, \\ \varsigma_2\nu_h^{*}, & \text{with probability } 1-p, \end{cases} \tag{7.5}$$

with ν_g^{*} and ν_h^{*} mutually independent with $(0, \sigma_{\nu}^2)$ over g and h, respectively. Similar to MWCB$_I$, the scaling factors ς_1 and ς_2 can be chosen as any nonzero finite constants for the studentized statistics. However, when

$$\varsigma_1 = p^{-1} \quad \text{and} \quad \varsigma_2 = (1-p)^{-1}, \tag{7.6}$$

the MWCB$_{II}$ bootstrap can be applied to consistently estimate the asymptotic variance of $\widehat{\boldsymbol{\beta}}$. The detailed proof can be found in the proof of Theorems 3.7 and 3.10 in two different scenarios.

Lemma 7.1. *Assume that the bootstrap scores $\boldsymbol{s}^*_{gh,i}$ are generated following the process in (3.5), the scaling terms χ_1 and χ_2 are defined as in (3.6). Moreover, suppose Assumptions 7-12 hold, strengthened by Assumption 10'. Then, no matter the DGP is clustered along at least one dimension, clustered only by intersections, or not clustered at all,*

$$\boldsymbol{V}^* = \sigma_\nu^2 \widehat{\boldsymbol{V}}_{CV} + \widehat{\boldsymbol{e}}_{GH},$$

where $\widehat{\boldsymbol{e}}_{GH}$ is a neglibible term, i.e., $\left(\sigma_\nu^2 \widehat{\boldsymbol{V}}_{CV}\right)^{-1} \widehat{\boldsymbol{e}}_{GH} = o_P(1)$ holds.

References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Arellano, M. (1987). Computing robust standard errors for within-groups estimators. *Oxford bulletin of Economics and Statistics 49*(4), 431–434.

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.

Bloom, N., J. Liang, J. Roberts, and Z. J. Ying (2015). Does working from home work? evidence from a chinese experiment. *The Quarterly Journal of Economics 130*(1), 165–218.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics 90*(3), 414–427.

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.

Canay, I. A., A. Santos, and A. M. Shaikh (2021). The wild bootstrap with a "small" number of "large" clusters. *Review of Economics and Statistics 103*(2), 346–363.

Card, D., A. R. Cardoso, and P. Kline (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics 131*(2), 633–686.

Carter, A. V., K. T. Schnepel, and D. G. Steigerwald (2017). Asymptotic behavior of at-test robust to cluster heterogeneity. *Review of Economics and Statistics 99*(4), 698–709.

Chen, K. and T. J. Vogelsang (2024). Fixed-b asymptotics for panel models with two-way clustering. *Journal of Econometrics 244*(1), 105831.

Chiang, H. D., B. E. Hansen, and Y. Sasaki (2024). Standard Errors for Two-Way Clustering with Serially Correlated Time Effects. *The Review of Economics and Statistics*, 1–40.

Chiang, H. D., K. Kato, and Y. Sasaki (2023). Inference for high-dimensional exchangeable arrays. *Journal of the American Statistical Association 118*(543), 1595–1605.

Conley, T. G., S. Gonçalves, M. S. Kim, and B. Perron (2023). Bootstrap inference under cross-sectional dependence. *Quantitative Economics 14*(2), 511–569.

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

Davidson, R. and J. G. MacKinnon (2010). Wild bootstrap tests for iv regression. *Journal of Business & Economic Statistics 28*(1), 128–144.

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.

Driscoll, J. C. and A. C. Kraay (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics 80*(4), 549–560.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics 7*(1), 1–26.

Fieler, A. C. (2011). Nonhomotheticity and bilateral trade: Evidence and a quantitative explanation. *Econometrica 79*(4), 1069–1101.

Gonçalves, S. and T. J. Vogelsang (2011). Block bootstrap hac robust tests: The sophistication of the naive bootstrap. *Econometric Theory 27*(4), 745–791.

Hansen, B. E. (1992). The likelihood ratio test under nonstandard conditions: testing the markov switching model of gnp. *Journal of Applied Econometrics 7*(S1), S61–S82.

Hansen, B. E. and S. Lee (2019). Asymptotic theory for clustered samples. *Journal of Econometrics 210*(2), 268–290.

Hounyo, U. (2023). A wild bootstrap for dependent data. *Econometric Theory 39*(2), 264–289.

Korovkin, V. and A. Makarin (2023). Conflict and intergroup trade: Evidence from the 2014 russia-ukraine crisis. *American Economic Review 113*(1), 34–70.

Liu, R. Y. (1988). Bootstrap procedures under some non-iid models. *Annals of Statistics 16*(4), 1696–1708.

MacKinnon, J. G. (2012). Thirty years of heteroskedasticity-robust inference. In *Recent advances and future directions in causality, prediction, and specification analysis: Essays in honor of Halbert L. White Jr*, pp. 437–461. Springer.

MacKinnon, J. G. (2023). Fast cluster bootstrap methods for linear regression models. *Econometrics and Statistics 26*, 52–71.

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.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2023a). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics 232*(2), 272–299.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2023b). Fast and reliable jackknife and bootstrap methods for cluster-robust inference. *Journal of Applied Econometrics 38*(5), 671–694.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2024). Jackknife inference with two-way clustering. *arXiv preprint arXiv:2406.08880*.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics 21*(1), 255–285.

McCullagh, P. (2000). Resampling and exchangeable arrays. *Bernoulli 6*(2), 285–301.

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

Newey, W. K. and K. D. West (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica 55*(3), 703–708.

Owen, A. B. (2007). The pigeonhole bootstrap. *The Annuals of Applied Statistics 1*(2), 386–411.

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

Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association 105*(489), 218–235.

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics 99*(1), 1–10.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817–838.

Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics 14*(4), 1261–1295.

Zeger, S. L. and K.-Y. Liang (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 121–130.