Jackknife Inference with Two-Way Clustering

James G. MacKinnon (Queen's University and ACE)
Morten Ørregaard Nielsen (Aarhus University, ACE)
Matthew D. Webb (Carleton University)

CFE-CMStatistics, December 15, 2025

Queen's University

QED
Queen's Economics Department

Introduction

Introduction

For models with cross-section or panel data, the disturbances may be clustered in two dimensions.

Introduction

For models with cross-section or panel data, the disturbances may be clustered in two dimensions.

Unfortunately, the finite-sample properties of two-way cluster-robust tests and confidence intervals are often poor.

Introduction

For models with cross-section or panel data, the disturbances may be clustered in two dimensions.

Unfortunately, the finite-sample properties of two-way cluster-robust tests and confidence intervals are often poor.

There can also be undefined standard errors when a cluster-robust variance estimator (CRVE) is not positive definite.

Introduction

For models with cross-section or panel data, the disturbances may be clustered in two dimensions.

Unfortunately, the finite-sample properties of two-way cluster-robust tests and confidence intervals are often poor.

There can also be undefined standard errors when a cluster-robust variance estimator (CRVE) is not positive definite.

We discuss ways to improve inference with two-way clustering:

Introduction

For models with cross-section or panel data, the disturbances may be clustered in two dimensions.

Unfortunately, the finite-sample properties of two-way cluster-robust tests and confidence intervals are often poor.

There can also be undefined standard errors when a cluster-robust variance estimator (CRVE) is not positive definite.

We discuss ways to improve inference with two-way clustering:

- Two existing methods for dealing with undefined standard errors.

Introduction

For models with cross-section or panel data, the disturbances may be clustered in two dimensions.

Unfortunately, the finite-sample properties of two-way cluster-robust tests and confidence intervals are often poor.

There can also be undefined standard errors when a cluster-robust variance estimator (CRVE) is not positive definite.

We discuss ways to improve inference with two-way clustering:

- Two existing methods for dealing with undefined standard errors.
- A new method that completely avoids undefined standard errors.

Introduction

For models with cross-section or panel data, the disturbances may be clustered in two dimensions.

Unfortunately, the finite-sample properties of two-way cluster-robust tests and confidence intervals are often poor.

There can also be undefined standard errors when a cluster-robust variance estimator (CRVE) is not positive definite.

We discuss ways to improve inference with two-way clustering:

- Two existing methods for dealing with undefined standard errors.
- A new method that completely avoids undefined standard errors.
- Several new two-way CRVEs based on the cluster jackknife.

Introduction

For models with cross-section or panel data, the disturbances may be clustered in two dimensions.

Unfortunately, the finite-sample properties of two-way cluster-robust tests and confidence intervals are often poor.

There can also be undefined standard errors when a cluster-robust variance estimator (CRVE) is not positive definite.

We discuss ways to improve inference with two-way clustering:

- Two existing methods for dealing with undefined standard errors.
- A new method that completely avoids undefined standard errors.
- Several new two-way CRVEs based on the cluster jackknife.

Simulations for models with two-way fixed effects suggest that a cluster-jackknife CRVE based on the new method often yields surprisingly accurate inferences.

Literature

Literature

- One-way CRVEs: Liang and Zeger (1986); Arellano (1987); Bester, Conley, and Hansen (2011); Hansen and Lee (2019); Djogbenou, MacKinnon, and Nielsen (2019).

Literature

- One-way CRVEs: Liang and Zeger (1986); Arellano (1987); Bester, Conley, and Hansen (2011); Hansen and Lee (2019); Djogbenou, MacKinnon, and Nielsen (2019).
- Two-way CRVEs: Cameron, Gelbach, and Miller (2011); Miglioretti and Heagerty (2006); Thompson (2011).

Literature

- One-way CRVEs: Liang and Zeger (1986); Arellano (1987); Bester, Conley, and Hansen (2011); Hansen and Lee (2019); Djogbenou, MacKinnon, and Nielsen (2019).
- Two-way CRVEs: Cameron, Gelbach, and Miller (2011); Miglioretti and Heagerty (2006); Thompson (2011).
- Theory for two-way CRVEs: Davezies, D'Haultfoeuille, and Guyonvarch (2021, 2025); Menzel (2021); Chiang, Kato, and Sasaki (2023), Yap (2025).

Literature

- One-way CRVEs: Liang and Zeger (1986); Arellano (1987); Bester, Conley, and Hansen (2011); Hansen and Lee (2019); Djogbenou, MacKinnon, and Nielsen (2019).
- Two-way CRVEs: Cameron, Gelbach, and Miller (2011); Miglioretti and Heagerty (2006); Thompson (2011).
- Theory for two-way CRVEs: Davezies, D'Haultfoeuille, and Guyonvarch (2021, 2025); Menzel (2021); Chiang, Kato, and Sasaki (2023), Yap (2025).
- Wild bootstrap methods: MacKinnon, Neilsen, and Webb (2021); Hounyo and Lin (2025). `boottest`. [Not entirely satisfactory]

Literature

- One-way CRVEs: Liang and Zeger (1986); Arellano (1987); Bester, Conley, and Hansen (2011); Hansen and Lee (2019); Djogbenou, MacKinnon, and Nielsen (2019).
- Two-way CRVEs: Cameron, Gelbach, and Miller (2011); Miglioretti and Heagerty (2006); Thompson (2011).
- Theory for two-way CRVEs: Davezies, D'Haultfoeuille, and Guyonvarch (2021, 2025); Menzel (2021); Chiang, Kato, and Sasaki (2023), Yap (2025).
- Wild bootstrap methods: MacKinnon, Neilsen, and Webb (2021); Hounyo and Lin (2025). `boottest`. [Not entirely satisfactory]
- Jackknife variance estimation: Tukey (1958); Efron (1981); Efron and Stein (1981); MacKinnon and White (1985); Bell and McCaffrey (2002); MacKinnon, Nielsen, and Webb (JAE 2023, SJ 2023); Hansen (2025 WP & JAE).

Linear Regression with Two-Way Clustering

Linear Regression with Two-Way Clustering

With two-way clustering, $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ can be written as

$$\boldsymbol{y}_{gh} = \boldsymbol{X}_{gh}\boldsymbol{\beta} + \boldsymbol{u}_{gh}, \quad g = 1, \ldots, G, \;\; h = 1, \ldots, H. \tag{2}$$

Linear Regression with Two-Way Clustering

With two-way clustering, $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ can be written as

$$\boldsymbol{y}_{gh} = \boldsymbol{X}_{gh}\boldsymbol{\beta} + \boldsymbol{u}_{gh}, \quad g = 1, \ldots, G, \;\; h = 1, \ldots, H. \tag{2}$$

The variance matrix of the score vector $\boldsymbol{s} = \boldsymbol{X}^\top \boldsymbol{u}$ is

$$\boldsymbol{\Sigma} = \sum_{g,g'=1}^{G} \sum_{h,h'=1}^{H} \mathrm{E}\big(\boldsymbol{s}_{gh}\boldsymbol{s}_{g'h'}^\top\big) = \sum_{g=1}^{G} \boldsymbol{\Sigma}_g + \sum_{h=1}^{H} \boldsymbol{\Sigma}_h - \sum_{g=1}^{G} \sum_{h=1}^{H} \boldsymbol{\Sigma}_{gh}. \tag{3}$$

Linear Regression with Two-Way Clustering

With two-way clustering, $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ can be written as

$$\boldsymbol{y}_{gh} = \boldsymbol{X}_{gh}\boldsymbol{\beta} + \boldsymbol{u}_{gh}, \quad g = 1, \ldots, G, \;\; h = 1, \ldots, H. \tag{2}$$

The variance matrix of the score vector $\boldsymbol{s} = \boldsymbol{X}^\top \boldsymbol{u}$ is

$$\boldsymbol{\Sigma} = \sum_{g,g'=1}^{G} \sum_{h,h'=1}^{H} \mathrm{E}\big(\boldsymbol{s}_{gh}\boldsymbol{s}_{g'h'}^\top\big) = \sum_{g=1}^{G} \boldsymbol{\Sigma}_g + \sum_{h=1}^{H} \boldsymbol{\Sigma}_h - \sum_{g=1}^{G} \sum_{h=1}^{H} \boldsymbol{\Sigma}_{gh}. \tag{3}$$

where

$$\boldsymbol{\Sigma}_g = \mathrm{E}(\boldsymbol{s}_g\boldsymbol{s}_g^\top), \;\; \boldsymbol{\Sigma}_h = \mathrm{E}(\boldsymbol{s}_h\boldsymbol{s}_h^\top), \;\; \text{and} \;\; \boldsymbol{\Sigma}_{gh} = \mathrm{E}(\boldsymbol{s}_{gh}\boldsymbol{s}_{gh}^\top). \tag{4}$$

Linear Regression with Two-Way Clustering

With two-way clustering, $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ can be written as

$$\boldsymbol{y}_{gh} = \boldsymbol{X}_{gh}\boldsymbol{\beta} + \boldsymbol{u}_{gh}, \quad g = 1, \ldots, G, \;\; h = 1, \ldots, H. \tag{2}$$

The variance matrix of the score vector $\boldsymbol{s} = \boldsymbol{X}^\top \boldsymbol{u}$ is

$$\boldsymbol{\Sigma} = \sum_{g,g'=1}^{G} \sum_{h,h'=1}^{H} \mathrm{E}\big(\boldsymbol{s}_{gh}\boldsymbol{s}_{g'h'}^\top\big) = \sum_{g=1}^{G} \boldsymbol{\Sigma}_g + \sum_{h=1}^{H} \boldsymbol{\Sigma}_h - \sum_{g=1}^{G}\sum_{h=1}^{H} \boldsymbol{\Sigma}_{gh}. \tag{3}$$

where

$$\boldsymbol{\Sigma}_g = \mathrm{E}(\boldsymbol{s}_g\boldsymbol{s}_g^\top), \;\; \boldsymbol{\Sigma}_h = \mathrm{E}(\boldsymbol{s}_h\boldsymbol{s}_h^\top), \;\; \text{and} \;\; \boldsymbol{\Sigma}_{gh} = \mathrm{E}(\boldsymbol{s}_{gh}\boldsymbol{s}_{gh}^\top). \tag{4}$$

The variance matrix of $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$ is

$$\boldsymbol{V}_\beta = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{\Sigma}\,(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = \boldsymbol{V}_G + \boldsymbol{V}_H - \boldsymbol{V}_I. \tag{5}$$

The first component of $\boldsymbol{V_\beta}$ is

$$\boldsymbol{V}_G = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\Bigg(\sum_{g=1}^{G}\boldsymbol{\Sigma}_g\Bigg)(\boldsymbol{X}^\top\boldsymbol{X})^{-1}. \tag{6}$$

The other two, $\boldsymbol{V}_H$ and $\boldsymbol{V}_I$, are defined similarly.

The first component of $\boldsymbol{V}_{\boldsymbol{\beta}}$ is

$$\boldsymbol{V}_G = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \left(\sum_{g=1}^{G} \boldsymbol{\Sigma}_g \right) (\boldsymbol{X}^\top \boldsymbol{X})^{-1}. \tag{6}$$

The other two, $\boldsymbol{V}_H$ and $\boldsymbol{V}_I$, are defined similarly.

The empirical analog of (5) is the three-term two-way CRVE

$$\hat{\boldsymbol{V}}_1^{(3)} = \hat{\boldsymbol{V}}_G + \hat{\boldsymbol{V}}_H - \hat{\boldsymbol{V}}_I, \tag{7}$$

The first component of $\boldsymbol{V}_{\boldsymbol{\beta}}$ is

$$\boldsymbol{V}_G = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \left(\sum_{g=1}^{G} \boldsymbol{\Sigma}_g \right) (\boldsymbol{X}^\top \boldsymbol{X})^{-1}. \tag{6}$$

The other two, $\boldsymbol{V}_H$ and $\boldsymbol{V}_I$, are defined similarly.

The empirical analog of (5) is the three-term two-way CRVE

$$\hat{\boldsymbol{V}}_1^{(3)} = \hat{\boldsymbol{V}}_G + \hat{\boldsymbol{V}}_H - \hat{\boldsymbol{V}}_I, \tag{7}$$

where

$$\hat{\boldsymbol{V}}_I = \frac{I(N-1)}{(I-1)(N-k)} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \left(\sum_{g=1}^{G} \sum_{h=1}^{H} \hat{\boldsymbol{s}}_{gh} \hat{\boldsymbol{s}}_{gh}^\top \right) (\boldsymbol{X}^\top \boldsymbol{X})^{-1}, \tag{8}$$

and likewise for $\boldsymbol{V}_G$ and $\boldsymbol{V}_H$. Here $I \leq GH$ is the number of *intersections*.

The first component of $\boldsymbol{V}_{\boldsymbol{\beta}}$ is

$$\boldsymbol{V}_G = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\Bigg(\sum_{g=1}^{G}\boldsymbol{\Sigma}_g\Bigg)(\boldsymbol{X}^\top\boldsymbol{X})^{-1}. \tag{6}$$

The other two, $\boldsymbol{V}_H$ and $\boldsymbol{V}_I$, are defined similarly.

The empirical analog of (5) is the three-term two-way CRVE

$$\hat{\boldsymbol{V}}_1^{(3)} = \hat{\boldsymbol{V}}_G + \hat{\boldsymbol{V}}_H - \hat{\boldsymbol{V}}_I, \tag{7}$$

where

$$\hat{\boldsymbol{V}}_I = \frac{I(N-1)}{(I-1)(N-k)}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\Bigg(\sum_{g=1}^{G}\sum_{h=1}^{H}\hat{\boldsymbol{s}}_{gh}\hat{\boldsymbol{s}}_{gh}^\top\Bigg)(\boldsymbol{X}^\top\boldsymbol{X})^{-1}, \tag{8}$$

and likewise for $\boldsymbol{V}_G$ and $\boldsymbol{V}_H$. Here $I \le GH$ is the number of *intersections*.

But $\hat{\boldsymbol{V}}_1^{(3)}$ may not be positive definite!

We could instead use the two-term estimator

$$\hat{\boldsymbol{V}}_1^{(2)} = \hat{\boldsymbol{V}}_G + \hat{\boldsymbol{V}}_H. \tag{9}$$

We could instead use the two-term estimator

$$\hat{\boldsymbol{V}}_1^{(2)} = \hat{\boldsymbol{V}}_G + \hat{\boldsymbol{V}}_H. \tag{9}$$

It is easier to compute than $\hat{\boldsymbol{V}}_1^{(3)}$ and must be positive semi-definite.

We could instead use the two-term estimator

$$\hat{\boldsymbol{V}}_1^{(2)} = \hat{\boldsymbol{V}}_G + \hat{\boldsymbol{V}}_H. \tag{9}$$

It is easier to compute than $\hat{\boldsymbol{V}}_1^{(3)}$ and must be positive semi-definite.

But $\hat{\boldsymbol{V}}_1^{(2)}$ is asymptotically invalid when the scores are independent or only correlated at the intersection level, so that $\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{V}_I$.

We could instead use the two-term estimator

$$\hat{\boldsymbol{V}}_1^{(2)} = \hat{\boldsymbol{V}}_G + \hat{\boldsymbol{V}}_H. \tag{9}$$

It is easier to compute than $\hat{\boldsymbol{V}}_1^{(3)}$ and must be positive semi-definite.

But $\hat{\boldsymbol{V}}_1^{(2)}$ is asymptotically invalid when the scores are independent or only correlated at the intersection level, so that $\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{V}_I$.

In that case, $\hat{\boldsymbol{V}}_G \approx \hat{\boldsymbol{V}}_H \approx \hat{\boldsymbol{V}}_I$. Therefore,

$$\hat{\boldsymbol{V}}_1^{(2)} = \hat{\boldsymbol{V}}_G + \hat{\boldsymbol{V}}_H \approx 2\hat{\boldsymbol{V}}_I, \tag{10}$$

We could instead use the two-term estimator

$$\hat{V}_1^{(2)} = \hat{V}_G + \hat{V}_H. \tag{9}$$

It is easier to compute than $\hat{V}_1^{(3)}$ and must be positive semi-definite.

But $\hat{V}_1^{(2)}$ is asymptotically invalid when the scores are independent or only correlated at the intersection level, so that $V_\beta = V_I$.

In that case, $\hat{V}_G \approx \hat{V}_H \approx \hat{V}_I$. Therefore,

$$\hat{V}_1^{(2)} = \hat{V}_G + \hat{V}_H \approx 2\hat{V}_I, \tag{10}$$

whereas

$$\hat{V}_1^{(3)} = \hat{V}_G + \hat{V}_H - \hat{V}_I \approx \hat{V}_I. \tag{11}$$

We could instead use the two-term estimator

$$\hat{V}_1^{(2)} = \hat{V}_G + \hat{V}_H. \tag{9}$$

It is easier to compute than $\hat{V}_1^{(3)}$ and must be positive semi-definite.

But $\hat{V}_1^{(2)}$ is asymptotically invalid when the scores are independent or only correlated at the intersection level, so that $V_\beta = V_I$.

In that case, $\hat{V}_G \approx \hat{V}_H \approx \hat{V}_I$. Therefore,

$$\hat{V}_1^{(2)} = \hat{V}_G + \hat{V}_H \approx 2\hat{V}_I, \tag{10}$$

whereas

$$\hat{V}_1^{(3)} = \hat{V}_G + \hat{V}_H - \hat{V}_I \approx \hat{V}_I. \tag{11}$$

Thus, in this case, $\hat{V}_1^{(2)}$ is approximately twice as large as $\hat{V}_1^{(3)}$, and twice as large as it should be.

We could instead use the two-term estimator

$$\hat{\boldsymbol{V}}_1^{(2)} = \hat{\boldsymbol{V}}_G + \hat{\boldsymbol{V}}_H. \tag{9}$$

It is easier to compute than $\hat{\boldsymbol{V}}_1^{(3)}$ and must be positive semi-definite.

But $\hat{\boldsymbol{V}}_1^{(2)}$ is asymptotically invalid when the scores are independent or only correlated at the intersection level, so that $\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{V}_I$.

In that case, $\hat{\boldsymbol{V}}_G \approx \hat{\boldsymbol{V}}_H \approx \hat{\boldsymbol{V}}_I$. Therefore,

$$\hat{\boldsymbol{V}}_1^{(2)} = \hat{\boldsymbol{V}}_G + \hat{\boldsymbol{V}}_H \approx 2\hat{\boldsymbol{V}}_I, \tag{10}$$

whereas

$$\hat{\boldsymbol{V}}_1^{(3)} = \hat{\boldsymbol{V}}_G + \hat{\boldsymbol{V}}_H - \hat{\boldsymbol{V}}_I \approx \hat{\boldsymbol{V}}_I. \tag{11}$$

Thus, in this case, $\hat{\boldsymbol{V}}_1^{(2)}$ is approximately twice as large as $\hat{\boldsymbol{V}}_1^{(3)}$, and twice as large as it should be.

Unfortunately, $\hat{\boldsymbol{V}}_1^{(3)}$ is not necessarily positive semi-definite, and its diagonal elements may be negative.

Cameron, Gelbach, and Miller (2011)suggested a work-around to avoid negative diagonals, which Stata 18+ now implements.

Cameron, Gelbach, and Miller (2011)suggested a work-around to avoid negative diagonals, which Stata 18+ now implements.

Compute the eigenvalues of $\hat{V}_1^{(3)}$, say $\lambda_1, \ldots, \lambda_k$.

Cameron, Gelbach, and Miller (2011)suggested a work-around to avoid negative diagonals, which Stata 18+ now implements.

Compute the eigenvalues of $\hat{\boldsymbol{V}}_1^{(3)}$, say $\lambda_1, \ldots, \lambda_k$.

When any of them is not positive, replace $\hat{\boldsymbol{V}}_1^{(3)}$ by

$$\hat{\boldsymbol{V}}_1^{(3+)} = \boldsymbol{U}\boldsymbol{\Lambda}^{+}\boldsymbol{U}^{\top},$$

where $\boldsymbol{U}$ is the $k \times k$ matrix of eigenvectors, and $\boldsymbol{\Lambda}^{+}$ is a diagonal matrix with typical diagonal element $\lambda_j^{+} = \max\{\lambda_j, 0\}$.

Cameron, Gelbach, and Miller (2011)suggested a work-around to avoid negative diagonals, which Stata 18+ now implements.

Compute the eigenvalues of $\hat{\boldsymbol{V}}_1^{(3)}$, say $\lambda_1, \ldots, \lambda_k$.

When any of them is not positive, replace $\hat{\boldsymbol{V}}_1^{(3)}$ by

$$\hat{\boldsymbol{V}}_1^{(3+)} = \boldsymbol{U}\boldsymbol{\Lambda}^{+}\boldsymbol{U}^{\top},$$

where $\boldsymbol{U}$ is the $k \times k$ matrix of eigenvectors, and $\boldsymbol{\Lambda}^{+}$ is a diagonal matrix with typical diagonal element $\lambda_j^{+} = \max\{\lambda_j, 0\}$.

Instead of 0, we use $\eta = 10^{-12}$, so that $\hat{\boldsymbol{V}}_1^{(3+)}$ is positive definite.

Cameron, Gelbach, and Miller (2011)suggested a work-around to avoid negative diagonals, which Stata 18+ now implements.

Compute the eigenvalues of $\hat{V}_1^{(3)}$, say $\lambda_1, \ldots, \lambda_k$.

When any of them is not positive, replace $\hat{V}_1^{(3)}$ by

$$\hat{V}_1^{(3+)} = \boldsymbol{U}\boldsymbol{\Lambda}^{+}\boldsymbol{U}^{\top},$$

where $\boldsymbol{U}$ is the $k \times k$ matrix of eigenvectors, and $\boldsymbol{\Lambda}^{+}$ is a diagonal matrix with typical diagonal element $\lambda_j^{+} = \max\{\lambda_j, 0\}$.

Instead of 0, we use $\eta = 10^{-12}$, so that $\hat{V}_1^{(3+)}$ is positive definite.

- Wald and t-statistics based on $\hat{V}_1^{(3+)}$ may be extremely large.

Cameron, Gelbach, and Miller (2011)suggested a work-around to avoid negative diagonals, which Stata 18+ now implements.

Compute the eigenvalues of $\hat{\boldsymbol{V}}_1^{(3)}$, say $\lambda_1, \ldots, \lambda_k$.

When any of them is not positive, replace $\hat{\boldsymbol{V}}_1^{(3)}$ by

$$\hat{\boldsymbol{V}}_1^{(3+)} = \boldsymbol{U}\boldsymbol{\Lambda}^{+}\boldsymbol{U}^{\top},$$

where $\boldsymbol{U}$ is the $k \times k$ matrix of eigenvectors, and $\boldsymbol{\Lambda}^{+}$ is a diagonal matrix with typical diagonal element $\lambda_j^{+} = \max\{\lambda_j, 0\}$.

Instead of 0, we use $\eta = 10^{-12}$, so that $\hat{\boldsymbol{V}}_1^{(3+)}$ is positive definite.

- Wald and t-statistics based on $\hat{\boldsymbol{V}}_1^{(3+)}$ may be extremely large.
- Replacing $\hat{\boldsymbol{V}}_1^{(3)}$ by $\hat{\boldsymbol{V}}_1^{(3+)}$ can change all the standard errors.

Cameron, Gelbach, and Miller (2011)suggested a work-around to avoid negative diagonals, which Stata 18+ now implements.

Compute the eigenvalues of $\hat{V}_1^{(3)}$, say $\lambda_1, \ldots, \lambda_k$.

When any of them is not positive, replace $\hat{V}_1^{(3)}$ by

$$\hat{V}_1^{(3+)} = \boldsymbol{U}\boldsymbol{\Lambda}^{+}\boldsymbol{U}^{\top},$$

where $\boldsymbol{U}$ is the $k \times k$ matrix of eigenvectors, and $\boldsymbol{\Lambda}^{+}$ is a diagonal matrix with typical diagonal element $\lambda_j^{+} = \max\{\lambda_j, 0\}$.

Instead of 0, we use $\eta = 10^{-12}$, so that $\hat{V}_1^{(3+)}$ is positive definite.

- Wald and t-statistics based on $\hat{V}_1^{(3+)}$ may be extremely large.
- Replacing $\hat{V}_1^{(3)}$ by $\hat{V}_1^{(3+)}$ can change all the standard errors.
- $\text{se}(\hat{\beta}_j)$ is not invariant to nonsingular transformations of the remaining columns of the matrix $\boldsymbol{X}$.

Cameron, Gelbach, and Miller (2011)suggested a work-around to avoid negative diagonals, which Stata 18+ now implements.

Compute the eigenvalues of $\hat{\boldsymbol{V}}_1^{(3)}$, say $\lambda_1, \ldots, \lambda_k$.

When any of them is not positive, replace $\hat{\boldsymbol{V}}_1^{(3)}$ by

$$\hat{\boldsymbol{V}}_1^{(3+)} = \boldsymbol{U}\boldsymbol{\Lambda}^+\boldsymbol{U}^\top,$$

where $\boldsymbol{U}$ is the $k \times k$ matrix of eigenvectors, and $\boldsymbol{\Lambda}^+$ is a diagonal matrix with typical diagonal element $\lambda_j^+ = \max\{\lambda_j, 0\}$.

Instead of 0, we use $\eta = 10^{-12}$, so that $\hat{\boldsymbol{V}}_1^{(3+)}$ is positive definite.

- Wald and t-statistics based on $\hat{\boldsymbol{V}}_1^{(3+)}$ may be extremely large.
- Replacing $\hat{\boldsymbol{V}}_1^{(3)}$ by $\hat{\boldsymbol{V}}_1^{(3+)}$ can change all the standard errors.
- $\text{se}(\hat{\beta}_j)$ is not invariant to nonsingular transformations of the remaining columns of the matrix $\boldsymbol{X}$.
- Thus precisely how fixed effects or other dummy variables are specified may affect $\text{se}(\hat{\beta}_j)$.

A Better Way to Avoid Non-Positive Test Statistics

A Better Way to Avoid Non-Positive Test Statistics

Compute three test statistics. Use the smallest one that is positive.
Also suggested in Davezies, D'Haultfoeuille, and Guyonvarch (2025).

A Better Way to Avoid Non-Positive Test Statistics

Compute three test statistics. Use the smallest one that is positive. Also suggested in Davezies, D'Haultfoeuille, and Guyonvarch (2025).

For the hypothesis that $\boldsymbol{R\beta} = \boldsymbol{r}$, the three Wald statistics are

$$\begin{aligned} W_3 &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_1^{(3)}\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}), \\ W_G &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_G\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}), \text{ and} \\ W_H &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_H\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}). \end{aligned} \tag{12}$$

A Better Way to Avoid Non-Positive Test Statistics

Compute three test statistics. Use the smallest one that is positive.
Also suggested in Davezies, D'Haultfoeuille, and Guyonvarch (2025).

For the hypothesis that $\boldsymbol{R\beta} = \boldsymbol{r}$, the three Wald statistics are

$$\begin{aligned} W_3 &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_1^{(3)}\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}), \\ W_G &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_G\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}), \text{ and} \\ W_H &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_H\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}). \end{aligned} \tag{12}$$

Our **max-se** procedure uses the statistic

$$W_{\min} = \min\big\{\max\{W_3, 0\}, W_G, W_H\big\}, \tag{13}$$

where $\max\{W_3, 0\} = 0$ if W_3 is either negative or undefined.

A Better Way to Avoid Non-Positive Test Statistics

Compute three test statistics. Use the smallest one that is positive. Also suggested in Davezies, D'Haultfoeuille, and Guyonvarch (2025).

For the hypothesis that $\boldsymbol{R\beta} = \boldsymbol{r}$, the three Wald statistics are

$$\begin{aligned} W_3 &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_1^{(3)}\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}), \\ W_G &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_G\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}), \text{ and} \\ W_H &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_H\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}). \end{aligned} \tag{12}$$

Our **max-se** procedure uses the statistic

$$W_{\min} = \min\big\{\max\{W_3, 0\}, W_G, W_H\big\}, \tag{13}$$

where $\max\{W_3, 0\} = 0$ if W_3 is either negative or undefined.

We denote the variance and standard error estimators based on $\hat{\boldsymbol{V}}_1^{(2)}$, $\hat{\boldsymbol{V}}_1^{(3)}$, and $\hat{\boldsymbol{V}}_1^{(3+)}$ as $\mathrm{CV}_1^{(2)}$, $\mathrm{CV}_1^{(3)}$, and $\mathrm{CV}_1^{(3+)}$, respectively, and the one that is implicit in (13) as the $\mathrm{CV}_1^{(\max)}$ estimator.

Two-Way Cluster Jackknife CRVEs

Two-Way Cluster Jackknife CRVEs

Let $J \in \{G, H, I\}$, and let j denote the corresponding lower-case letter. The OLS estimates of $\boldsymbol{\beta}$ when each cluster in the J dimension is omitted in turn are

$$\hat{\boldsymbol{\beta}}^{(j)} = (\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}_j^\top \boldsymbol{X}_j)^{-1} (\boldsymbol{X}^\top \boldsymbol{y} - \boldsymbol{X}_j^\top \boldsymbol{y}_j), \quad j = 1, \ldots, J. \tag{14}$$

Two-Way Cluster Jackknife CRVEs

Let $J \in \{G, H, I\}$, and let j denote the corresponding lower-case letter. The OLS estimates of $\boldsymbol{\beta}$ when each cluster in the J dimension is omitted in turn are

$$\hat{\boldsymbol{\beta}}^{(j)} = (\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}_j^\top \boldsymbol{X}_j)^{-1} (\boldsymbol{X}^\top \boldsymbol{y} - \boldsymbol{X}_j^\top \boldsymbol{y}_j), \quad j = 1, \ldots, J. \tag{14}$$

Then the component cluster jackknife variance matrix estimators are

$$\hat{\boldsymbol{V}}_J^{\mathrm{JK}} = \frac{J-1}{J} \sum_{j=1}^{J} (\hat{\boldsymbol{\beta}}^{(j)} - \hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}^{(j)} - \hat{\boldsymbol{\beta}})^\top \quad \text{for } \{j, J\} = \{g, G\}, \{h, H\}, \{i, I\}. \tag{15}$$

Two-Way Cluster Jackknife CRVEs

Let $J \in \{G, H, I\}$, and let j denote the corresponding lower-case letter. The OLS estimates of $\boldsymbol{\beta}$ when each cluster in the J dimension is omitted in turn are

$$\hat{\boldsymbol{\beta}}^{(j)} = (\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}_j^\top \boldsymbol{X}_j)^{-1} (\boldsymbol{X}^\top \boldsymbol{y} - \boldsymbol{X}_j^\top \boldsymbol{y}_j), \quad j = 1, \ldots, J. \tag{14}$$

Then the component cluster jackknife variance matrix estimators are

$$\hat{\boldsymbol{V}}_J^{\mathrm{JK}} = \frac{J-1}{J} \sum_{j=1}^{J} (\hat{\boldsymbol{\beta}}^{(j)} - \hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}^{(j)} - \hat{\boldsymbol{\beta}})^\top \quad \text{for } \{j, J\} = \{g, G\}, \{h, H\}, \{i, I\}. \tag{15}$$

Thus the three-term jackknife CRVE is

$$\hat{\boldsymbol{V}}_3^{(3)} = \hat{\boldsymbol{V}}_G^{\mathrm{JK}} + \hat{\boldsymbol{V}}_H^{\mathrm{JK}} - \hat{\boldsymbol{V}}_I^{\mathrm{JK}}, \tag{16}$$

which is analogous to (7). Notation is based on HC_3.

Computation

Computation

First, calculate the cluster-level matrices and vectors

$$X_j^\top X_j \;\text{ and }\; X_j^\top y_j, \;\; j = 1, \ldots, J, \;\text{ for } \{j, J\} = \{g, G\}, \{h, H\}, \{i, I\}. \quad (17)$$

The ones for the intersections can be computed in a single pass over the N observations. The others are just sums of some of them.

Computation

First, calculate the cluster-level matrices and vectors

$$\boldsymbol{X}_j^\top \boldsymbol{X}_j \text{ and } \boldsymbol{X}_j^\top \boldsymbol{y}_j, \quad j = 1, \ldots, J, \quad \text{for } \{j, J\} = \{g, G\}, \{h, H\}, \{i, I\}. \tag{17}$$

The ones for the intersections can be computed in a single pass over the N observations. The others are just sums of some of them.

With two-way fixed effects in the G and H dimensions,

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\beta}_p + \boldsymbol{D}^G \gamma + \boldsymbol{D}^H \delta + \boldsymbol{u}. \tag{18}$$

Now $\boldsymbol{X} = [\boldsymbol{Z} \;\; \boldsymbol{D}^G \;\; \boldsymbol{D}^H]$, and $k = p + G + H - 1$, and the matrices $\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}_g^\top \boldsymbol{X}_g$ and $\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}_h^\top \boldsymbol{X}_h$ cannot be inverted.

Computation

First, calculate the cluster-level matrices and vectors

$$X_j^\top X_j \text{ and } X_j^\top y_j, \;\; j = 1, \ldots, J, \;\; \text{for } \{j, J\} = \{g, G\}, \{h, H\}, \{i, I\}. \quad (17)$$

The ones for the intersections can be computed in a single pass over the N observations. The others are just sums of some of them.

With two-way fixed effects in the G and H dimensions,

$$y = Z\beta_p + D^G\gamma + D^H\delta + u. \quad (18)$$

Now $X = [Z \;\; D^G \;\; D^H]$, and $k = p + G + H - 1$, and the matrices $X^\top X - X_g^\top X_g$ and $X^\top X - X_h^\top X_h$ cannot be inverted.

Simplest approach is to replace the inverse in (14) by a generalized inverse. Then $\hat{V}^{(3)}_{\text{JK}}$ in (16) can only be calculated as a $p \times p$ matrix.

Computation

First, calculate the cluster-level matrices and vectors

$$\boldsymbol{X}_j^\top \boldsymbol{X}_j \;\text{ and }\; \boldsymbol{X}_j^\top \boldsymbol{y}_j, \;\; j = 1, \ldots, J, \;\; \text{for } \{j, J\} = \{g, G\}, \{h, H\}, \{i, I\}. \tag{17}$$

The ones for the intersections can be computed in a single pass over the N observations. The others are just sums of some of them.

With two-way fixed effects in the G and H dimensions,

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\beta}_p + \boldsymbol{D}^G \boldsymbol{\gamma} + \boldsymbol{D}^H \boldsymbol{\delta} + \boldsymbol{u}. \tag{18}$$

Now $\boldsymbol{X} = [\boldsymbol{Z} \;\; \boldsymbol{D}^G \;\; \boldsymbol{D}^H]$, and $k = p + G + H - 1$, and the matrices $\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}_g^\top \boldsymbol{X}_g$ and $\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}_h^\top \boldsymbol{X}_h$ cannot be inverted.

Simplest approach is to replace the inverse in (14) by a generalized inverse. Then $\hat{\boldsymbol{V}}_{\text{JK}}^{(3)}$ in (16) can only be calculated as a $p \times p$ matrix.

Computing $\text{CV}_3^{(3)}$ and friends for (18) can be costly when G and H are not fairly small.

Critical Values

Critical Values

Student's t distribution with $\min\{G, H\} - 1$ degrees of freedom is normally used for t-statistics based on CV_1.

Critical Values

Student's t distribution with $\min\{G, H\} - 1$ degrees of freedom is normally used for t-statistics based on CV_1.

We do this for t-statistics based on CV_3 as well.

Critical Values

Student's t distribution with $\min\{G, H\} - 1$ degrees of freedom is normally used for t-statistics based on CV_1.

We do this for t-statistics based on CV_3 as well.

There are at least two possible alternatives:

Critical Values

Student's t distribution with $\min\{G, H\} - 1$ degrees of freedom is normally used for t-statistics based on CV_1.

We do this for t-statistics based on CV_3 as well.

There are at least two possible alternatives:

1. Employ a t-distribution with an estimated degrees-of-freedom parameter, and maybe an estimated scale parameter, as in Bell and McCaffrey (2002), Imbens and Kolesár (2016), and Hansen (2023).

Critical Values

Student's t distribution with $\min\{G, H\} - 1$ degrees of freedom is normally used for t-statistics based on CV_1.

We do this for t-statistics based on CV_3 as well.

There are at least two possible alternatives:

1. Employ a t-distribution with an estimated degrees-of-freedom parameter, and maybe an estimated scale parameter, as in Bell and McCaffrey (2002), Imbens and Kolesár (2016), and Hansen (2023).
2. Use the wild cluster bootstrap (MacKinnon, Nielsen, and Webb 2021) or the pigeonhole bootstrap (Owen, 2007).

Critical Values

Student's t distribution with $\min\{G, H\} - 1$ degrees of freedom is normally used for t-statistics based on CV_1.

We do this for t-statistics based on CV_3 as well.

There are at least two possible alternatives:

1. Employ a t-distribution with an estimated degrees-of-freedom parameter, and maybe an estimated scale parameter, as in Bell and McCaffrey (2002), Imbens and Kolesár (2016), and Hansen (2023).
2. Use the wild cluster bootstrap (MacKinnon, Nielsen, and Webb 2021) or the pigeonhole bootstrap (Owen, 2007).
   - For two-way clustering, no version of the wild cluster bootstrap can replicate the intra-cluster covariances in the residuals.

Critical Values

Student's t distribution with $\min\{G, H\} - 1$ degrees of freedom is normally used for t-statistics based on CV_1.

We do this for t-statistics based on CV_3 as well.

There are at least two possible alternatives:

1. Employ a t-distribution with an estimated degrees-of-freedom parameter, and maybe an estimated scale parameter, as in Bell and McCaffrey (2002), Imbens and Kolesár (2016), and Hansen (2023).
2. Use the wild cluster bootstrap (MacKinnon, Nielsen, and Webb 2021) or the pigeonhole bootstrap (Owen, 2007).
   - For two-way clustering, no version of the wild cluster bootstrap can replicate the intra-cluster covariances in the residuals.
   - The pigeonhole bootstrap is an ingenious generalization of the ordinary pairs (resampling) bootstrap for one-way clustering.

Critical Values

Student's t distribution with $\min\{G, H\} - 1$ degrees of freedom is normally used for t-statistics based on CV_1.

We do this for t-statistics based on CV_3 as well.

There are at least two possible alternatives:

1. Employ a t-distribution with an estimated degrees-of-freedom parameter, and maybe an estimated scale parameter, as in Bell and McCaffrey (2002), Imbens and Kolesár (2016), and Hansen (2023).
2. Use the wild cluster bootstrap (MacKinnon, Nielsen, and Webb 2021) or the pigeonhole bootstrap (Owen, 2007).
   - For two-way clustering, no version of the wild cluster bootstrap can replicate the intra-cluster covariances in the residuals.
   - The pigeonhole bootstrap is an ingenious generalization of the ordinary pairs (resampling) bootstrap for one-way clustering.
   - But pairs bootstrap typically performs worse than WCR bootstrap (MacKinnon and Webb, TPM 2017; MacKinnon, 2023).

Consistency of the Cluster-Jackknife CRVE

Consistency of the Cluster-Jackknife CRVE

Theorem 1.

Let $\hat{V}_3$ denote any of the three jackknife CRVEs — $\text{CV}_3^{(2)}$, $\text{CV}_3^{(3)}$, and $\text{CV}_3^{(\max)}$ — and let $\text{Var}(\hat{\boldsymbol{\beta}})$ be given in (5).

Consistency of the Cluster-Jackknife CRVE

Theorem 1.

Let $\hat{V}_3$ denote any of the three jackknife CRVEs — ${CV_3}^{(2)}$, ${CV_3}^{(3)}$, and ${CV_3}^{(\max)}$ — and let $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ be given in (5).

Then, under suitable assumptions, $(\mathrm{Var}(\hat{\boldsymbol{\beta}}))^{-1}\hat{V}_3 \xrightarrow{P} 1$.

Consistency of the Cluster-Jackknife CRVE

Theorem 1.

Let $\hat{V}_3$ denote any of the three jackknife CRVEs — ${\mathrm{CV}_3}^{(2)}$, ${\mathrm{CV}_3}^{(3)}$, and ${\mathrm{CV}_3}^{(\mathrm{max})}$ — and let $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ be given in (5).

Then, under suitable assumptions, $(\mathrm{Var}(\hat{\boldsymbol{\beta}}))^{-1}\hat{\boldsymbol{V}}_3 \xrightarrow{P} 1$.

It follows that

$$(\mathrm{Var}(\hat{\boldsymbol{\beta}}))^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{I}_k). \tag{19}$$

The assumptions and proof follow Yap (2025).

Consistency of the Cluster-Jackknife CRVE

Theorem 1.

Let $\hat{V}_3$ denote any of the three jackknife CRVEs — ${\rm CV_3}^{(2)}$, ${\rm CV_3}^{(3)}$, and ${\rm CV_3}^{(\rm max)}$ — and let $\text{Var}(\hat{\boldsymbol{\beta}})$ be given in (5).

Then, under suitable assumptions, $(\text{Var}(\hat{\boldsymbol{\beta}}))^{-1}\hat{V}_3 \xrightarrow{P} 1$.

It follows that

$$(\text{Var}(\hat{\boldsymbol{\beta}}))^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \text{N}(\mathbf{0}, \mathbf{I}_k). \tag{19}$$

The assumptions and proof follow Yap (2025).

He demonstrates the consistency of $\hat{V}_1$ without assuming that the disturbances are generated as separately exchangeable arrays, as Davezies, D'Haultfoeuille, and Guyonvarch (2018, 2025) do.

Consistency of the Cluster-Jackknife CRVE

Theorem 1.

Let $\hat{V}_3$ denote any of the three jackknife CRVEs — ${\rm CV_3}^{(2)}$, ${\rm CV_3}^{(3)}$, and ${\rm CV_3}^{(\max)}$ — and let $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ be given in (5).

Then, under suitable assumptions, $(\mathrm{Var}(\hat{\boldsymbol{\beta}}))^{-1}\hat{V}_3 \xrightarrow{P} 1$.

It follows that

$$(\mathrm{Var}(\hat{\boldsymbol{\beta}}))^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{I}_k). \tag{19}$$

The assumptions and proof follow Yap (2025).

He demonstrates the consistency of $\hat{V}_1$ without assuming that the disturbances are generated as separately exchangeable arrays, as Davezies, D'Haultfoeuille, and Guyonvarch (2018, 2025) do.

Yap's assumptions are weaker, and his method of proof is simpler.

Simulation Experiments

Simulation Experiments

The disturbances are generated so that cluster fixed effects do not eliminate intra-cluster correlation. We use factor models of the form

$$
\begin{aligned}
z_{ghi} &= \sigma_g \xi_g^1 + \sigma_h \xi_h^1 + \sigma_\epsilon \zeta_{ghi} \quad \text{if } i \text{ is odd,} \\
z_{ghi} &= \sigma_g \xi_g^2 + \sigma_h \xi_h^2 + \sigma_\epsilon \zeta_{ghi} \quad \text{if } i \text{ is even.}
\end{aligned} \tag{20}
$$

Simulation Experiments

The disturbances are generated so that cluster fixed effects do not eliminate intra-cluster correlation. We use factor models of the form

$$\begin{aligned} z_{ghi} &= \sigma_g \xi_g^1 + \sigma_h \xi_h^1 + \sigma_\epsilon \zeta_{ghi} \quad \text{if } i \text{ is odd}, \\ z_{ghi} &= \sigma_g \xi_g^2 + \sigma_h \xi_h^2 + \sigma_\epsilon \zeta_{ghi} \quad \text{if } i \text{ is even}. \end{aligned} \tag{20}$$

- ξ_g^1 and ξ_g^2 are $\mathrm{N}(0,1)$ random effects which apply to the odd-numbered and even-numbered observations within the g^{th} cluster.

Simulation Experiments

The disturbances are generated so that cluster fixed effects do not eliminate intra-cluster correlation. We use factor models of the form

$$\begin{aligned} z_{ghi} &= \sigma_g \xi_g^1 + \sigma_h \xi_h^1 + \sigma_\epsilon \zeta_{ghi} \quad \text{if } i \text{ is odd,} \\ z_{ghi} &= \sigma_g \xi_g^2 + \sigma_h \xi_h^2 + \sigma_\epsilon \zeta_{ghi} \quad \text{if } i \text{ is even.} \end{aligned} \tag{20}$$

- ξ_g^1 and ξ_g^2 are $\mathrm{N}(0,1)$ random effects which apply to the odd-numbered and even-numbered observations within the g^{th} cluster.
- ξ_h^1 and ξ_h^2 are $\mathrm{N}(0,1)$ random effects which apply to the odd-numbered and even-numbered observations within the h^{th} cluster. The ζ_{ghi} are independent standard normals.

Simulation Experiments

The disturbances are generated so that cluster fixed effects do not eliminate intra-cluster correlation. We use factor models of the form

$$\begin{aligned} z_{ghi} &= \sigma_g \xi_g^1 + \sigma_h \xi_h^1 + \sigma_\epsilon \zeta_{ghi} \quad \text{if } i \text{ is odd,} \\ z_{ghi} &= \sigma_g \xi_g^2 + \sigma_h \xi_h^2 + \sigma_\epsilon \zeta_{ghi} \quad \text{if } i \text{ is even.} \end{aligned} \tag{20}$$

- ξ_g^1 and ξ_g^2 are $\mathrm{N}(0,1)$ random effects which apply to the odd-numbered and even-numbered observations within the g^{th} cluster.
- ξ_h^1 and ξ_h^2 are $\mathrm{N}(0,1)$ random effects which apply to the odd-numbered and even-numbered observations within the h^{th} cluster. The ζ_{ghi} are independent standard normals.
- σ_g and σ_h are specified as functions of correlations ρ_g and ρ_h, with $\sigma_j = (\rho_j/(1-\rho_j))^{1/2}$ for $j = g, h$.

Simulation Experiments

The disturbances are generated so that cluster fixed effects do not eliminate intra-cluster correlation. We use factor models of the form

$$\begin{aligned} z_{ghi} &= \sigma_g \xi_g^1 + \sigma_h \xi_h^1 + \sigma_\epsilon \zeta_{ghi} \quad \text{if } i \text{ is odd,} \\ z_{ghi} &= \sigma_g \xi_g^2 + \sigma_h \xi_h^2 + \sigma_\epsilon \zeta_{ghi} \quad \text{if } i \text{ is even.} \end{aligned} \tag{20}$$

- ξ_g^1 and ξ_g^2 are $\mathrm{N}(0,1)$ random effects which apply to the odd-numbered and even-numbered observations within the g^{th} cluster.
- ξ_h^1 and ξ_h^2 are $\mathrm{N}(0,1)$ random effects which apply to the odd-numbered and even-numbered observations within the h^{th} cluster. The ζ_{ghi} are independent standard normals.
- σ_g and σ_h are specified as functions of correlations ρ_g and ρ_h, with $\sigma_j = (\rho_j/(1-\rho_j))^{1/2}$ for $j = g, h$.
- The value of σ_ϵ is $(1 - \sigma_g^2 - \sigma_h^2)^{1/2}$, so that $\mathrm{Var}(z_{ghi}) = 1$.

The cluster sizes in the G dimension are given by

$$N_g = \left[N \frac{\exp(\gamma g/G)}{\sum_{j=1}^{G} \exp(\gamma j/G)} \right], \quad g = 1, \ldots, G-1, \tag{21}$$

where $[x]$ denotes the integer part of x. Then $N_G = N - \sum_{g=1}^{G-1} N_g$.

The cluster sizes in the G dimension are given by

$$N_g = \left[N \frac{\exp(\gamma g/G)}{\sum_{j=1}^{G} \exp(\gamma j/G)} \right], \quad g = 1, \ldots, G-1, \tag{21}$$

where $[x]$ denotes the integer part of x. Then $N_G = N - \sum_{g=1}^{G-1} N_g$.

(21), perhaps with a different γ, is also used in the H dimension.

The cluster sizes in the G dimension are given by

$$N_g = \left[N \frac{\exp(\gamma g/G)}{\sum_{j=1}^{G} \exp(\gamma j/G)} \right], \quad g = 1, \ldots, G-1, \tag{21}$$

where $[x]$ denotes the integer part of x. Then $N_G = N - \sum_{g=1}^{G-1} N_g$.

(21), perhaps with a different γ, is also used in the H dimension.

Assuming that the distributions are independent, $N_{gh} \approx N_g N_h / N$.

The cluster sizes in the G dimension are given by

$$N_g = \left[N \frac{\exp(\gamma g/G)}{\sum_{j=1}^{G} \exp(\gamma j/G)} \right], \quad g = 1, \ldots, G-1, \tag{21}$$

where $[x]$ denotes the integer part of x. Then $N_G = N - \sum_{g=1}^{G-1} N_g$.

(21), perhaps with a different γ, is also used in the H dimension.

Assuming that the distributions are independent, $N_{gh} \approx N_g N_h / N$.

In a final step, the cluster sizes are adjusted to ensure that they are all integers with $N = \sum_{g=1}^{G} N_g = \sum_{h=1}^{H} N_h = \sum_{g=1}^{G} \sum_{h=1}^{H} N_{gh}$.

The cluster sizes in the G dimension are given by

$$N_g = \left[N \frac{\exp(\gamma g/G)}{\sum_{j=1}^{G} \exp(\gamma j/G)} \right], \quad g = 1, \ldots, G-1, \tag{21}$$

where $[x]$ denotes the integer part of x. Then $N_G = N - \sum_{g=1}^{G-1} N_g$.

(21), perhaps with a different γ, is also used in the H dimension.

Assuming that the distributions are independent, $N_{gh} \approx N_g N_h / N$.

In a final step, the cluster sizes are adjusted to ensure that they are all integers with $N = \sum_{g=1}^{G} N_g = \sum_{h=1}^{H} N_h = \sum_{g=1}^{G} \sum_{h=1}^{H} N_{gh}$.

The experiments currently involve **normally distributed regressors**, which follow the factor model (20).

The cluster sizes in the G dimension are given by

$$N_g = \left[N \frac{\exp(\gamma g/G)}{\sum_{j=1}^{G} \exp(\gamma j/G)} \right], \quad g = 1, \ldots, G-1, \tag{21}$$

where $[x]$ denotes the integer part of x. Then $N_G = N - \sum_{g=1}^{G-1} N_g$.

(21), perhaps with a different γ, is also used in the H dimension.

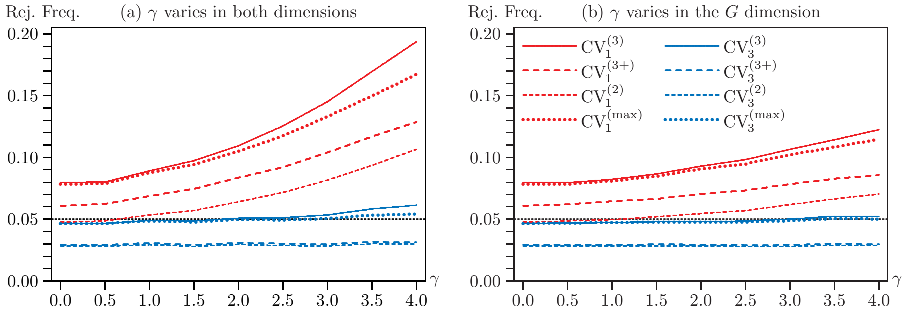
Assuming that the distributions are independent, $N_{gh} \approx N_g N_h / N$.

In a final step, the cluster sizes are adjusted to ensure that they are all integers with $N = \sum_{g=1}^{G} N_g = \sum_{h=1}^{H} N_h = \sum_{g=1}^{G} \sum_{h=1}^{H} N_{gh}$.

The experiments currently involve **normally distributed regressors**, which follow the factor model (20).

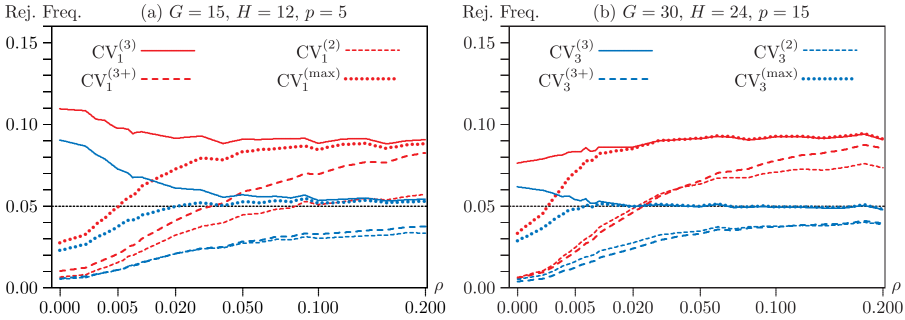
In most experiments, we set $\rho_g^x = \rho_h^x = 0.2$ for the regressors and $\rho_g = \rho_h = 0.1$ for the disturbances.

Figure 1. Rejection frequencies as functions of cluster size variation



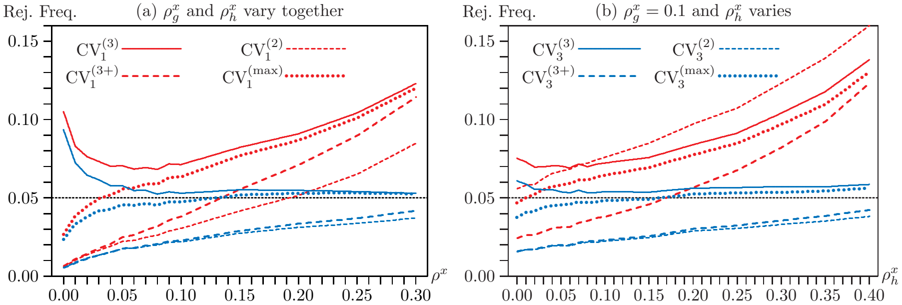
- $N = 10{,}000$, $G = 15$, $H = 12$, $I = 180$, $p = 10$, $k = 36$.
- Regressors are from factor model (20), with $\rho_g^x = \rho_h^x = 0.2$.
- Disturbances are from factor model (20), with $\rho_g = \rho_h = 0.1$.
- Results are based on 100,000 replications.

Figure 2. Rejection frequencies as functions of disturbance correlations



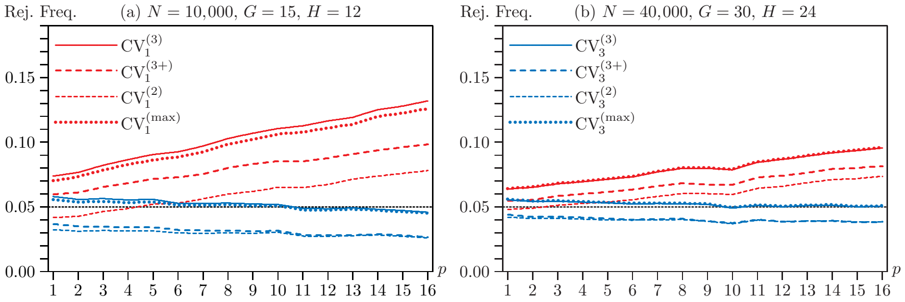
- (a) $N = 10{,}000$, $G = 15$, $H = 12$, $I = 180$, $p = 5$, $\gamma = 2$.
- (b) $N = 40{,}000$, $G = 30$, $H = 24$, $I = 720$, $p = 15$, $\gamma = 2$.
- Regressors are from factor model (20), with $\rho_g^x = \rho_h^x = 0.2$.
- Disturbances are from factor model (20), with $\rho_g = \rho_h$ that vary.
- Results are based on 100,000 replications.

Figure 3. Rejection frequencies as functions of regressor correlations



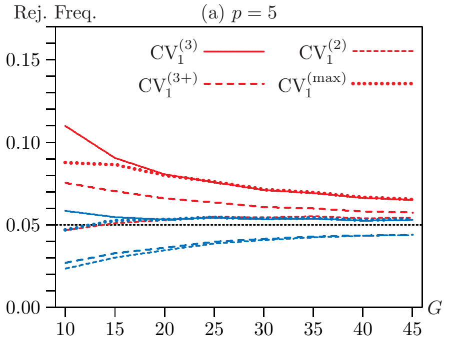
- $N = 10{,}000$, $G = 15$, $H = 12$, $I = 180$, $p = 5$, $\gamma = 2$.
- Disturbances are from factor model (20), with $\rho_g = \rho_h = 0.1$.
- Regressors are from factor model; one or both values of ρ^x vary.
- Results are based on 100,000 replications.

Figure 4. Rejection frequencies as functions of number of regressors



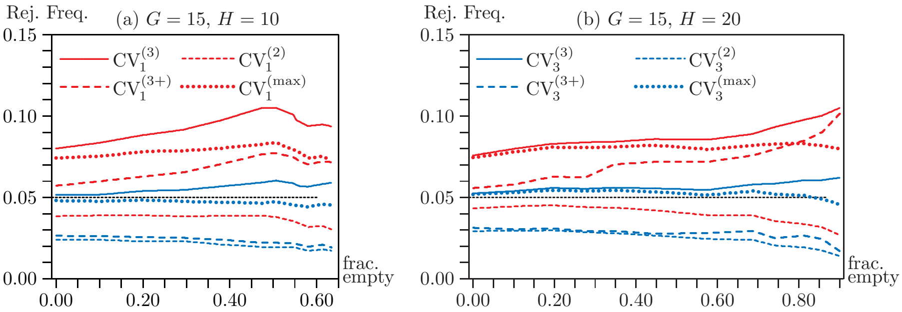
- (a) $N = 10{,}000$, $G = 15$, $H = 12$, $I = 180$, $\gamma = 2$.
- (b) $N = 40{,}000$, $G = 30$, $H = 24$, $I = 720$, $\gamma = 2$.
- Regressors are from factor model (20), with $\rho_g^x = \rho_h^x = 0.2$.
- Disturbances are from factor model (20), with $\rho_g = \rho_h = 0.1$.

Figure 5. Rejection frequencies as functions of numbers of clusters



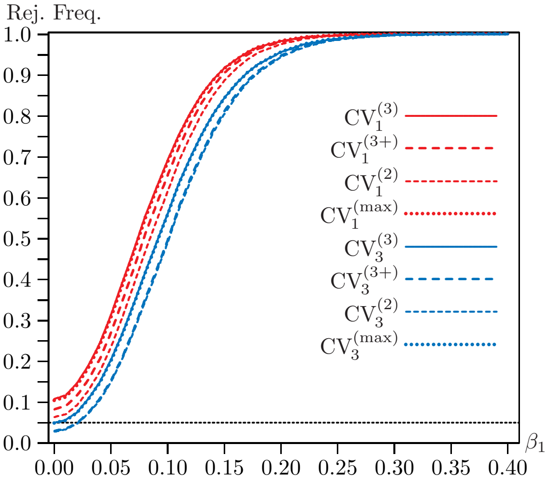
- The value of G varies from 5 to 45 by 5, with $H = 4G/5$.
- The value of N varies from 1,111 to 90,000.
- Regressors are from factor model (20), with $\rho_g^x = \rho_h^x = 0.2$.
- Disturbances are from factor model (20), with $\rho_g = \rho_h = 0.1$.

Figure 6. Rejection frequencies as functions of fraction of empty intersections



- $N = 6000$ in Panel (a) and $N = 12000$ in Panel (b).
- The first 5 regressors are from the model (20), with $\rho_g^x = \rho_h^x = 0.2$.
- The extra 5 regressors are binary and equal 1 with probability 0.25.
- Disturbances are from factor model (20), with $\rho_g = \rho_h = 0.1$.

Figure 7. Power functions for eight tests



- $N = 10{,}000$, $G = 15$, $H = 12$, $I = 180$, $p = 5$, $\gamma = 2$.
- Regressors and disturbances are from factor model (20).
- All coefficients except β_1 equal 0.

Empirical Examples

Empirical Examples

The first example is based on Alsan (2015), which studies the impact of the tsetse fly on economic development in Africa.

Empirical Examples

The first example is based on Alsan (2015), which studies the impact of the tsetse fly on economic development in Africa.

- The main explanatory variable is the TSI or "Tsetse Suitability Index." It measures how suitable an area is to support the tsetse fly, which carries a parasite that affects humans and livestock.

Empirical Examples

The first example is based on Alsan (2015), which studies the impact of the tsetse fly on economic development in Africa.

- The main explanatory variable is the TSI or "Tsetse Suitability Index." It measures how suitable an area is to support the tsetse fly, which carries a parasite that affects humans and livestock.
- The paper tests the extent to which the tsetse fly inhibited political and agriculatural development in parts of Africa.

Empirical Examples

The first example is based on Alsan (2015), which studies the impact of the tsetse fly on economic development in Africa.

- The main explanatory variable is the TSI or "Tsetse Suitability Index." It measures how suitable an area is to support the tsetse fly, which carries a parasite that affects humans and livestock.
- The paper tests the extent to which the tsetse fly inhibited political and agriculatural development in parts of Africa.
- There are seven different outcome variables. These are regressed on TSI and various controls.

Empirical Examples

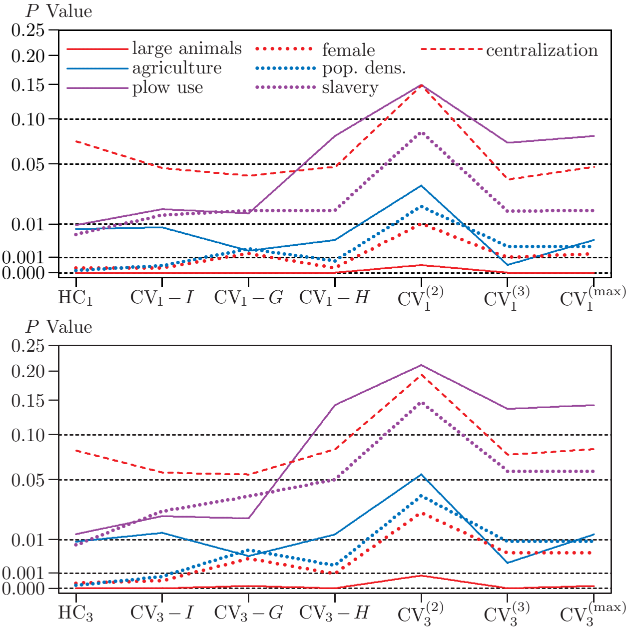
The first example is based on Alsan (2015), which studies the impact of the tsetse fly on economic development in Africa.

- The main explanatory variable is the TSI or "Tsetse Suitability Index." It measures how suitable an area is to support the tsetse fly, which carries a parasite that affects humans and livestock.
- The paper tests the extent to which the tsetse fly inhibited political and agriculatural development in parts of Africa.
- There are seven different outcome variables. These are regressed on TSI and various controls.
- Sample sizes vary from 315 to 485. There are two clustering dimensions, country and "cultural province." Most results use one-way clustering by the latter.

Empirical Examples

The first example is based on Alsan (2015), which studies the impact of the tsetse fly on economic development in Africa.

- The main explanatory variable is the TSI or "Tsetse Suitability Index." It measures how suitable an area is to support the tsetse fly, which carries a parasite that affects humans and livestock.
- The paper tests the extent to which the tsetse fly inhibited political and agriculatural development in parts of Africa.
- There are seven different outcome variables. These are regressed on TSI and various controls.
- Sample sizes vary from 315 to 485. There are two clustering dimensions, country and "cultural province." Most results use one-way clustering by the latter.
- There are 44 countries, 43 or 44 provinces, and between 112 and 142 non-empty intersections. Since $44^2 = 1936$, $I << GH$.


P Value
0.25
0.20
0.15
0.10
0.05
0.01
0.001
0.000
large animals
agriculture
plow use
female
pop. dens.
slavery
centralization
HC1
CV1 − I
CV1 − G
CV1 − H
CV1(2)
CV1(3)
CV1(max)
P Value
0.25
0.20
0.15
0.10
0.05
0.01
0.001
0.000
HC3
CV3 − I
CV3 − G
CV3 − H
CV3(2)
CV3(3)
CV3(max)

The second example studies the effects of minimum wages on earnings for young people in Canada, using Labour Force Survey data.

The second example studies the effects of minimum wages on earnings for young people in Canada, using Labour Force Survey data.

- Individuals aged 18–24 who have been in Canada less than ten years. 28,599 observations in 10 provinces for 2008 to 2019.

The second example studies the effects of minimum wages on earnings for young people in Canada, using Labour Force Survey data.

- Individuals aged 18–24 who have been in Canada less than ten years. 28,599 observations in 10 provinces for 2008 to 2019.
- The dependent variable is the log of weekly earnings, and the regressor of interest is the log of the minimum wage, which varies by province. There are 63 distinct values.

The second example studies the effects of minimum wages on earnings for young people in Canada, using Labour Force Survey data.

- Individuals aged 18–24 who have been in Canada less than ten years. 28,599 observations in 10 provinces for 2008 to 2019.
- The dependent variable is the log of weekly earnings, and the regressor of interest is the log of the minimum wage, which varies by province. There are 63 distinct values.

$$\begin{aligned} \text{logearn}_{ipmt} &= \alpha + \beta\,\text{logmw}_{pmt} + \gamma\,\text{big city}_{ipmt} + \delta\,\text{older}_{ipmt} \\ &\quad + \text{year}_t + \text{month}_m + \text{prov}_p + \epsilon_{ipmt} \end{aligned}$$

The second example studies the effects of minimum wages on earnings for young people in Canada, using Labour Force Survey data.

- Individuals aged 18–24 who have been in Canada less than ten years. 28,599 observations in 10 provinces for 2008 to 2019.
- The dependent variable is the log of weekly earnings, and the regressor of interest is the log of the minimum wage, which varies by province. There are 63 distinct values.

$$\text{logearn}_{ipmt} = \alpha + \beta\,\text{logmw}_{pmt} + \gamma\,\text{big city}_{ipmt} + \delta\,\text{older}_{ipmt} + \text{year}_t + \text{month}_m + \text{prov}_p + \epsilon_{ipmt}$$

- Observations per province $(H = 10)$ vary from 163 to 6554.

The second example studies the effects of minimum wages on earnings for young people in Canada, using Labour Force Survey data.

- Individuals aged 18–24 who have been in Canada less than ten years. 28,599 observations in 10 provinces for 2008 to 2019.
- The dependent variable is the log of weekly earnings, and the regressor of interest is the log of the minimum wage, which varies by province. There are 63 distinct values.

$$\text{logearn}_{ipmt} = \alpha + \beta\,\text{logmw}_{pmt} + \gamma\,\text{big city}_{ipmt} + \delta\,\text{older}_{ipmt} + \text{year}_t + \text{month}_m + \text{prov}_p + \epsilon_{ipmt}$$

- Observations per province $(H = 10)$ vary from 163 to 6554.
- Observations per year $(G = 12)$ vary from 2051 to 2723.

The second example studies the effects of minimum wages on earnings for young people in Canada, using Labour Force Survey data.

- Individuals aged 18–24 who have been in Canada less than ten years. 28,599 observations in 10 provinces for 2008 to 2019.
- The dependent variable is the log of weekly earnings, and the regressor of interest is the log of the minimum wage, which varies by province. There are 63 distinct values.

$$\begin{aligned}\text{logearn}_{ipmt} &= \alpha + \beta\,\text{logmw}_{pmt} + \gamma\,\text{big city}_{ipmt} + \delta\,\text{older}_{ipmt}\\ &\quad + \text{year}_t + \text{month}_m + \text{prov}_p + \epsilon_{ipmt}\end{aligned}$$

- Observations per province $(H = 10)$ vary from 163 to 6554.
- Observations per year $(G = 12)$ vary from 2051 to 2723.
- Observations per intersection $(I = 120)$ vary from 3 to 710.

The second example studies the effects of minimum wages on earnings for young people in Canada, using Labour Force Survey data.

- Individuals aged 18–24 who have been in Canada less than ten years. 28,599 observations in 10 provinces for 2008 to 2019.
- The dependent variable is the log of weekly earnings, and the regressor of interest is the log of the minimum wage, which varies by province. There are 63 distinct values.

$$\begin{aligned}\text{logearn}_{ipmt} &= \alpha + \beta\,\text{logmw}_{pmt} + \gamma\,\text{big city}_{ipmt} + \delta\,\text{older}_{ipmt} \\ &\quad + \text{year}_t + \text{month}_m + \text{prov}_p + \epsilon_{ipmt}\end{aligned}$$

- Observations per province $(H = 10)$ vary from 163 to 6554.
- Observations per year $(G = 12)$ vary from 2051 to 2723.
- Observations per intersection $(I = 120)$ vary from 3 to 710.
- Coefficient estimate is 0.2934, with standard errors between 0.0254 for HC_1 and 0.1663 for $\text{CV}_3^{(2)}$.

We also run **placebo regressions**, which generalize the idea of "placebo laws" proposed in Bertrand, Duflo, and Mullainathan (2004).

We also run **placebo regressions**, which generalize the idea of "placebo laws" proposed in Bertrand, Duflo, and Mullainathan (2004).

- The idea is to add a randomly generated regressor that looks similar to the minimum wage to the actual regression.

We also run **placebo regressions**, which generalize the idea of "placebo laws" proposed in Bertrand, Duflo, and Mullainathan (2004).

- The idea is to add a randomly generated regressor that looks similar to the minimum wage to the actual regression.
- This is done many times (100,000 in our simulations). Placebo regressor changes across simulations, but not the regressand.

We also run **placebo regressions**, which generalize the idea of "placebo laws" proposed in Bertrand, Duflo, and Mullainathan (2004).

- The idea is to add a randomly generated regressor that looks similar to the minimum wage to the actual regression.
- This is done many times (100,000 in our simulations). Placebo regressor changes across simulations, but not the regressand.
- Since the placebo regressors are random, they should have no explanatory power if the regression is specified correctly.

We also run **placebo regressions**, which generalize the idea of "placebo laws" proposed in Bertrand, Duflo, and Mullainathan (2004).

- The idea is to add a randomly generated regressor that looks similar to the minimum wage to the actual regression.
- This is done many times (100,000 in our simulations). Placebo regressor changes across simulations, but not the regressand.
- Since the placebo regressors are random, they should have no explanatory power if the regression is specified correctly.
- If a test at the .05 level rejects much more or less often than 5%, then we should not trust results of that test.

We also run **placebo regressions**, which generalize the idea of "placebo laws" proposed in Bertrand, Duflo, and Mullainathan (2004).

- The idea is to add a randomly generated regressor that looks similar to the minimum wage to the actual regression.
- This is done many times (100,000 in our simulations). Placebo regressor changes across simulations, but not the regressand.
- Since the placebo regressors are random, they should have no explanatory power if the regression is specified correctly.
- If a test at the .05 level rejects much more or less often than 5%, then we should not trust results of that test.
- We model the placebo minimum wage as a two-stage process at the province-year level.

We also run **placebo regressions**, which generalize the idea of "placebo laws" proposed in Bertrand, Duflo, and Mullainathan (2004).

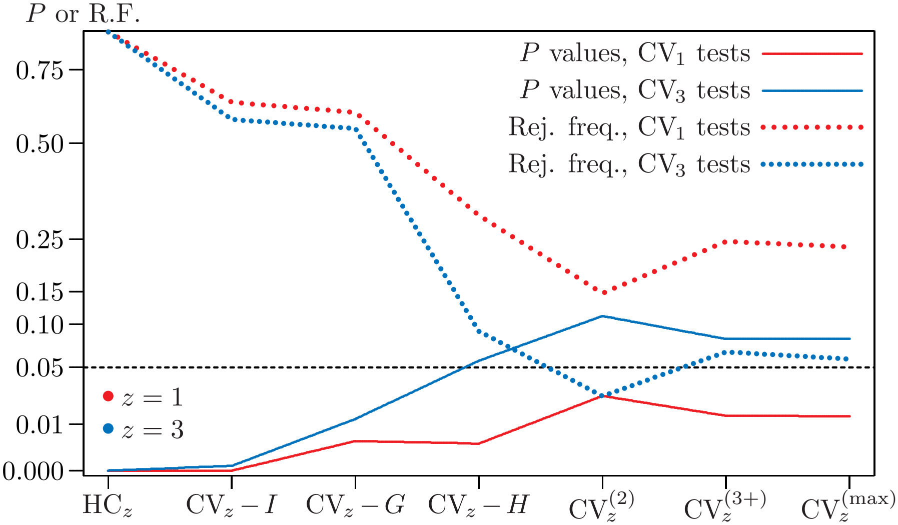
- The idea is to add a randomly generated regressor that looks similar to the minimum wage to the actual regression.
- This is done many times (100,000 in our simulations). Placebo regressor changes across simulations, but not the regressand.
- Since the placebo regressors are random, they should have no explanatory power if the regression is specified correctly.
- If a test at the .05 level rejects much more or less often than 5%, then we should not trust results of that test.
- We model the placebo minimum wage as a two-stage process at the province-year level.
- It increases if random variables ν_p and ν_y both exceed threshold values. The increase is a random amount of 0.25, 0.50, 0.75, or 1.00.

We also run **placebo regressions**, which generalize the idea of "placebo laws" proposed in Bertrand, Duflo, and Mullainathan (2004).

- The idea is to add a randomly generated regressor that looks similar to the minimum wage to the actual regression.
- This is done many times (100,000 in our simulations). Placebo regressor changes across simulations, but not the regressand.
- Since the placebo regressors are random, they should have no explanatory power if the regression is specified correctly.
- If a test at the .05 level rejects much more or less often than 5%, then we should not trust results of that test.
- We model the placebo minimum wage as a two-stage process at the province-year level.
- It increases if random variables ν_p and ν_y both exceed threshold values. The increase is a random amount of 0.25, 0.50, 0.75, or 1.00.

The details matter, and current results are preliminary.

Figure 9. P Values and Placebo Regression Rejection Frequencies



Conclusions

Conclusions

1. We propose two-way cluster jackknife CRVEs. New $\mathrm{CV}_3^{(3)}$ and $\mathrm{CV}_3^{(\max)}$ estimators often work very well indeed. They can over-reject or under-reject, but usually quite modestly.

Conclusions

1. We propose two-way cluster jackknife CRVEs. New $\mathrm{CV}_3^{(3)}$ and $\mathrm{CV}_3^{(\max)}$ estimators often work very well indeed. They can over-reject or under-reject, but usually quite modestly.
2. In contrast, t-statistics based on the widely-used $\mathrm{CV}_1^{(3)}$ CRVE for OLS with two-way clustering often over-reject severely.

Conclusions

1. We propose two-way cluster jackknife CRVEs. New $\mathrm{CV}_3^{(3)}$ and $\mathrm{CV}_3^{(\max)}$ estimators often work very well indeed. They can over-reject or under-reject, but usually quite modestly.
2. In contrast, t-statistics based on the widely-used $\mathrm{CV}_1^{(3)}$ CRVE for OLS with two-way clustering often over-reject severely.
3. Using an eigen-decomposition when $\boldsymbol{V}_1^{(3)}$ is not positive definite yields $\mathrm{CV}_1^{(3+)}$ (Stata default). It is parametrization-dependent!

Conclusions

1. We propose two-way cluster jackknife CRVEs. New $\mathrm{CV}_3^{(3)}$ and $\mathrm{CV}_3^{(\max)}$ estimators often work very well indeed. They can over-reject or under-reject, but usually quite modestly.
2. In contrast, *t*-statistics based on the widely-used $\mathrm{CV}_1^{(3)}$ CRVE for OLS with two-way clustering often over-reject severely.
3. Using an eigen-decomposition when $\boldsymbol{V}_1^{(3)}$ is not positive definite yields $\mathrm{CV}_1^{(3+)}$ (Stata default). It is parametrization-dependent!
4. Fixed effects must be handled with care when computing cluster-jackknife (CV_3) CRVEs for two-way clustering.

Conclusions

1. We propose two-way cluster jackknife CRVEs. New $\mathrm{CV}_3^{(3)}$ and $\mathrm{CV}_3^{(\max)}$ estimators often work very well indeed. They can over-reject or under-reject, but usually quite modestly.
2. In contrast, t-statistics based on the widely-used $\mathrm{CV}_1^{(3)}$ CRVE for OLS with two-way clustering often over-reject severely.
3. Using an eigen-decomposition when $\boldsymbol{V}_1^{(3)}$ is not positive definite yields $\mathrm{CV}_1^{(3+)}$ (Stata default). It is parametrization-dependent!
4. Fixed effects must be handled with care when computing cluster-jackknife (CV_3) CRVEs for two-way clustering.
5. The $\mathrm{CV}_3^{(2)}$ CRVE is cheaper but usually under-rejects, sometimes severely. So does $\mathrm{CV}_3^{(3+)}$ in some cases with fixed effects.

Conclusions

1. We propose two-way cluster jackknife CRVEs. New $\mathrm{CV}_3^{(3)}$ and $\mathrm{CV}_3^{(\max)}$ estimators often work very well indeed. They can over-reject or under-reject, but usually quite modestly.
2. In contrast, *t*-statistics based on the widely-used $\mathrm{CV}_1^{(3)}$ CRVE for OLS with two-way clustering often over-reject severely.
3. Using an eigen-decomposition when $\boldsymbol{V}_1^{(3)}$ is not positive definite yields $\mathrm{CV}_1^{(3+)}$ (Stata default). It is parametrization-dependent!
4. Fixed effects must be handled with care when computing cluster-jackknife (CV_3) CRVEs for two-way clustering.
5. The $\mathrm{CV}_3^{(2)}$ CRVE is cheaper but usually under-rejects, sometimes severely. So does $\mathrm{CV}_3^{(3+)}$ in some cases with fixed effects.
6. The number of regressors and their features matter!