Leave-out Estimation of Variance Components

Patrick Kline[1], Raffaele Saggio[2], Mikkel Sølvsten[3]

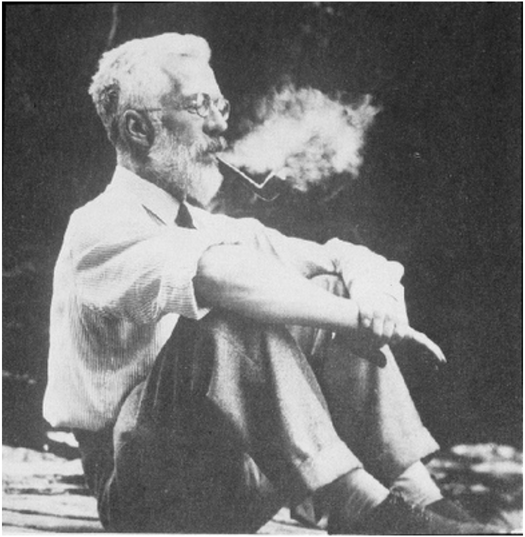
[1]Department of Economics, University of California, Berkeley

[2]Department of Economics, University of British Columbia

[3]Department of Economics, University of Wisconsin-Madison

NBER Labor Studies, July 2018

Mo' data, mo' problems



-

Overview

- As our data grow so does the complexity of our models

- Classic tool: ANOVA (Fisher, 1925) provides low dimensional summary of heavily parameterized models in terms of "variance components"

Overview

- As our data grow so does the complexity of our models

- Classic tool: ANOVA (Fisher, 1925) provides low dimensional summary of heavily parameterized models in terms of "variance components"

- Along with a framework for testing large numbers of linear restrictions (F-test)

Overview

- As our data grow so does the complexity of our models

- Classic tool: ANOVA (Fisher, 1925) provides low dimensional summary of heavily parameterized models in terms of "variance components"

- Along with a framework for testing large numbers of linear restrictions (F-test)

- Extensions: Hierarchical Linear Models (HLM), Multi-way Fixed Effect Models

Partying like it's 1929..

Recent applications of two-way FE (AKM) models to wage data:

Card, Heining, Kline (2013); Song, Price, Guvenen, Bloom, von Wachter (2015); Card, Cardoso, Kline (2016); Macis and Schivardi (2016); Lavetti and Schmutte (2016); Sorkin (2018); Lachowska, Mas, Woodbury (2018).

Related applications involving ANOVA, HLM, and/or Multi-way FE:

Graham (2008); Chetty, Friedman, Hilger, Saez, Schanzenbach, Yagan (2011); Arcidiacono, Foster, Goodpaster, Kinsler (2012); Chetty, Friedman, Rockoff (2014); Finkelstein, Gentzkow, Williams (2016); Silver (2016); Angrist, Hull, Pathak, Walters (2017); Best, Hjort, Szakonyi (2017); Chetty and Hendren (2018); Altonji and Mansfield (2018).

Today

- Extend the classic toolkit to develop new estimator of quadratic forms that accomodates heteroscedasticity

Today

- Extend the classic toolkit to develop new estimator of quadratic forms that accomodates heteroscedasticity

- Develop feasible inference procedure that adapts to different data designs (including cases where variance components are *weakly* identified)

Today

- Extend the classic toolkit to develop new estimator of quadratic forms that accomodates heteroscedasticity

- Develop feasible inference procedure that adapts to different data designs (including cases where variance components are *weakly* identified)

- Application: Two-way fixed effects on weakly connected network of firms

Framework

Consider a linear model

$$y_i = x_i'\beta + \varepsilon_i \qquad (i = 1, \cdots, n),$$

with the following features:

- *Many* non-random regressors (dim$(x_i) = k \propto n$)

- Potentially *heteroscedastic* mean-zero error terms ($\mathbb{E}[\varepsilon_i^2] = \sigma_i^2$)

Framework

Consider a linear model

$$y_i = x_i'\beta + \varepsilon_i \qquad (i = 1, \cdots, n),$$

with the following features:

- *Many* non-random regressors ($\dim(x_i) = k \propto n$)
- Potentially *heteroscedastic* mean-zero error terms ($\mathbb{E}[\varepsilon_i^2] = \sigma_i^2$)

Object of interest is $\theta = \beta' A \beta$ where A is known and has rank r.

Motivating Example: AKM

Example I (Two-way fixed effects, AKM)

Our leading application is

$$\text{log-wage}_{gt} = \alpha_g + \psi_{j(g,t)} + x_{gt}'\delta + \varepsilon_{gt} \qquad (g = 1, \cdots, N,\ t = 1, ..., T_g),$$

where $j(\cdot,\cdot)$ assigns each employee to one of $J+1$ employers in each period.

Motivating Example: AKM

Example I (Two-way fixed effects, AKM)

Our leading application is

$$\text{log-wage}_{gt} = \alpha_g + \psi_{j(g,t)} + x_{gt}'\delta + \varepsilon_{gt} \qquad (g = 1, \cdots, N,\ t = 1, ..., T_g),$$

where $j(\cdot,\cdot)$ assigns each employee to one of $J+1$ employers in each period. Objects of interest are σ_α^2, σ_ψ^2, and $\sigma_{\alpha,\psi}$ where, e.g.,

$$\sigma_\psi^2 = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T_g}(\psi_{j(g,t)} - \bar{\psi})^2, \qquad \bar{\psi} = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T_g}\psi_{j(g,t)}.$$

Motivating Example: AKM

Example I (Two-way fixed effects, AKM)

Our leading application is

$$\text{log-wage}_{gt} = \alpha_g + \psi_{j(g,t)} + x_{gt}'\delta + \varepsilon_{gt} \qquad (g = 1, \cdots, N,\ t = 1, ..., T_g),$$

where $j(\cdot,\cdot)$ assigns each employee to one of $J+1$ employers in each period. Objects of interest are σ_α^2, σ_ψ^2, and $\sigma_{\alpha,\psi}$ where, e.g.,

$$\sigma_\psi^2 = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T_g}(\psi_{j(g,t)} - \bar{\psi})^2, \qquad \bar{\psi} = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T_g}\psi_{j(g,t)}.$$

- $\sigma_\psi^2 = \beta' A \beta$ where the rank of A is J (often on the order of 1M!).

Motivating Example: AKM

Example I (Two-way fixed effects, AKM)

Our leading application is

$$\text{log-wage}_{gt} = \alpha_g + \psi_{j(g,t)} + x'_{gt}\delta + \varepsilon_{gt} \qquad (g = 1, \cdots, N,\ t = 1, ..., T_g),$$

where $j(\cdot,\cdot)$ assigns each employee to one of $J+1$ employers in each period. Objects of interest are σ^2_α, σ^2_ψ, and $\sigma_{\alpha,\psi}$ where, e.g.,

$$\sigma^2_\psi = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T_g}(\psi_{j(g,t)} - \bar{\psi})^2, \qquad \bar{\psi} = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T_g}\psi_{j(g,t)}.$$

- $\sigma^2_\psi = \beta' A \beta$ where the rank of A is J (often on the order of 1M!).
- Dimensionality presents substantial obstacles to estimation and inference

Outline

Related methods / theoretical results

Variance Components (R^2, ANOVA, HLM, Two-way FEs): Wright (1921); Fisher (1925); Theil (1961); Akritas and Papadatos (2004); Akritas and Wang (2011); Dicker (2014); Andrews, Gill, Schank, Upward (2008); Verdier (2016); **Jochmans and Weidner (2016); Bonhomme, Lamadon, Manresa (2017); Borovičková and Shimer (2017)**.

Related methods / theoretical results

Variance Components (R^2, ANOVA, HLM, Two-way FEs): Wright (1921); Fisher (1925); Theil (1961); Akritas and Papadatos (2004); Akritas and Wang (2011); Dicker (2014); Andrews, Gill, Schank, Upward (2008); Verdier (2016); **Jochmans and Weidner (2016); Bonhomme, Lamadon, Manresa (2017); Borovičková and Shimer (2017)**.

Leave-out or cross-fitting: Hahn and Newey (2004); Dhaene and Jochmans (2015); **Phillips and Hale (1977); Powell, Stock, Stoker (1989); Angrist, Imbens, Krueger (1999); Hausman et al. (2012); Kolesár (2013); Newey and Robins (2018)**.

Related methods / theoretical results

Variance Components (R^2, ANOVA, HLM, Two-way FEs): Wright (1921); Fisher (1925); Theil (1961); Akritas and Papadatos (2004); Akritas and Wang (2011); Dicker (2014); Andrews, Gill, Schank, Upward (2008); Verdier (2016); **Jochmans and Weidner (2016); Bonhomme, Lamadon, Manresa (2017); Borovičková and Shimer (2017)**.

Leave-out or cross-fitting: Hahn and Newey (2004); Dhaene and Jochmans (2015); **Phillips and Hale (1977); Powell, Stock, Stoker (1989); Angrist, Imbens, Krueger (1999); Hausman et al. (2012); Kolesár (2013); Newey and Robins (2018)**.

Inference with heteroskedasticity and/or many regressors: Anatolyev (2012); Karoui and Purdom (2016); Lei, Bickel, Karoui (2016); **Cattaneo, Jansson, Newey (2017)**.

Related methods / theoretical results

Variance Components (R^2, ANOVA, HLM, Two-way FEs): Wright (1921); Fisher (1925); Theil (1961); Akritas and Papadatos (2004); Akritas and Wang (2011); Dicker (2014); Andrews, Gill, Schank, Upward (2008); Verdier (2016); **Jochmans and Weidner (2016); Bonhomme, Lamadon, Manresa (2017); Borovičková and Shimer (2017)**.

Leave-out or cross-fitting: Hahn and Newey (2004); Dhaene and Jochmans (2015); **Phillips and Hale (1977); Powell, Stock, Stoker (1989); Angrist, Imbens, Krueger (1999); Hausman et al. (2012); Kolesár (2013); Newey and Robins (2018)**.

Inference with heteroskedasticity and/or many regressors: Anatolyev (2012); Karoui and Purdom (2016); Lei, Bickel, Karoui (2016); **Cattaneo, Jansson, Newey (2017)**.

Inference in non-standard problems: Staiger and Stock (1997); Andrews and Cheng (2012); Elliott, Müller, Watson (2015); **Andrews and Mikusheva (2016)**.

Outline

Literature

Model and Estimator

Consistency

Distribution Theory

Application

Model

Linear regression

$$y_i = x_i'\beta + \varepsilon_i \qquad (i = 1, \cdots, n),$$

with

- $x_i \in \mathbb{R}^k$ non-random and $S_{xx} = \sum_{i=1}^n x_i x_i'$ of full rank $(k \leq n)$,

Model

Linear regression

$$y_i = x_i'\beta + \varepsilon_i \qquad (i = 1, \cdots, n),$$

with

- $x_i \in \mathbb{R}^k$ non-random and $S_{xx} = \sum_{i=1}^n x_i x_i'$ of full rank $(k \le n)$,
- $\{\varepsilon_i\}_{i=1}^n$ mutually independent, $\mathbb{E}[\varepsilon_i] = 0$ and $\mathbb{E}[\varepsilon_i^2] = \sigma_i^2$,

Model

Linear regression

$$y_i = x_i'\beta + \varepsilon_i \qquad (i = 1, \cdots, n),$$

with

- $x_i \in \mathbb{R}^k$ non-random and $S_{xx} = \sum_{i=1}^n x_i x_i'$ of full rank $(k \le n)$,
- $\{\varepsilon_i\}_{i=1}^n$ mutually independent, $\mathbb{E}[\varepsilon_i] = 0$ and $\mathbb{E}[\varepsilon_i^2] = \sigma_i^2$,
- $\max_i P_{ii} < 1$ where $P_{ii} = x_i' S_{xx}^{-1} x_i$ is the i'th leverage.

Model

Linear regression

$$y_i = x_i'\beta + \varepsilon_i \qquad (i = 1, \cdots, n),$$

with

- $x_i \in \mathbb{R}^k$ non-random and $S_{xx} = \sum_{i=1}^n x_i x_i'$ of full rank $(k \leq n)$,
- $\{\varepsilon_i\}_{i=1}^n$ mutually independent, $\mathbb{E}[\varepsilon_i] = 0$ and $\mathbb{E}[\varepsilon_i^2] = \sigma_i^2$,
- $\max_i P_{ii} < 1$ where $P_{ii} = x_i' S_{xx}^{-1} x_i$ is the i'th leverage.

Object of interest: $\theta = \beta' A \beta$ where A is known, non-random, and symmetric with rank r.

Limits are taken as $n \to \infty$

Linear regression

$$y_{i,n} = x'_{i,n}\beta_n + \varepsilon_{i,n} \qquad (i = 1, \cdots, n),$$

with

- $x_{i,n} \in \mathbb{R}^{k_n}$ non-random and $S_{xx,n} = \sum_{i=1}^{n} x_{i,n}x'_{i,n}$ of full rank $(k_n \leq n)$,
- $\{\varepsilon_{i,n}\}_{i=1}^{n}$ mutually independent, $\mathbb{E}[\varepsilon_{i,n}] = 0$ and $\mathbb{E}[\varepsilon_{i,n}^2] = \sigma_{i,n}^2$,
- $\max_i P_{ii,n} < 1$ where $P_{ii,n} = x'_{i,n}S_{xx,n}^{-1}x_{i,n}$ is the i'th leverage.

Object of interest: $\theta_n = \beta'_n A_n \beta_n$ where A_n is known, non-random, and symmetric with rank r_n.

The problem w/ plugging in..

Sampling variability in $\hat{\beta}$ generates bias in plug-in estimator $\hat{\theta}_{\mathsf{PI}} = \hat{\beta}' A \hat{\beta}$:

$$\mathbb{E}[\hat{\theta}_{\mathsf{PI}} - \theta] = \text{trace}\left(A \mathbb{V}[\hat{\beta}]\right) = \sum_{i=1}^{n} B_{ii} \sigma_i^2$$

for $\quad B_{ii} = x_i' S_{xx}^{-1} A S_{xx}^{-1} x_i.$

The problem w/ plugging in..

Sampling variability in $\hat{\beta}$ generates bias in plug-in estimator $\hat{\theta}_{\mathsf{PI}} = \hat{\beta}' A \hat{\beta}$:

$$\mathbb{E}[\hat{\theta}_{\mathsf{PI}} - \theta] = \text{trace}\left(A \mathbb{V}[\hat{\beta}]\right) = \sum_{i=1}^{n} B_{ii} \sigma_i^2$$

$$\text{for} \quad B_{ii} = x_i' S_{xx}^{-1} A S_{xx}^{-1} x_i.$$

- B_{ii} closely related to leverage P_{ii}

- Special case (*ESS*): $A = S_{xx} \Rightarrow B_{ii} = P_{ii}$

Estimating the bias

The plug-in estimator $\hat{\theta}_{\mathsf{PI}} = \hat{\beta}' A \hat{\beta}$ has a bias of

$$\text{trace}\left(A\mathbb{V}[\hat{\beta}]\right) = \sum_{i=1}^{n} B_{ii}\sigma_i^2 \quad \text{where} \quad B_{ii} = x_i' S_{xx}^{-1} A S_{xx}^{-1} x_i.$$

Basic insight: an unbiased "cross-fit" estimator of σ_i^2 is

$$\begin{aligned}\hat{\sigma}_i^2 &= y_i(y_i - x_i'\hat{\beta}_{-i}) \\ &= \left(\varepsilon_i + x_i'\beta\right)\left(\varepsilon_i + x_i'(\beta - \hat{\beta}_{-i})\right),\end{aligned}$$

where $\hat{\beta}_{-i} = \left(\sum_{\ell \neq i} x_\ell x_\ell'\right)^{-1} \sum_{\ell \neq i} x_\ell y_\ell$.

Leave-out Estimator

Thus, we propose the bias corrected estimator of θ:

$$\hat{\theta} = \hat{\beta}' A \hat{\beta} - \sum_{i=1}^{n} B_{ii} \hat{\sigma}_i^2.$$

Leave-out Estimator

Thus, we propose the bias corrected estimator of θ:

$$\hat{\theta} = \hat{\beta}' A \hat{\beta} - \sum_{i=1}^{n} B_{ii} \hat{\sigma}_i^2.$$

A "leave-out" representation:

$$\hat{\theta} = \sum_{i=1}^{n} y_i \tilde{x}_i' \hat{\beta}_{-i} \quad \text{where} \quad \tilde{x}_i = A S_{xx}^{-1} x_i \in \mathbb{R}^k,$$

Leave-out Estimator

Thus, we propose the bias corrected estimator of θ:

$$\hat{\theta} = \hat{\beta}' A \hat{\beta} - \sum_{i=1}^{n} B_{ii} \hat{\sigma}_i^2.$$

A "leave-out" representation:

$$\begin{aligned} \hat{\theta} &= \sum_{i=1}^{n} y_i \tilde{x}_i' \hat{\beta}_{-i} \quad \text{where} \quad \tilde{x}_i = A S_{xx}^{-1} x_i \in \mathbb{R}^k, \\ &= \sum_{i=1}^{n} \sum_{\ell \neq i} C_{i\ell} y_i y_\ell \quad \text{for} \quad C_{i\ell} = B_{i\ell} - 2^{-1} M_{i\ell} \left(M_{ii}^{-1} B_{ii} + M_{\ell\ell}^{-1} B_{\ell\ell} \right) \end{aligned}$$

Highlights the connection with existing leave-one-out ideas in parametric and non-parametric models, e.g., JIVE and weighted average derivatives.

"Fixing" HC2 in high dimensions..

Recall the HC2 variance estimator of Mackinnon and White (1985):

$$\hat{\mathbb{V}}_{HC2} = S_{xx}^{-1} \left(\sum_{i=1}^{n} x_i x_i' \frac{(y_i - x_i'\hat{\beta})^2}{1 - P_{ii}} \right) S_{xx}^{-1}$$

- HC2 inconsistent when $k \propto n$ (Cattaneo, Jansson, Newey, 2017)

"Fixing" HC2 in high dimensions..

Recall the HC2 variance estimator of Mackinnon and White (1985):

$$\hat{\mathbb{V}}_{HC2} = S_{xx}^{-1} \left(\sum_{i=1}^{n} x_i x_i' \frac{(y_i - x_i'\hat{\beta})^2}{1 - P_{ii}} \right) S_{xx}^{-1}$$

- HC2 inconsistent when $k \propto n$ (Cattaneo, Jansson, Newey, 2017)

A cross-fit replacement:

$$\hat{\mathbb{V}}_{KSS} = S_{xx}^{-1} \left(\sum_{i=1}^{n} x_i x_i' \hat{\sigma}_i^2 \right) S_{xx}^{-1}$$

"Fixing" HC2 in high dimensions..

Recall the HC2 variance estimator of Mackinnon and White (1985):

$$\hat{\mathbb{V}}_{HC2} = S_{xx}^{-1} \left(\sum_{i=1}^{n} x_i x_i' \frac{(y_i - x_i'\hat{\beta})^2}{1 - P_{ii}} \right) S_{xx}^{-1}$$

- HC2 inconsistent when $k \propto n$ (Cattaneo, Jansson, Newey, 2017)

A cross-fit replacement:

$$\hat{\mathbb{V}}_{KSS} = S_{xx}^{-1} \left(\sum_{i=1}^{n} x_i x_i' \hat{\sigma}_i^2 \right) S_{xx}^{-1}$$
$$= S_{xx}^{-1} \left(\sum_{i=1}^{n} x_i x_i' \frac{y_i (y_i - x_i'\hat{\beta})}{1 - P_{ii}} \right) S_{xx}^{-1}$$

- Will show that this enables testing "a few" linear restrictions..

The "Homoscedastic-only" correction

A commonly applied estimator based on homoscedasticity is (adjusted-R^2, bias-corrected 2SLS, ANOVA, ...)

$$\hat{\theta}_{\mathsf{HO}} = \hat{\theta}_{\mathsf{PI}} - \sum_{i=1}^{n} B_{ii}\hat{\sigma}^2 \quad \text{where} \quad \hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2.$$

The "Homoscedastic-only" correction

A commonly applied estimator based on homoscedasticity is (adjusted-R^2, bias-corrected 2SLS, ANOVA, ...)

$$\hat{\theta}_{\mathsf{HO}} = \hat{\theta}_{\mathsf{PI}} - \sum_{i=1}^{n} B_{ii}\hat{\sigma}^2 \quad \text{where} \quad \hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2.$$

- In general, biased when P_{ii} or B_{ii} correlate with σ_i^2.

- Special case (*balanced design*): (B_{ii}, P_{ii}) do not vary w/ i.

Example II (Uncentered R^2)

$$R^2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i'\beta)^2}{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[y_i^2\right]}$$

- Numerator targeted by choosing $A = S_{xx}/n$

Example II (Uncentered R^2)

$$R^2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i'\beta)^2}{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[y_i^2\right]}$$

- Numerator targeted by choosing $A = S_{xx}/n$
- Plug-in estimator $\tilde{R}^2$ (Wright, 1921) uses

$$\frac{1}{n}\sum_{i=1}^{n}(x_i'\hat{\beta})^2$$

Example II (Uncentered R^2)

$$R^2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i'\beta)^2}{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[y_i^2\right]}$$

- Numerator targeted by choosing $A = S_{xx}/n$
- Plug-in estimator $\tilde{R}^2$ (Wright, 1921) uses

$$\frac{1}{n}\sum_{i=1}^{n}(x_i'\hat{\beta})^2$$

- Homoscedasticity corrected estimator is $\hat{R}^2_{adj}$ (Theil,1961)

$$\frac{1}{n}\sum_{i=1}^{n}(x_i'\hat{\beta})^2 - \frac{k}{n-k}\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2$$

Example II (Uncentered R^2)

$$R^2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i'\beta)^2}{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[y_i^2\right]}$$

- HO adjustment relies on Degrees of freedom correction:

$$(1-\hat{R}^2_{adj})/(1-\tilde{R}^2) = n/(n-k)$$

Example II (Uncentered R^2)

$$R^2 = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i'\beta\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[y_i^2\right]}$$

- HO adjustment relies on Degrees of freedom correction:

$$(1-\hat{R}^2_{adj})/(1-\tilde{R}^2) = n/(n-k)$$

- Contrast w/ leave out estimator $\hat{\theta}$, which can be written:

$$\frac{1}{n}\sum_{i=1}^{n} y_i x_i' \hat{\beta}_{-i}$$

Outline

Assumption 1

(a) $\max_i \mathbb{E}[\varepsilon_i^4] + \sigma_i^{-2} = O(1)$,
(b) $\max_i P_{ii} \leq c < 1$,
(c) $\max_i (x_i'\beta)^2 = O(1)$.

(a) ensures thin tails of ε_i.

(b) + (c) implies that $\hat{\sigma}_i^2$ has bounded variance.

(c) can be relaxed (technical condition).

An important matrix

Eigenvalues $(\lambda_1, \ldots, \lambda_r)$ of following matrix govern properties of $\hat{\theta}$:

$$\tilde{A} = S_{xx}^{-1/2} A S_{xx}^{-1/2}$$

- A defines target parameter
- S_{xx}^{-1} summarizes regressor design / difficulty of estimating each coefficient

An important matrix

Eigenvalues $(\lambda_1, \ldots, \lambda_r)$ of following matrix govern properties of $\hat{\theta}$:

$$\tilde{A} = S_{xx}^{-1/2} A S_{xx}^{-1/2}$$

- A defines target parameter
- S_{xx}^{-1} summarizes regressor design / difficulty of estimating each coefficient
- Special case (*orthogonal regressors*): $S_{xx} = I \Rightarrow \tilde{A} = A$

Lemma 1 (Consistency)

Let $\tilde{A} = S_{xx}^{-1/2} A S_{xx}^{-1/2}$.

1. If A *is positive semi-definite, (i)* $\theta = O(1)$*, and*

$$\textit{(ii) trace}(\tilde{A}^2) = o(1),$$

then $\hat{\theta} - \theta \xrightarrow{p} 0$.

Lemma 1 (Consistency)

Let $\tilde{A} = S_{xx}^{-1/2} A S_{xx}^{-1/2}$.

1. If A *is positive semi-definite, (i)* $\theta = O(1)$*, and*

$$\text{(ii) } \textit{trace}(\tilde{A}^2) = o(1),$$

then $\hat{\theta} - \theta \xrightarrow{p} 0$.

2. If A *is non-definite then write* $A = A_1' A_2$ *for some* A_1, A_2*. If* $\theta_k = \beta' A_k' A_k \beta$ *satisfies (i) and (ii) for* $k = 1, 2$*, then* $\hat{\theta} - \theta \xrightarrow{p} 0$.

Lemma 1 (Consistency)

Let $\tilde{A} = S_{xx}^{-1/2} A S_{xx}^{-1/2}$.

1. If A *is positive semi-definite, (i)* $\theta = O(1)$*, and*

$$\textit{(ii) trace}(\tilde{A}^2) = o(1),$$

then $\hat{\theta} - \theta \xrightarrow{p} 0$.

2. If A *is non-definite then write* $A = A_1' A_2$ *for some* A_1, A_2*. If* $\theta_k = \beta' A_k' A_k \beta$ *satisfies (i) and (ii) for* $k = 1, 2$*, then* $\hat{\theta} - \theta \xrightarrow{p} 0$.

- For "leave out R^2" we have trace$(\tilde{A}^2) = k/n^2 \to 0$.

  $\Rightarrow \hat{\theta}$ is consistent

Lemma 1 (Consistency)

Let $\tilde{A} = S_{xx}^{-1/2} A S_{xx}^{-1/2}$.

1. If A *is positive semi-definite, (i)* $\theta = O(1)$*, and*

$$\text{(ii) trace}(\tilde{A}^2) = o(1),$$

then $\hat{\theta} - \theta \xrightarrow{p} 0$.

2. If A *is non-definite then write* $A = A_1' A_2$ *for some* A_1, A_2*. If* $\theta_k = \beta' A_k' A_k \beta$ *satisfies (i) and (ii) for* $k = 1, 2$*, then* $\hat{\theta} - \theta \xrightarrow{p} 0$.

- For "leave out R^2" we have trace$(\tilde{A}^2) = k/n^2 \to 0$.

  $\Rightarrow \hat{\theta}$ is consistent

- Next: Verify (ii) analytically in some stylized examples (ANOVA and HLM).

Lemma 1 (Consistency)

Let $\tilde{A} = S_{xx}^{-1/2} A S_{xx}^{-1/2}$.

1. If A *is positive semi-definite, (i)* $\theta = O(1)$, *and*

$$\text{(ii) } \mathit{trace}(\tilde{A}^2) = o(1),$$

then $\hat{\theta} - \theta \xrightarrow{p} 0$.

2. If A *is non-definite then write* $A = A_1' A_2$ *for some* A_1, A_2. *If* $\theta_k = \beta' A_k' A_k \beta$ *satisfies (i) and (ii) for* $k = 1, 2$, *then* $\hat{\theta} - \theta \xrightarrow{p} 0$.

- For "leave out R^2" we have trace$(\tilde{A}^2) = k/n^2 \to 0$.

  $\Rightarrow \hat{\theta}$ is consistent

- Next: Verify (ii) analytically in some stylized examples (ANOVA and HLM).

- Can assess (ii) empirically in cases where analytically intractable.

Example III (ANOVA)

Consider

$$y_{gt} = \alpha_g + \varepsilon_{gt} \qquad (g = 1, \ldots, N,\ t = 1, \ldots, T_g),$$

where the object of interest is

$$\sigma_\alpha^2 = \frac{1}{n} \sum_{g=1}^{N} T_g \alpha_g^2.$$

Example III (ANOVA)

Consider

$$y_{gt} = \alpha_g + \varepsilon_{gt} \qquad (g = 1, \ldots, N,\ t = 1, \ldots, T_g),$$

where the object of interest is

$$\sigma_\alpha^2 = \frac{1}{n} \sum_{g=1}^{N} T_g \alpha_g^2.$$

- Chetty et al. (2011): $\sigma_\alpha^2 =$ variance of "classroom effects" in STAR

Example III (ANOVA)

Consider

$$y_{gt} = \alpha_g + \varepsilon_{gt} \qquad (g = 1, \ldots, N,\ t = 1, \ldots, T_g),$$

where the object of interest is

$$\sigma_\alpha^2 = \frac{1}{n} \sum_{g=1}^{N} T_g \alpha_g^2.$$

- Chetty et al. (2011): $\sigma_\alpha^2 =$ variance of "classroom effects" in STAR

- $\max_i P_{ii} < 1$ is equivalent to $\min_g T_g \geq 2$.

Example III (ANOVA)

Consider

$$y_{gt} = \alpha_g + \varepsilon_{gt} \qquad (g = 1, \ldots, N,\ t = 1, \ldots, T_g),$$

where the object of interest is

$$\sigma_\alpha^2 = \frac{1}{n} \sum_{g=1}^{N} T_g \alpha_g^2.$$

- Chetty et al. (2011): $\sigma_\alpha^2 =$ variance of "classroom effects" in STAR
- $\max_i P_{ii} < 1$ is equivalent to $\min_g T_g \geq 2$.
- Here, $P_{ii} = nB_{ii} = \frac{1}{T_{g(i)}} \Rightarrow \hat{\theta}_{HO}$ biased when σ_i^2 vary w/ group size

Example III (ANOVA)

Consider

$$y_{gt} = \alpha_g + \varepsilon_{gt} \qquad\qquad (g = 1, \ldots, N,\ t = 1, \ldots, T_g),$$

where the object of interest is

$$\sigma_\alpha^2 = \frac{1}{n} \sum_{g=1}^{N} T_g \alpha_g^2.$$

Leave out estimator can be written:

$$\hat{\sigma}_\alpha^2 = \frac{1}{n} \sum_{g=1}^{N} \left(T_g \hat{\alpha}_g^2 - \hat{\sigma}_g^2 \right)$$

where $\hat{\alpha}_g = \frac{1}{T_g} \sum_{t=1}^{T_g} y_{gt}$ and $\hat{\sigma}_g^2 = \frac{1}{T_g - 1} \sum_{t=1}^{T_g} (y_{gt} - \hat{\alpha}_g)^2$

Example III (ANOVA)

Consider

$$y_{gt} = \alpha_g + \varepsilon_{gt} \qquad (g = 1, \ldots, N,\ t = 1, \ldots, T_g),$$

where the object of interest is

$$\sigma_\alpha^2 = \frac{1}{n} \sum_{g=1}^{N} T_g \alpha_g^2.$$

$\tilde{A}$ is diagonal with N non-zero entries of $\frac{1}{n}$ so

$$\text{trace}\left(\tilde{A}^2\right) = \frac{N}{n^2} \leq \frac{1}{n} = o(1).$$

Example IV (Hierarchical Linear Model (HLM))

Consider

$$y_{gt} = \alpha_g + x_{gt}\delta_g + \varepsilon_{gt} \qquad (g = 1, \ldots, N,\ t = 1, \ldots, T_g),$$

where $\sum_{t=1}^{T_g} x_{gt} = 0$ and the object of interest is

$$\sigma_\delta^2 = \frac{1}{n} \sum_{g=1}^{N} T_g \delta_g^2.$$

Example IV (Hierarchical Linear Model (HLM))

Consider

$$y_{gt} = \alpha_g + x_{gt}\delta_g + \varepsilon_{gt} \qquad (g = 1, \ldots, N,\ t = 1, \ldots, T_g),$$

where $\sum_{t=1}^{T_g} x_{gt} = 0$ and the object of interest is

$$\sigma_\delta^2 = \frac{1}{n} \sum_{g=1}^{N} T_g \delta_g^2.$$

- Raudenbush and Bryk (1986): $\sigma_\delta^2 =$ student-wgt'd var of slopes wrt SES

Example IV (Hierarchical Linear Model (HLM))

Consider

$$y_{gt} = \alpha_g + x_{gt}\delta_g + \varepsilon_{gt} \qquad (g = 1, \ldots, N,\ t = 1, \ldots, T_g),$$

where $\sum_{t=1}^{T_g} x_{gt} = 0$ and the object of interest is

$$\sigma_\delta^2 = \frac{1}{n} \sum_{g=1}^{N} T_g \delta_g^2.$$

- Raudenbush and Bryk (1986): $\sigma_\delta^2 =$ student-wgt'd var of slopes wrt SES
- $\max_i P_{ii} < 1$ is implied by $\min_g T_g \geq 3$ and $x_{gt_1} \neq x_{gt_2} \neq x_{gt_3} \neq x_{gt_1}$.

Example IV (Hierarchical Linear Model (HLM))

Consider

$$y_{gt} = \alpha_g + x_{gt}\delta_g + \varepsilon_{gt} \qquad (g = 1, \ldots, N,\ t = 1, \ldots, T_g),$$

where $\sum_{t=1}^{T_g} x_{gt} = 0$ and the object of interest is

$$\sigma_\delta^2 = \frac{1}{n} \sum_{g=1}^{N} T_g \delta_g^2.$$

- Raudenbush and Bryk (1986): $\sigma_\delta^2 =$ student-wgt'd var of slopes wrt SES
- $\max_i P_{ii} < 1$ is implied by $\min_g T_g \geq 3$ and $x_{gt_1} \neq x_{gt_2} \neq x_{gt_3} \neq x_{gt_1}$.
- $\tilde{A}$ is diagonal with N non-zero entries of $\frac{1}{n} \frac{T_g}{\sum_{t=1}^{T_g} x_{gt}^2}$, so

$$\text{trace}\left(\tilde{A}^2\right) = o(1) \qquad \text{if} \qquad \min_g \frac{n}{T_g} \sum_{t=1}^{T_g} x_{gt}^2 \to \infty.$$

Example I (Two-way fixed effects, AKM)

Consider ($T_g = 2$ and no X_{gt})

$$y_{gt} = \alpha_g + \psi_{j(g,t)} + \varepsilon_{gt} \qquad (i = g, \cdots, N,\ t = 1, 2),$$

and $\sigma_\psi^2 = \frac{1}{n} \sum_{g=1}^{N} \sum_{t=1}^{2} (\psi_{j(g,t)} - \bar{\psi})^2$.

Example I (Two-way fixed effects, AKM)

Consider ($T_g = 2$ and no X_{gt})

$$y_{gt} = \alpha_g + \psi_{j(g,t)} + \varepsilon_{gt} \qquad (i = g, \cdots, N,\ t = 1, 2),$$

and $\sigma_\psi^2 = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{2}(\psi_{j(g,t)} - \bar{\psi})^2$.

$\tilde{A}$ is not diagonal, but ℓ'th largest eigenvalue given by:

$$\lambda_\ell = \frac{1}{n}\frac{1}{\dot{\lambda}_{J+1-\ell}(E^{1/2}\mathcal{L}E^{1/2})}$$

where E is a diagonal matrix of employer specific "churn rates", $\mathcal{L}$ is the normalized Laplacian for the worker-firm mobility network, and $\dot{\lambda}_\ell(\cdot)$ gives the ℓ'th largest eigenvalue of argument.

Example I (Two-way fixed effects, AKM)

Consider ($T_g = 2$ and no X_{gt})

$$y_{gt} = \alpha_g + \psi_{j(g,t)} + \varepsilon_{gt} \qquad (i = g, \cdots, N, \ t = 1, 2),$$

and $\sigma^2_\psi = \frac{1}{n} \sum_{g=1}^{N} \sum_{t=1}^{2} (\psi_{j(g,t)} - \bar{\psi})^2$.

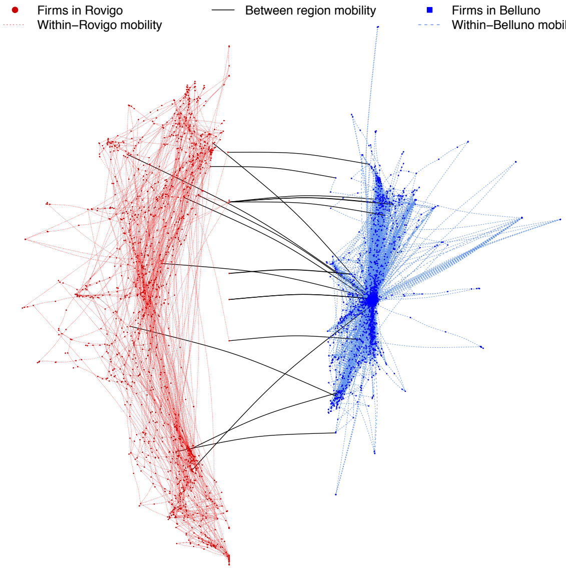
- Sufficient condition for consistency: *strong connectivity*

$$\sqrt{J}\mathcal{C} \to \infty$$

  where $\mathcal{C} \in (0, 1]$ is Cheeger's constant

- Intepretation: no "bottlenecks" in mobility network

Rovigo and Belluno – Employer Mobility Network



Outline

Notation / Overview

We can represent the plug-in estimator $\hat{\theta}_{\text{PI}}$ as

$$\hat{\beta}' A \hat{\beta} = \hat{\beta}' S_{xx}^{1/2} \tilde{A} S_{xx}^{1/2} \hat{\beta}$$

where we write

- $\tilde{A} = S_{xx}^{-1/2} A S_{xx}^{-1/2}$.

Notation / Overview

We can represent the plug-in estimator $\hat{\theta}_{\mathsf{PI}}$ as

$$\hat{\beta}' A \hat{\beta} = \hat{\beta}' S_{xx}^{1/2} \tilde{A} S_{xx}^{1/2} \hat{\beta}$$

where we write

- $\tilde{A} = S_{xx}^{-1/2} A S_{xx}^{-1/2}$.
- $\tilde{A} = QDQ'$ for $D = diag(\lambda_1, \ldots, \lambda_r)$, $\lambda_1^2 \geq \cdots \geq \lambda_r^2 > 0$, and $Q'Q = I_r$,
- $\hat{b} = Q' S_{xx}^{1/2} \hat{\beta}$

Notation / Overview

We can represent the plug-in estimator $\hat{\theta}_{\mathsf{PI}}$ as

$$\hat{\beta}' A \hat{\beta} = \hat{\beta}' S_{xx}^{1/2} \tilde{A} S_{xx}^{1/2} \hat{\beta} = \hat{b}' D \hat{b} = \sum_{\ell=1}^{r} \lambda_\ell \hat{b}_\ell^2$$

where we write

- $\tilde{A} = S_{xx}^{-1/2} A S_{xx}^{-1/2}$.
- $\tilde{A} = QDQ'$ for $D = diag(\lambda_1, \dots, \lambda_r)$, $\lambda_1^2 \geq \cdots \geq \lambda_r^2 > 0$, and $Q'Q = I_r$,
- $\hat{b} = Q' S_{xx}^{1/2} \hat{\beta}$

Notation / Overview

We can represent the plug-in estimator $\hat{\theta}_{\text{PI}}$ as

$$\hat{\beta}' A \hat{\beta} = \hat{\beta}' S_{xx}^{1/2} \tilde{A} S_{xx}^{1/2} \hat{\beta} = \hat{b}' D \hat{b} = \sum_{\ell=1}^{r} \lambda_\ell \hat{b}_\ell^2$$

where we write

- $\tilde{A} = S_{xx}^{-1/2} A S_{xx}^{-1/2}$.
- $\tilde{A} = QDQ'$ for $D = diag(\lambda_1, \ldots, \lambda_r)$, $\lambda_1^2 \geq \cdots \geq \lambda_r^2 > 0$, and $Q'Q = I_r$,
- $\hat{b} = Q' S_{xx}^{1/2} \hat{\beta}$

"Warmup" result: Distribution of infeasible estimator when $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$

$$\theta^* = \hat{\beta}' A \hat{\beta} - \sum_{i=1}^{n} B_{ii} \sigma_i^2$$

Lemma 1 (Finite Sample)

If $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$*, then*

$$\theta^* = \sum_{\ell=1}^{r} \lambda_\ell \left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell] \right) \quad \textit{and} \quad \hat{b} \sim \mathcal{N}\left(b, \mathbb{V}[\hat{b}] \right)$$

where $b = Q' S_{xx}^{1/2} \beta$*.*

Lemma 1 (Finite Sample)

If $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, *then*

$$\theta^* = \sum_{\ell=1}^{r} \lambda_\ell \left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell]\right) \quad \textit{and} \quad \hat{b} \sim \mathcal{N}\left(b, \mathbb{V}[\hat{b}]\right)$$

where $b = Q' S_{xx}^{1/2} \beta$.

- Sums of squares of uncentered normals $\Rightarrow$ non-central χ^2

- Noncentrality governed by b

Building intuition..

$$\theta^* = \sum_{\ell=1}^{r} \lambda_\ell \left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell]\right) \quad \text{and} \quad \hat{b} \sim \mathcal{N}\left(b, \mathbb{V}[\hat{b}]\right)$$

Seek asymptotic approximations that simplify computation and relax assumptions.

Building intuition..

$$\theta^* = \sum_{\ell=1}^{r} \lambda_\ell \left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell] \right) \quad \text{and} \quad \hat{b} \sim \mathcal{N}\left(b, \mathbb{V}[\hat{b}] \right)$$

Seek asymptotic approximations that simplify computation and relax assumptions.

Note: can write $\hat{b}$ as weighted sum $\sum_{i=1}^{n} w_i y_i$

- Weights are $w_i = Q' S_{xx}^{-1/2} x_i$ and obey $\sum_{i=1}^{n} w_i w_i' = I_r$.

- $\max_i w_i' w_i$ provides inverse measure of eff sample size

- Plausible that elements of $\hat{b}$ are approx normal even when ε_i is not..

Building intuition..

$$\theta^* = \sum_{\ell=1}^{r} \lambda_\ell \left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell] \right) \quad \text{and} \quad \hat{b} \sim \mathcal{N}\left(b, \mathbb{V}[\hat{b}] \right)$$

Preview of asymptotic results:

1) When r small (e.g. testing a single linear restriction) and $\hat{b}$ is approximately normally distributed, we obtain non-central χ^2

2) When r large (e.g., testing LOTS of linear restrictions) and eigenvalues same order of magnitude, can invoke a CLT to get normal approximation

3) When r large and eigenvalues different orders of magnitude (weak-id), get a combination of χ^2 and normal components

The "low rank" case

Proposition 1 (Low Rank)

If Assumption 1 *holds, (i)* $\max_i w_i' w_i = o(1)$, *and (ii)* r *is fixed, then*

$$\hat{\theta} = \sum_{\ell=1}^{r} \lambda_\ell \left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell] \right) + o_p(\mathbb{V}[\hat{\theta}]^{1/2}) \quad \textit{and} \quad \mathbb{V}[\hat{b}]^{-1/2}(\hat{b} - b) \xrightarrow{d} \mathcal{N}\left(0, I_r\right).$$

Recall that $\hat{b} = \sum_{i=1}^{n} w_i y_i$ where $w_i = Q' S_{xx}^{-1/2} x_i$ and $\sum_{i=1}^{n} w_i w_i' = I_r$.

The “low rank” case

Proposition 1 (Low Rank)

If Assumption 1 *holds, (i)* $\max_i w_i' w_i = o(1)$*, and (ii)* r *is fixed, then*

$$\hat{\theta} = \sum_{\ell=1}^{r} \lambda_\ell \left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell] \right) + o_p(\mathbb{V}[\hat{\theta}]^{1/2}) \quad \textit{and} \quad \mathbb{V}[\hat{b}]^{-1/2}(\hat{b} - b) \xrightarrow{d} \mathcal{N}\left(0, I_r\right).$$

Recall that $\hat{b} = \sum_{i=1}^{n} w_i y_i$ where $w_i = Q' S_{xx}^{-1/2} x_i$ and $\sum_{i=1}^{n} w_i w_i' = I_r$.

The Lindeberg condition *(i)* ensures that

- no observation is too influential

- sampling error in the bias correction can be ignored.

Application: testing a linear restriction

Suppose we are interested in testing

$$H_0 : v'\beta = 0 \quad \text{for} \quad v \in \mathbb{R}^{k\times 1}$$

Example 1: testing for regional diffs in firm FEs

Example 2: std err on projection of firm FEs onto firm characteristics

Application: testing a linear restriction

Suppose we are interested in testing

$$H_0 : v'\beta = 0 \quad \text{for} \quad v \in \mathbb{R}^{k \times 1}$$

Example 1: testing for regional diffs in firm FEs

Example 2: std err on projection of firm FEs onto firm characteristics

Prop 1 implies that, under H_0, choosing $A = vv'$ yields

$$\mathbb{V}[v'\hat{\beta}]^{-1}\hat{\theta} \xrightarrow{d} \chi^2(1) - 1$$

Application: testing a linear restriction

Suppose we are interested in testing

$$H_0 : v'\beta = 0 \quad \text{for} \quad v \in \mathbb{R}^{k\times 1}$$

Example 1: testing for regional diffs in firm FEs

Example 2: std err on projection of firm FEs onto firm characteristics

Prop 1 implies that, under H_0, choosing $A = vv'$ yields

$$\mathbb{V}[v'\hat{\beta}]^{-1}\hat{\theta} \xrightarrow{d} \chi^2(1) - 1$$

Eicker-White style variance estimator for inference:

$$\hat{\mathbb{V}}[v'\hat{\beta}] = v' S_{xx}^{-1} \left(\sum_{i=1}^{n} x_i x_i' \hat{\sigma}_i^2 \right) S_{xx}^{-1} v$$

Proposition 2 (High Rank, Strong Id)

If Assumption 1 *holds, (i)* $\mathbb{V}[\hat{\theta}]^{-1}\max_i\left((\tilde{x}_i'\beta)^2+(\check{x}_i'\beta)^2\right)=o(1)$*, and*

$$\text{(ii)}\ \frac{\lambda_1^2}{\sum_{\ell=1}^r\lambda_\ell^2}=o(1),$$

then $\mathbb{V}[\hat{\theta}]^{-1/2}(\hat{\theta}-\theta)\xrightarrow{d}\mathcal{N}(0,1)$*.*

Objects appearing in *(i)* are:

- $\tilde{x}_i=AS_{xx}^{-1}x_i$ where $\theta=\sum_{i=1}^n\mathbb{E}[y_i\tilde{x}_i'\beta]$.
- $\check{x}_i=\sum_{\ell=1}^n M_{i\ell}\frac{B_{\ell\ell}}{1-P_{\ell\ell}}x_\ell$ stems from bias correction.

Proposition 2 (High Rank, Strong Id)

If Assumption 1 *holds, (i)* $\mathbb{V}[\hat{\theta}]^{-1}\max_i\left((\tilde{x}_i'\beta)^2+(\check{x}_i'\beta)^2\right)=o(1)$*, and*

$$\text{(ii)}\ \frac{\lambda_1^2}{\sum_{\ell=1}^{r}\lambda_\ell^2}=o(1),$$

then $\mathbb{V}[\hat{\theta}]^{-1/2}(\hat{\theta}-\theta)\xrightarrow{d}\mathcal{N}(0,1)$.

Objects appearing in *(i)* are:

- $\tilde{x}_i = AS_{xx}^{-1}x_i$ where $\theta=\sum_{i=1}^{n}\mathbb{E}[y_i\tilde{x}_i'\beta]$.
- $\check{x}_i=\sum_{\ell=1}^{n}M_{i\ell}\frac{B_{\ell\ell}}{1-P_{\ell\ell}}x_\ell$ stems from bias correction.

- **Intuition**: Averaging $r\to\infty$ terms yields normality under (ii), but estimation of the bias can not be ignored ($\check{x}_i$ is present in $\mathbb{V}[\hat{\theta}]$).

Application: testing *many* linear restrictions

Suppose we are interested in testing

$$H_0 : R\beta = 0 \quad \text{for} \quad R \in \mathbb{R}^{r \times k}$$

- Example: testing block of FEs=0

- Traditional "F-test" would require homoscedasticity

Application: testing *many* linear restrictions

Suppose we are interested in testing

$$H_0 : R\beta = 0 \quad \text{for} \quad R \in \mathbb{R}^{r \times k}$$

- Example: testing block of FEs=0
- Traditional "F-test" would require homoscedasticity

Prop 2 implies that, under H_0, choosing $A = \frac{1}{r} R'(RS_{xx}^{-1}R')^{-1}R$ yields

$$\mathbb{V}[\hat{\theta}]^{-1/2}\hat{\theta} \xrightarrow{d} \mathcal{N}(0, 1)$$

Application: testing *many* linear restrictions

Suppose we are interested in testing

$$H_0 : R\beta = 0 \quad \text{for} \quad R \in \mathbb{R}^{r \times k}$$

- Example: testing block of FEs=0

- Traditional "F-test" would require homoscedasticity

Prop 2 implies that, under H_0, choosing $A = \frac{1}{r} R'(RS_{xx}^{-1}R')^{-1}R$ yields

$$\mathbb{V}[\hat{\theta}]^{-1/2}\hat{\theta} \xrightarrow{d} \mathcal{N}(0, 1)$$

Consistent estimator of $\mathbb{V}[\hat{\theta}]$ provided in paper

Assumption 2

Suppose there exist a known and fixed $q \in \{1, \ldots, r-1\}$ such that

$$\frac{\lambda_{q+1}^2}{\sum_{\ell=1}^r \lambda_\ell^2} = o(1) \quad \text{and} \quad \frac{\lambda_q^2}{\sum_{\ell=1}^r \lambda_\ell^2} \geq c \quad \forall n.$$

Assumption 2

Suppose there exist a known and fixed $q \in \{1, \ldots, r-1\}$ such that

$$\frac{\lambda_{q+1}^2}{\sum_{\ell=1}^{r} \lambda_\ell^2} = o(1) \quad \text{and} \quad \frac{\lambda_q^2}{\sum_{\ell=1}^{r} \lambda_\ell^2} \geq c \quad \forall n.$$

Decomposition:

$$\hat{\mathsf{b}}_q = (\hat{b}_1, \ldots, \hat{b}_q)' = \sum_{i=1}^{n} \mathsf{w}_{iq} y_i, \qquad \mathsf{w}_{iq} = (w_{i1}, \ldots, w_{iq})',$$

$$\hat{\theta}_q = \hat{\theta} - \sum_{\ell=1}^{q} \lambda_\ell (\hat{b}_\ell^2 - \hat{\mathbb{V}}[\hat{b}_\ell]), \qquad \hat{\mathbb{V}}[\hat{b}] = \sum_{i=1}^{n} w_i w_i' \hat{\sigma}_i^2.$$

Theorem 1 (High Rank, Weak Id)

If $\max_i \mathsf{w}_{iq}'\mathsf{w}_{iq} = o(1)$, $\mathbb{V}[\hat{\theta}_q]^{-1} \max_i \left((\tilde{x}_{iq}'\beta)^2 + (\check{x}_{iq}'\beta)^2 \right) = o(1)$, *and Assumption 2 holds, then*

$$\hat{\theta} = \sum_{\ell=1}^{q} \lambda_\ell \left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell] \right) + \hat{\theta}_q + o_p(\mathbb{V}[\hat{\theta}]^{1/2})$$

and

$$\mathbb{V}[(\hat{\mathsf{b}}_q', \hat{\theta}_q)']^{-1/2} \left((\hat{\mathsf{b}}_q', \hat{\theta}_q)' - \mathbb{E}[(\hat{\mathsf{b}}_q', \hat{\theta}_q)'] \right) \xrightarrow{d} \mathcal{N}\left(0, I_{q+1}\right).$$

Theorem 1 (High Rank, Weak Id)

If $\max_i \mathsf{w}_{iq}'\mathsf{w}_{iq} = o(1)$, $\mathbb{V}[\hat{\theta}_q]^{-1} \max_i \left((\tilde{x}_{iq}'\beta)^2 + (\check{x}_{iq}'\beta)^2 \right) = o(1)$, *and Assumption 2 holds, then*

$$\hat{\theta} = \sum_{\ell=1}^{q} \lambda_\ell \left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell] \right) + \hat{\theta}_q + o_p(\mathbb{V}[\hat{\theta}]^{1/2})$$

and

$$\mathbb{V}[(\hat{\mathsf{b}}_q', \hat{\theta}_q)']^{-1/2} \left((\hat{\mathsf{b}}_q', \hat{\theta}_q)' - \mathbb{E}[(\hat{\mathsf{b}}_q', \hat{\theta}_q)'] \right) \xrightarrow{d} \mathcal{N}\left(0, I_{q+1}\right).$$

- Result: q non-central χ^2 terms + a normal
- When $q \ll r$: major simplification relative to finite sample dist.
- But still need to deal w/ q-dimensional nuisance parameter $\mathbb{E}[\hat{\mathsf{b}}_q]$

Weak-id Robust Confidence Interval

To construct a confidence interval we invert a minimum distance statistic:

$$\hat{C}_q^{\theta} = \left[\min_{(\dot{b}_1,\ldots,\dot{b}_q,\dot{\theta}_q)' \in \mathsf{B}_q} \sum_{\ell=1}^{q} \lambda_\ell \dot{b}_\ell^2 + \dot{\theta}_q, \max_{(\dot{b}_1,\ldots,\dot{b}_q,\dot{\theta}_q)' \in \mathsf{B}_q} \sum_{\ell=1}^{q} \lambda_\ell \dot{b}_\ell^2 + \dot{\theta}_q \right]$$

where

$$\mathsf{B}_q = \left\{ (\mathsf{b}_q', \theta_q)' \in \mathbb{R}^{q+1} : \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix}' \hat{\Sigma}_q^{-1} \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix} \leq z_{\hat{\kappa}}^2 \right\}$$

Weak-id Robust Confidence Interval

To construct a confidence interval we invert a minimum distance statistic:

$$\hat{C}_q^{\theta} = \left[\min_{(\dot{b}_1,\ldots,\dot{b}_q,\dot{\theta}_q)' \in \mathsf{B}_q} \sum_{\ell=1}^{q} \lambda_\ell \dot{b}_\ell^2 + \dot{\theta}_q, \max_{(\dot{b}_1,\ldots,\dot{b}_q,\dot{\theta}_q)' \in \mathsf{B}_q} \sum_{\ell=1}^{q} \lambda_\ell \dot{b}_\ell^2 + \dot{\theta}_q \right]$$

where

$$\mathsf{B}_q = \left\{ (\mathsf{b}_q', \theta_q)' \in \mathbb{R}^{q+1} : \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix}' \hat{\Sigma}_q^{-1} \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix} \leq z_{\hat{\kappa}}^2 \right\}$$

- $\hat{\Sigma} = \hat{\mathbb{V}}[(\hat{b}_q', \hat{\theta}_q)']$ and $\hat{\kappa} = \kappa(\hat{\Sigma})$,

Weak-id Robust Confidence Interval

To construct a confidence interval we invert a minimum distance statistic:

$$\hat{C}_q^{\theta} = \left[\min_{(\dot{b}_1,\ldots,\dot{b}_q,\dot{\theta}_q)'\in \mathsf{B}_q} \sum_{\ell=1}^{q} \lambda_\ell \dot{b}_\ell^2 + \dot{\theta}_q, \max_{(\dot{b}_1,\ldots,\dot{b}_q,\dot{\theta}_q)'\in \mathsf{B}_q} \sum_{\ell=1}^{q} \lambda_\ell \dot{b}_\ell^2 + \dot{\theta}_q \right]$$

where

$$\mathsf{B}_q = \left\{ (\mathsf{b}_q', \theta_q)' \in \mathbb{R}^{q+1} : \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix}' \hat{\Sigma}_q^{-1} \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix} \leq z_{\hat{\kappa}}^2 \right\}$$

- $\hat{\Sigma} = \hat{\mathbb{V}}[(\hat{b}_q', \hat{\theta}_q)']$ and $\hat{\kappa} = \kappa(\hat{\Sigma})$,

- z_κ is the critical value proposed in Andrews and Mikusheva (2016).

- κ measures the curvature (non-linearity) of the problem.

Outline

An application to Italian data

Wage and employment data on 2 provinces within the Veneto region of Italy.

Years: 1999 and 2001

Number of movers: 3,531 and 6,414.

Number of employers: 1,282 and 1,684

Example I (Two-way fixed effects, AKM)

Model ($T_g = 2$ and no X_{gt}):

$$\text{log-wage}_{gt} = \alpha_g + \psi_{j(g,t)} + \varepsilon_{gt} \qquad (g = 1, \cdots, N,\ t = 1, 2).$$

The Provinces of Veneto

Leave-out sample preserves first two moments

Table 1: Comparing Samples and Places

	Rovigo [1]	Belluno [2]	Rovigo - Belluno [3]
Largest Connected Set			
Number of Observations	43,330	63,462	106,964
Number of Movers	5,061	7,921	13,022
Number of Firms	2,579	3,131	5,732
Mean Log Daily Wage	4.6089	4.7482	4.6917
Variance Log Daily Wage	0.1560	0.1256	0.1427
Leave Out Sample (Pruned)			
Number of Observations	32,848	56,044	89,666
Number of Movers	3,531	6,414	9,972
Number of Firms	1,282	1,684	2,974
Mean Log Daily Wage	4.6015	4.7636	4.7047
Variance Log Daily Wage	0.1674	0.1245	0.1465
Maximum Leverage (P_{ii})	0.9241	0.9085	0.9236

High leverage ⇒ low-dimensional methods inappropriate

Table 1: Comparing Samples and Places

	Rovigo [1]	Belluno [2]	Rovigo - Belluno [3]
Largest Connected Set			
Number of Observations	43,330	63,462	106,964
Number of Movers	5,061	7,921	13,022
Number of Firms	2,579	3,131	5,732
Mean Log Daily Wage	4.6089	4.7482	4.6917
Variance Log Daily Wage	0.1560	0.1256	0.1427
Leave Out Sample (Pruned)			
Number of Observations	32,848	56,044	89,666
Number of Movers	3,531	6,414	9,972
Number of Firms	1,282	1,684	2,974
Mean Log Daily Wage	4.6015	4.7636	4.7047
Variance Log Daily Wage	0.1674	0.1245	0.1465
Maximum Leverage (P_{ii})	0.9241	0.9085	0.9236

HO adjustment *under*-corrects

(Evidence of substantial heteroscedasticity)

Table 2: Variance Decomposition

	Rovigo [1]	Belluno [2]	Rovigo - Belluno [3]
Variance of Log Wages	0.1674	0.1245	0.1465
Variance of Firm Effects			
Plug in (AKM)	0.0831	0.0198	0.0607
Homoscedatic Correction	0.0722	0.0136	0.0538
Leave Out	0.0609	0.0103	0.0442
	(0.0083)	(0.0011)	(0.0110)
Variance of Worker Effects			
Plug in (AKM)	0.0926	0.1035	0.1032
Homoscedatic Correction	0.0758	0.0883	0.0859
Leave Out	0.0647	0.0853	0.0792
	(0.0043)	(0.0011)	(0.0038)

HO adjustment *under*-corrects

(Evidence of substantial heteroscedasticity)

Table 2: Variance Decomposition

	Rovigo [1]	Belluno [2]	Rovigo - Belluno [3]
Variance of Log Wages	0.1674	0.1245	0.1465
Variance of Firm Effects			
Plug in (AKM)	0.0831	0.0198	0.0607
Homoscedatic Correction	0.0722	0.0136	0.0538
Leave Out	0.0609	0.0103	0.0442
	(0.0083)	(0.0011)	(0.0110)
Variance of Worker Effects			
Plug in (AKM)	0.0926	0.1035	0.1032
Homoscedatic Correction	0.0758	0.0883	0.0859
Leave Out	0.0647	0.0853	0.0792
	(0.0043)	(0.0011)	(0.0038)

Covariance flips sign!

Table 2: Variance Decomposition

	Rovigo [1]	Belluno [2]	Rovigo - Belluno [3]
Variance of Log Wages	0.1674	0.1245	0.1465
Covariance Firm, Worker Effects			
Plug in (AKM)	-0.0072	-0.0039	-0.0126
Homoscedatic Correction	0.0030	0.0018	-0.0038
Leave Out	0.0138	0.0046	0.0028
	(0.0043)	(0.0009)	(0.0076)
Correlation of Worker, Firm Effects			
Plug in (AKM)	-0.0821	-0.0863	-0.1593
Homoscedatic Correction	0.0409	0.0511	-0.0555
Leave Out	0.2202	0.1538	0.0469
Coefficient of Determination			
Plug in (AKM)	0.9637	0.9280	0.9463
Homoscedatic Correction	0.9213	0.8490	0.8850
Leave Out	0.9153	0.8414	0.8797

Leave out finds substantial PAM

Table 2: Variance Decomposition

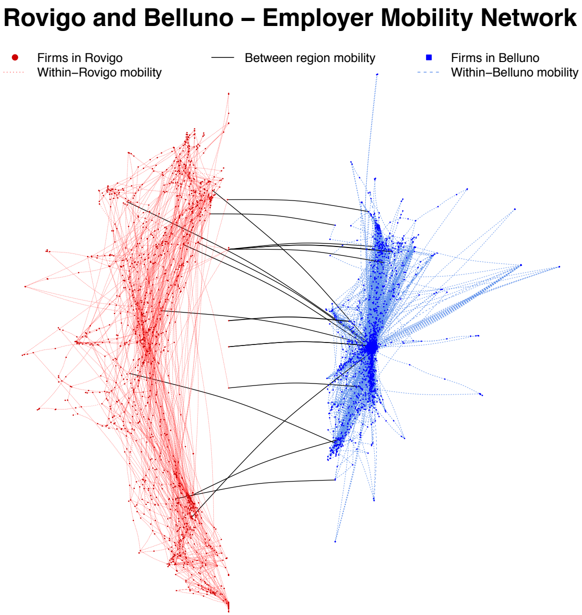
	Rovigo	Belluno	Rovigo - Belluno
	[1]	[2]	[3]
Variance of Log Wages	0.1674	0.1245	0.1465
Covariance Firm, Worker Effects			
Plug in (AKM)	-0.0072	-0.0039	-0.0126
Homoscedatic Correction	0.0030	0.0018	-0.0038
Leave Out	0.0138	0.0046	0.0028
	(0.0043)	(0.0009)	(0.0076)
Correlation of Worker, Firm Effects			
Plug in (AKM)	-0.0821	-0.0863	-0.1593
Homoscedatic Correction	0.0409	0.0511	-0.0555
Leave Out	0.2202	0.1538	0.0469
Coefficient of Determination			
Plug in (AKM)	0.9637	0.9280	0.9463
Homoscedatic Correction	0.9213	0.8490	0.8850
Leave Out	0.9153	0.8414	0.8797

AKM model exhibits very strong explanatory power

(Even after adjustment for "over-fitting")

Table 2: Variance Decomposition

	Rovigo [1]	Belluno [2]	Rovigo - Belluno [3]
Variance of Log Wages	0.1674	0.1245	0.1465
Covariance Firm, Worker Effects			
Plug in (AKM)	-0.0072	-0.0039	-0.0126
Homoscedatic Correction	0.0030	0.0018	-0.0038
Leave Out	0.0138	0.0046	0.0028
	(0.0043)	(0.0009)	(0.0076)
Correlation of Worker, Firm Effects			
Plug in (AKM)	-0.0821	-0.0863	-0.1593
Homoscedatic Correction	0.0409	0.0511	-0.0555
Leave Out	0.2202	0.1538	0.0469
Coefficient of Determination			
Plug in (AKM)	0.9637	0.9280	0.9463
Homoscedatic Correction	0.9213	0.8490	0.8850
Leave Out	0.9153	0.8414	0.8797


Rovigo and Belluno – Employer Mobility Network
Firms in Rovigo
Within−Rovigo mobility
Between region mobility
Firms in Belluno
Within−Belluno mobility

Firm effects higher in Belluno

Appendix Table A.1: Provincial Differences in Mean Effects

Firm Effects	
Avg. Firm Effects (Belluno)	-0.0189
Avg. Firm Effects (Rovigo)	-0.2787
Difference	0.2598
	(0.0941)
Lindeberg Condition $(\max_i w_{i1}^2)$	0.0381
Person Effects	
Avg. Person Effects (Belluno)	4.7823
Avg. Person Effects (Rovigo)	4.8854
Difference	-0.1020
	(0.0941)
Lindeberg Condition $(\max_i w_{i1}^2)$	0.0381

But person effects seem lower

(Hard to tell b/c of limited mobility!)

Appendix Table A.1: Provincial Differences in Mean Effects

Firm Effects	
Avg. Firm Effects (Belluno)	-0.0189
Avg. Firm Effects (Rovigo)	-0.2787
Difference	0.2598
	(0.0941)
Lindeberg Condition $(\max_i w_{i1}^2)$	0.0381
Person Effects	
Avg. Person Effects (Belluno)	4.7823
Avg. Person Effects (Rovigo)	4.8854
Difference	-0.1020
	(0.0941)
Lindeberg Condition $(\max_i w_{i1}^2)$	0.0381

Pooling *increases* the std error!

Table 3: Inference on Variance of Firm Effects

	Rovigo	Belluno	Rovigo - Belluno
	[1]	[2]	[3]
Variance of Firm Effects			
Leave out estimate	0.0609	0.0103	0.0442
	(0.0083)	(0.0011)	(0.0110)
95% Confidence Intervals - Strong id (q=0)	[0.0446; 0.0771]	[0.0081; 0.0125]	[0.0226; 0.0658]
95% Confidence Intervals - Weak id (q=1)	[0.0455; 0.0795]	[0.0081; 0.0127]	[0.0288; 0.0786]
Curvature ($\hat{\kappa}$)	0.1792	0.1372	1.4448
Diagnostics			
Eigenvalue Ratio - 1	0.1062	0.0465	0.8866
Eigenvalue Ratio - 2	0.0623	0.0439	0.0132
Eigenvalue Ratio - 3	0.0499	0.0348	0.0081
Lindeberg Condition ($\max_i \mathrm{w}_{i1}^2$)	0.0200	0.2681	0.0378
Sum of Squared Eigenvalues	0.0006	0.0002	0.0001

Consistent estimates

Table 3: Inference on Variance of Firm Effects

	Rovigo [1]	Belluno [2]	Rovigo - Belluno [3]
Variance of Firm Effects			
Leave out estimate	0.0609	0.0103	0.0442
	(0.0083)	(0.0011)	(0.0110)
95% Confidence Intervals - Strong id (q=0)	[0.0446; 0.0771]	[0.0081; 0.0125]	[0.0226; 0.0658]
95% Confidence Intervals - Weak id (q=1)	[0.0455; 0.0795]	[0.0081; 0.0127]	[0.0288; 0.0786]
Curvature ($\hat{\kappa}$)	0.1792	0.1372	1.4448
Diagnostics			
Eigenvalue Ratio - 1	0.1062	0.0465	0.8866
Eigenvalue Ratio - 2	0.0623	0.0439	0.0132
Eigenvalue Ratio - 3	0.0499	0.0348	0.0081
Lindeberg Condition ($\max_i w_{i1}^2$)	0.0200	0.2681	0.0378
Sum of Squared Eigenvalues	0.0006	0.0002	0.0001

Confidence interval adapts to bottleneck

Table 3: Inference on Variance of Firm Effects

	Rovigo [1]	Belluno [2]	Rovigo - Belluno [3]
Variance of Firm Effects			
Leave out estimate	0.0609	0.0103	0.0442
	(0.0083)	(0.0011)	(0.0110)
95% Confidence Intervals - Strong id (q=0)	[0.0446; 0.0771]	[0.0081; 0.0125]	[0.0226; 0.0658]
95% Confidence Intervals - Weak id (q=1)	[0.0455; 0.0795]	[0.0081; 0.0127]	[0.0288; 0.0786]
Curvature ($\hat{\kappa}$)	0.1792	0.1372	1.4448
Diagnostics			
Eigenvalue Ratio - 1	0.1062	0.0465	0.8866
Eigenvalue Ratio - 2	0.0623	0.0439	0.0132
Eigenvalue Ratio - 3	0.0499	0.0348	0.0081
Lindeberg Condition ($\max_i \mathrm{w}_{i1}^2$)	0.0200	0.2681	0.0378
Sum of Squared Eigenvalues	0.0006	0.0002	0.0001

Strong curvature / big top eig share in pooled sample

(But Lindeberg condition is satisfied)

Table 3: Inference on Variance of Firm Effects

	Rovigo [1]	Belluno [2]	Rovigo - Belluno [3]
Variance of Firm Effects			
Leave out estimate	0.0609	0.0103	0.0442
	(0.0083)	(0.0011)	(0.0110)
95% Confidence Intervals - Strong id (q=0)	[0.0446; 0.0771]	[0.0081; 0.0125]	[0.0226; 0.0658]
95% Confidence Intervals - Weak id (q=1)	[0.0455; 0.0795]	[0.0081; 0.0127]	[0.0288; 0.0786]
Curvature ($\hat{\kappa}$)	0.1792	0.1372	1.4448
Diagnostics			
Eigenvalue Ratio - 1	0.1062	0.0465	0.8866
Eigenvalue Ratio - 2	0.0623	0.0439	0.0132
Eigenvalue Ratio - 3	0.0499	0.0348	0.0081
Lindeberg Condition ($\max_i \mathrm{w}_{i1}^2$)	0.0200	0.2681	0.0378
Sum of Squared Eigenvalues	0.0006	0.0002	0.0001

Simulations condition on observed mobility network

Table 4: Montecarlo Results for the Variance of Firm Effects

	[1] Rovigo	[2] Belluno	[3] Rovigo - Belluno
True Variance of the Firm Effects	0.0609	0.0103	0.0442
Mean, Standard deviation across Simulations			
Leave Out	0.0608	0.0103	0.0443
	(0.0073)	(0.0010)	(0.0116)
Plug-in (AKM)	0.0841	0.0196	0.0619
	(0.0073)	(0.0010)	(0.0116)
Homoscedatic Correction	0.0735	0.0134	0.0524
	(0.0073)	(0.0010)	(0.0116)
Mean estimated Standard Error	0.0074	0.0010	0.0108
Coverage Rate at 95%			
Leave Out - Strong Id (q=0)	0.9479	0.9469	0.8535
Leave Out - Weak Id (q=1)	0.9634	0.9701	0.9736

Leave-out estimator is unbiased

Table 4: Montecarlo Results for the Variance of Firm Effects

	[1] **Rovigo**	[2] **Belluno**	[3] **Rovigo - Belluno**
True Variance of the Firm Effects	0.0609	0.0103	0.0442
Mean, Standard deviation across Simulations			
Leave Out	0.0608	0.0103	0.0443
	(0.0073)	(0.0010)	(0.0116)
Plug-in (AKM)	0.0841	0.0196	0.0619
	(0.0073)	(0.0010)	(0.0116)
Homoscedatic Correction	0.0735	0.0134	0.0524
	(0.0073)	(0.0010)	(0.0116)
Mean estimated Standard Error	0.0074	0.0010	0.0108
Coverage Rate at 95%			
Leave Out - Strong Id (q=0)	0.9479	0.9469	0.8535
Leave Out - Weak Id (q=1)	0.9634	0.9701	0.9736

Plug-in / HO severely biased

Table 4: Montecarlo Results for the Variance of Firm Effects

	[1] Rovigo	[2] Belluno	[3] Rovigo - Belluno
True Variance of the Firm Effects	0.0609	0.0103	0.0442
Mean, Standard deviation across Simulations			
Leave Out	0.0608	0.0103	0.0443
	(0.0073)	(0.0010)	(0.0116)
Plug-in (AKM)	0.0841	0.0196	0.0619
	(0.0073)	(0.0010)	(0.0116)
Homoscedatic Correction	0.0735	0.0134	0.0524
	(0.0073)	(0.0010)	(0.0116)
Mean estimated Standard Error	0.0074	0.0010	0.0108
Coverage Rate at 95%			
Leave Out - Strong Id (q=0)	0.9479	0.9469	0.8535
Leave Out - Weak Id (q=1)	0.9634	0.9701	0.9736

Leave out standard error is consistent

Table 4: Montecarlo Results for the Variance of Firm Effects

	[1] **Rovigo**	[2] **Belluno**	[3] **Rovigo - Belluno**
True Variance of the Firm Effects	0.0609	0.0103	0.0442
Mean, Standard deviation across Simulations			
Leave Out	0.0608	0.0103	0.0443
	(0.0073)	(0.0010)	(0.0116)
Plug-in (AKM)	0.0841	0.0196	0.0619
	(0.0073)	(0.0010)	(0.0116)
Homoscedatic Correction	0.0735	0.0134	0.0524
	(0.0073)	(0.0010)	(0.0116)
Mean estimated Standard Error	0.0074	0.0010	0.0108
Coverage Rate at 95%			
Leave Out - Strong Id (q=0)	0.9479	0.9469	0.8535
Leave Out - Weak Id (q=1)	0.9634	0.9701	0.9736

Invalid normal approximation

Table 4: Montecarlo Results for the Variance of Firm Effects

	[1] Rovigo	[2] Belluno	[3] Rovigo - Belluno
True Variance of the Firm Effects	0.0609	0.0103	0.0442
Mean, Standard deviation across Simulations			
Leave Out	0.0608	0.0103	0.0443
	(0.0073)	(0.0010)	(0.0116)
Plug-in (AKM)	0.0841	0.0196	0.0619
	(0.0073)	(0.0010)	(0.0116)
Homoscedatic Correction	0.0735	0.0134	0.0524
	(0.0073)	(0.0010)	(0.0116)
Mean estimated Standard Error	0.0074	0.0010	0.0108
Coverage Rate at 95%			
Leave Out - Strong Id (q=0)	0.9479	0.9469	0.8535
Leave Out - Weak Id (q=1)	0.9634	0.9701	0.9736

Weak-id interval slightly conservative

Table 4: Montecarlo Results for the Variance of Firm Effects

	[1] Rovigo	[2] Belluno	[3] Rovigo - Belluno
True Variance of the Firm Effects	0.0609	0.0103	0.0442
Mean, Standard deviation across Simulations			
Leave Out	0.0608	0.0103	0.0443
	(0.0073)	(0.0010)	(0.0116)
Plug-in (AKM)	0.0841	0.0196	0.0619
	(0.0073)	(0.0010)	(0.0116)
Homoscedatic Correction	0.0735	0.0134	0.0524
	(0.0073)	(0.0010)	(0.0116)
Mean estimated Standard Error	0.0074	0.0010	0.0108
Coverage Rate at 95%			
Leave Out - Strong Id (q=0)	0.9479	0.9469	0.8535
Leave Out - Weak Id (q=1)	0.9634	0.9701	0.9736

Summary

We proposed an unbiased and consistent estimator of any variance component in a heteroscedastic linear model w/ many regressors.

Robust inference procedure can be used to

- Test linear restrictions ("het consistent F-test")
- Build weak-id robust confidence intervals for variance components
- Eigenvalue based diagnostics for weak identification – in practice, $q = 1$ appears to provide good coverage even with very weak connectivity

MATLAB code available at:
https://github.com/rsaggio87/LeaveOutTwoWay.