

Estimation and exclusion restrictions in clustered linear models

Anna Mikusheva (MIT and ACE)

Mikkel Sølvsten (Aarhus University, ACE)

Baiyun Jing (Harvard University)



Intro: Clustered Data

- Observations $\ell \in \{1, \dots, n\}$, grouped in clusters $S_i, i \in \{1, \dots, N\}$
 - Independence across clusters; arbitrary dependence within clusters
 - Common examples:
 - Panel data
 - Group/family data
 - Spatial/network data
- Challenge: need to control for unobserved heterogeneity results in many controls (high-dimensional nuisance parameters)
- New challenge: How to handle *reasonable exogeneity conditions*

Model and Exogeneity

- Linear model:

$$y_\ell = x_\ell\beta + w_\ell'\delta + e_\ell$$

- Parameter of interest: β ; w_ℓ are fixed high-dimensional controls
- Key challenge: Formulating exogeneity
 - $\mathbb{E}[x_\ell e_\ell] = 0$: Too weak (no identifying power with fixed effects)
 - $\mathbb{E}[e_\ell | x] = 0$: Strict exogeneity—unrealistic in many applications
 - Our framework: Define an exclusion matrix \mathcal{E} with

$$\mathcal{E}_{\tilde{\ell}\ell} = 1 \quad \text{whenever} \quad \mathbb{E}[x_{\tilde{\ell}}e_\ell] = 0$$

- Maintained assumption: $\mathcal{E}_{\ell\ell} = 1$
- From independence: $\mathcal{E}_{\tilde{\ell}\ell} = 1$ when $\tilde{\ell}$ and ℓ are in different clusters

Example: Dynamic Panel

$$y_{it} = x_{it}\beta + w'_{it}\delta + e_{it}$$

- Controls may include entity and time fixed effects, personal trends
- Clustering may be at the individual or group (family) level
- Weak exogeneity assumptions:
 - $\mathbb{E}[e_{it} | x_{is}, s \leq t] = 0$: Allows for feedback
 - $\mathbb{E}[e_{it} | x_{is}, s \leq t \text{ or } s > t + m] = 0$: Limited feedback
 - May allow or restrict feedback from siblings (if clustered at the family level)

Example: Spatial Data/Network data

$$y_\ell = x_\ell\beta + w_\ell'\delta + e_\ell$$

- Spatial data (Egger et al. (2022): RCT in Kenya):
 - x_ℓ, y_ℓ : Policy and outcome for village ℓ in RCT
 - Villages belong to geographic areas (clusters S_i)
 - Interference concerns: direct effect from x_ℓ to neighboring $y_{\tilde{\ell}}$
 - One cannot impose exclusion assumption on close villages, but can on distant ones
- Network data (Paluck et al. (2016): RCT in middle schools):
 - x_ℓ, y_ℓ : treatment and outcome for individual ℓ in RCT
 - Clusters: different schools
 - Friend's treatment may have a direct hard to model effect on student's outcome

Consequences of Partial Exogeneity

$$y_\ell = x_\ell\beta + w_\ell'\delta + e_\ell$$

- OLS estimator: $\hat{\beta}^{\text{LS}} = \frac{x'My}{x'Mx}$, M : projects controls out
- Under strict exogeneity ($\mathbb{E}[e | x] = 0$):
 - Can treat x as fixed
 - Estimator is linear in data (y)
 - LLN and CLT yield standard inference (with clustered standard errors)
- Under partial exogeneity (\mathcal{E}): Estimator is a **ratio of quadratic forms**

Challenges under Partial Exogeneity

$$\hat{\beta}^{\text{LS}} = \frac{x' M y}{x' M x}$$

- Estimator is a **ratio of quadratic forms**
 - Is the numerator properly centered? May have asymptotic bias (Nickell-type)
 - How to quantify uncertainty? Standard CLT not applicable
 - Denominator is random \Rightarrow potential weak identification

Summary of Contributions

$$y_\ell = x_\ell\beta + w_\ell'\delta + e_\ell$$

Let \mathcal{E} be the exclusion matrix with $\mathbb{E}[x_{\tilde{\ell}}e_\ell] = 0$ if $\mathcal{E}_{\tilde{\ell}\ell} = 1$

- **Estimator:** Internal-instrument IV
 - Automatically adapts to specified set of exclusion restrictions
 - Closest to OLS in a certain metric
 - Has desirable efficiency properties
- **Inference:**
 - Quadratic CLT for clustered data
 - Robust standard errors
- **Weak Identification:**
 - Inference robust to sampling noise in the denominator

Plan for This Talk

- **Estimator construction**
 - Formulation as an optimization problem
 - Solution and intuitive interpretation
 - Efficiency costs of relaxing exogeneity assumptions
- **Quantifying uncertainty**
 - Why clustered standard errors may fail
 - Quadratic central limit theorem (CLT)
- **Weak identification**
- **Variance estimation**

Estimation: Asymptotic Bias

$$y_\ell = x_\ell \beta + w_\ell' \delta + e_\ell$$

Theorem (Nickell Bias)

Under Assumption 1,

$$\hat{\beta}^{\text{LS}} = \beta + \frac{\frac{1}{n} \sum_{\ell, \tilde{\ell}} M_{\tilde{\ell}\ell} \mathbb{E}[x_{\tilde{\ell}} e_\ell]}{Q} + o_p(1)$$

- Assumption 1 (standard):

- $\frac{1}{n} \sum_{\ell=1}^n \mathbb{E}[\tilde{x}_\ell^2] \rightarrow Q > 0$, where $\tilde{x} = Mx$
- Clustered data, finite fourth moments
- $K/n < 1 - c$, $K = \dim(w_\ell)$, and $\max_i T_i/n \rightarrow 0$ ($T_i = |S_i|$)

Estimation: Nickell Bias in Literature

- Nickell (1981): $y_{it} = \alpha_i + y_{i,t-1}\beta + e_{it}$, with $\mathbb{E}[e_{it} | \{y_{is} : s < t\}] = 0$
- If $T_i = T$ is fixed, OLS is inconsistent; bias is $O(1/T)$ as $T \rightarrow \infty$
- Source of bias: projection mixes timing (past and future)
- Solutions in literature:
 - Analytical corrections: Hahn and Kuersteiner (2002), Hahn and Kuersteiner (2011), Breitung et al. (2022), Dhaene and Jochmans (2016)
 - Sample-splitting: Dhaene and Jochmans (2015)
 - Internal instruments: Anderson and Hsiao (1981), Arellano and Bond (1991), Ahn and Schmidt (1995), Alvarez and Arellano (2003), Blundell and Bond (1998)

Estimation: Correct centering

- Take W, \mathcal{E} and cluster identities S_i fixed and known. Consider any stochastic data (x_ℓ, y_ℓ) :
 - $y_\ell = x_\ell \beta + w_\ell' \delta + e_\ell$
 - $\mathbb{E}[x_{\tilde{\ell}} e_\ell] = 0$ if $\mathcal{E}_{\tilde{\ell}\ell} = 1$
- Then NO unbiased estimator of β exists
- No unbiased estimator of β exists even in a cross-section $y_\ell = \beta x_\ell + e_\ell$ when x_ℓ are random with $\mathbb{E}[e_\ell] = \mathbb{E}[x_\ell e_\ell] = 0$

Definition

An estimator $\hat{\beta} = \frac{C_1(x, y)}{C_2(x)}$ is said to be *correctly centered* if for all $F \in \mathcal{F}$:

$$\mathbb{E}_F[C_1(x, y)] = \beta \mathbb{E}_F[C_2(x)].$$

Estimation: Correctly Centered Estimator

Theorem

Let A be an $n \times n$ matrix. Consider estimators of the form $\hat{\beta}^A = \frac{x' Ay}{x' Ax}$. Estimator $\hat{\beta}^A$ is correctly centered **if and only if** it satisfies

- *Partialling-out property (POP):* $AM = A$
- *Correct centering (CC):* $A\tilde{\epsilon}_\ell = 0$ when $\mathcal{E}\tilde{\epsilon}_\ell = 0$

$$\hat{\beta}^A = \frac{x' Ay}{x' Ax} = \frac{x' A(x\beta + W\delta + e)}{x' Ax} = \beta + \frac{x' Ae}{x' Ax}$$

Estimation: Internal Instruments

Interpretation of $\hat{\beta}^A$ satisfying (POP) and (CC): internal instrument $z = A'x$

- Condition (CC) implies $\mathbb{E}[z_\ell e_\ell] = 0$
- Apply IV to the regression of y on x using instrument z_ℓ and controls w_ℓ

$$\hat{\beta}^{\text{IV}} = \frac{z'My}{z'Mx} = \frac{x'AMy}{x'AMx} = \frac{x'Ay}{x'Ax} = \hat{\beta}^A$$

All linear just identified internal instruments are in this class

Estimation: Optimization Problem

- Class \mathcal{A} of valid matrices A defined by:

(POP) $AM = A$

(CC) $A_{\tilde{\ell}\ell} = 0$ if $\mathcal{E}_{\tilde{\ell}\ell} = 0$

- Valid choice closest to OLS

$$A^* = \arg \min_{A \in \mathcal{A}} \|A - I_n\|_F = \arg \min_{A \in \mathcal{A}} \|A - M\|_F$$

Lemma

Assume $\mathbb{E}[ee' | x] = \sigma^2 I_n$, $\mathbb{E}[xx'] = \lambda I_n$. Then, the asymptotic variance functional

$$V(A) = \frac{\text{Var}(x' Ae)}{(\mathbb{E}[x' Ax])^2} = \frac{\text{Var}(z' e)}{(\mathbb{E}[z' x])^2}$$

is minimized over \mathcal{A} at $A = A^*$.

Estimation: Efficiency Considerations

- Efficiency loss as price of robustness (assumptions imply strict exogeneity, class \mathcal{A} is robust to weak exogeneity)
- Efficiency depends on distribution of x and e
 - role of conditional heteroskedasticity as in Chamberlain (1987)
 - any information about x may be used to improve efficiency via re-weighting of GLS-transformation
- If $\mathbb{E}[xx'] = \Omega$, then

$$A_{\Omega}^* = \arg \min_{A \in \mathcal{A}} \|\Omega^{1/2}(A - I_n)\|_F.$$

- Potential danger: estimation of Ω may bring back bias (as in Arellano and Bond (1991): *many instrument bias*)

Estimation: Solution (1) - Linear Projection

$$A^* = \arg \min_A \|A - M\|_F$$

subject to (POP) : $AM = A$, (CC) : $A_{\tilde{\ell}\ell} = 0$ if $\mathcal{E}_{\tilde{\ell}\ell} = 0$

- Quadratic minimization with linear constraints
- Let (CC) constraint be $\mathcal{L}\text{vec}(A) = 0$, then:

$$\text{vec}(A^*) = (I_{n^2} - \mathcal{L}'(\mathcal{L}\mathcal{L}')^+ \mathcal{L})\text{vec}(M)$$

- Computationally feasible (especially if solved cluster-by-cluster)

Estimation: Solution (2) - Projection Interpretation

$$A^* = \arg \min_{A \in \mathcal{A}} \|A - I_n\|_F = \arg \min_{A \in \mathcal{A}} \|A - M\|_F.$$

$$A^* = \text{proj}_F(M, \mathcal{A})$$

- Let B be a sparse matrix: $B_{\tilde{\ell}\ell} = 0$ if $\mathcal{E}_{\tilde{\ell}\ell} = 1$ ($\Leftrightarrow \mathbb{E}[x_{\tilde{\ell}}e_{\ell}] = 0$)
- Then $BM \perp\!\!\!\perp \mathcal{A}$ in Frobenius inner product
- We can find B with this property and $(I - B)M \in \mathcal{A}$, then

$$A^* = (I - B)M$$

- Matrix B is sparse (block-clustered); estimation can be done cluster-by-cluster

Estimation: Solution (3) - Leave-Out Interpretation

- For each ℓ :
 - Drop observations $\tilde{\ell}$ s.t. $\mathcal{E}_{\ell\tilde{\ell}} = 0$ (i.e., x_ℓ correlated with $e_{\tilde{\ell}}$)
 - Define $M^{(\ell)}$: projection using only remaining data
 - Transformed equation: $y_\ell^* = \beta x_\ell^* + e_\ell^*$, where e_ℓ^* is residual from leave-out projection $M^{(\ell)}$
- By construction, $\mathbb{E}[x_\ell e_\ell^*] = 0$, so can use x_ℓ as instrument

$$\hat{\beta}^{\text{IV}} = \frac{\sum_\ell x_\ell y_\ell^*}{\sum_\ell x_\ell x_\ell^*} = \frac{x' A^* y}{x' A^* x} = \hat{\beta}^{A^*}$$

Estimation: Dynamic Panel Data

$$y_{it} = \alpha_i + \beta x_{it} + e_{it}, \quad \text{with} \quad \mathbb{E}[e_{it} | x_{it}, x_{i,t-1}, \dots] = 0$$

- The optimal matrix A^* corresponds to forward-differencing:

$$y_{it}^* = y_{it} - \frac{1}{T_i - t + 1} \sum_{s=t}^{T_i} y_{is}, \quad x_{it}^* = x_{it} - \frac{1}{T_i - t + 1} \sum_{s=t}^{T_i} x_{is}$$

- Estimate β from transformed model (Arellano and Bover (1995)):

$$y_{it}^* = \beta x_{it}^* + e_{it}^*, \quad \text{using } x_{it} \text{ as instrument}$$

- Remark: If the exogeneity assumption is only $\mathbb{E}[e_{it} | x_{it}] = 0$, then $A^* = 0$, i.e., no identifying variation

Plan for This Talk

- **Estimator construction**
 - Formulation as an optimization problem
 - Solution and intuitive interpretation
 - Efficiency costs of relaxing exogeneity assumptions
- **Quantifying uncertainty**
 - Why clustered standard errors may fail
 - Quadratic central limit theorem (CLT)
- **Weak identification**
- **Variance estimation**

Uncertainty Quantification: Core Setup

- Data organized into **clusters** S_i , with size $T_i = |S_i|$
For each cluster, define $(x_i, e_i) = (x_\ell, e_\ell : \ell \in S_i)$

- **Independence across clusters:**

$$(x_i, e_i) \perp\!\!\!\perp (x_j, e_j) \quad \text{for } i \neq j$$

but **arbitrary dependence within** each cluster

- Estimator:

$$\hat{\beta}^{A^*} - \beta = \frac{x' A^* e}{x' A^* x} = \frac{\sum_{i=1}^N z_i' e_i}{\sum_{i=1}^N z_i' x_i} = \frac{\sum_{i=1}^N x_i' e_i^*}{\sum_{i=1}^N x_i' x_i^*}$$

where $z = (A^*)'x$, and $z_i = (A^*)'_i x$

- Components z_i, e_i^* typically exhibit **cross-cluster dependence**

Estimator Structure and Dependence

- Full expression:

$$\hat{\beta}^{A^*} - \beta = \frac{\sum_{i,j=1}^N x_i' A_{ij}^* e_j}{\sum_{i,j=1}^N x_i' A_{ij}^* x_j}$$

- A_{ij}^* is the $T_i \times T_j$ block of matrix A^*
- Even with independent clusters (x_i, e_i) , A^* can induce cross-cluster dependence
- **LLN and CLT may not apply directly**; inference is nontrivial
- Exceptions:
 - (a) $A_{ij} = 0$ for all $i \neq j$ (only cluster fixed effects or trends)
 - (b) strict exogeneity

Motivating Example: Teacher-Student Panel

$$y_{it} = x_{it}\beta + \alpha_i + \gamma_{g(i,t)} + e_{it}$$

- α_i : student fixed effect
- $\gamma_{g(i,t)}$: teacher fixed effect assigned to student i at time t
- **Setup:**
 - Period 1: random assignment to teachers $g = 1, 2$
 - Period 2: reassignment to $g = 3, 4$
- Students with teacher sequence (1,3):

$$\Delta y_i = y_{i2} - y_{i1} = \gamma_3 - \gamma_1 + \beta \Delta x_i + \Delta e_i$$

Motivating Example: Teacher-Student Panel

- OLS on transformed outcomes:

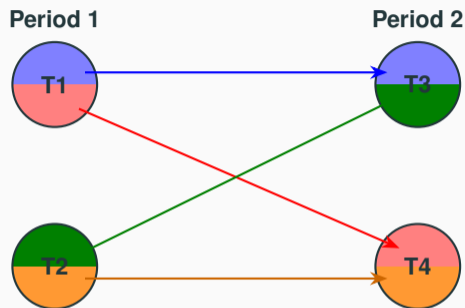
$$y_i^* = \Delta y_i - \left\{ \frac{3}{4} \overline{\Delta y}^{(1,3)} + \frac{1}{4} \overline{\Delta y}^{(1,4)} + \frac{1}{4} \overline{\Delta y}^{(2,3)} - \frac{1}{4} \overline{\Delta y}^{(2,4)} \right\}$$

- Estimator:

$$y_i^* = \beta x_i^* + e_i^* \quad (\text{IV using } x_{i1} \text{ as instrument})$$

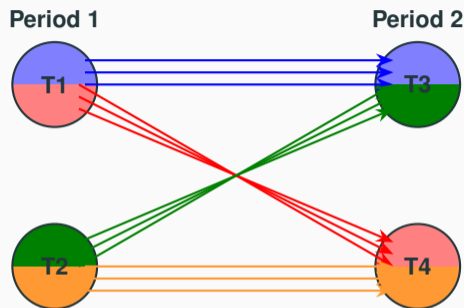
- $\hat{\beta}^{A*} = \frac{\sum_{i=1}^N x_{i1} y_i^*}{\sum_{i=1}^N x_{i1} x_i^*} = \beta + \frac{\sum_{i=1}^N x_{i1} e_i^*}{\sum_{i=1}^N x_{i1} x_i^*}$
- But: e_i^* are **not independent**
- LLN/CLT do not apply directly; difference between e_i^* and Δe_i is due to teacher effect estimation error

Alternative Strategy: Super-clusters



- **Idea 1:** Estimate $\gamma_3 - \gamma_1$ using only students in group (1,3)
- Less efficient, but forms **independent super-clusters**
- If number of such clusters increases, CLT becomes valid

Alternative Strategy: Pairwise Differencing



- **Idea 2:** Pair students with identical assignment histories $y_i^* = \Delta y_i - \Delta y_j$, run IV using x_{i1}
- Related to strategy in Verdier (2020): considers design with teachers fixed effects for data in North Carolina
 - Pairing students with similar educational history increases SEs
 - Clustered standard errors of OLS capture 80 percent of uncertainty

Our Strategy: Quadratic CLT

- Do **not modify estimator**; instead, analyze distribution
- Numerator: $\sum_{i,j=1}^N x_i' A_{ij} e_j$ is a **quadratic form**
- Establish CLT for quadratic forms under cluster dependence

Lemma

Let ξ_i be independent with $\mathbb{E}[\xi_i] = 0$, $\text{Var}(\xi_i) = I_{T_i}$, and $\mathbb{E}|\tau' \xi_i|^4 \leq C \|\tau\|^4$. Let Γ be symmetric block matrix with $\Gamma_{ii} = 0$.

Then (under negligibility condition) and as $N \rightarrow \infty$:

$$\frac{1}{\sqrt{V}} \sum_{i \neq j} \xi_i' \Gamma_{ij} \xi_j \Rightarrow \mathcal{N}(0, 1), \quad V = 2 \|\Gamma\|_F^2$$

Asymptotics of the Estimator's Numerator

- Let $\lambda_i = \mathbb{E}[x_i]$ and $v_i = x_i - \lambda_i$

$$x' Ae = \sum_{i,j} x'_i A_{ij} e_j = \sum_{i,j} (\lambda_i + v_i)' A_{ij} e_j$$

Hoeffding decomposition:

$$x' Ae = \sum_{j=1}^N \left((\lambda' A)_j + v'_j A_{jj} \right) e_j + \sum_{j=1}^N \sum_{i \neq j} v'_i A_{ij} e_j$$

- First-order term: linear, standard CLT applies
- Second-order term: quadratic, use Quadratic CLT
- **Theorem:** Asymptotically Gaussian under many-cluster setup

Asymptotics of the Estimator's Numerator

Assumptions for Gaussianity come in two flavors:

(1) Assumptions on within-cluster dependence

▶ Assumption 2

- Permits $\|\Sigma_i\| \propto CT_i$, where $\Sigma_i = \mathbb{V}(e_i, v_i)$
- Requires that higher-order dependence is controlled by second-moment dependence

(2) Assumptions on the matrix A^*

▶ Assumption 3

- Ensures an effective sample size: $\text{tr}(A^*) \geq cn$
- Imposes restrictions to limit the influence of any single cluster
- A key requirement is $\frac{\max_i T_i^4}{n} \rightarrow 0$
- In contrast, the linear CLT typically requires only $\frac{\max_i T_i^2}{n} \rightarrow 0$

If $\|\Sigma_i\| \leq C$, then we can weaken assumption to $\frac{\max_i T_i^2}{n} \rightarrow 0$

Inference: Weak Identification

- Our estimator $\hat{\beta}^{A^*} = \frac{x' A^* y}{x' A^* x} = \frac{z' y}{z' x}$ is an IV estimator
- Weak IV arises if stochastic variation in $x' A^* x$ is comparable to $\mathbb{E}[x' A^* x]$
- Weak identification is common in dynamic panel models (Arellano and Bond (1991), Bun and Windmeijer (2010))
- For inference on $H_0 : \beta = \beta_0$, we use the Anderson-Rubin (AR) test:

$$AR(\beta_0) = \frac{(x' A^* (y - x\beta_0))^2}{\hat{V}(\beta_0)}, \quad \text{critical value: } \chi_{1,1-\alpha}^2$$

- Confidence set for β constructed via test inversion:
 - Numerator and denominator are quadratic in β_0 ; allows for analytical inversion
 - $\hat{\beta}^{A^*}$ is always contained in the confidence set

Variance Estimation

$$y_\ell = x_\ell\beta + w_\ell'\delta + e_\ell$$

- Goal: Estimate the variance of $x' Ae$ when y, x are observed and under the null $H_0 : \beta = \beta_0$
- Challenge: The implied error $U = y - \beta_0 x = w'\delta + e$ differs from e
- Partialling out w : residuals $\hat{e} = MU$ differs non-negligibly from e when w is high-dimensional

Jackknife Variance Estimator

- The statistic $\mathcal{Z} = x' A^* U = f(D_1, \dots, D_N)$, where $D_i = (x_i, U_i)$ is data from cluster i .
- Define $\mathcal{Z}_{(i)} = f(D_1, \dots, D_{i-1}, 0, D_{i+1}, \dots, D_N)$,
- The jackknife variance estimator is:

$$\hat{V} = \sum_{i=1}^N (\mathcal{Z} - \mathcal{Z}_{(i)})^2 = \sum_{i=1}^N \left(\sum_{j \neq i} x'_j A_{ji}^* U_i + \sum_{j=1}^N x'_i A_{ij}^* U_j \right)^2.$$

Lemma

The jackknife variance estimator is conservative: $\mathbb{E}[\hat{V}] \geq \text{Var}(\mathcal{Z})$.

If $A_{ij}^ = 0$ for all $i \neq j$, then it is unbiased: $\mathbb{E}[\hat{V}] = \text{Var}(\mathcal{Z})$.*

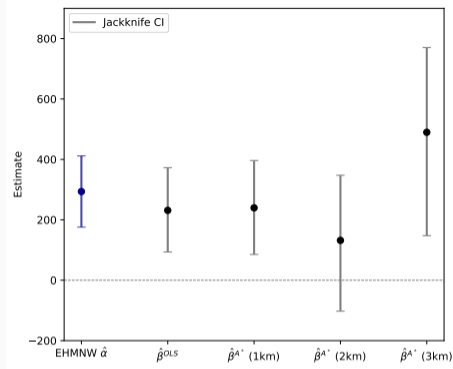
Interpretation and Literature

- The jackknife variance estimator was introduced by Efron and Stein (1981) for symmetric statistics based on i.i.d. data.
- For symmetric statistics under i.i.d. sampling, the jackknife estimator is conservative due to over-counting higher-order variance terms in Hoeffding's decomposition.
- Applications to non-symmetric and non-i.i.d. settings: Bousquet et al. (2004); MacKinnon et al. (2023).
- In our setting, conservativeness is further amplified by non-centrality of $\mathcal{Z}_{(i)}$:

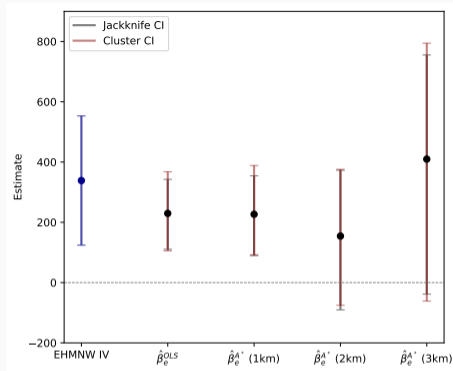
$$\mathbb{E}[\mathcal{Z}_{(i)}] \neq 0 \quad \text{while} \quad \mathbb{E}[\mathcal{Z}] = 0.$$

Empirical Application

- Based on Egger et al. (2022), who study a large-scale fiscal intervention in rural Kenya.
- RCT: one-time cash transfers equivalent to approximately 15 percent of local GDP were randomly assigned to 653 villages.
 - Rich spatial variation in treatment (two-stage randomization procedure). Cluster structure: 68 sublocations.
 - Spillover effects of not clear structural form. Spillover most likely stops after 2km
- Outcome- average household consumption in a village- depends on treatment of neighbors in complicated way
- Our approach: design approach: $x_\ell = w'_\ell \delta_x + v_\ell$
- $\mathbb{E}[v_{\tilde{\ell}}(y_\ell - \beta x_\ell)] = 0$ whenever villages ℓ and $\tilde{\ell}$ are 2km apart



Treatment effect on average household consumption; Jackknife confidence intervals are computed by inverting the respective AR test statistics. The estimates EHMNW is based on Specification (1) from EHMNW Table 1 (row 1, column 1).



Effect on average household consumption among eligible households scaled by the average transfer amount. Jackknife and cluster-robust confidence intervals inverting the AR test. EHMNW is the IV estimate reported in Table 1 (row 1, column 2).

Conclusion

- We propose a comprehensive empirical strategy for linear models with many controls and clustered data, covering:
 - Estimation under weak exogeneity
 - Quantification of uncertainty
 - Inference robust to weak identification
- Emphasis: randomness in x cannot be ignored
- Trade-off: stronger exogeneity assumptions yield higher precision, but lower credibility
- Our estimator:
 - Has a simple, leave-out interpretation
 - Is easy to compute
 - Achieves efficiency under standard conditions

Thank you!

Negligibility Condition

- For CLT to hold:

$$\frac{\max_i \left(\sum_{j \neq i} \|\Gamma_{ij}\|_F^2 \right)}{\sum_i \sum_{j \neq i} \|\Gamma_{ij}\|_F^2} \rightarrow 0 \text{ and } \frac{\|\Gamma\|^2}{\|\Gamma\|_F^2} \rightarrow 0.$$

- Interpretation:

- No large clusters
- Contributions are diffuse across blocks

▶ [Back to Quadratic CLT](#)

Asymptotic Gaussianity

Theorem

Let $x = \lambda + v$ with $\lambda = \mathbb{E}[x]$, and suppose Assumptions 2 and 3 hold. Then, as $N \rightarrow \infty$, $\frac{x' A e}{\omega} \xrightarrow{d} N(0, 1)$, where

$$\omega^2 = \sum_{i=1}^n \mathbb{V}(\omega_i e_i) + \text{tr}(\tilde{A} \tilde{A}') + \text{tr}(\tilde{A}^2), \text{ with } \omega_i = (\lambda' A)_i + v_i' A_{ii},$$

and there exists $c > 0$ such that $\omega^2 \geq cn$.

▶ [Back to Asymptotics](#)

Assumptions for Gaussianity I

- Define $\xi_i = (e_i', v_i')' \in \mathbb{R}^{2T_i}$ with $\mathbb{E}[\xi_i] = 0$
- Covariances:

$$\Sigma_i = \text{Var}(\xi_i), \quad \Phi_i = \text{Var}(v_i \otimes e_i), \quad \Psi_i = \text{Cov}(\xi_i, v_i \otimes e_i)$$

- Assumptions:

(i) Σ_i invertible

(ii) $\frac{\tau' \Phi_i \tau}{\tau' (\Sigma_{v,i} \otimes \Sigma_{e,i}) \tau} > c$

(iii) $\left\| \Sigma_{e,i}^{-1/2} \Psi_i \Phi_i^{-1/2} \right\| \leq 1 - c$

(iv) Higher moment bounds on $\xi_i, v_i \otimes e_i$

$$\mathbb{E}|\tau' \xi_i|^4 \leq C \|\tau\|^4, \quad \text{and} \quad \mathbb{E}|\tau'(v_i \otimes e_i)|^{2+2\delta} \leq C (\tau' \Phi_i \tau)^{1+\delta},$$

Assumptions for Gaussianity II

- Effective sample size: $\text{tr}(A^*) \geq cn$
- Block size: $\frac{\max_i T_i^4}{n} \rightarrow 0$
- Leverage condition:

$$\frac{\max_i (A'\lambda)'_i \Sigma_{e,i} (A'\lambda)_i}{n} \rightarrow 0$$

- Sufficient condition: $\|\lambda\|_\infty \leq C, \frac{(\max_i T_i)^4 \|M\|_\infty^2}{n} \rightarrow 0$
- Leave-out condition: $\|M_{\mathcal{X}_\ell^c}^{-1}\| \leq C$
 - $M_{\mathcal{X}_\ell^c}$ is submatrix formed from entries M_{ℓ_1, ℓ_2} with $\mathcal{E}_{\ell, \ell_j} = 0$ (condition that leave-out projection is close to full projection)