Inference for Regression with Clustered and Spatially Correlated Data

A. Colin Cameron and Douglas L. Miller .
Univ. of California - Davis, Dept. of Economics
Cornell University, Brooks School of Public Policy and Dept. of Economics

INCOMPLETE SLIDES
These slides are based on a survey article that is in preparation.
When complete, the survey will be posted at cameron.econ.ucdavis.edu/ in the Papers section.

April 7, 2025

Introduction

- Consider inference for regression with data that are correlated across some measure of distance
  - most often geographic distance.
- Positive correlation in distance leads to an information loss relative to independent data.
- Failure to adequately account for this loss in precision can lead to
  - greatly under-estimated standard errors for model parameters
  - confidence intervals that are too narrow
  - hypothesis tests that greatly over-reject.

- First consider inference with clustering
  - appropriate when spatial correlation disappears when a threshold is crossed.
  - **the discussion is actually relevant for any clustered setting**
  - key surveys are Cameron and Miller (2005, JHR) and MacKinnon, Nielson and Webb (2023a, JE).
- Second consider spatial HAC inference
  - appropriate when spatial correlation is dampening in distance
  - key reference is Conley (1999, JE).

- Topics covered include
  - standard asymptotic results
    - essential to at least use these
  - adjustment to finite samples
    - asymptotics can provide poor finite-sample approximation
  - briefly: design based inference for binary treatment.
- We focus on OLS
- Most asymptotic results extend to IV, probit, logit and GMM.
  - though many finite sample improvements are specific to OLS.

- Nunn and Wantchekon (AER 2011)
  - effect of historical slave trade on current levels of trust
  - sample of individuals in 17 African countries.
- y: *trust*: individual level of trust in their locally-elected council
  - scored on a four point-scale
    - 0 ("not at all"), 1 ("just a little"), 2 ("somewhat"), 3 ("a lot").
- Key x: *exports* is a measure of slave exports
  - natural logarithm of one plus (total slave exports / area in kilometers^2)
  - **collected at the ethnicity level**.
- What to cluster on?
  - *ethnicity* as x invariant within ethnicity
  - distance
    - cluster by town, district, region, country?
    - spatial HAC by ethnicity distance.

Outline

1. Introduction
2. One-way Clustering
3. Beyond One-way Clustering
4. Spatially Dampening Correlation
5. Regression with Spatial Weights
6. Conclusion

2.1 How to Form Clusters

- It is not always clear how to form clusters.
- For OLS: the within cluster correlation of $\mathbf{x}_i u_i$ matters.
- For $\widehat{\beta}_k$ the cluster-robust variance is a multiple τ_k^2 of the default OLS variance under i.i.d. errors where

$$\tau_k^2 \simeq 1 + (\bar{N}_g - 1) \times \rho_{x_k u} \text{ (very approximately).}$$

- Here $\rho_{x_k u}$ is the intracluster correlation of $x_k u$ and $\bar{N}_g$ is the average number of observations per cluster
  - so large if N_g large even if $\rho_{x_k u}$ is very small.

How to Form Clusters (Continued)

- In many studies interest lies in a specific regressor, say x_k
  - form clusters on groupings with high within-group correlation of x_k
  - in many policy applications treatment is grouped.
- Possibly cluster on groupings of y with high within-group correlation
  - since this might lead to nontrivial within group correlation of u
    - ★ even after controlling for the other regressors.
- For complex surveys one should at least cluster at the level of the primary sampling unit.

How to Form Clusters (continued)

- If potential clustering is nested such as individuals in families in towns in states
  - common practice: cluster at increasingly aggregated levels
    - stop when there is relatively little increase in standard errors.
  - MacKinnon, Nielson and Webb (2023c, JE) provide formal tests
  - possible downside is noisier inference due to few clusters.
- In some cases there are two (or more) nonnested ways to cluster
  - e.g. in an individual wage regression potentially cluster on occupation and on industry.
  - extensions to one-way clustering are presented in section 3.

2.2 One-way Cluster Variance Estimators

- Traditional **model-based approach** - randomness from the error
  - recent **design-based approach** presented later.
- OLS estimates model for individual i in cluster g

$$\begin{aligned} y_{ig} &= \mathbf{x}_{ig}'\beta + u_{ig}, \;\; i = 1, ..., N_g, \; g = 1, ..., G, \; N = \textstyle\sum_{g=1}^{G} N_g \\ \mathbf{y}_g &= \mathbf{X}_g'\beta + \mathbf{u}_g, \quad g = 1, ..., G \\ \mathbf{y} &= \mathbf{X}\beta + \mathbf{u}_g \end{aligned}$$

- Clustered errors: u_{ig} independent over g and arbitrarily correlated within g

$$\mathrm{E}[u_{ig}u_{jg'}|\mathbf{x}_{ig}, \mathbf{x}_{jg'}] = 0, \text{ unless } g = g'.$$

- Then OLS estimator $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ has (conditional on $\mathbf{X}$)

$$\begin{aligned} \mathrm{Var}[\widehat{\beta}] &= (\mathbf{X}'\mathbf{X})^{-1}\mathrm{E}[\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}](\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\textstyle\sum_{g=1}^{G} \mathrm{E}[\mathbf{X}_g'\mathbf{u}_g\mathbf{u}_g'\mathbf{X}_g])(\mathbf{X}'\mathbf{X})^{-1}. \end{aligned}$$

Aside: Intuition for why cluster leads to larger variance

- Consider regression model $y_i = \mu + u_i$
  where $u_i \sim (\mu, \sigma^2)$ and $\mathrm{Cov}(u_i, u_j) = \rho\sigma^2$ for $i \neq j$
- OLS yields the sample mean $\widehat{\mu} = \bar{y}$ and

$$\mathrm{Var}[\widehat{\mu}] = \mathrm{Var}[\bar{y}] = \mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^N y_i\right] = \frac{1}{N^2}\left[\sum_{i=1}^N \sum_{j=1}^N \mathrm{Cov}(y_i, y_j)\right].$$

- Here $\mathrm{Cov}(y_i) = \sigma^2$ and $\mathrm{Cov}(y_i, y_j) = \rho\sigma^2$ for $i \neq j$

So $\mathrm{Var}[\mathbf{y}] = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{bmatrix}$

$$\begin{aligned} \text{and } \mathrm{Var}[\bar{y}] &= \frac{1}{N^2}\left[\sum_{i=1}^N \mathrm{Var}(y_i) + \sum_{i=1}^N \sum_{j=1; j\neq i}^N \mathrm{Cov}(y_i, y_j)\right] \\ &= \frac{1}{N^2}[N\sigma^2 + N(N-1)\rho\sigma^2] = \frac{1}{N}\sigma^2\{1 + (N-1)\rho\}. \end{aligned}$$

Aside: continued

- So $\text{Var}[\bar{y}] = \{1 + (N-1)\rho\}$ times usual $\frac{1}{N}\sigma^2$.
- The multiplier grows linearly in N and ρ
- The multiplier can be very large even when ρ is small
    - e.g. $\rho = 0.1$ and $N = 81$ then $\text{Var}[\bar{y}] = 9 \times (\frac{1}{N}\sigma^2)$.
- More generally with regressors and clustering
    - let $\bar{N}_g$ be average number of observations per cluster
    - let ρ_{xu} be the within cluster correlation of the product xu
    - then very approximately cluster variance is τ_j^2 times that of default OLS

$$\tau_j^2 \simeq 1 + \rho_{xu}(\bar{N}_g - 1)$$

One-way Cluster-Robust Variance Matrix Estimators

- For OLS with independent clustered errors

$$\text{Var}[\widehat{\beta}] = (\mathbf{X}'\mathbf{X})^{-1}(\textstyle\sum_{g=1}^{G} \text{E}[\mathbf{X}_g'\mathbf{u}_g\mathbf{u}_g'\mathbf{X}_g])(\mathbf{X}'\mathbf{X})^{-1}$$

- A (heteroskedastic- and) cluster-robust variance estimate (CRVE) is

$$\begin{aligned} \widehat{\text{V}}_{\text{CR}}[\widehat{\beta}] &= c_N(\mathbf{X}'\mathbf{X})^{-1}(\textstyle\sum_{g=1}^{G} \mathbf{X}_g'\widetilde{\mathbf{u}}_g\widetilde{\mathbf{u}}_g'\mathbf{X}_g)(\mathbf{X}'\mathbf{X})^{-1} \\ \widetilde{\mathbf{u}}_g \text{ is } \widehat{\mathbf{u}}_g &= \mathbf{y}_g - \mathbf{X}_g'\widehat{\beta} \text{ or a modification of } \widehat{\mathbf{u}}_g \\ c_N &\geq 1 \text{ is a finite sample correction} \end{aligned}$$

- Key for consistency is $G \rightarrow \infty$ and no single cluster too dominant.
- Due to Liang and Zeger (1986, *JASA*) and Arellano (1987, *JE*)
  - Natural generalization of White (1980) heteroskedastic-robust
    - special case where $N_g = 1$ for all g.

Finite Sample Adjustments

- OLS residuals overfit: $\widehat{\mathbf{u}} = \mathbf{M}\mathbf{u}$ where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
- This leads to adjustments that generalize ones proposed for independent heteroskedastic errors by MacKinnon and White (1985)
  - Bell and McCaffrey (2002) is a key reference
    - CV2 is called bias-reduced linearization (BLR)
    - CV3 is equivalent to a jackknife (if all $\mathbf{M}_g$ invertible).

	c_N	$\widetilde{\mathbf{u}}_g$	Formulas
CV0	1	$\widehat{\mathbf{u}}_g$	$\widehat{\mathbf{u}}_g = \mathbf{y}_g - \mathbf{X}_g'\widehat{\beta}$
CV1a	$\frac{G}{G-1}$	$\widehat{\mathbf{u}}_g$	"
CV1b	$\frac{G}{G-1} \times \frac{N-1}{N-K}$	$\widehat{\mathbf{u}}_g$	"
CV2	1	$\mathbf{M}_g^{-1/2}\widehat{\mathbf{u}}_g$	$\mathbf{M}_g = \mathbf{I} - \mathbf{X}_g(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_g'$
CV3/CV_{Jack}	$\frac{G}{G-1}$	$\mathbf{M}_g^{-1}\widehat{\mathbf{u}}_g$	"

- **Key: Recent research strongly supports use CV2 or CV3.**

CV3 and Jackknife

- Let $\widehat{\boldsymbol{\beta}}_{-g}$ be OLS when delete cluster g, $g = 1, ..., G$

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{-g} &= \left[\sum\nolimits_{h \neq g}^{G} \mathbf{X}_h' \mathbf{X}_h\right]^{-1} \sum\nolimits_{h \neq g}^{G} \mathbf{X}_h' \mathbf{y}_h \\ &= \left[\mathbf{X}'\mathbf{X} - \mathbf{X}_h'\mathbf{X}_h\right]^{-1} \left[\mathbf{X}'\mathbf{y} - \mathbf{X}_h'\mathbf{y}_h\right].\end{aligned}$$

- The leave-one-out cluster jackknife uses

$$\text{CV}_{Jack} = \frac{G}{G-1} \sum\nolimits_{g=1}^{G} (\widehat{\boldsymbol{\beta}}_{-g} - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_{-g} - \widehat{\boldsymbol{\beta}})'.$$

- $\text{CV}_{Jack} = \text{CV3}$ if $\widehat{\boldsymbol{\beta}}_{-g}$ can be computed for all G
  - i.e. if $\mathbf{X}'\mathbf{X} - \mathbf{X}_g'\mathbf{X}_g$ is invertible for all G.
- Otherwise
  - drop clusters where can't compute $\widehat{\boldsymbol{\beta}}_{-g}$
  - better: use a Moore-Penrose generalized inverse

2.3 Application

- Nunn and Wantchekon (AER 2011)
  - effect of historical slave trade on current levels of trust
  - sample of individuals in 17 African countries.
- y: *trust*: individual level of trust in their locally-elected council
  - scored on a four point-scale
    - 0 ("not at all"), 1 ("just a little"), 2 ("somewhat"), 3 ("a lot").
- Key x: *exports* is a measure of slave exports
  - natural logarithm of one plus (total slave exports / area in kilometers^2)
  - **collected at the ethnicity level**.
- Controls: age, age-squared, gender, urban indicator, five living conditions, ten educational-level indicators, 25 occupation indicators, ethnic fractionalization within district, percentage of district with the same ethnicity as the current individual, and 17 country fixed effects.

OLS Regression

- Summary statistics for y and key x

Variable	Obs	Mean	Std. dev.	Min	Max
trust3	19,733	1.65996	1.100935	0	3
exports	19,733	.5357018	.9511826	0	3.65603

- Focus on OLS regression reported in their Table 2 Column 3
  - $N = 17,773$ individuals in 1,257 districts, 171 regions and 17 countries.
- Estimated coefficient of trust is -0.1106
  - a 10% change in intensity of slave exports is associated with a decline of -0.01106 in the trust score.
- Estimated standard error is 0.021
  - authors two-way cluster on *ethnicity* and *district*
    - *ethnicity* as *trust* does not vary for individuals of same *ethnicity*
    - *district* due to possible geographical correlation in the error here errors can be correlated if live in the same district.
  - authors also consider spatial HAC on *ethnicity* location.

Different Clustering Schemes and CV1b Standard Errors

- CV1b is Stata `vce(cluster)` with $c_N = \frac{G}{G-1} \times \frac{N-1}{N-K}$ and $\widetilde{\mathbf{u}}_g = \widehat{\mathbf{u}}_g$.
- Different clustering schemes reveal for $\widehat{\beta} = \widehat{\beta}_{exports}$
  - larger standard errors when cluster (*None* is heteroskedastic-robust)
  - largest when cluster on country (0.292)
    - ★ note: then country FEs did not sop all the error correlation.
-

Cluster	*None*	*Ethnicity*	*Town*	*District*	*Region*	*Countr*
$\widehat{\beta}$	-0.1106	-0.1106	-0.1106	-0.1106	-0.1106	-0.110
$se(\widehat{\beta})$	0.0125	0.0216	0.0150	0.0156	0.0239	0.0292
G	19,733	185	2,766	1,257	171	17
$\overline{N_g}$	1	106.7	7.1	15.7	115.4	1160.8
$\widehat{\rho}_x$	–	1.00	0.881	0.880	0.787	0.615

Difference in the CV1, CV2 and CV3 Estimates

- Here cluster on country and find
  - CV3 > CV2 > CV1b > CV0 > None

Cluster :	None	CV0	CV1b	CV2	CV3
$\widehat{\beta}$	-0.1106	-0.1106	-0.1106	-0.1106	-0.1106
$se(\widehat{\beta})$	0.0125	0.0283	0.0292	0.0330	0.0382

- CV2 can be computed in Stata using vce(cluster, hc2) after xtset country
- CV3 in Stata has problems in this example.
  - Use Stata add-on `summclust` (MacKinnon, Nielsen and Webb (2023b)).
- Use CV2 or CV3: details below.

2.4 Subsequent Inference

- For a single coefficient β, asymptotic theory gives

$$t = \frac{\widehat{\beta}-\beta_0}{\sqrt{\text{se}_{\text{CV}}[\widehat{\beta}]}} \sim N[0,1].$$

- Standard ad hoc adjustment uses the $T(G-1)$ distribution
  - better as $T(G-1)$ distribution has fatter tails than $N[0,1]$
  - ad hoc (Bester, Conley and Hansen (2009, *JE*) derive for a special case)
- But in practice with finite samples and the usual CV1a or CV1b
  - tests based on $T(G-1)$ over-reject
  - confidence intervals based on $T(G-1)$ undercover.

Finite Sample Problems

- Problem 1: G is too small.
- Problem 2: $\mathbf{X}_g'\mathbf{X}_g$ varies between clusters (unbalanced clusters)
  - likely if N_g varies greatly between clusters
  - likely for regressors that take nonzero values in few clusters (or in many clusters)
    - ★ e.g. few clusters are treated.
- Solution 1: Use CV2 or CV3
- Solution 2: Use $t(G^*)$ for data determined $G^* < (G-1)$.
- Solution 3: Combine Solutions 1 and 2.
- MacKinnon, Nielsen and (2023b) Stata add-on `summclust` considers various issues related to cluster balance.

T with Data-Determined Degrees of Freedom

- Use the usual t statistic but use $t(G^*)$ distribution
- Bell and McCaffrey (2002) have cluster generalization of Satterthwaite (1946)
  - requires a pilot matrix
  - they use i.i.d. errors and CV2 or CV3.
- Imbens and Kolesar (2016, *REStat*).
  - use equicorrelated errors (random effects model) within cluster and CV2.
- Pustejovsky and Tipton (2017, *JBES*)
  - extend Imbens and Kolesar to joint hypothesis tests
- Hansen (2024) prefers to use Imbens and Kolesar and CV3 and a modified t statistic
  - this can be very conservative
  - e.g. simulations have 95% confidence intervals with 99% coverage.

Bottom Line for Inference

- At least use CV2 or CV3 with $t(G-1)$.
- Most extreme is to use CV3 with $t(G^*)$
  - or Hansen (2024) modification
  - can be very conservative.
- Not clear which is more conservative
  - CV3 with $t(G-1)$.
  - CV2 with $t(G^*)$.
- Jackknife (CV3) has advantage of being applicable to nonlinear estimators and two-way clustering.

2.5 Cluster Balance, Leverage and Influential Observations

- Unbalanced data not only makes inference more challenging.
- It can lead e.g. to $\widehat{\beta}$ being determined by just a few clusters!
- MacKinnon, Nielsen and Matthew D. Webb (2023a, Sections 7 and 8) present and illustrate
  - cluster leverage measures based on $\mathbf{X}_g(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_g$
  - cluster influence measures based on $\widehat{\beta}_{-g}$ that omits cluster G
- MacKinnon, Nielsen and Matthew D. Webb (2023b, SJ)
  - Stata `summclust` command for cluster leverage and influence
  - includes the measure G^* of the effective number of clusters due to Carter, Schnepel and Steigerwald (2017, *REStat*).
- Young (2019, *QJE*) shows that leverage can lead to great over-rejection using CV1.

2.6 Placebo Test

- Suppose interest lies in a particular x_k.
- Then 10,000 times
  - randomly generate a dataset of $z_k's$ that have the same properties as the $x_k's$
  - run OLS regression with z_k as a regressor
  - perform a 5% test of the statistical significance of z_k
  - this test should reject 5% of the time (500 times out of 10,000)
- Bertrand, Duflo and Mullainathan (2004, QJE)
- See MacKinnon, Nielsen and Matthew D. Webb (2023b, SJ)
  - sections 3.5 and 8.2.

2.7 Wild Cluster Bootstrap with Asymptotic Refinement

- There are several ways to bootstrap
  - different resampling methods
  - different ways to then use for inference
    - in some cases can get an asymptotic refinement.
- A fairly general procedure to get an asymptotic refinement is
  - percentile-t (or "studentized") bootstrap that bootstraps the t statistic
  - with cluster-pairs resampling that resamples with replacement $(\mathbf{y}_g, \mathbf{X}_g)$.
- Cameron, Gelbach and Miller (2008) in simulations find better performance with finite G if instead
  - resample residuals $\widehat{\mathbf{u}}_g$ holding $\mathbf{X}_g$ fixed ("wild" cluster bootstrap)
  - impose H_0 in getting the residuals.

Wild Restricted Cluster Bootstrap

1. Obtain the restricted LS estimator $\widehat{\beta}$ that imposes H_0.
   Compute the residuals $\widehat{\mathbf{u}}_g$, $g = 1, ..., G$.
2. Do B iterations of this step. On the b^{th} iteration:
   1. For each cluster $g = 1, ..., G$:
      Form $\widehat{\mathbf{u}}_g^* = d_g \times \widehat{\mathbf{u}}_g$ where $d_g = -1$ or 1 each with probability 0.5
      Hence form $\widehat{\mathbf{y}}_g^* = \mathbf{X}_g' \widehat{\beta} + \widehat{\mathbf{u}}_g^*$.
      This yields wild cluster bootstrap resample $\{(\widehat{\mathbf{y}}_1^*, \mathbf{X}_1), ..., (\widehat{\mathbf{y}}_G^*, \mathbf{X}_G)\}$.
   2. Calculate the OLS estimate $\widehat{\beta}_{1,b}^*$ and its standard error $s_{\widehat{\beta}_{1,b}^*}$.
      Hence form the Wald test statistic $w_b^* = (\widehat{\beta}_{1,b}^* - \widehat{\beta}_1)/s_{\widehat{\beta}_{1,b}^*}$.
3. Reject H_0 at level α if and only if

$$w < w_{[\alpha/2]}^* \text{ or } w > w_{[1-\alpha/2]}^*,$$

where $w_{[q]}^*$ denotes the q^{th} quantile of $w_1^*, ..., w_B^*$.

Wild Restricted Cluster Bootstrap (continued)

- Implementation is fast and easy for practitioners.
- Roodman, MacKinnon, Nielsen and Webb (2019, *SJ*)
  - `boottest` add-on command to Stata is very fast
  - implements wild and score bootstrap of Wald or score test for many estimators
  - provides confidence intervals by test inversion
  - update includes using CV3
  - has many possible variations.
- MacKinnon (2022, E&S)
  - further computational savings using sums of products and cross-products of observations within each cluster.
- May handle small G and often unbalanced clusters
  - but problems if there are few treated or untreated clusters.

Application

- Apply to OLS regression with clustering on country
- `reg trust3 exports $base_regressors, vce(cluster country)`

trust3	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interv
exports	-.110552	.0292151	-3.78	0.002	-.1724853 -.0486

- `boottest exports`
  - Wild bootstrap-t, null imposed, 999 replications, Wald test, bootstrap clustering by country, Rademacher weights

```
exports

                 t(16) =    -3.7841
              Prob>|t| =     0.0731

95% confidence set for null hypothesis expression: [-.1801, .05535]
```

Wild Restricted Cluster Bootstrap (continued)

- Webb (2014, QED WP 1315) proposed a 6-point distribution for d_g in $\widehat{\mathbf{u}}_g^* = d_g \widehat{\mathbf{u}}_g$
  - better when $G < 10$.
- MacKinnon and Webb (2017, *JAE*)
  - unbalanced cluster sizes worsens poor test size using $\mathrm{V}_{CR}[\widehat{\beta}]$.
  - wild cluster bootstrap does well.
- Djogbenou, MacKinnon, Nielsen (2019, *JE*)
  - prove that the Wild cluster bootstrap provides an asymptotic refinement (using Edgeworth expansions).
- Canay, Santos and Shaikh (2021, *REStat*)
  - provides randomization inference theory for the wild bootstrap when $N_g \rightarrow \infty$ and symmetry holds
  - considers both studentized and unstudentized test statistics.

2.8 Cluster-Specific Fixed Effects Models: Summary

- Now $y_{ig} = \mathbf{x}'_{ig}\boldsymbol{\beta} + \alpha_g + u_{ig} = \mathbf{x}'_{ig}\boldsymbol{\beta} + \sum_{h=1}^{G} \alpha_g dh_{ig} + u_{ig}$.
- 1. **FE's do not in practice absorb all within–cluster correlation**
  - **still need to use cluster-robust VCE.**
- 2. Cluster-robust VCE is still okay with FE's (if $G \rightarrow \infty$)
  - Arellano (1987, JE) for N_g small; Hansen (2007a, JE p.600) for $N_g \rightarrow \infty$
- 3. If N_g is small use `xtreg, fe` not `reg i.id_clu` or `areg`
  - as `reg` or `areg` uses wrong degrees of freedom.
- 4. FGLS with fixed effects needs to bias-adjust for $\widehat{\alpha}_g$ inconsistent.
- 5. Need to do a modified Hausman test for fixed effects.
- 6. Modify with `idcluster` option if bootstrapping.
- 7. Can mean-difference out FEs to save computation time.
- 8. Several ways of dealing with many two-way fixed effects
  - `reg2hdfe`, `felsdvreg`, McCaffrey et al. (SJ, 2012) review.

2.9 Panel Data

- For state-year panel data it is standard to one-way cluster on state
  - this is valid if #states$\rightarrow\infty$ regardless of whether there are few or many time periods
- We may consider two-way cluster on both state and time
  - in many microeconomics applications it is enough to use time dummies and cluster on state
    - ★ but for panel data on firms there may be more reason to two-way
  - **note: do not one-way cluster on state-year pair**
- For individual panel data y_{it} it is standard to one-way cluster on i or some higher level such as state if policy variable is at state level.
- If $T \rightarrow \infty$ can use Driscoll-Kraay (1998) that generalizes time series HAC
  - and allows clusters to be correlated.

Time series robust HAC Standard Errors

- Panel data with $T \rightarrow \infty$ and errors are correlated only up to m periods apart.
- Driscoll and Kraay (1998) generalize Newey-West HAC standard errors for pure time series to allow for errors to be spatially correlated across individual units.
- OLS in the model $y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}$, $i = 1, .., N$, $t = 1, ..., T$
  - $T \rightarrow \infty$ and $\text{Cor}[u_{it}, u_{j,t-k}] = 0$ for $k > m$.
- Then panel HAC variance estimate

$$\widehat{\text{V}}_{panelHAC}[\widehat{\boldsymbol{\beta}}_{OLS}] = (\mathbf{X}'\mathbf{X})^{-1}\text{E}\left[\widehat{\Omega}_0 + \sum_{k=1}^{m}(\widehat{\Omega}_k + \widehat{\Omega}_k')\right](\mathbf{X}'\mathbf{X})^{-1},$$

where $\widehat{\Omega}_k = (1 - \frac{k}{m+1})\sum_t \sum_i \sum_j \widehat{u}_{it}\widehat{u}_{j,t-k}\mathbf{x}_{it}\mathbf{x}_{j,t-k}'$.
- Stata add-on command `xtscc`, due to Hoechle (2007).

2.10 Panel Application

- To come.

2.11 Design-based inference

- Let $Y = f(D, Z, \varepsilon)$ where
  - D is treatment variable
  - Z is other variables (called "attributes" rather than "controls")
  - ε is error.
- Randomness may potentially come from U, Z, ε and from sample S from the population.
- Traditional approaches
  - randomness is due to model errors ε (called "model" approach)
  - randomness is due to selection of sample S from the population
    - ★ problem if sample is the population e.g. states
    - ★ model-based approach presumes a superpopulation of states.

Design-based inference summary

- Design-based approach (newer)
  - randomness is due to assignment of treatment D
  - no role for model error
  - also consider sampling.
- Abadie, Athey, Imbens, Wooldridge (2020, Ecta) consider heteroskedastic case.
- Abadie, Athey, Imbens, Wooldridge (2023, QJE) consider clustered case
  - binary treatment
  - potential outcomes framework with potential outcomes not random
  - heterogeneous effects
  - no attributes.

Design-based inference summary

- AAIW (2023, QJE) clustered case is too complicated to detail here.
- They propose two asymptotically equivalent variance estimators
  - analytical
  - two-stage bootstrap
- Can implement using the Stata add-ons `ccv` and `tcsb` of Clarke and Pailañir (2024).

Design-based inference summary (continued)

- The design-based cluster method can lead too much smaller standard errors than CV1-CV3 when
  - most clusters are sampled
  - treatment varies within cluster
  - treatment effects vary across clusters
  - there are many observations per cluster.
- Big decision
  - is there no role for a model error?
- Also generalizability
  - current work generalize to binary treatment with attribute variables.

3. Beyond one-way clustering

- Richer forms of clustering than one-way
  - Multi-way clustering
  - Dyadic clustering
- Other topics for one-way
  - Few treated clusters
  - Feasible GLS
  - Instrumental Variables
  - m-estimators and GMM

3.1 Two-way Clustering

- What if have two non-nested reasons for clustering?
  - e.g. regress individual wages on job injury rate in industry and on job injury rate on occupation
  - e.g. matched employer - employee data.
- Obtain three different cluster-robust "variance" matrices by
  - cluster-robust in (1) first dimension, (2) second dimension, and (3) intersection of the first and second dimensions
  - add the first two variance matrices and, to account for double-counting, subtract the third.

$$\widehat{\mathrm{V}}_{\text{two-way}}[\widehat{\boldsymbol{\beta}}] = \widehat{\mathrm{V}}_{G}[\widehat{\boldsymbol{\beta}}] + \widehat{\mathrm{V}}_{H}[\widehat{\boldsymbol{\beta}}] - \widehat{\mathrm{V}}_{G \cap H}[\widehat{\boldsymbol{\beta}}]$$

- A simpler more conservative estimate drops the third term
  - this guarantees that $\widehat{\mathrm{V}}_{\text{two-way}}[\widehat{\boldsymbol{\beta}}]$ is positive definite.

Two-way Application

- Two-way on country and ethnicity (murdock_name)
- CV1-based standard error is 0.0297
  - close to 0.0292 for CV1 one-way on country
  - larger than 0.0216 for CV1 one-way on ethnicity (murdock_name)

```
.  reg trust3 exports $base_regressors, vce(cluster country murdock_name)
note: multiway-cluster variance-covariance matrix is not positive semidefinite.

Linear regression                             Number of obs = 19,733
Clusters per comb.:                           Cluster comb. =      3
  min =  17                                   F(49, 16)     =      .
  avg = 143                                   Prob > F      =      .
  max = 226                                   R-squared     = 0.1960
                                              Adj R-squared = 0.1928
                                              Root MSE      = 0.9891

                          (Std. err. adjusted for multiway clustering)
```

trust3	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
exports	-.110552	.0296875	-3.72	0.002	-.1734868	-.0476173

Two-way Clustering

- Independently proposed by
  - Cameron, Gelbach, and Miller (2006; 2011, *JBES*) in econometrics
  - Miglioretti and Heagerty (2006, *AJE*) in biostatistics
  - Thompson (2006; 2011, *JFE*) in finance
  - Extends to multi-way clustering.
- Theory provided by Davezies, D'Haultfoeuille and Guyonvarch (2021, *AS*), Menzel (2021, *Ecta*), and MacKinnon, Nielsen and Matthew Webb (2021, *JBES*).
- MacKinnon, Nielsen and Matthew Webb (2024, WP)
  - good results using the jackknife (CV3) rather than CV1
  - have Stata add-on `twowayjack` at http://qed.econ.queensu.ca/pub/faculty/mackinnon/twowayjack/
- Chiang, Hansen and Sasaki (2024, *REStat*) panel add extra terms.

3.2 Dyadic Clustering

- A dyad is a pair. An example is country pairs.
- The errors for two pairs are correlated with each other if they have one person in common.
  - Call the pairs (g, h) and (g', h')
  - Two-way picks up error correlation for cases with $g = g'$ and $h = h'$
  - Dyadic-robust additionally picks up $g = h'$ and $h = g'$.
- Fafchamps and Gubert (2007, *JDE*)
  - provide variance matrix
  - apply to a sparse network where it makes little difference.
- Cameron and Miller (2014, WP)
  - apply to international trade data where the network is dense and find it makes a big difference
  - we are currently finishing an updated paper.

Dyadic Clustering (continued)

- Aronow and Assenova (2015, Political Analysis)
  - prove variance estimate but not asymptotic normal distribution.
- Tabord-Meehan (2018, *JBES*)
  - use a central limit theorem for dependency graphs (S. Jannson (1988)).
- Davezies, D'Haultfoeuille and Guyonvarch (2021, *AS*)
  - provides empirical process theory that assumes exchangeability and propose a pigeonhole bootstrap.
- Still not used much
  - especially for gravity model of trade.

Networks

- Dyadic data are network data.
- Graham (2022a, 2022b) provides theory for dyadic data under exchangeability from a network perspective.
- Many (but not all) networks are sparse.
- One issue is Data-determined clusters
  - Cao et al (2022, WP) and Leung (2023, *Ecta*).

3.3 Few treated clusters

- Few treated clusters
  - often arises especially in differences-in-differences settings
  - basic cluster-robust inference can work poorly.
- MacKinnon and Webb (2018, *PM*)
  - extreme problem if only one treated cluster as then the OLS residuals in that cluster sum to zero
  - this leads to too small a variance estimate.
- Solutions often require strong assumptions such as
  - exchangeability within cluster
  - homogeneity across cluster
  - symmetry
  - identification can be obtained using only within-cluster estimates.

Few treated clusters (continued)

- Wild cluster bootstrap with few (treated) clusters
  - MacKinnon and Webb (2018, *EJ*)
- T distribution for t statistics from cluster-level estimates
  - Ibragimov and Müller (2010, *JBES*)
    - only within-group variation is relevant, separately estimate $\widehat{\boldsymbol{\beta}}_g s$ and average, G small and $N_g \longrightarrow \infty$.
    - rules out $y_{ig} = \mathbf{x}'_{ig}\boldsymbol{\beta} + \mathbf{z}'_g\gamma + u_{ig}$.
  - Bester, Conley and Hansen (2011, JE) closely related.
  - Ibragimov and Müller (2016, *REStat*)
    - extend to allow treated and untreated groups.
- Difference in difference settings
  - Conley and Taber (2011) assume exchangeability and have fixed T, fixed treated clusters, number of control clusters $\longrightarrow \infty$
  - Ferman and Pinto (2019) extend this to (known) heteroskedastic errors.

Randomization inference

- A permutation test (Fisher) provides a test of exact size.
- For settings where data are exchangeable under the null hypothesis
  - e.g. two-sample difference in means test with two samples from the same distribution
- The procedure:
  - 1. Compute the test statistic using the original sample.
  - 2. Recompute this test statistic for every permutation of the data.
  - 3. p-value $=$ fraction of times permuted test statistic $\geq$ original sample test statistic.

Randomization inference (continued)

- Extends to a regressor of interest that is uncorrelated with other regressors
  - e.g. if the regressor is a randomly assigned treatment.
- Young (2019, *QJE*) does this and compares to conventional methods and bootstrap.
- MacKinnon and Webb (2020, *JE*) consider when treatment is not randomly assigned.
- MacKinnon and Webb (2019, book chapter) adjust when there are few possible randomizations.

Randomization inference (continued)

- Canay, Romano and Shaikh (2017, *Ecta*)
  - extend to symmetric limiting distribution of a function of the data under H_0
  - covers DinD with few clusters and many observations per cluster.
- Cai, Kim and Shaikh (2021)
  - Stata and R packages to implement in linear models with few clusters.
- Hagemann (2019, *JE*)
  - assigns placebo treatments to untreated clusters to get nearly exact sharp test of no effect of a binary treatment.
- Hagemann (2020)
  - a rearrangement test for a single treated cluster with a finite number of heterogeneous clusters.
- Hagemann (2021)
  - adjusts permutation inference to get non-sharp test on binary treatment with finitely many heterogeneous clusters.

3.4 Feasible GLS

- Potential efficiency gains for feasible GLS compared to OLS.
- And for one-way clustering there is a cluster-robust VCE (as $G \rightarrow \infty$)

$$\widehat{V}_{CR}[\widehat{\beta}_{FGLS}] = \left(\mathbf{X}'\widehat{\Omega}^{-1}\mathbf{X}\right)^{-1}\left(\sum_{g=1}^{G}\mathbf{X}_g'\widehat{\Omega}_g^{-1}\widehat{\mathbf{u}}_g\widehat{\mathbf{u}}_g'\widehat{\Omega}_g^{-1}\mathbf{X}_g\right)\left(\mathbf{X}'\widehat{\Omega}^{-1}\mathbf{X}\right)^{-1}$$

- Stata offers many FGLS estimators with CR standard errors.
- Yet this is not done much in economics.
- Brewer and Crossley (2018, *JEM*)
  - panel data with cluster-specific fixed effects and AR(2) error and bias-adjust
  - find much better test size performance using BDM data.

3.5 Instrumental variables

- Cluster-robust variance generalizes immediately
  - main focus is on cluster-robust inference with weak instruments.
- Chernozhukov and Hansen (2008, *EL*)
  - Cluster-robust version of Anderson-Rubin test is immediate
  - AR has no power loss in just-identified single endogenous regressor case.
- Weak instruments diagnostics
  - First-stage F-statistic should be cluster-robust.
- Current feeling is that if concerned about weak instruments then do Anderson-Rubin directly and skip the first stage F-statistic as screen.
- Young (2021) considers leverage and clustering in IV applications.

3.6 Nonlinear m-estimators and GMM

- Cluster-robust methods extend to nonlinear estimators
  - e.g. logit and nonlinear GMM.
  - e.g. generalized estimating equations (Liang and Zeger 1986)
  - replace $\mathbf{X}_g' \widehat{\mathbf{u}}_g$ with the score for the cluster.
- Kline and Santos (2012, *EM*)
  - wild score bootstrap
  - this extends to nonlinear models such as logit and probit
  - Stata addon `boottest` includes this.

GMM

- Cluster-robust extends to GMM.
- Hansen and Lee (2019, *JE*)
  - provide very general asymptotic theory for clustered samples
- Hansen and Lee (2021, *Ecta*)
  - inference for Iterated GMM under misspecification
  - consider heteroskedastic errors (journal dropped clustering).
- Hansen and Lee (2020, WP)
  - also has clustered errors.
- Hwang (2019, *JE*)
  - two-step GMM fixed-G asymptotics with recentering of the CRVE used at the second step.

3.7 Quantile regression

- Parente and Silva (2016, *JEM*)
  - quantile regression with clustered data.
- Yoon and Galvao (2020, *QE*)
  - cluster-robust inference for panel quantile regression models with individual fixed effects and serial correlation.
- Hagemann (2017, *JASA*)
  - Cluster-robust bootstrap inference.

3.8 Machine learning prediction and clustering

- Cameron and Trivedi (2022, chapter 28) provide an accessible introduction to machine learning.
- Leading ML methods used by econometricians in order of current usage
  - lasso (and to a lesser extent ridge)
  - random forests (collections of regression trees)
  - neural networks (including deep nets).
- For lasso linear regression with independent data choose $\boldsymbol{\beta}$ to minimize
  - $Q_\lambda(\beta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 + \lambda\sum_{j=1}^{p}\kappa_j|\beta_j|$
    - where in the simplest case the regressors are standardized and $\kappa_j = 1$.
- With clustered data we could use the same objective function.
- Stata instead uses a weighted average
  - $Q_\lambda(\beta) = \frac{1}{G}\sum_{g=1}^{G}\left\{\frac{1}{N_g}\sum_{i=1}^{N_g}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2\right\} + \lambda\sum_{j=1}^{p}\kappa_j|\beta_j|$
  - same as simple unweighted in the case of balanced clusters.

Causal machine learning

- A key general paper for double/debiased ML is Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins, J. (2018, EJ).
- A leading example is the partial linear model with scalar regressor of interest d and many potential controls $\mathbf{x}_c$
  - $y = \alpha d_i + g(\mathbf{x}_c) + u$ where $g(\cdot)$ is unspecified.
- Then
  - a machine learner is used to approximate $g(\mathbf{x}_c)$
  - estimation of α is based on an "orthogonalized" moment condition that enables standard inference on α despite the first-stage use of a machine learner
  - performance is improved by using cross fitting
    - ★ a bigger part of the data is used in the ML stage and the smaller remainder is used in second stage estimation of α.

Causal machine learning and clustered data

- With clustering the cross fitting needs to be adapted.
- For one-way clustering (such as panel data)
  - Belloni, Chernozhukov, Hansen, and Damien Kozbur (2016, *JBES*)
  - cross fitting keeps clusters intact.
- For two-way clustering (such as panel data)
  - Chiang, Kato, Ma and Sasaki (2022, *JBES*)
  - cross fitting in simplest case splits sample in each direction in half giving $2^2 = 4$ distinct groups.
- For dyadic clustering (such as panel data)
  - Chiang, Kato, Ma and Sasaki (2022, WP)
  - a more complex cross fitting is proposed.

4. Inference with Spatially Dampening Correlation

- The clustering approach assumes observations can be grouped into blocks, with errors uncorrelated across blocks.
  - e.g. region-level clustering assumes independence across regions.
- A richer approach may allow correlation with other regions, with correlation dampening in distance
  - such correlation is called spatial correlation.
- Theory for spatial correlation is an extension of that for time series.
- For time series correlation the unit is in time.
- By contrast are many ways to model spatial correlation
  - e.g. For geospatial data one could use inverse distance or inverse distance squared.

4.1 How to Measure Spatial Distance

- Most often geospatial distance
  - using latitude and longitude tricky due to curvature of the earth.
- But could be e.g. economic distance.

4.2 Spatial-HAC Standard Errors

- Suppose that any error correlation disappears for observations more than distance δ apart.
- Then we use (d_{ij} is the distance between i and j)

$$\widehat{\mathrm{V}}_{spatial}[\widehat{\boldsymbol{\beta}}] = (\mathbf{X}'\mathbf{X})^{-1}\Big(\sum_{d_{ij}<\delta}\sum \kappa_{ij}\mathbf{x}_i\mathbf{x}_j'\widehat{u}_i\widehat{u}_j\Big)(\mathbf{X}'\mathbf{X})^{-1}.$$

- The obvious variance matrix estimator sets $\kappa_{ij} = \mathbf{1}[d_{ij} \leq \delta]$
  - this can lead to a variance matrix that is not positive semi-definite.
- So Conley (1999) proposed using $\kappa_{ij} = k(d_{ij}, \delta) \times \mathbf{1}[d_{ij} \leq \delta]$ where $k(\cdot)$ are kernel weights such as Bartlett weights $k(d_{ij}, \delta) = (1 - d_{ij}/\delta)$.
  - similar to Newey-West HAC for time series.
- Consistency requires that spatial correlation disappears in a large enough fraction of the $N \times (N-1)$ error correlations
  - this clearly depends on the application.

5.2 Measuring Distance

- Distance can be geographic distance, economic distance, inclusion (or not) in the same group (region or peer group or ....), whether or not a border is shared (contiguity), ...
- Geographic location is recorded using latitude and longitude.
  - Lines of longitude are vertical lines that pass through the poles
  - Lines of latitude are horizontal rings around the globe.
- If we move one degree of latitude, such as from 31^0 north to 32^0 north, then we move approximately 69 miles.
- If instead we move one degree of longitude east or west, such as from 31^0 east to 32^0 east, then the distance ranges greatly from 69 miles at the equator to 0 miles at a pole.
- To correctly calculate Euclidean distance between two points, data need to be converted from Cartesian coordinates (latitude and longitude) for a globe to coordinates on a plane, called planar (or rectangular) coordinates.
- There are N^2 distances to compute, so use specialized software.

4.3 Subsequent Inference

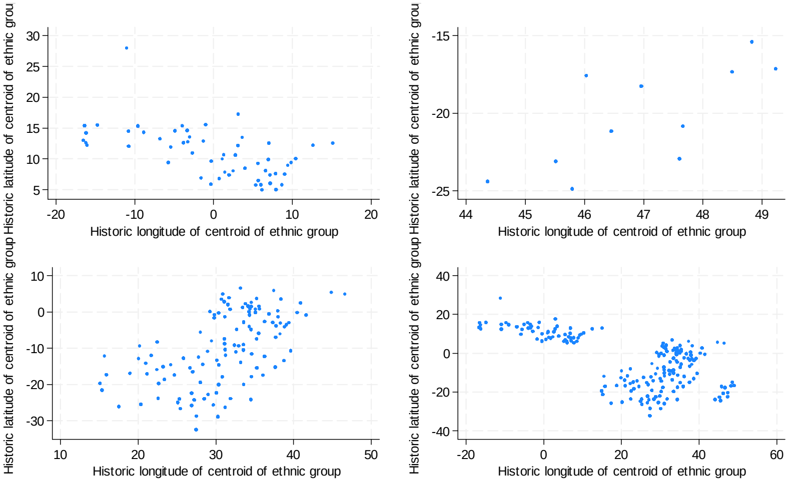
- Wald test is asymptotically normal.
- Asymptotic theory may not kick in if spatial correlation is slow to disappear as distance increases.
- See finite-sample adjustments below.

4.3 Application

- Use Nunn and Watchekon data example.
- Use Stata add-on command acreg due to Colella, Lalive, Sakalli, and Thoenig (2020)
- `trust3 exports, bartlett spatial distcutoff(500) /// latitude(centroid_lat) longitude(centroid_long)`
- This yields same standard errors as just cluster on *ethnicity*.
- Reason is that people of the same ethnicity have the same centroid_lat and centroid_long
- And these range greatly: latitude -32.7^0 to 27.8^0 and longitude -16.4^0 to 49.2^0.

Application (continued)

- Plot of latitude on longitude shows three clusters.



4.4 Finite-Sample Adjustments

- Key paper is Conley and Kelly (2025, JIE).
- Consider studies with spatial correlation that does not dampen sufficiently for Conley spatial HAC to work well.
- Then should prewhiten the data
  - use a spatial basis rather than e.g. quadratic in latitude and longitude
  - use k-medoids (a generalization of k-means) to select just a few clusters
  - use Bester, Conley and Taber (2011) to do inference with a few clusters and many observations per cluster.
  - also has placebo tests.
- Important paper that shows many studies greatly overstate precision of estimates.

Other Work

- Müller and Watson (2022, 2023) propose an alternative approach to spatial correlation robust inference.
  - using spatial-correlation principal components.
- Müller and Watson (2024) consider spatial unit roots.
- Xu and Wooldridge (2022) consider a design-based approach to spatial correlation.

5.1 Spatial Autoregressive Models

- The preceding sections consider inference for regression of y on $\mathbf{x}$ under relatively weak assumptions.
- Much of the spatial literature instead considers more parametric models that specify the relationship between observations and/or between errors, called spatial autocorrelation.
  - qualitatively similar to autoregressive models for time series.
- Standard econometrics references are Anselin (1988) and LeSage and Pace (2009).
- A standard statistics text is Cressie and Wilke (2011).
- Stata `spregress` and related commands estimate these models
  - Cameron and Trivedi (2022, chapter 26) provide a detailed summary.

Spatial autoregressive Models

- The simplest Cliff-Ord spatial model, due to Cliff and Ord (1973),
  - specifies $y_i = \lambda \sum_{j=1}^{N} w_{ij} y_j + u_i$
  - where the w_{ij} are specified by the researcher (with $w_{ii} = 0$)
  - and λ is a parameter to be estimated.
- More generally we introduce other regressors.
- SAR(1): spatial autoregressive in the mean of order one
  - specifies $y_i = \mathbf{x}_i' \boldsymbol{\beta} + \lambda \times \sum_{j=1}^{N} w_{ij} y_j + u_i$ (with $w_{ii} = 0$)
  - or $\mathbf{y} = \lambda \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{u}$.
- Alternatively spatial dependence may be in the error
  - specify $u_i = \rho \sum_{j=1}^{N} w_{ij} u_j + \varepsilon_i$, with $w_{ii} = 0$, so $\mathbf{u} = \rho \mathbf{W} \mathbf{u} + \boldsymbol{\varepsilon}$.
- Combining an (SAR1,1) model
  - $\mathbf{y} = \lambda \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{u}$ and $\mathbf{u} = \rho \mathbf{W}_2 \mathbf{u} + \boldsymbol{\varepsilon}$.

6. Conclusion

- Where clustering is present it is important to control for it.
- Most work is for OLS and one-way clustering.
- Often clusters are very unbalanced / heterogeneous and/or "few" clusters
  - then use CV3 or CV2.
- And many spatial applications are ones for which spatial dependence does not disappear fast enough to use Conley spatial HAC.

Some Key References

- Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge (2023), "When should you adjust standard errors for clustering?" Quarterly Journal of Economics, 138(1), 1-35.
- Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge (2020), "Sampling-based versus design-based uncertainty in regression analysis," Econometrica, 88(1), 265–296.
- Bell, R.M., and McCaffrey, D.F. (2002), "Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples", Survey Methodology, 28(2), pp. 169-181.
- Bester, Conley and Taber (2011), "Inference with dependent data using cluster covariance estimators," Journal of Econometrics, 165, 137-151.
- Cameron, A. Colin, Jonah G. Gelbach, and Douglas L. Miller (2008), "Bootstrap-Based Improvements for Inference with Clustered Errors," Review of Economics and Statistics 90, 414-427.

- Cameron, A. Colin, Jonah G. Gelbach, and Douglas L. Miller (2011), "Robust Inference with Multi-way Clustering," Journal of Business and Economic Statistics, 29, 238-249.
- Cameron, A. Colin, and Douglas L. Miller (2014), "Robust Inference for Dyadic Data," https://cameron.econ.ucdavis.edu/research/dyadic_cameron_miller_december2014_
- Cameron, A. Colin, and Douglas L. Miller (2015), "A Practitioner's Guide to Cluster-Robust Inference, Journal of Human Resources, 50(2), 317-373.
- Clarke, Damian and Daniel Pailañir (2024), "TSCB-CCV", https://github.com/Daniel-Pailanir/TSCB-CCV.
- Colella, Fabrizio; Lalive, Rafael; Sakalli, Seyhun Orcan; Thoenig, Mathias. (2019), "Inference with Arbitrary Clustering, IZA Discussion Paper n. 12584.
- Colella, Fabrizio; Lalive, Rafael; Sakalli, Seyhun Orcan; Thoenig, Mathias. (2020), "Acreg: arbitrary correlation regression."
- Conley, Timothy G. (1999), "GMM Estimation with Cross Sectional Dependence", Journal of Econometrics, 92, 1-45.

- Conley, Timothy G. and Morgan Kelly (2025), “The standard errors of persistence”, Journal of International Economics.
- Conley, Timothy G., Morgan Kelly and Damein Kozbur (2024) “Improved Spatial Dependence Robust Inference via’Pre-Whitening,” Working paper ETH Zurich.
- Hansen, Bruce E, (2024), “Jackknife Standard Errors for Clustered Regression,” working paper.
- Ibragimov, Rustam, and Ulrich K. Müller (2010), “T-Statistic Based Correlation and Heterogeneity Robust Inference,” Journal of Business and Economics Statistics, 28, 453-468.
- Ibragimov, Rustam and Ulrich K. Müller (2016), “Inference with few heterogeneous clusters”, Review of Economics and Statistics, 98(1), 83-96.
- Imbens, Guido W., and Michal Kolesar (2016), “Robust Standard Errors in Small Samples: Some Practical Advice,” Review of Economics and Statistics, 98, 701-712.

- MacKinnon, James G., Morten Ø. Nielsen and Matthew D. Webb (2023a), "Cluster-robust inference: A guide to empirical practice", Journal of Econometrics, 232(2), 272-299.
- MacKinnon, James G., Morten Ø. Nielsen and Matthew D. Webb Webb (2023b), "Leverage, influence, and the jackknife in clustered regression models: Reliable inference using summclust" The Stata Journal, 23(4), 942-982.
- MacKinnon, James G., Morten Ø. Nielsen and Matthew D. Webb (2023c), "Fast and reliable jackknife and bootstrap methods for cluster-robust inference," Journal of Applied Econometrics, 38(5), 671-694.
- MacKinnon, James G., Morten Ø. Nielsen and Matthew D. Webb Webb (2024), "Jackknife Inference with Two-Way Clustering." https://arxiv.org/pdf/2406.08880.
- Müller, Ulrich K. and Mark W. Watson (2022), "Spatial Correlation Robust Inference," Econometrica, 90, 2901–2935.
- Müller, Ulrich K., and Mark W. Watson (2023), "Spatial Correlation Robust Inference in Linear Regression and Panel Models," Journal of Business and Economics Statistics, 41(4), 1050-1064.

- Nunn, Nathan and Leonard Wantchekon (2011), "The Slave Trade and the Origins of Mistrust in Africa," American Economic Review, 101(7), 3221-3252.
- Roodman, David, Morten Ø. Nielsen, James G. MacKinnon, and Matthew D. Webb (2019), "Fast and wild: Bootstrap inference in Stata using boottest," Stata Journal, 19(1), 4-60.
- Jeffrey M Wooldridge, Jeffrey M. (2023), "What is a standard error?(And how should we compute it?)", Journal of Econometrics, 237(2).
- Xu, Ruonan and Jeffrey M Wooldridge (2022), "A Design-based Approach to Spatial Correlation".