Jackknife inference for multiway clustering and CSDID in Stata: twowayjack and csdidjack

James G. MacKinnon [1] Morten Ørregaard Nielsen [2]
Matthew D. Webb [3] Sunny Karim [3] Yunhan Liu [3]

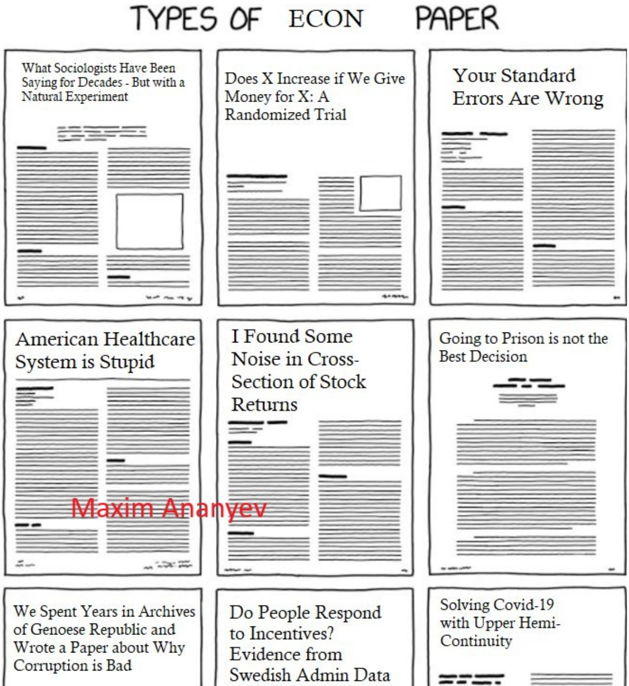
[1]Queen's University and Aarhus Centre for Econometrics

[2]Aarhus Centre for Econometrics

[3]Carleton University

October 3rd, 2025

2025 Canadian Stata Conference


TYPES OF ECON PAPER
What Sociologists Have Been Saying for Decades - But with a Natural Experiment
Does X Increase if We Give Money for X: A Randomized Trial
Your Standard Errors Are Wrong
American Healthcare System is Stupid
I Found Some Noise in Cross-Section of Stock Returns
Going to Prison is not the Best Decision
We Spent Years in Archives of Genoese Republic and Wrote a Paper about Why Corruption is Bad
Do People Respond to Incentives? Evidence from Swedish Admin Data
Solving Covid-19 with Upper Hemi-Continuity

Overview

This talk will cover:

- Overview of two new `Stata` packages for jackknife clustering
- `csdidjack`
  - Available here Liu (2025)
  - Based on the paper MacKinnon, Nielsen, Webb, and Karim (2025)
  - Jackknife clustering for Callaway and Sant'Anna (2021)
- `twowayjack`
  - Available here Webb (2025)
  - based on the paper MacKinnon, Nielsen, and Webb (2024)
  - two-way cluster jackknife

Staggered Adoption - 6 $\times$ 3 Rollout Matrix

- Setup: 6 regions, 3 time periods
- Each region is either treated or not
- Treatment timing is staggered:

$$\begin{array}{ccccccc} & A & B & C & D & E & F \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 & 0 \\ 3 & 1 & 1 & 0 & 1 & 1 & 0 \end{array} \tag{1}$$

Estimands from 6-Region Rollout

- Follow the setup above
- 6 regions: A, B, C, D, E, F
- Treatment cohorts:
  - Group I: regions A, D (treated at $t = 2$)
  - Group II: regions B, E (treated at $t = 3$)
  - Group III: regions C, F (never treated)
- Define overall ATT:

$$\text{ATT} = \frac{1}{3}(\text{ATT}(\text{I}, 2) + \text{ATT}(\text{I}, 3) + \text{ATT}(\text{II}, 3)) \tag{2}$$

ATT Formulas and CSDID Estimator

- $\text{ATT}(g, t)$ defined via difference-in-differences:

$$\begin{aligned}
\text{ATT}(\text{I}, 2) &= (\mathrm{E}[Y_{\text{I},2}] - \mathrm{E}[Y_{\text{I},1}]) - (\mathrm{E}[Y_{\text{III},2}] - \mathrm{E}[Y_{\text{III},1}]), \\
\text{ATT}(\text{I}, 3) &= (\mathrm{E}[Y_{\text{I},3}] - \mathrm{E}[Y_{\text{I},1}]) - (\mathrm{E}[Y_{\text{III},3}] - \mathrm{E}[Y_{\text{III},1}]), \\
\text{ATT}(\text{II}, 3) &= (\mathrm{E}[Y_{\text{II},3}] - \mathrm{E}[Y_{\text{II},2}]) - (\mathrm{E}[Y_{\text{III},3}] - \mathrm{E}[Y_{\text{III},2}]).
\end{aligned}$$

- CSDID with 'simple' weights estimates:

$$\widehat{\text{ATT}} = \frac{\widehat{\text{ATT}}(\text{I}, 2) + \widehat{\text{ATT}}(\text{I}, 3) + \widehat{\text{ATT}}(\text{II}, 3)}{3}. \tag{3}$$

Clustering in CSDID

- Default SEs in Callaway and Sant'Anna (2021) use influence functions
- Expressions are long (see Theorem 2), but derivation is not our focus
- Key result:

$$\sqrt{n}(\widehat{\mathrm{ATT}}^{dr,nev}_{t\geq(g-\delta)} - \mathrm{ATT}_{t\geq(g-\delta)}) \xrightarrow{d} \mathrm{N}(0, \Sigma), \tag{4}$$

- Where $\Sigma = \mathrm{E}[\Psi(\boldsymbol{W}_i)\Psi(\boldsymbol{W}_i)']$ is from doubly robust estimator
- Normal approximation is used to compute P-values and CIs
- Callaway and Sant'Anna (2021, Remark 13): inference requires many clusters

Problems with Clustering in CSDID

There are three main issues with clustering, by region, in CSDID on account of estimating the $\text{ATT}(g, t)$ terms first

- Observations from the same region can appear in multiple separate $\text{ATT}(g, t)$ estimates, such as region I in $\text{ATT}(\text{I}, 2)$ and $\text{ATT}(\text{I}, 3)$
- Untreated observations from the same region and period can appear in multiple separate $\text{ATT}(g, t)$ estimates, such as region III in $\text{ATT}(\text{I}, 3)$ and $\text{ATT}(\text{II}, 3)$
- When not yet treated control units are specified, treated observations from the same region and period can be used twice as both a pre-period for a treated region, and a post-period for an untreated region

Notation for the Cluster Jackknife

- $\widehat{\text{ATT}}$: full-sample CSDID estimate
- $\widehat{\text{ATT}}^{(h)}$: leave-one-cluster-out estimate
- H: total number of clusters
- CV_3 generalizes HC_3 from MacKinnon and White (1985)
- Attractive feature:
  - No need for $\widehat{\mathrm{Var}}[\widehat{\text{ATT}}(g, t)]$
- Standard error of $\widehat{\text{ATT}}$ remains well-defined

The jackknife has received a lot of attention for clustering recently, see MacKinnon et al. (2023a,b); Hansen (2025a,b); MacKinnon et al. (2025); Hounyo and Lin (2025) among others.

Cluster Jackknife for CSDID

- Alternative to Callaway and Sant'Anna (2021) standard errors
- We use the cluster jackknife for inference on $\widehat{\mathrm{ATT}}$
- Jackknife omits one cluster h at a time:

$$\widehat{\mathrm{ATT}}^{(h)} \text{ omits cluster } h$$

- Resulting SE is:

$$\mathrm{CV}_3(\mathrm{ATT}) = \frac{H-1}{H} \sum_{h=1}^{H} \left(\widehat{\mathrm{ATT}}^{(h)} - \widehat{\mathrm{ATT}}\right)^2 \tag{5}$$

Monte Carlo Design

- Simulations based on CPS MORG data, 1979–1999
- Outcome: log of weekly earnings for women
- Sample: 547,818 observations; earnings $\geq$ \$20
- Placebo laws design inspired by Bertrand, Duflo, and Mullainathan (2004)
  - State-level staggered adoption in 2 cohorts
  - Early cohort (J states): treated in year 4
  - Late cohort (L states): treated in year 6
  - $J = L$; total treated states $= J + L$
- Treatment is absorptive and state-wide
- Each replication samples H states and 8 years
- Recent simulations in Mizushima and Powell (2025)
  - They use state-level unemployment rates
  - We use individual-level earnings
- Inference procedures: `csdid`, `wboot`, `csdidjack`

Monte Carlo Design

Table: Values of H and $J = L$ studied in our experiments

	$H = 8$	$H = 16$	$H = 24$	$H = 32$
$J = L = 1$	✓	✓	✓	✓
$J = L = 2$	✓	✓	✓	✓
$J = L = 3$	✓	✓	✓	✓
$J = L = 4$		✓	✓	✓
$J = L = 6$			✓	✓
$J = L = 8$			✓	✓
$J = L = 10$				✓
$J = L = 12$				✓

Rejection Frequencies: RIF (csdid)

	$H = 8$	$H = 16$	$H = 24$	$H = 32$
$J = L = 1$	0.3046	0.3333	0.3558	0.3800
$J = L = 2$	0.1779	0.1867	0.1750	0.1979
$J = L = 3$	0.2146	0.1245	0.1371	0.1388
$J = L = 4$		0.1132	0.1204	0.1187
$J = L = 6$			0.0829	0.0954
$J = L = 8$			0.0950	0.0850
$J = L = 10$				0.0779
$J = L = 12$				0.0925

Notes: Nominal level is 5%.

Rejection Frequencies: Multiplier Bootstrap (`csdid`)

	$H = 8$	$H = 16$	$H = 24$	$H = 32$
$J = L = 1$	0.3079	0.3458	0.3725	0.3950
$J = L = 2$	0.1808	0.1887	0.1762	0.2004
$J = L = 3$	0.2125	0.1233	0.1408	0.1388
$J = L = 4$		0.1161	0.1200	0.1204
$J = L = 6$			0.0846	0.0967
$J = L = 8$			0.0929	0.0842
$J = L = 10$				0.0787
$J = L = 12$				0.0938

Notes: Nominal level is 5%.

Rejection Frequencies: Jackknife (`csdidjack`)

	$H = 8$	$H = 16$	$H = 24$	$H = 32$
$J = L = 1$	0.0938	0.1279	0.1450	0.1642
$J = L = 2$	0.0379	0.0763	0.0796	0.0929
$J = L = 3$	0.0533	0.0488	0.0600	0.0688
$J = L = 4$		0.0505	0.0621	0.0571
$J = L = 6$			0.0450	0.0571
$J = L = 8$			0.0425	0.0483
$J = L = 10$				0.0450
$J = L = 12$				0.0500

Notes: Nominal level is 5%.

Minimum Wage and Teen Employment

- Adapted from Callaway and Sant'Anna (2021)
- County-level QWI data (2001–2007), 29 states
- Outcome: log of teen employment (`lemp`)
- Covariates: demographics + economics
  - Region dummies, % white, % HS diploma, poverty rate
  - County pop. and median income (quadratics)
- Estimation via `csdid`, calendar aggregation

Table: Effect of minimum wage on teen employment

Method	ATT	Std. error	P value	CI lower	CI upper
`csdid`	-0.0309	0.0076	0.000	-0.0457	-0.0160
`csdidjack`	-0.0309	0.0166	0.073	-0.0649	0.0031

Software Packages

- Open-source implementations of CV_3 jackknife inference
- Available for both `Stata` and `R` users

`Stata`: `csdidjack` **(post-estimation for `csdid`)**

- Supports `agg(simple)`, `agg(group)`, and `agg(calendar)`
- Install via:
```
net install csdidjack,
from("https://raw.githubusercontent.com/liu-yunhan/csdidjac
replace
```
- After estimating a model with `csdid` (Rios-Avila, Callaway, and Sant'Anna, 2021), run:
```
csdidjack
```
- Returns CV_3 SEs, t-stats, P values, and CIs
- Docs: `help csdidjack`
- Source code: `https://github.com/liu-yunhan/csdidjack`

Two-Way Cluster Jackknife

- Multiway clustering is very popular, introduced by Cameron et al. (2011) and Thompson (2011)
- Shown to be over-sized with few clusters in either dimension in MacKinnon et al. (2021)
- A (sort of) two-way version of the wild cluster bootstrap is proposed in that paper
- MacKinnon et al. (2024) proposes a two-way jackknife
- The paper also proposes a max(se) alternative to eigen-value corrections
- The paper describes `twowayjack` which will be the focus here
- That package basically makes three calls to `summclust` described in MacKinnon et al. (2023c)

Two-Way Cluster-Robust Variance

- OLS estimator: $\hat{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$
- Residuals: $\hat{\boldsymbol{u}}$, with cluster slices $\hat{\boldsymbol{u}}_g$, $\hat{\boldsymbol{u}}_h$, $\hat{\boldsymbol{u}}_{gh}$
- From sandwich formula:

$$\mathrm{Var}(\hat{\beta}) = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \Sigma (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \boldsymbol{V}_G + \boldsymbol{V}_H - \boldsymbol{V}_I \tag{6}$$

$$\boldsymbol{V}_G = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \left(\sum_{g=1}^{G} \Sigma_g \right) (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \tag{7}$$

$$\boldsymbol{V}_H = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \left(\sum_{h=1}^{H} \Sigma_h \right) (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \tag{8}$$

$$\boldsymbol{V}_I = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \left(\sum_{g=1}^{G} \sum_{h=1}^{H} \Sigma_{gh} \right) (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \tag{9}$$

Two-Way Cluster-Jackknife CRVEs

- Standard CRVEs (e.g., $\hat{\boldsymbol{V}}_G$) follow CV_1 form
- Jackknife estimators improve performance in small samples
- For $J \in \{G, H, I\}$, define leave-one-out estimator:

$$\hat{\beta}^{(j)} = (\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}_j^\top \boldsymbol{X}_j)^{-1}(\boldsymbol{X}^\top \boldsymbol{y} - \boldsymbol{X}_j^\top \boldsymbol{y}_j) \tag{10}$$

- Cluster-jackknife variance for dimension J:

$$\hat{\boldsymbol{V}}_J^{\mathrm{JK}} = \frac{J-1}{J} \sum_{j=1}^{J} (\hat{\beta}^{(j)} - \hat{\beta})(\hat{\beta}^{(j)} - \hat{\beta})^\top \tag{11}$$

- Three-term jackknife CRVE:

$$\hat{\boldsymbol{V}}_3^{(3)} = \hat{\boldsymbol{V}}_G^{\mathrm{JK}} + \hat{\boldsymbol{V}}_H^{\mathrm{JK}} - \hat{\boldsymbol{V}}_I^{\mathrm{JK}} \tag{12}$$

- Related CRVEs: $\text{CV}_3^{(2)}$, $\text{CV}_3^{(3)}$, $\text{CV}_3^{(\text{max})}$

Modified Wald Statistic: $W_{\min}$

- Problem: $\hat{\boldsymbol{V}}_1^{(3)}$ may not be positive definite
- We propose:
  - Compute 3 Wald statistics: W_3, W_G, W_H
  - Use smallest valid one: $W_{\min}$

$$
\begin{aligned}
W_3 &= (\boldsymbol{R}\hat{\beta} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_1^{(3)}\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\beta} - \boldsymbol{r}) \\
W_G &= (\boldsymbol{R}\hat{\beta} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_G\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\beta} - \boldsymbol{r}) \\
W_H &= (\boldsymbol{R}\hat{\beta} - \boldsymbol{r})^\top (\boldsymbol{R}\hat{\boldsymbol{V}}_H\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\beta} - \boldsymbol{r})
\end{aligned} \tag{13}
$$

$$
W_{\min} = \min\{\max(W_3, 0),\ W_G,\ W_H\} \tag{14}
$$

Tsetse Fly and African Development

- Alsan (2015) examines long-run effects of tsetse fly exposure
- Dependent variables: agricultural + technological outcomes
- Two-Way clustered inference using multiple CRVEs

Table: Selected P-values for Conventional and Jackknife CRVEs

Dependent Var.	$\hat{\beta}$	HC_1	$\text{CV}_{(1)}^{(I)}$	$\text{CV}_{(1)}^{(G)}$	$\text{CV}_{(3)}$	$\text{CV}_{(3)}^{(3)}$	$\text{CV}_{(3)}^{(\max)}$
Panel A: Conventional							
Large animals	−0.231	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Intensive agriculture	−0.090	0.0080	0.0087	0.0020	0.0321	0.0003	0.0045
Plow use	−0.057	0.0096	0.0171	0.0149	0.1496	0.0715	0.0791
Panel B: Jackknife							
Large animals	−0.231	0.0000	0.0000	0.0000	0.0007	0.0000	0.0000
Intensive agriculture	−0.090	0.0092	0.0130	0.0044	0.0549	0.0027	0.0123
Plow use	−0.057	0.0124	0.0220	0.0207	0.2112	0.1365	0.1422

Syntax: `twowayjack` Command

```
twowayjack depvar indepvar controls,
    cluster(var1 var2) [fevar(varlist) sample(cond)]
```

- `depvar`: dependent variable
- `indepvar`: main regressor of interest
- `controls`: other covariates (binary or continuous)
- `cluster(varlist)` (required): exactly two cluster vars
- `fevar(varlist)`: fixed effects (like `i.var`)
  - Uses a generalized inverse (works with singular subsamples)
- `sample(string)`: subset selection, e.g., `sample(female==1)`

More info:
`http://qed.econ.queensu.ca/pub/faculty/mackinnon/twowayjack/`

Bibliography I

Alsan, M. (2015). The effect of the tsetse fly on African development. American Economic Review 105, 382–410.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? Quarterly Journal of Economics 119, 249–275.

Callaway, B. and P. H. C. Sant'Anna (2021). Difference-in-differences with multiple time periods. Journal of Econometrics 225, 200–230.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. Journal of Business & Economic Statistics 29, 238–249.

Hansen, B. E. (2025a). Jackknife standard errors for clustered regression. Working paper, University of Wisconsin.

Hansen, B. E. (2025b). Standard errors for difference-in-difference regression. Journal of Applied Econometrics 40, 291–309.

Hounyo, U. and J. Lin (2025). Jackknife variances for two-way clustering with serially correlated time effects. Working paper, SUNY Albany.

Liu, Y. (2025). csdidjack: Cluster jackknife inference for callaway and sant'anna difference-in-differences (stata package). https://github.com/liu-yunhan/csdidjack. GitHub repository; Stata package.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2021). Wild bootstrap and asymptotic inference with multiway clustering. Journal of Business & Economic Statistics 39, 509–519.

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

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

Bibliography III

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2023c). Leverage, influence, and the jackknife in clustered regression models: Reliable inference using summclust. Stata Journal 23, 942–982.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2025). Cluster-robust jackknife and bootstrap inference for logistic regression models. Econometric Reviews. forthcoming.

MacKinnon, J. G. and H. White (1985). Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics 29, 305–325.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2024). Jackknife inference with two-way clustering. Technical report.

MacKinnon, J. G., M. Ø. Nielsen, M. D. Webb, and S. Karim (2025). Improving inferences for callaway and sant'anna DiD using the cluster jackknife. Mimeo.

Bibliography IV

Mizushima, Y. and D. Powell (2025). Inference with modern difference-in-differences methods. Technical report, SSRN 5221387.

Rios-Avila, F., B. Callaway, and P. H. C. Sant'Anna (2021). csdid: Difference-in-differences with multiple time periods in Stata. In Stata Conference, pp. 47.

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

Webb, M. D. (2025). twowayjack: Stata module for two-way cluster jackknife variance estimation. https://github.com/mattdwebb/twowayjack. GitHub repository; Stata module.