Inference with a single treated cluster

Andreas Hagemann

Department of Economics
University of Michigan

Why is this interesting?

- Natural experiments are often limited to a single state.
- Standard and bootstrap inference with a single treated cluster relies on strong homogeneity assumptions.
- Methods like Conley-Taber assume “infinitely many” control clusters and strong homogeneity. If heterogeneity is allowed, it must be known exactly.
- Permutation inference has intuitive appeal but limited to testing sharp nulls.

Contribution of this paper

- I establish validity of test with one treated cluster and finitely many control clusters. Tests conventional null hypotheses.
- Needs 15 or so untreated clusters. Clusters are large and can exhibit heterogeneity of unknown form with mild restriction in one direction.
- No standard errors or knowledge of the structure of the within-cluster dependence are required.
- Underlying result: new bound on comparison of two samples of heterogenous normals where one sample has single observation.

Typical identification strategy

- One treated and some untreated clusters.
- Parameter of interest is δ.
- Treated cluster identifies $\theta_1 = \theta_0 + \delta$.
- Each untreated cluster identifies θ_0.
- Between-cluster comparisons identify $\delta = \theta_1 - \theta_0$.
- Want to test $H_0 : \delta = 0$.

Examples

- Clusters indexed by k. One treated cluster and q untreated clusters.
- Difference in differences (D_k indicates treatment, $I_t = 1$ is after intervention):

$$Y_{i,t,k} = \theta_0 I_t + \delta I_t D_k + \beta_k' Z_{i,t,k} + \zeta_{i,k} + U_{i,t,k}$$

- In first differences: $\Delta Y_{i,k} = \theta_0 + \delta D_k + \beta_k' \Delta Z_{i,k} + \Delta U_{i,k}$
- If $D_{q+1} = 1$, constant term in treated cluster becomes $\theta_1 = \theta_0 + \delta$. Write this as

$$\Delta Y_{i,k} = \begin{cases} \theta_0 + \beta_k' \Delta Z_{i,k} + \Delta U_{i,k}, & 1 \leq k \leq q \\ \theta_1 + \beta_k' \Delta Z_{i,k} + \Delta U_{i,k}, & k = q+1 \end{cases}$$

- Same idea works in levels and for probit, quantile regression, etc.

Regression with cluster-level treatment, cont.

- Recall

$$\Delta Y_{i,k} = \begin{cases} \theta_0 + \beta_k' \Delta Z_{i,k} + \Delta U_{i,k}, & 1 \leq k \leq q \\ \theta_1 + \beta_k' \Delta Z_{i,k} + \Delta U_{i,k}, & k = q+1 \end{cases}$$

- Get $q+1$ estimates $(\hat{\theta}_1, \hat{\theta}_{0,1}, \ldots, \hat{\theta}_{0,q})$. Can expect for some diagonal Σ

$$\sqrt{n}(\hat{\theta}_1 - \theta_1, \hat{\theta}_{0,1} - \theta_0, \ldots, \hat{\theta}_{0,q} - \theta_0) \rightsquigarrow N(0, \Sigma)$$

- If $H_0 : \delta = 0$ is true, then $\theta_1 = \theta_0$ but we don't know θ_0 and Σ.

- Asymptotically like testing if mutually independent $X_1 \sim N(\mu_1, \sigma^2)$ and $X_{0,k} \sim N(\mu_0, \sigma_k^2)$, $1 \leq k \leq q$, satisfy $\mu_1 = \mu_0$.

The problem

- Test $H_0 : \mu_1 = \mu_0$ with indep. $X_1 \sim N(\mu_1, \sigma^2)$ and $X_{0,k} \sim N(\mu_0, \sigma_k^2), 1 \leq k \leq q$.
- If X_1 is much larger than $X_{0,1}, \ldots, X_{0,q}$, this should be evidence against H_0.
- Don't know $\mu_0, \sigma^2, \sigma_1^2, \ldots, \sigma_q^2$ and can't estimate them.
- If we knew $\sigma^2 = \sigma_1^2 = \cdots = \sigma_q^2$, could use comparison of means and permutation test.
- Data are not exchangeable, so permutation test can be arbitrarily bad.

Towards a solution

- If X_1 is much larger than $X_{0,1}, \ldots, X_{0,q}$, this should be evidence against H_0.
- Use $\bar{X}_0 = q^{-1} \sum_{k=1}^{q} X_{0,k}$ and $w \in (0,1)$ to center and scale (more later) the data.
- $S = S(X, w) = \big((1+w)(X_1 - \bar{X}_0), (1-w)(X_1 - \bar{X}_0), X_{0,1} - \bar{X}_0, \ldots, X_{0,q} - \bar{X}_0\big)$
- Pretend S is the data and plug into comparison of means

$$s = (s_1, \ldots, s_{q+2}) \mapsto T(s) = \frac{s_1 + s_2}{2} - \frac{1}{q} \sum_{k=1}^{q} s_{k+2}$$

- $T(S) = X_1 - \bar{X}_0$. Independent of w but this changes if we permute S.

Worst-case permutations

- $S(X, w) = S = (S_1, \ldots, S_{q+2})$ and $T(S) = \dfrac{S_1 + S_2}{2} - \dfrac{1}{q} \sum_{k=1}^{q} S_{k+2}$
- Permute S to find critical value. $T(S)$ can't be larger than largest permutation statistic. Check if $T(S)$ equals largest permutation statistic.
- Rearrange S from largest to smallest element and call this $S^{\triangledown}$. Largest permutation statistic for this particular statistic T is $T(S^{\triangledown})$.

Rearrangement test

- $S = S(X, w) = \big((1+w)(X_1 - \bar{X}_0), (1-w)(X_1 - \bar{X}_0), X_{0,1} - \bar{X}_0, \ldots, X_{0,q} - \bar{X}_0\big)$
- Call the test function
$$\varphi(X, w) = 1\big\{T(S) = T(S^{\triangledown})\big\}$$
a **rearrangement test**.
- Need to account for the fact that we don't know $\mu_0, \sigma, \sigma_1, \ldots, \sigma_q$
- Nuisance parameters are under the null are, for some $\bar{\sigma} < \infty$ and $\underline{\sigma} > 0$,
$$\Lambda = \{(\mu_0, \sigma, \sigma_1, \ldots, \sigma_q) \in \mathbb{R} \times (0, \infty)^{q+1} : \sigma \leq \bar{\sigma} \text{ and } \sigma_k \geq \underline{\sigma} \text{ for all } k \text{ but one}\}.$$

- Rearrangement test $\varphi(X,w) = 1\{T(S) = T(S^{\triangledown})\}$
- $\lambda \in \Lambda = \{(\mu_0, \sigma, \sigma_1, \ldots, \sigma_q) \in \mathbb{R} \times (0,\infty)^{q+1} : \sigma \leq \bar{\sigma}$ and $\sigma_k \geq \underline{\sigma}$ for all k but one$\}$
- Under alternative, $\delta = \mu_1 - \mu_0$ and λ determine $\mathrm{E}_{\lambda,\delta}\,\varphi(X,w)$. Let $\varrho = \bar{\sigma}/\underline{\sigma}$.

THEOREM (SIZE)

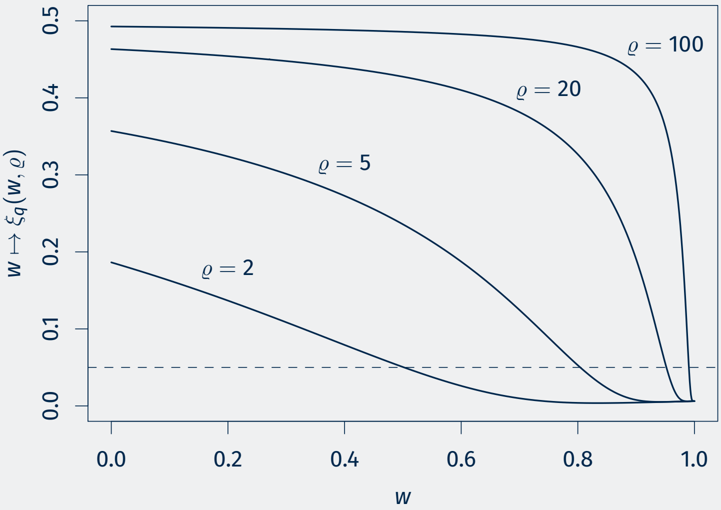
Under $H_0 : \delta = 0$ with indep. $X_1 \sim N(\mu_1, \sigma^2)$ and $X_{0,k} \sim N(\mu_0, \sigma_k^2), 1 \leq k \leq q$,

$$\sup_{\lambda\in\Lambda} \mathrm{E}_{\lambda,0}\,\varphi(X,w) \leq \int_0^\infty \Phi\big((1-w)\varrho y\big)^{q-1}\phi(y)dy + \frac{1}{2^{q+1}} + \min_{t>0}\big(\Phi(\sqrt{q-1}wt)^{q-1} + 2\Phi(-qt)\big)\,\textbf{nea}$$

Implications

- Controls size: $\sup_{\lambda \in \Lambda} \mathrm{E}_{\lambda,0}\, \varphi(X, w) \leq \alpha$ if $\xi_q(w, \varrho) \leq \alpha$
- $\xi_q(w, \varrho) \leq \alpha$ is decreasing in q and (essentially) in w, increasing in ϱ.
- $\Lambda = \{(\mu_0, \sigma, \sigma_1, \ldots, \sigma_q) \in \mathbb{R} \times (0, \infty)^{q+1} : \sigma \leq \bar{\sigma}$ and $\sigma_k \geq \underline{\sigma}$ for all k but one$\}$
- $\varrho = \bar{\sigma}/\underline{\sigma}$ is how much more variable X_1 can be than second least variable $X_{0,k}$. Can be infinitely more variable than least variable $X_{0,k}$.
- $\varrho = 2$ means variance can be up to four times larger. No other restrictions.

Size control is possible



How to test

- Pick ϱ and find $w = w_q(\alpha, \varrho)$ such that $\xi_q(w, \varrho) = \alpha$.
- Let $x \mapsto \varphi_\alpha(x) = \varphi\big(x, w_q(\alpha, \varrho)\big)$. Then
$$\sup_{\lambda \in \Lambda} \mathrm{E}_{\lambda,0}\, \varphi_\alpha(X) \leq \alpha.$$
- Available for wide range of ϱ and $\alpha = .05$ starting from $q = 15$.
- Essentially no restrictions starting from $q = 25$. Tables and R code available.
- Alternative: **robustness check** where we increase ϱ until we no longer reject H_0.

- Suppose, with fixed q and unknown $\theta_1, \theta_0, \sigma, \sigma_1, \dots, \sigma_q$,

$$\sqrt{n}\left(\frac{\hat{\theta}_1 - \theta_1}{\sigma}, \frac{\hat{\theta}_{0,1} - \theta_0}{\sigma_1}, \dots, \frac{\hat{\theta}_{0,q} - \theta_0}{\sigma_q}\right) \rightsquigarrow N(0, I_{q+1}) \tag{AN}$$

- Satisfied under unknown spatio-temporal dependence.

- Let $\hat{\theta}_n = (\hat{\theta}_1, \hat{\theta}_{0,1}, \dots, \hat{\theta}_{0,q})$. The test satisfies $\varphi_\alpha(\hat{\theta}_n) = \varphi_\alpha\left(\sqrt{n}(\hat{\theta}_n - \theta_0 1_{q+1})\right)$.

THEOREM (CONSISTENCY)

Suppose (AN) holds. If $\theta_1 = \theta_0$, then

$$\lim_{n\to\infty} \mathrm{E}\,\varphi_\alpha(\hat{\theta}_n) = \mathrm{E}\,\varphi_\alpha(X) \le \alpha.$$

REARRANGEMENT TEST

1. Choose w for given q, α, and maximal tolerance for heterogeneity, e.g., $\varrho = 2$.
2. Compute for each untreated cluster k an estimate $\hat{\theta}_{0,k}$ of θ_0 and compute an estimate $\hat{\theta}_1$ of θ_1 from the treated cluster so that the difference $\theta_1 - \theta_0$ is the treatment effect of interest. Compute $S = S(\hat{\theta}_n, w)$.
3. Reorder S from largest to smallest. Call this by $S^{\triangledown}$. Compute $T(S)$ and $T(S^{\triangledown})$.
4. Reject $H_0 \colon \theta_1 = \theta_0$ in favor of $H_1 \colon \theta_1 > \theta_0$ if $T(S) = T(S^{\triangledown})$.

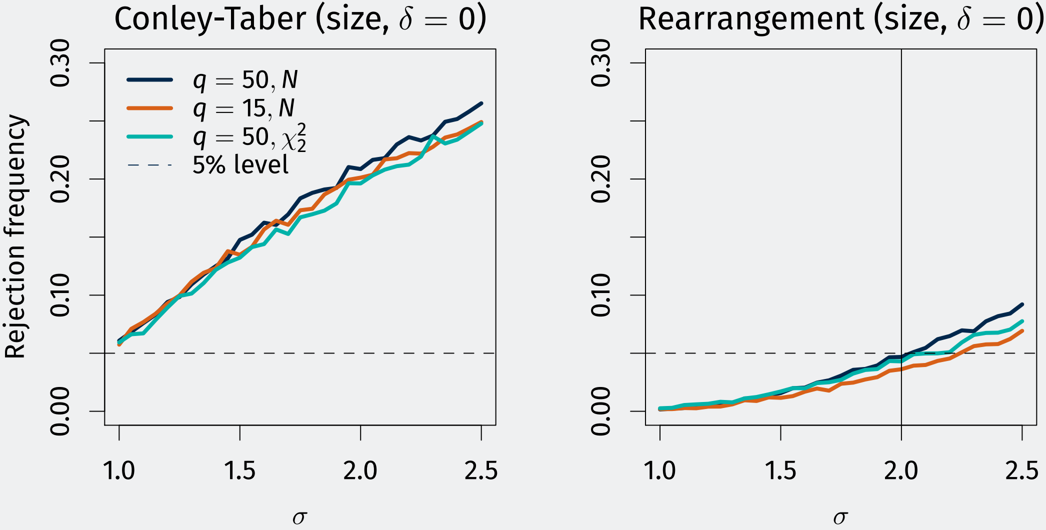
- No bootstrap, simulation, matching. Can do two-sided or one-sided in either direction, and is able to detect all fixed and $1/\sqrt{n}$-local alternatives.

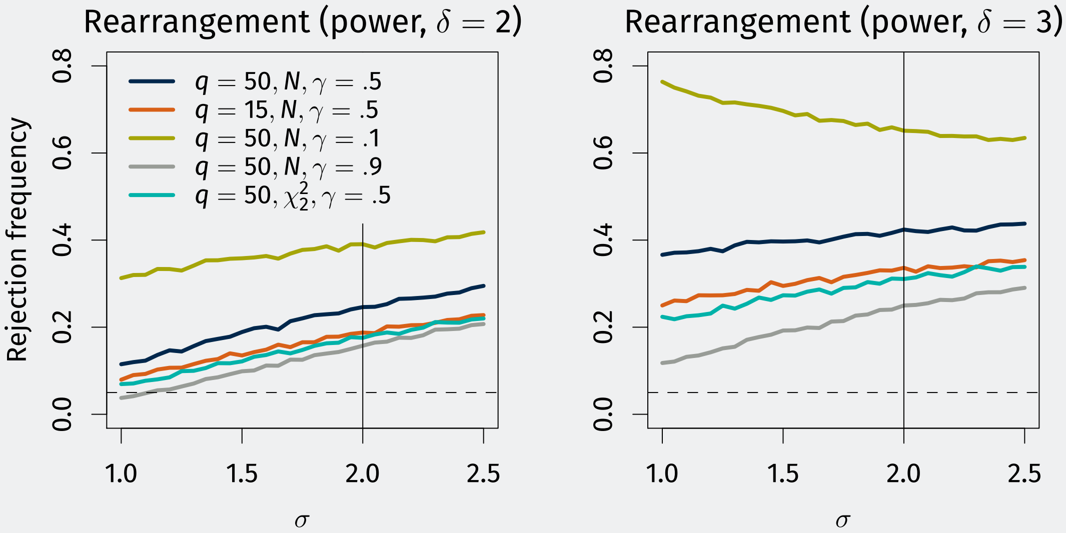
State of the literature

- Conley and Taber (2011) allows for single treated cluster but all clusters have to behave like i.i.d. copies of one another in absence of treatment.
- Ferman and Pinto (2019) and Ferman (2020) extend Conley-Taber approach to situations where heterogeneity is known exactly.
- Placebo exercises pioneered by Buchmueller, DiNardo, and Valetta (2011) require exchangeability and test sharp nulls.

Conley and Taber (2011)

- Two-way fixed effects $Y_{t,k} = \delta I_t D_k + \eta_t + \zeta_k + U_{t,k}$.
- Errors are serially correlated $U_{t,k} = \gamma U_{t-1,k} + \sigma^{1\{k=q+1\}} V_{t,k}$.
- Strength of time series correlation driven by γ. Heterogeneity driven by σ.
- Innovations $V_{t,k}$ are either standard normal or recentered/scaled chi-squared.
- Conley-Taber needs $\sigma = 1$ and number of control clusters $q \to \infty$. Rearrangement valid with fixed q and unknown σ, uses $\varrho = 2$ here.
- Record outcome of 5% level test (10,000 per coordinate) as σ changes.


Conley-Taber (size, $\delta = 0$)
Rearrangement (size, $\delta = 0$)
Rejection frequency
0.00
0.10
0.20
0.30
1.0
1.5
2.0
2.5
σ
$q = 50, N$
$q = 15, N$
$q = 50, \chi^2_2$
5% level


Rearrangement (power, $\delta = 2$)
Rearrangement (power, $\delta = 3$)
Rejection frequency
$q = 50, N, \gamma = .5$
$q = 15, N, \gamma = .5$
$q = 50, N, \gamma = .1$
$q = 50, N, \gamma = .9$
$q = 50, \chi^2_2, \gamma = .5$
0.0
0.2
0.4
0.6
0.8
1.0
1.5
2.0
2.5
σ

Garthwaite, Gross, and Notowidigdo (2014)

- Garthwaite, Gross, and Notowidigdo (2014) use DiD to study effects of large-scale disruption of public heath insurance on labor supply.
- In 2005, $\sim$170,000 adults in Tennessee (roughly 4% of non-elderly, adult population) lost access state's public health insurance system TennCare.
- Health insurance and work status for 2000-2007 from 2001-2008 March CPS. Comparison groups for Tennessee are the 16 other Southern states.
- $Y_{t,k} = \theta_0 I_t + \delta I_t D_k + \zeta_k + U_{t,k}$, where $Y_{t,k}$ is state-by-year mean of outcome for state k in year t with $17 \times 8 = 136$ state-by-year means in total.
- Bootstrap standard errors and t_{16} critical values. Their bootstrap draws states and then individuals within states with replacement.

TennCare results and robustness check

	(1)	(2)	(3)	(4)	(5)	(6)
	Has public health insurance	Employed	Employed working $<$20 hours per week	Employed working $\geq$20 hours per week	Employed working 20-35 hours per week	Employed working $\geq$35 hours per week
$\hat{\delta}$	-0.046	0.025	-0.001	0.026	0.001	0.025
s.e.	(0.010)	(0.011)	(0.004)	(0.010)	(0.007)	(0.011)
p-val.	[0.000]	[0.019]	[0.621]	[0.011]	[0.453]	[0.020]

α	Rearrangement test: largest ϱ^2 at which $H_0: \delta = 0$ is rejected ("$\times$" indicates that $H_0: \delta = 0$ cannot be rejected for any $\varrho \geq 0$)					
.10	5.434	1.793	$\times$	2.208	$\times$	$\times$
.05	2.914	0.972	$\times$	1.195	$\times$	$\times$

Concluding remarks

- Valid conventional (non-sharp) inference with one treated and finitely many control clusters. Clusters can be heterogeneous of unknown form.
- Robust inference with $\sim$15 control clusters.
- Easy to use because weights are tabulated or can be computed instantly. No matching, simulation, or resampling.