Inference with Difference-in-Differences Revisited

M. Brewer, T- F. Crossley and R. Joyce

Journal of Econometric Methods, 2018

presented by
Federico Curci

February 22nd, 2018

This paper

What we know

- **Misleading inference in D-D** because of serial correlation in the errors

This paper

What we know

- **Misleading inference in D-D** because of serial correlation in the errors

What has been proposed

- **Cluster-robust** standard errors can deal with error correlation
- Not possible with **small** number of groups

This paper

What we know

- **Misleading inference in D-D** because of serial correlation in the errors

What has been proposed

- **Cluster-robust** standard errors can deal with error correlation
- Not possible with **small** number of groups

What they propose

- Combination of **feasible GLS** with **cluster-robust** inference solve this problem even with small number of groups

Presentation

- **Inference problem in D-D: Bertrand et al. (2004)**
- Cluster-based solutions: Cameron and Miller (2015)
- Brewer et al. (2018) solution

Previously on

Bertrand et al. (2004)

Standard D-D model

$$y_{igt} = \alpha_g + \delta_t + \beta T_{gt} + \gamma w_{igt} + v_{igt} \tag{1}$$

Bertrand et al. (2004)

Standard D-D model

$$y_{igt} = \alpha_g + \delta_t + \beta T_{gt} + \gamma w_{igt} + v_{igt} \tag{1}$$

Standard errors might be inconsistent (even if estimator unbiased)

- v_{igt} may not be iid within group: $v_{igt} = \epsilon_{gt} + u_{igt}$

Bertrand et al. (2004)

Standard D-D model

$$y_{igt} = \alpha_g + \delta_t + \beta T_{gt} + \gamma w_{igt} + v_{igt} \tag{1}$$

Standard errors might be inconsistent (even if estimator unbiased)

- v_{igt} may not be iid within group: $v_{igt} = \epsilon_{gt} + u_{igt}$

...because of **serial correlation**

- DD estimation usually relies on fairly long time-series
- Most commonly used dependent variables in DD are serially correlated
- Treatment variable changes itself very little within a state over time

Bertrand et al. (2004)

Standard D-D model

$$y_{igt} = \alpha_g + \delta_t + \beta T_{gt} + \gamma w_{igt} + v_{igt} \tag{1}$$

Standard errors might be inconsistent (even if estimator unbiased)

- v_{igt} may not be iid within group: $v_{igt} = \epsilon_{gt} + u_{igt}$

...because of **serial correlation**

- DD estimation usually relies on fairly long time-series
- Most commonly used dependent variables in DD are serially correlated
- Treatment variable changes itself very little within a state over time

...and **cross sectional correlation**

- Use micro-data but estimate effect treatment which varies only at group level

Bertrand et al. (2004)

Wrong inference in terms of both **size and power**

- Size: **type I error**
  - Probability to reject null hypothesis when the null hypothesis is true
  - Probability to assert something that is absent
- Power: **type II error**
  - Probability to not reject null hypothesis when the null hypothesis is false
  - Ability of a test to detect an effect, if the effect actually exist

To solve cross-sectional variation

First-stage aggregation

- Regress using micro data y_{igt} on w_{igt} and take the mean residual within each group-time cell ($\hat{Y}_{gt}$)
- Regress $\hat{Y}_{igt}$ on fixed effects and treatment

To solve cross-sectional variation

First-stage aggregation

- Regress using micro data y_{igt} on w_{igt} and take the mean residual within each group-time cell ($\hat{Y}_{gt}$)
- Regress $\hat{Y}_{igt}$ on fixed effects and treatment

We **still** have a problem of **serial correlation**

Bertrand et al. (2004)

Assess the extent of serial correlation problem in DD

- Examine how DD performs on placebo laws: treated states and year of passage are chosen at random
- Since law are fictitious a significant effect at the 5 percent level should be found roughly 5 percent of the time
- Use Monte Carlo simulations

Assess magnitude of serial correlation problem

Use women's wages from **Current Population Survey**

- Years 1979-1999
- Women between 25 and 50 with positive earnings
- 50*21 state-years couples (1050)

Assess magnitude of serial correlation problem

Use women's wages from **Current Population Survey**

- Years 1979-1999
- Women between 25 and 50 with positive earnings
- 50*21 state-years couples (1050)

Randomly generate laws that affect some states and not other

- Draw a year at random from a uniform distribution between 85-95
- Select half the states at random
- Create treatment dummy

Assess magnitude of serial correlation problem

Use women's wages from **Current Population Survey**

- Years 1979-1999
- Women between 25 and 50 with positive earnings
- 50*21 state-years couples (1050)

Randomly generate laws that affect some states and not other

- Draw a year at random from a uniform distribution between 85-95
- Select half the states at random
- Create treatment dummy

Estimate $w_{igt} = \alpha_g + \delta_t + \beta T_{gt} + \gamma w_{igt} + v_{igt}$

- Generate estimate $\hat{\beta}$ and standard error
- Repeat this exercise large number of times each time drawing new laws at random

Bertrand et al.

A. CPS DATA

Data	$\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$	Modifications	Rejection rate: No effect	Rejection rate: 2% effect
1) CPS micro, log wage			.675 (.027)	.855 (.020)
2) CPS micro, log wage		Cluster at state-year level	.44 (.029)	.74 (.025)
3) CPS agg, log wage	.509, .440, .332		.435 (.029)	.72 (.026)
4) CPS agg, log wage	.509, .440, .332	Sampling w/replacement	.49 (.025)	.663 (.024)
5) CPS agg, log wage	.509, .440, .332	Serially uncorrelated laws	.05 (.011)	.988 (.006)
6) CPS agg, employment	.470, .418, .367		.46 (.025)	.88 (.016)
7) CPS agg, hours worked	.151, .114, .063		.265 (.022)	.280 (.022)
8) CPS agg, changes in log wage	−.046, .032, .002		0	.978 (.007)

200 independent draws of placebo laws

Bertrand et al.

A. CPS DATA

Data	$\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$	Modifications	Rejection rate: No effect	Rejection rate: 2% effect
1) CPS micro, log wage			.675 (.027)	.855 (.020)
2) CPS micro, log wage		Cluster at state-year level	.44 (.029)	.74 (.025)
3) CPS agg, log wage	.509, .440, .332		.435 (.029)	.72 (.026)
4) CPS agg, log wage	.509, .440, .332	Sampling w/replacement	.49 (.025)	.663 (.024)
5) CPS agg, log wage	.509, .440, .332	Serially uncorrelated laws	.05 (.011)	.988 (.006)
6) CPS agg, employment	.470, .418, .367		.46 (.025)	.88 (.016)
7) CPS agg, hours worked	.151, .114, .063		.265 (.022)	.280 (.022)
8) CPS agg, changes in log wage	−.046, .032, .002		0	.978 (.007)

Reject null of no effect 67.5 percent of the time

Bertrand et al.

A. CPS DATA

Data	$\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$	Modifications	Rejection rate: No effect	Rejection rate: 2% effect
1) CPS micro, log wage			.675 (.027)	.855 (.020)
2) CPS micro, log wage		Cluster at state-year level	.44 (.029)	.74 (.025)
3) CPS agg, log wage	.509, .440, .332		.435 (.029)	.72 (.026)
4) CPS agg, log wage	.509, .440, .332	Sampling w/replacement	.49 (.025)	.663 (.024)
5) CPS agg, log wage	.509, .440, .332	Serially uncorrelated laws	.05 (.011)	.988 (.006)
6) CPS agg, employment	.470, .418, .367		.46 (.025)	.88 (.016)
7) CPS agg, hours worked	.151, .114, .063		.265 (.022)	.280 (.022)
8) CPS agg, changes in log wage	−.046, .032, .002		0	.978 (.007)

Power: reject null of no effect against alternative of 2 percent 66 times

Presentation

- Inference problem in D-D: Bertrand et al. (2004)
- **Cluster-based solutions: Cameron and Miller (2015)**
- Brewer et al. (2018) solution

Cameron and Miller (2015)

Cluster-robust standard errors can deal with error correlation

- Cluster-Robust Standard Error (CRSE):
  $V_{clu}\left[\hat{\beta}\right] = \left(X'X\right)^{-1}\left(\sum_{g=1}^{G} X_g v_g v_g' X_g'\right)\left(X'X\right)^{-1}$
- Consistent and Wald statistics based on it asymptotical normal, as $G \to \infty$

Cameron and Miller (2015)

Cluster-robust standard errors can deal with error correlation

- Cluster-Robust Standard Error (CRSE):
  $V_{clu}\left[\hat{\beta}\right] = \left(X'X\right)^{-1}\left(\sum_{g=1}^{G} X_g v_g v_g' X_g'\right)\left(X'X\right)^{-1}$
- Consistent and Wald statistics based on it asymptotical normal, as $G \to \infty$

Wald t-statistics

- $w = \frac{\hat{\beta}-\beta_0}{s_{\hat{\beta}}}$
- If $G \to \infty$, then $w \sim N(0,1)$ under $H_0 : \beta = \beta_0$
- Finite G: unknown distribution of w. Common to use $w \sim T(G-1)$

Cameron and Miller (2015)

Problem with Few Clusters

- Despite reasonable precision in estimating β, $\hat{V}_{clu}(\hat{\beta})$ can be **downwards-biased**
  - "Overfitting": estimated residuals smaller than true errors
- Small-G bias larger when **distribution of regressors is skewed** (Mackinnon and Webb, 2017)
  - Imbalance between number of treated and control

Cameron and Miller (2015)

Solution 1: Bias-Corrected CRSE

- $\tilde{v}_g = \sqrt{\frac{G(N-1)}{(G-1)(N-K)}}\hat{v}_g$
- Reduce but not eliminate over-rejection when there are few clusters

Cameron and Miller (2015)

Solution 1: Bias-Corrected CRSE

- $\tilde{v}_g = \sqrt{\frac{G(N-1)}{(G-1)(N-K)}}\hat{v}_g$
- Reduce but not eliminate over-rejection when there are few clusters

Solution 2: Wild Cluster Bootstrap

- Idea
  - Estimate main model imposing null hypothesis that you wish to test to give estimate of $\tilde{\beta}_{H_0}$
- Example: test statistical significance of one OLS regressor.
  - Regress y_{ig} on all components of x_{ig} except regressor with coefficient 0
  - Obtain residuals $\tilde{v}_{ig} = y_{ig} - x'_{ig}\tilde{\beta}_{H_0}$
- Less trivial to implement and computationally more intensive

Montecarlo

Presentation

- Inference problem in D-D: Bertrand et al. (2004)
- Cluster-based solutions: Cameron and Miller (2015)
- **Brewer et al. (2018) solution**

Brewer et al. (2018)

Solution: Combination of **feasible GLS** with **cluster-robust** inference

Brewer et al. (2018)

Solution: Combination of **feasible GLS** with **cluster-robust** inference

- Tests of **correct size** can be obtained with standard statistical software
  - vce(cluster *clustervar*)
  - Even with few groups

Brewer et al. (2018)

Solution: Combination of **feasible GLS** with **cluster-robust** inference

- Tests of **correct size** can be obtained with standard statistical software
  - vce(cluster *clustervar*)
  - Even with few groups
- Real problem is **power**
  - Gains in power can be achieved using feasible GLS even with small groups

Brewer et al. (2018)

Solution: Combination of **feasible GLS** with **cluster-robust** inference

- Tests of **correct size** can be obtained with standard statistical software
  - vce(cluster *clustervar*)
  - Even with few groups
- Real problem is **power**
  - Gains in power can be achieved using feasible GLS even with small groups
- **Combination** feasible GLS and cluster-robust inference can also control test size

Structure of the paper

Replicate Monte Carlo simulations of Bertrand et al. (2004)

- **CRSE and wild-bootstrap have low size distortion**
- CRSE and wild-bootstrap have low power to detect the real effects
- Increasing Power with Feasible GLS

Rejection rates when the Null is True

Table 1: Rejection Rates for Tests of Nominal 5% Size with Placebo Treatments in Log-Earnings Data.

	G = 50	G = 20	G = 10	G = 6
Assume iid	0.366	0.394	0.396	0.398
	(0.007)	(0.007)	(0.007)	(0.007)
CRSE, N(0,1) critical values	0.058	0.078	0.127	0.228
	(0.003)	(0.004)	(0.005)	(0.006)
$\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.039	0.039	0.053
	(0.003)	(0.003)	(0.003)	(0.003)
Wild cluster bootstrap-t	0.038	0.060	0.044	0.060
	(0.001)	(0.002)	(0.003)	(0.003)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample groups (US states) from CPS-MORG data, having imposed the sampling restrictions described in the text. The treatment parameter has a true coefficient of zero. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The different inference methods used are discussed in the text.

Rejection rates when the Null is True

Table 1: Rejection Rates for Tests of Nominal 5% Size with Placebo Treatments in Log-Earnings Data.

	G = 50	G = 20	G = 10	G = 6
Assume iid	0.366	0.394	0.396	0.398
	(0.007)	(0.007)	(0.007)	(0.007)
CRSE, N(0,1) critical values	0.058	0.078	0.127	0.228
	(0.003)	(0.004)	(0.005)	(0.006)
$\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.039	0.039	0.053
	(0.003)	(0.003)	(0.003)	(0.003)
Wild cluster bootstrap-t	0.038	0.060	0.044	0.060
	(0.001)	(0.002)	(0.003)	(0.003)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample groups (US states) from CPS-MORG data, having imposed the sampling restrictions described in the text. The treatment parameter has a true coefficient of zero. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The different inference methods used are discussed in the text.

Assuming iid errors, reject more than 40 %

Rejection rates when the Null is True

Table 1: Rejection Rates for Tests of Nominal 5% Size with Placebo Treatments in Log-Earnings Data.

	G = 50	G = 20	G = 10	G = 6
Assume iid	0.366	0.394	0.396	0.398
	(0.007)	(0.007)	(0.007)	(0.007)
CRSE, N(0,1) critical values	0.058	0.078	0.127	0.228
	(0.003)	(0.004)	(0.005)	(0.006)
$\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.039	0.039	0.053
	(0.003)	(0.003)	(0.003)	(0.003)
Wild cluster bootstrap-t	0.038	0.060	0.044	0.060
	(0.001)	(0.002)	(0.003)	(0.003)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample groups (US states) from CPS-MORG data, having imposed the sampling restrictions described in the text. The treatment parameter has a true coefficient of zero. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The different inference methods used are discussed in the text.

Cluster-robust SE solve serial correlation with big G

Rejection rates when the Null is True

Table 1: Rejection Rates for Tests of Nominal 5% Size with Placebo Treatments in Log-Earnings Data.

	G = 50	G = 20	G = 10	G = 6
Assume iid	0.366	0.394	0.396	0.398
	(0.007)	(0.007)	(0.007)	(0.007)
CRSE, N(0,1) critical values	0.058	0.078	0.127	0.228
	(0.003)	(0.004)	(0.005)	(0.006)
$\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.039	0.039	0.053
	(0.003)	(0.003)	(0.003)	(0.003)
Wild cluster bootstrap-t	0.038	0.060	0.044	0.060
	(0.001)	(0.002)	(0.003)	(0.003)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample groups (US states) from CPS-MORG data, having imposed the sampling restrictions described in the text. The treatment parameter has a true coefficient of zero. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The different inference methods used are discussed in the text.

Problem cluster-robust SE with small G

Rejection rates when the Null is True

Table 1: Rejection Rates for Tests of Nominal 5% Size with Placebo Treatments in Log-Earnings Data.

	G = 50	G = 20	G = 10	G = 6
Assume iid	0.366	0.394	0.396	0.398
	(0.007)	(0.007)	(0.007)	(0.007)
CRSE, N(0,1) critical values	0.058	0.078	0.127	0.228
	(0.003)	(0.004)	(0.005)	(0.006)
$\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.039	0.039	0.053
	(0.003)	(0.003)	(0.003)	(0.003)
Wild cluster bootstrap-t	0.038	0.060	0.044	0.060
	(0.001)	(0.002)	(0.003)	(0.003)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample groups (US states) from CPS-MORG data, having imposed the sampling restrictions described in the text. The treatment parameter has a true coefficient of zero. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The different inference methods used are discussed in the text.

Bias-correction of cluster-robust SE have low size distortion, even with small number of groups

Rejection rates when the Null is True

Table 1: Rejection Rates for Tests of Nominal 5% Size with Placebo Treatments in Log-Earnings Data.

	G = 50	G = 20	G = 10	G = 6
Assume iid	0.366	0.394	0.396	0.398
	(0.007)	(0.007)	(0.007)	(0.007)
CRSE, N(0,1) critical values	0.058	0.078	0.127	0.228
	(0.003)	(0.004)	(0.005)	(0.006)
$\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.039	0.039	0.053
	(0.002)	(0.003)	(0.003)	(0.003)
Wild cluster bootstrap-t	0.038	0.060	0.044	0.060
	(0.001)	(0.002)	(0.003)	(0.003)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample groups (US states) from CPS-MORG data, having imposed the sampling restrictions described in the text. The treatment parameter has a true coefficient of zero. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The different inference methods used are discussed in the text.

Similar for wild cluster bootstrap

Rejection rates when the Null is True

Results **robust**

- Binary dependent variable
- Wide range of error process

Rejection rates when the Null is True

Results **robust**

- Binary dependent variable
- Wide range of error process

Bias-corrected CRSE **not robust**

- Large **imbalance** between treatment and control groups
- Different from bootstrap

Structure of the paper

Replicate Monte Carlo simulations of Bertrand et al. (2004)

- CRSE and wild-bootstrap have low size distortion
- **CRSE and wild-bootstrap have low power to detect the real effects**
- Increasing Power with Feasible GLS

Power to detect the real effects

Table 6: Rejection Rates for Tests of True 5% Size with Different Treatment Effects (β) in Log-Earnings Data.

	G = 50	G = 20	G = 10	G = 6
$\beta = 0.02$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.378	0.162	0.095	0.072
	(0.007)	(0.005)	(0.004)	(0.004)
$\beta = 0.02$: wild cluster bootstrap-t	0.405	0.131	0.095	0.078
	(0.004)	(0.003)	(0.004)	(0.004)
$\beta = 0.05$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.954	0.625	0.338	0.178
	(0.003)	(0.007)	(0.007)	(0.005)
$\beta = 0.05$: wild cluster bootstrap-t	0.954	0.510	0.309	0.176
	(0.002)	(0.005)	(0.006)	(0.005)
$\beta = 0.10$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	1.000	0.963	0.782	0.482
	(.)	(0.003)	(0.006)	(0.007)
$\beta = 0.10$: wild cluster bootstrap-t	1.000	0.902	0.737	0.439
	(.)	(0.003)	(0.006)	(0.007)
$\beta = 0.15$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	1.000	0.999	0.933	0.776
	(.)	(0.000)	(0.004)	(0.006)
$\beta = 0.15$: wild cluster bootstrap-t	1.000	0.992	0.906	0.707
	(.)	(0.001)	(0.004)	(0.006)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample states from CPS-MORG data, having imposed the sampling restrictions described in the text. β is the true value of the treatment parameter. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The inference methods used are discussed in the text. We adjust for test size when making power comparisons using the procedure outlined by Davidson and MacKinnon (1998). See text for details.

Power to detect the real effects

Table 6: Rejection Rates for Tests of True 5% Size with Different Treatment Effects (β) in Log-Earnings Data.

	G = 50	G = 20	G = 10	G = 6
$\beta = 0.02$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.378	0.162	0.095	0.072
	(0.007)	(0.005)	(0.004)	(0.004)
$\beta = 0.02$: wild cluster bootstrap-t	0.405	0.131	0.095	0.078
	(0.004)	(0.003)	(0.004)	(0.004)
$\beta = 0.05$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.954	0.625	0.338	0.178
	(0.003)	(0.007)	(0.007)	(0.005)
$\beta = 0.05$: wild cluster bootstrap-t	0.954	0.510	0.309	0.176
	(0.002)	(0.005)	(0.006)	(0.005)
$\beta = 0.10$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	1.000	0.963	0.782	0.482
	(.)	(0.003)	(0.006)	(0.007)
$\beta = 0.10$: wild cluster bootstrap-t	1.000	0.902	0.737	0.439
	(.)	(0.003)	(0.006)	(0.007)
$\beta = 0.15$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	1.000	0.999	0.933	0.776
	(.)	(0.000)	(0.004)	(0.006)
$\beta = 0.15$: wild cluster bootstrap-t	1.000	0.992	0.906	0.707
	(.)	(0.001)	(0.004)	(0.006)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample states from CPS-MORG data, having imposed the sampling restrictions described in the text. β is the true value of the treatment parameter. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The inference methods used are discussed in the text. We adjust for test size when making power comparisons using the procedure outlined by Davidson and MacKinnon (1998). See text for details.

Low power for both methods: 2 % effect detected only 40 % times

Power to detect the real effects

Table 6: Rejection Rates for Tests of True 5% Size with Different Treatment Effects (β) in Log-Earnings Data.

	G = 50	G = 20	G = 10	G = 6
$\beta = 0.02$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.378	0.162	0.095	0.072
	(0.007)	(0.005)	(0.004)	(0.004)
$\beta = 0.02$: wild cluster bootstrap-t	0.405	0.131	0.095	0.078
	(0.004)	(0.003)	(0.004)	(0.004)
$\beta = 0.05$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.954	0.625	0.338	0.178
	(0.003)	(0.007)	(0.007)	(0.005)
$\beta = 0.05$: wild cluster bootstrap-t	0.954	0.510	0.309	0.176
	(0.002)	(0.005)	(0.006)	(0.005)
$\beta = 0.10$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	1.000	0.963	0.782	0.482
	(.)	(0.003)	(0.006)	(0.007)
$\beta = 0.10$: wild cluster bootstrap-t	1.000	0.902	0.737	0.439
	(.)	(0.003)	(0.006)	(0.007)
$\beta = 0.15$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	1.000	0.999	0.933	0.776
	(.)	(0.000)	(0.004)	(0.006)
$\beta = 0.15$: wild cluster bootstrap-t	1.000	0.992	0.906	0.707
	(.)	(0.001)	(0.004)	(0.006)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample states from CPS-MORG data, having imposed the sampling restrictions described in the text. β is the true value of the treatment parameter. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The inference methods used are discussed in the text. We adjust for test size when making power comparisons using the procedure outlined by Davidson and MacKinnon (1998). See text for details.

Higher power with higher effect

Power to detect the real effects

Table 6: Rejection Rates for Tests of True 5% Size with Different Treatment Effects (β) in Log-Earnings Data.

	G = 50	G = 20	G = 10	G = 6
$\beta = 0.02$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.378	0.162	0.095	0.072
	(0.007)	(0.005)	(0.004)	(0.004)
$\beta = 0.02$: wild cluster bootstrap-t	0.405	0.131	0.095	0.078
	(0.004)	(0.003)	(0.004)	(0.004)
$\beta = 0.05$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.954	0.625	0.338	0.178
	(0.003)	(0.007)	(0.007)	(0.005)
$\beta = 0.05$: wild cluster bootstrap-t	0.954	0.510	0.309	0.176
	(0.002)	(0.005)	(0.006)	(0.005)
$\beta = 0.10$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	1.000	0.963	0.782	0.482
	(.)	(0.003)	(0.006)	(0.007)
$\beta = 0.10$: wild cluster bootstrap-t	1.000	0.902	0.737	0.439
	(.)	(0.003)	(0.006)	(0.007)
$\beta = 0.15$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	1.000	0.999	0.933	0.776
	(.)	(0.000)	(0.004)	(0.006)
$\beta = 0.15$: wild cluster bootstrap-t	1.000	0.992	0.906	0.707
	(.)	(0.001)	(0.004)	(0.006)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample states from CPS-MORG data, having imposed the sampling restrictions described in the text. β is the true value of the treatment parameter. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The inference methods used are discussed in the text. We adjust for test size when making power comparisons using the procedure outlined by Davidson and MacKinnon (1998). See text for details.

Problem of power higher with lower groups

Power to detect the real effects

Table 6: Rejection Rates for Tests of True 5% Size with Different Treatment Effects (β) in Log-Earnings Data.

	G = 50	G = 20	G = 10	G = 6
$\beta = 0.02$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.378	0.162	0.095	0.072
	(0.007)	(0.005)	(0.004)	(0.004)
$\beta = 0.02$: wild cluster bootstrap-t	0.405	0.131	0.095	0.078
	(0.004)	(0.003)	(0.004)	(0.004)
$\beta = 0.05$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.954	0.625	0.338	0.178
	(0.003)	(0.007)	(0.007)	(0.005)
$\beta = 0.05$: wild cluster bootstrap-t	0.954	0.510	0.309	0.176
	(0.002)	(0.005)	(0.006)	(0.005)
$\beta = 0.10$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	1.000	0.963	0.782	0.482
	(.)	(0.003)	(0.006)	(0.007)
$\beta = 0.10$: wild cluster bootstrap-t	1.000	0.902	0.737	0.439
	(.)	(0.003)	(0.006)	(0.007)
$\beta = 0.15$: $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	1.000	0.999	0.933	0.776
	(.)	(0.000)	(0.004)	(0.006)
$\beta = 0.15$: wild cluster bootstrap-t	1.000	0.992	0.906	0.707
	(.)	(0.001)	(0.004)	(0.006)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample states from CPS-MORG data, having imposed the sampling restrictions described in the text. β is the true value of the treatment parameter. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The inference methods used are discussed in the text. We adjust for test size when making power comparisons using the procedure outlined by Davidson and MacKinnon (1998). See text for details.

Problem of power higher with lower groups even for large true effects

Structure of the paper

Replicate Monte Carlo simulations of Bertrand et al. (2004)

- CRSE and wild-bootstrap have low size distortion
- CRSE and wild-bootstrap have low power to detect the real effects
- **Increasing Power with Feasible GLS**

Feasible GLS

Way to increase efficiency by exploiting knowledge of serial correlation

- $y = X\beta + v, \qquad Cov[v|X] = \Omega, \qquad \hat{\beta}^{GLS} = \left(X'\Omega X\right)^{-1} X'\Omega y$
- Assume AR(k) process for group-time shocks
- Estimate model
- Estimate k parameters using OLS residuals
- Apply GLS transformation
- Estimate model on transformed variables

Feasible GLS

Problems

- Estimate k lag parameters inconsistent: Hansen's bias correction may not work for small G
- Misspecification of the error process

Feasible GLS

Problems

- Estimate k lag parameters inconsistent: Hansen's bias correction may not work for small G
- Misspecification of the error process

Not affect consistency and (likely) more efficient than OLS estimator, but **affect test size**

- Possible to use cluster-robust inference to improve test size
- Plug FGLS residuals into the CRSE formula

Increasing power with Feasible GLS

Table 7: Rejection Rates for Tests of True 5% Size with Treatment Effects of Zero and +0.05 in Log-Earnings Data.

	G = 50		G = 20		G = 6	
	effect = 0	**effect = 0.05**	**effect = 0**	**effect = 0.05**	**effect = 0**	**effect = 0.05**
OLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.954	0.039	0.625	0.053	0.178
	(0.003)	(0.003)	(0.003)	(0.007)	(0.003)	(0.005)
FGLS	0.060	0.996	0.092	0.800	0.114	0.292
	(0.003)	(0.001)	(0.004)	(0.006)	(0.004)	(0.006)
FGLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.994	0.044	0.799	0.058	0.272
	(0.003)	(0.001)	(0.003)	(0.006)	(0.003)	(0.006)
BC-FGLS	0.042	0.995	0.064	0.798	0.089	0.291
	(0.003)	(0.001)	(0.003)	(0.006)	(0.004)	(0.006)
BC-FGLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.045	0.991	0.051	0.784	0.063	0.267
	(0.003)	(0.001)	(0.003)	(0.006)	(0.003)	(0.006)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample states from CPS-MORG data, having imposed the sampling restrictions described in the text. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The inference methods used are discussed in the text. We adjust for test size when making power comparisons using the procedure outlined by Davidson and MacKinnon (1998). See text for details.

MDE

Increasing power with Feasible GLS

Table 7: Rejection Rates for Tests of True 5% Size with Treatment Effects of Zero and +0.05 in Log-Earnings Data.

	G = 50		G = 20		G = 6	
	effect = 0	effect = 0.05	effect = 0	effect = 0.05	effect = 0	effect = 0.05
OLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.954	0.039	0.625	0.053	0.178
	(0.003)	(0.003)	(0.003)	(0.007)	(0.003)	(0.005)
FGLS	0.060	0.996	0.092	0.800	0.114	0.292
	(0.003)	(0.001)	(0.004)	(0.006)	(0.004)	(0.006)
FGLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.994	0.044	0.799	0.058	0.272
	(0.003)	(0.001)	(0.003)	(0.006)	(0.003)	(0.006)
BC-FGLS	0.042	0.995	0.064	0.798	0.089	0.291
	(0.003)	(0.001)	(0.003)	(0.006)	(0.004)	(0.006)
BC-FGLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.045	0.991	0.051	0.784	0.063	0.267
	(0.003)	(0.001)	(0.003)	(0.006)	(0.003)	(0.006)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample states from CPS-MORG data, having imposed the sampling restrictions described in the text. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The inference methods used are discussed in the text. We adjust for test size when making power comparisons using the procedure outlined by Davidson and MacKinnon (1998). See text for details.

FGLS improves power wrt OLS, even with small G

Increasing power with Feasible GLS

Table 7: Rejection Rates for Tests of True 5% Size with Treatment Effects of Zero and +0.05 in Log-Earnings Data.

	G = 50		G = 20		G = 6	
	effect = 0	effect = 0.05	effect = 0	effect = 0.05	effect = 0	effect = 0.05
OLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.954	0.039	0.625	0.053	0.178
	(0.003)	(0.003)	(0.003)	(0.007)	(0.003)	(0.005)
FGLS	0.060	0.996	0.092	0.800	0.114	0.292
	(0.003)	(0.001)	(0.004)	(0.006)	(0.004)	(0.006)
FGLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.046	0.994	0.044	0.799	0.058	0.272
	(0.003)	(0.001)	(0.003)	(0.006)	(0.003)	(0.006)
BC-FGLS	0.042	0.995	0.064	0.798	0.089	0.291
	(0.003)	(0.001)	(0.003)	(0.006)	(0.004)	(0.006)
BC-FGLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.045	0.991	0.051	0.784	0.063	0.267
	(0.003)	(0.001)	(0.003)	(0.006)	(0.003)	(0.006)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample states from CPS-MORG data, having imposed the sampling restrictions described in the text. G is the number of sampled states. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The inference methods used are discussed in the text. We adjust for test size when making power comparisons using the procedure outlined by Davidson and MacKinnon (1998). See text for details.

FGLS combined with bias-corrected CRSE improves size, even with small G

Increasing power with Feasible GLS

Results not robust

- Parametric assumption about serial correlation (power advantage FGLS if process is not MA)
- Number of time periods (power advantage FGLS with high T)

What we learn from this paper

Over-rejection null of no effect not a problem

- Both with simple bias-correction CRSE and bootstrap
- Even with small groups
- Bias-correction CRSE: vce(cluster *clustervar*)
- We know it already from Cameron and Miller (2015)

What we learn from this paper

Over-rejection null of no effect not a problem

- Both with simple bias-correction CRSE and bootstrap
- Even with small groups
- Bias-correction CRSE: vce(cluster *clustervar*)
- We know it already from Cameron and Miller (2015)

Bias-correction CRSE **not robust** in case of imbalances treatment-control and reduce (but not eliminate) over-rejection

$\Downarrow$

Bootstrap

What we learn from this paper

Problem to detect real effect

- **Combining FGLS and CRSE** possible to solve power with correct test size
- Even with small groups

What we learn from this paper

Problem to detect real effect

- **Combining FGLS and CRSE** possible to solve power with correct test size
- Even with small groups

FGLS+CRSE **not robust** in case to small T periods and parametric assumption of serial correlation + **improve but not solve** power problem

$\Downarrow$

Bootstrap

Solution 2: Wild Cluster Bootstrap, cont'd

Implementation

1. Obtain b^{th} resample
   1. Obtain $\tilde{u}_{ig} = y_{ig} - x'_{ig}\tilde{\beta}_{H_0}$
   2. Randomly assign cluster g the weight $d_g = \begin{cases} -1 & \text{w.p. } 0.5 \\ 1 & \text{w.p. } 0.5 \end{cases}$
   3. Generate new pseudo-residuals $u^*_{ig} = d_g \times \tilde{u}_{ig}$ and new outcome variables $y^*_{ig} = x'_{ig}\tilde{\beta}_{H_0} + u^*_{ig}$
2. Compute OLS estimate $\hat{\beta}^*_b$
3. Calculate the Wald test statistics $w^*_b = \frac{\hat{\beta}^*_b - \hat{\beta}}{s_{\hat{\beta}^*_b}}$

Essentially

- Replacing y_g in each resample with $y^*_g = X\tilde{\beta}_{H_0} + \tilde{u}_g$ or $y^*_g = X\tilde{\beta}_{H_0} - \tilde{u}_g$
- Obtain 2^G unique values of $w^*_1, ..., w^*_B$

Back to Solutions

Cameron and Miller (2015): Montecarlo

Table 2 - Cross-section individual level data
Monte Carlo rejection rates of true null hypothesis (slope = 0) with different number of clusters and different rejection methods
Nominal 5% rejection rates

	Estimation Method	Numbers of Clusters				
		6	10	20	30	50
	Wald Tests					
1	White Robust, T(N-k) for critical value	0.165	0.174	0.172	0.181	0.176
2	Cluster on state, N(0,1) for critical value	0.213	0.130	0.091	0.098	0.080
3	Cluster on state, T(G-1) for critical value	0.124	0.094	0.075	0.080	0.070
4	Cluster on state, T(G-2) for critical value	0.108	0.090	0.075	0.079	0.070
5	Cluster on state, CR2 bias correction, T(G-1) for critical value	0.089	0.075	0.066	0.071	0.065
6	Cluster on state, CR3 bias correction, T(G-1) for critical value	0.051	0.058	0.047	0.061	0.063
7	Cluster on state, CR2 bias correction, IK degrees of freedom	0.060	0.056	0.045	0.056	0.055
8	Cluster on state, CR2 bias correction, T(CSS effective # clusters)	0.118	0.077	0.055	0.062	0.060
9	Pairs cluster bootstrap for standard error, T(G-1) for critical value	0.090	0.063	0.066	0.070	0.072
	Bootstrap Percentile-T methods					
10	Pairs cluster bootstrap	0.019	0.037	0.043	0.069	0.057
11	Wild cluster bootstrap, Rademacher 2 point distribution	0.081	0.062	0.050	0.068	0.055
12	Wild cluster bootstrap, Webb 6 point distribution	0.087	0.063	0.058	0.064	0.055
13	Wild cluster bootstrap, Rademacher 2 pt, do not impose null hypothesis	0.085	0.076	0.060	0.073	0.062
14	IK effective DOF (mean)	3.3	5.5	9.6	12.6	17.1
15	IK effective DOF (5th percentile)	2.7	4.1	5.3	6.7	9.7
16	IK effective DOF (95th percentile)	3.8	6.9	14.3	20.3	30.4
17	CSS effective # clusters (mean)	4.6	6.6	10.2	12.9	17.1

Notes: Data drawn from March 2012 CPS data, 3% sample from IPUMS download (later version to use a larger data set). 1000 Monte Carlo replications (later version to have more reps). 399 Bootstrap replications. "IK effective DOF" from Imbens and Kolesar (2013), and "CSS effective # clusters" from Carter, Schnepel and Steigerwald (2013), see section x.x.

Back to Solutions Back to Conclusion

Size robustness to error specification

Simulate state-time shocks, changing degree of serial correlation and non-normality. Assume AR(1) process

$$\epsilon_{gt} = \rho \epsilon_{gt-1} + \sqrt{\frac{0.004\left(1-0.4^2\right)(d-2)}{d}} w_{gt} \tag{2}$$

$$\epsilon_{g1} = \sqrt{\frac{0.004d}{d-1}} w_{g1} \tag{3}$$

Back to Robustness

Size robustness to error specification

Table 5: Rejection Rates for Tests of Nominal 5% Size Using $\sqrt{G/(G-1)}$-CRSEs and t_{G-1} Critical Values with 10 Groups (Simulated Data).

	$\rho = 0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	ρ varies with g
d = 4	0.049	0.053	0.049	0.053	0.047	0.050
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
d = 20	0.055	0.052	0.050	0.053	0.050	0.048
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
d = 60	0.055	0.053	0.051	0.055	0.052	0.050
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
d = 120	0.055	0.048	0.050	0.051	0.054	0.052
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 10,000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample groups (US states) from CPS-MORG data, having imposed the sampling restrictions described in the text. The treatment parameter has a true coefficient of zero. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. Simulated log-earnings are generated by effectively replacing the empirical regression residuals with a simulated error term generated according to an AR(1) process. Each cell in the table represents a different AR(1) process. ρ denotes the AR(1) parameter. In the final column the AR(1) parameter is drawn separately for each group, from a uniform distribution between 0 and 1. d denotes the degrees of freedom of the scaled t distribution from which the white noise is drawn (hence it controls the degree of non-normality). See text for full details.

Back to Robustness

Size robustness to imbalance groups

Table 2: Rejection Rates for Tests of Nominal 5% Size with Placebo Treatments in Log-Earnings Data with 10 Groups.

	G1 = 5	G1 = 4	G1 = 3	G1 = 2
$\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.039	0.046	0.071	0.137
	(0.003)	(0.003)	(0.004)	(0.005)
Wild cluster bootstrap-t	0.044	0.052	0.053	0.024
	(0.003)	(0.003)	(0.003)	(0.002)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample groups (US states) from CPS-MORG data, having imposed the sampling restrictions described in the text. The treatment parameter has a true coefficient of zero. G1 denotes the number of treated groups. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The different inference methods used are discussed in the text.

Back to Robustness

Power MDE

Minimum Detectable Effects (MDE)

- Smallest effect that would lead to a rejection of the null hypothesis of no effect with given probabilities

$$MDE\left(\underbrace{x}_{\text{Power}}\right) = \hat{se}_{clu}\left(\hat{\beta}\right)\left[\underbrace{c_u}_{\substack{\text{Upper critical} \\ \text{value } t_{G-1}}} - \underbrace{p^t_{1-x}}_{\substack{\text{(1-x)th percentile of } t_{G-1} \\ \text{under null of no effect}}}\right] \tag{4}$$

Back to Power

Power MDE

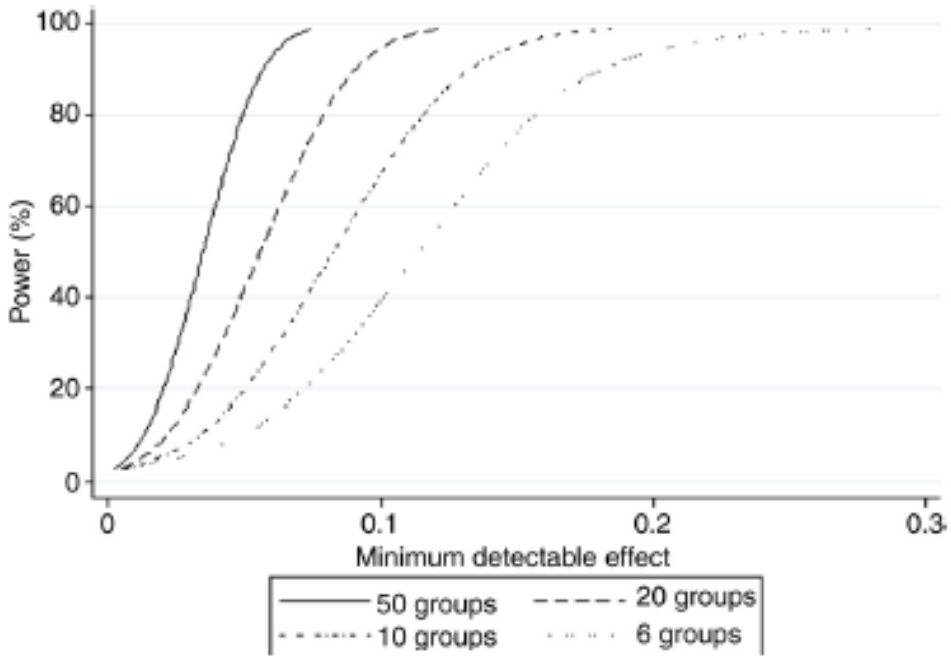

Figure 1: Minimum Detectable Effects on Log-Earnings Using $\sqrt{G/(G-1)}$-CRSEs and t_{G-1} Critical Values and Tests of Size 0.05. The figure shows the proportion of the time that the null hypothesis of no treatment effect is rejected when the treatment parameter has a true coefficient ranging from 0 to 0.3. Numbers are computed using the results of 100,000 Monte Carlo simulations combined with equation 4, as described in the text. The simulations resample states from CPS-MORG data, having imposed the sampling restrictions described in the text. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data.

Back to Power

Power MDE with FGLS

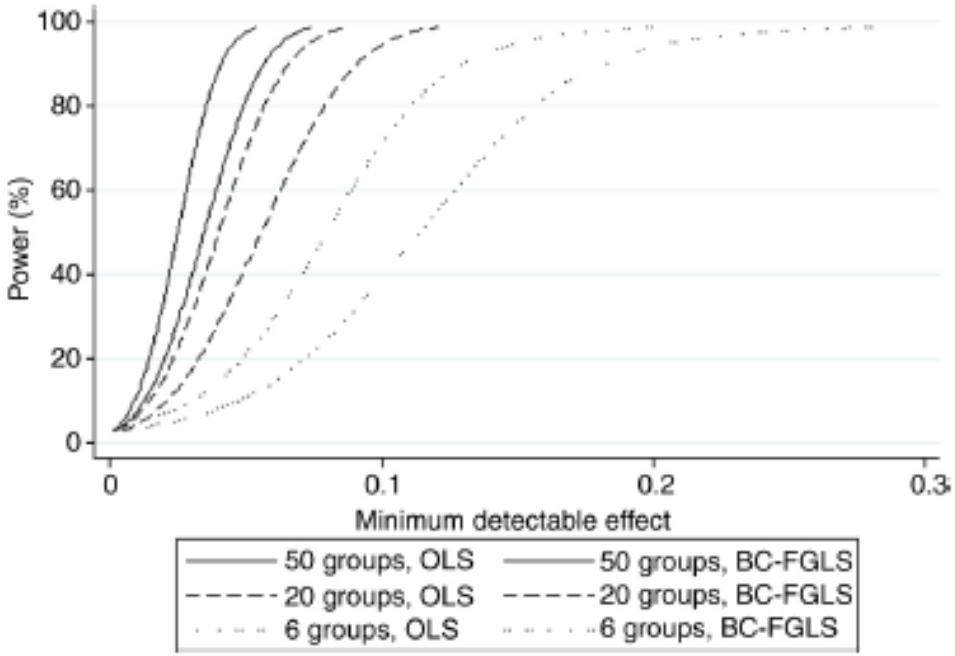

Figure 2: Minimum Detectable Effects on Log-Earnings Using $\sqrt{G/(G-1)}$-CRSEs and t_{G-1} Critical Values and Tests of Size 0.05. The figure shows the proportion of the time that the null hypothesis of no treatment effect is rejected when the treatment parameter has a true coefficient ranging from 0 to 0.3. Numbers are computed using the results of 100,000 Monte Carlo simulations combined with equation 4, as described in the text. The simulations resample groups (US states) from CPS-MORG data, having imposed the sampling restrictions described in the text. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Estimation of the treatment effect is conducted on aggregated state-year data either by OLS (reproducing part of Figure 1), or by feasible GLS assuming a AR(2) error process (homogeneous across states) and using bias-corrected AR parameter estimates as in Hansen (2007) (denoted "BC-FGLS"). See text for full details.

Back to Power FGLS

Power robustness to error specification

Table 8: Rejection Rates with 5%-Level Tests and Treatment Effects of Zero and +0.05 in Simulated Log-Earnings Data (10 Groups).

	Heterogenous AR(2)		MA(1)	
	effect = 0	**effect = 0.05**	**effect = 0**	**effect = 0.05**
OLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.056	0.520	0.056	0.619
	(0.003)	(0.007)	(0.003)	(0.007)
FGLS	0.106	0.782	0.088	0.713
	(0.004)	(0.006)	(0.004)	(0.006)
FGLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.065	0.712	0.061	0.606
	(0.003)	(0.006)	(0.003)	(0.007)
BC-FGLS	0.069	0.807	0.072	0.710
	(0.004)	(0.006)	(0.004)	(0.006)
BC-FGLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.063	0.727	0.059	0.603
	(0.003)	(0.006)	(0.003)	(0.007)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample states from CPS-MORG data, having imposed the sampling restrictions described in the text. Data from 1979 to 2008 inclusive are sampled (i.e. T = 30). Regressions are run on aggregated state-year data. The underlying data effectively replaces empirical CPS regression residuals with a simulated error term, generated according to an AR(2) process which varies across groups and an MA(1) process respectively. The inference methods used are discussed in the text. We adjust for test size when making power comparisons using the procedure outlined by Davidson and MacKinnon (1998). See text for details.

Back to Robustness

Power robustness to number of time periods

Table 9: Rejection Rates with 5%-Level Tests and Treatment Effects of Zero and +0.05 in Log-Earnings Data (10 Groups).

	T = 30		T = 20		T = 10	
	effect = 0	**effect = 0.05**	**effect = 0**	**effect = 0.05**	**effect = 0**	**effect = 0.05**
OLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.039	0.338	0.034	0.380	0.037	0.447
	(0.003)	(0.007)	(0.003)	(0.007)	(0.003)	(0.007)
FGLS	0.101	0.449	0.107	0.436	0.091	0.415
	(0.004)	(0.007)	(0.004)	(0.007)	(0.004)	(0.007)
FGLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.044	0.484	0.039	0.478	0.040	0.479
	(0.003)	(0.007)	(0.003)	(0.007)	(0.003)	(0.007)
BC-FGLS	0.078	0.453	0.078	0.431	0.078	0.413
	(0.004)	(0.007)	(0.004)	(0.007)	(0.004)	(0.007)
BC-FGLS, $\sqrt{G/(G-1)}$-CRSEs, t(G − 1) critical values	0.049	0.479	0.043	0.481	0.038	0.486
	(0.003)	(0.007)	(0.003)	(0.007)	(0.003)	(0.007)

The table shows the proportion of the time that the null hypothesis of no treatment effect was rejected in 5000 Monte Carlo simulations. Simulation standard errors are reported in parentheses. The simulations resample states from CPS-MORG data, having imposed the sampling restrictions described in the text. T is the number of (consecutive) time periods (years). The first year of data is chosen from a uniform distribution between 1979 and (2009-T) in each Monte Carlo simulation. Regressions are run on aggregated state-year data. The inference methods used are discussed in the text. We adjust for test size when making power comparisons using the procedure outlined by Davidson and MacKinnon (1998). See text for details.

Back to Robustness