Corrected Standard Errors with Clustered Data: Online Appendices

John E. Jackson*

A Monte Carlo Numerical Results

A.1 Homogenous Clusters[a]

		AMSE		Rejection Rates	
# Clusters	r_v^b	CRSE	$CESE_2$	CRSE	$CESE_2$
12	0.1	-0.0406	0.0331	13.3	4.35
	0.5	-0.0572	0.0236	12.6	3.81
24	0.1	-0.054	0.0165	11.6	5.03
	0.5	-0.058	0.0114	11.3	4.86
48	0.1	-0.025	0.007	8.64	4.85
	0.5	-0.036	0.003	8.59	4.81
72	0.1	-0.016	0.005	7.05	5.23
	0.5	-0.02	0.002	7.36	5.24
96	0.1	-0.019	-0.006	6.81	5.29
	0.5	-0.024	-0.011	6.52	5.65

a. Explanatory variables and stochastic terms drawn from identical distributions.
b. r_v is the expected within cluster stochastic term correlation.

*Dept. of Political Science, University of Michigan. email: jjacksn@umich.edu.

A.2 Heterogeneous Explanatory Variables

		Average Mean Standardized Error			
		Homo., Normal		Hetero., Expon.	
G	covariance clustering[a]	CRSE	$CESE_2$	CRSE	$CESE_3$
12	0.0	-0.081	0.016	-0.120	0.008
	0.3	-0.210	0.003	-0.253	-0.028
	0.6	-0.312	-0.013	-0.353	-0.057
	0.9	-0.396	-0.030	-0.437	-0.087
24	0.0	-0.064	0.012	-0.092	0.007
	0.3	-0.084	-0.001	-0.120	-0.017
	0.6	-0.179	-0.005	-0.218	-0.036
	0.9	-0.270	-0.009	-0.309	-0.049
48	0.0	-0.026	0.005	-0.051	-0.004
	0.3	-0.055	0.006	-0.090	-0.014
	0.6	-0.100	0.006	-0.145	-0.026
	0.9	-0.134	0.001	-0.180	-0.034
72	0.0	-0.015	0.012	-0.042	-0.003
	0.3	-0.050	-0.001	-0.070	-0.008
	0.6	-0.073	-0.002	-0.102	-0.017
	0.9	-0.090	0.000	-0.123	-0.019

A.2 Heterogeneous Explanatory Variables, Cont.

		Rejection Rates, $\alpha = 0.05$			
		Homo., Normal		Hetero., Expon	
G	covariance clustering[b]	CRSE	CESE_2	CRSE	CESE_3
12	0.0	15.00	4.80	12.78	4.14
	0.3	28.16	5.88	26.24	6.88
	0.6	43.00	8.01	38.38	9.49
	0.9	50.70	11.14	46.63	12.88
24	0.0	11.90	4.64	10.36	4.46
	0.3	11.50	4.69	9.53	4.98
	0.6	18.96	5.85	15.95	6.71
	0.9	29.41	7.31	25.12	8.72
48	0.0	7.49	4.73	6.87	5.07
	0.3	9.75	4.98	8.50	5.43
	0.6	13.11	5.52	12.07	6.07
	0.9	15.88	6.17	13.66	6.90
72	0.0	7.10	4.50	6.55	4.68
	0.3	8.57	5.31	7.37	5.34
	0.6	10.04	5.32	8.59	5.76
	0.9	12.00	5.53	9.53	6.25

a. Average between variance/total variance.

A.3 CESE with Different Stochastic Term Distributions

		Average Mean Standardized Error					
		Homoskedastic			Heteroskedastic		
G	covariance clustering[b]	Normal	χ^2	Expon	Normal	χ^2	Expon
12	0.0	0.016	0.040	-0.008	0.029	0.034	0.008
	0.3	0.002	-0.017	-0.030	0.009	0.003	-0.028
	0.6	-0.013	-0.038	-0.056	-0.012	-0.028	-0.057
	0.9	-0.030	-0.059	-0.080	-0.036	-0.062	-0.087
24	0.0	0.012	0.005	0.001	0.020	0.010	0.007
	0.3	-0.001	-0.006	-0.008	0.019	-0.005	-0.017
	0.6	-0.005	-0.013	-0.019	0.004	-0.014	-0.036
	0.9	-0.009	-0.019	-0.032	-0.012	-0.026	-0.049
48	0.0	0.005	0.001	0.001	0.012	0.007	-0.004
	0.3	0.006	-0.005	-0.006	0.000	-0.013	-0.014
	0.6	0.006	-0.006	-0.011	-0.008	-0.017	-0.026
	0.9	0.001	-0.006	-0.015	-0.012	-0.021	-0.034
72	0.0	0.012	-0.004	-0.006	0.007	0.004	-0.003
	0.3	-0.001	-0.007	-0.017	0.009	-0.004	-0.008
	0.6	-0.002	-0.009	-0.019	0.003	-0.006	-0.017
	0.9	0.000	-0.009	-0.019	0.001	-0.006	-0.019

A.3 CESE with Different Stochastic Term Distributions, Cont.

		Rejection Rate, $\alpha = 0.05$					
		Homoskedastic			Heteroskedastic		
G	covariance clustering[a]	Normal	χ^2	Expon	Normal	χ^2	Expon
12	0.0	4.80	4.77	4.97	4.32	3.94	4.14
	0.3	5.88	6.50	7.11	6.10	6.25	6.88
	0.6	8.01	9.17	9.96	8.96	9.18	9.49
	0.9	11.14	12.87	13.67	12.54	12.74	12.88
24	0.0	4.64	4.38	4.88	4.32	4.39	4.46
	0.3	4.69	5.31	5.31	4.96	5.19	4.98
	0.6	5.85	6.38	6.79	6.27	6.74	6.71
	0.9	7.31	8.16	8.53	8.16	8.68	8.72
48	0.0	4.73	4.61	5.09	5.05	4.77	5.07
	0.3	4.98	5.15	4.99	5.24	5.24	5.43
	0.6	5.52	5.72	5.79	5.92	5.79	6.07
	0.9	6.17	6.32	6.79	6.36	6.41	6.90
72	0.0	4.50	5.13	4.77	4.80	4.42	4.68
	0.3	5.31	5.27	5.17	4.44	4.80	5.34
	0.6	5.32	5.52	5.48	5.10	5.15	5.76
	0.9	5.53	5.92	5.99	5.65	5.74	6.25

a. Average between variance/total variance.

A.4 Error Term Interdependence w/ Heterogeneous Clusters, Homoskedastic Error Terms[a]

	r_v^b		
# Clusters	0.10	0.50	0.75
		AMSE	
12	0.001	-0.013	-0.018
24	0.004	-0.005	-0.007
48	0.010	0.006	0.003
72	-0.002	-0.002	-0.001
96	0.001	0.001	-0.000
		Rejection Rate	
12	6.76	8.01	8.65
24	5.53	5.85	5.96
48	5.30	5.52	5.48
72	5.56	5.32	5.31
96	5.31	5.19	5.07

a. Covariate clustering equals 0.6.
b. r_v is the expected within cluster error term correlation.

A.5 Bootstrapping

# Clusters	Scenario A	B	C	D	E
			AMSE		
12	-0.054	-0.068	-0.087	-0.107	-0.101
24	0.000	-0.033	-0.112	-0.115	-0.106
48	-0.022	-0.015	-0.073	-0.087	-0.085
72	-0.019	-0.008	-0.046	-0.061	-0.061
			Rejection Rates		
12	11.44	12.30	11.24	10.44	11.34
24	8.20	9.02	12.24	9.76	9.50
48	7.70	7.04	11.60	9.60	10.18
72	6.94	6.20	8.22	8.18	8.06

Scenarios:

A. Homogenous clusters, homoskedastic normal errors.
B. Scenario A with unequal # of observations per cluster.
C. Scenario B with covariate clustering equal 0.6.
D. Scenario C with heteroskedastic errors.
E. Scenario C with χ^2 distributed errors.

B Variations in # of Clusters and Sample Size

The simulations with heterogeneous clusters all use the same distribution of numbers of observations per cluster, e.g. five, ten and fifteen in equal proportions. A consequence of this design is that the total sample size increases as the number of clusters increases, with $N = \sum_{g=1}^{G} n_g = 10G$. Eq. 22 implies that CRSE performance varies with G but not with N, which is consistent with MacKinnon and Webb (2017, fn. 3, p. 237) and Esarey & Menger (2018). The latter attribute this condition to the fact that, "...the center summation [in eq. 6] happens over the number of clusters G and not over the number of observations N." If this statement is correct it should apple to the CESE as well, making all of the simulation results responsive to the changing numbers of clusters, but not to the fact that N is also increasing. This appendix reports some simulation results examining this prediction.

The examination expands the number of observations per cluster in the simulations with twelve clusters to create samples with $N = 480$. This means that clusters now have either twenty, forty or sixty observations. In the simulations the values for X in the additional observations are simply duplicated from the original sample to expand the size of the clusters. The advantage of this method is that it guarantees that the distributions of the explanatory variables both within and across clusters are identical to those in the first simulations.[1]

The simulations are done with homoskedastically normally distributed stochastic terms, as in the left panel of Fig. 2.[2] Fig. B.1 and Table B.1 show the results for the original and new simulations with both the CRSE and CESE methods. The results of the simulations with $G = 48, N = 480$ are also shown. The results show little difference in the simulations with N = 120 and N = 480. (The results for $N = 120$ and $G = 48$ are taken from Table A.2, column 2.) The performance with N = 480 when $G = 12$ for both methods is actually slightly worse. The average

[1] Simulations were also done with a new random sample of values for X drawn from the same distributions as the original X. The simulation results are identical to the third significant digit.

[2] As in figure 2 the CESE results use the hc_2 adjustment.

mean standardized errors (amse) are about -0.03 more negative for the CRSE and -0.02 more negative for the CESE while the rejection rates average three percent higher for the CRSE and one percent higher for the CESE method. The performance plots with $G = 48$ but the same number of observations, $N = 480$ demonstrate increasing the numbers of clusters substantially improves the methods' performance, particularly for the CRSE.

Collectively the evidence is consistent with the initial propositions that it is the number of clusters, G, not the total sample size, N, that affects the performance of the methods. This conclusion means that the evidence from all the simulations varying the number of clusters is credible even though the total sample size is increasing as the number of clusters increases.

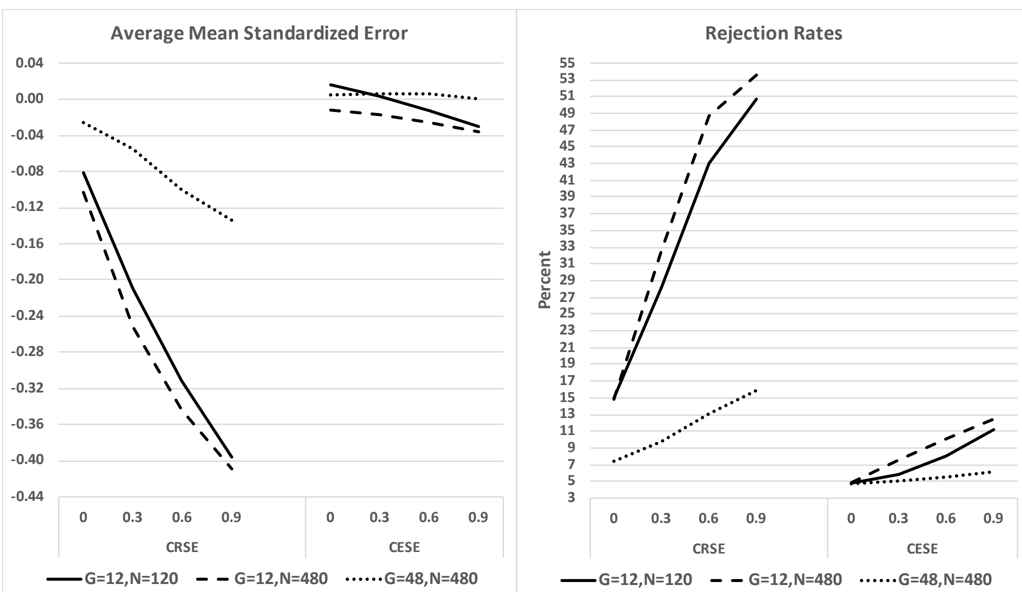

Fig. B.1: Comparing Different Sample Sizes with G = 12

Covariate	*AMSE*		Rej. Rate	
Clustering	CRSE	CESE	CRSE	CESE
0.0	-0.103	-0.012	14.84	4.86
0.3	-0.252	-0.017	32.59	7.59
0.6	-0.343	-0.026	48.60	10.04
0.9	-0.409	-0.036	53.57	12.39

Table B.1: CRSE & CESE with $G = 12$ & $N = 480$

C Standard Errors with Clustered Data: An Alternative Estimator

This appendix presents and evaluates an alternative way of estimating the cluster error term variance-covariance matrix. This method is from Baltagi and Chang (1994) based on a paper by Swamy and Arora (1972), denoted as the BCSE method. Their method models V_{g_i} as a cluster specific random effect plus an observation specific stochastic term, $V_{g_i} = U_g + \varepsilon_{g_i}$. Their estimate for Σ_g assumes homogeneity for all stochastic terms, $\sigma^2_{v_{g_i}} = \sigma^2_v = \sigma^2_u + \sigma^2_\varepsilon$ and $\rho_{g_{ij}} = \rho = \sigma^2_u$, for all g, i and j.

C.1 Baltagi-Chang Method for Estimating Σ_g

The BCSE method uses residuals from a combination of within and between group OLS regressions to estimate σ^2_u and σ^2_ε. The residuals from a regression of group mean centered variables, the within regression, which eliminates U_g estimates σ^2_ε. Write this equation as,

$$y_{g_i} = Y_{g_i} - \bar{Y}_g = X_{g_i}\beta + U_g + \varepsilon_{g_i} - (\bar{X}_g\beta + U_g + \bar{\varepsilon}_g) = x^*_{g_i}\beta^* + (\varepsilon_{g_i} - \bar{\varepsilon}_g), \tag{1}$$

where $x^*_{g_i}$ excludes variables that do not vary within clusters, such as a country having a parliamentary or presidential system. This regression gives unbiased and consistent estimates for β^*.[3] Denote the residuals from estimating eq. 1 as $\tilde{e}$. Baltagi and Chang (1994) propose the following estimate for σ^2_ε,

$$\hat{\sigma}^2_\varepsilon = \frac{\sum_{g=1}^{G}\sum_{i=1}^{n_g}\tilde{e}^2_{g_i}}{(N-G-K^*)} = \frac{tr(\tilde{e}\tilde{e}')}{(N-G-K^*)}, \tag{2}$$

where K^* is the number of variables in x^*_{gi}.

In their second step Baltagi and Chang (1994) use the residuals from a weighted between groups regression $\bar{Y}_g = \bar{X}_g\bar{\beta} + U_g + \bar{\varepsilon}_g$ to estimate σ^2_u. The weights are the number of observations

[3] These estimates for β^* are not efficient because the presence of $\bar{\varepsilon}_g$ in the error term for each observation means the stochastic terms in the within regression are not independent within clusters as $E[(\varepsilon_{gi} - \bar{\varepsilon}_g)(\varepsilon_{gj} - \bar{\varepsilon}_g)_{i\neq j} = -\sigma^2_\varepsilon/n_g \neq 0$. Also, with unbalanced panels where n_g varies by group the stochastic terms are heteroskedastic even when σ^2_ε is identical for all clusters.

within each group, n_g, so that $\bar{b} = \left[\sum_{g=1}^{G}(n_g \bar{X}_g^{'} \bar{X}_g)\right]^{-1} \left[\sum_{g=1}^{G}(n_g \bar{X}_g^{'} \bar{Y}_g)\right]$. Their estimate for σ_u^2 uses the following notation; ι_g is a n_g x 1 column vector of ones, $Z_g = \iota_g \iota_g^{'}$, and $W_g = (1/n_g)Z_g$. Create the block diagonal matrices Z and W where the blocks are Z_g and W_g respectively. The weighted sum of squared residuals from the between regression is $e^{*^{'}} We^*$, where $e^* = \bar{Y} - \bar{X}\bar{b}$. From the expected value of $e^{*^{'}} We^*$ the estimate for σ_u^2 is,

$$\hat{\sigma}_u^2 = \frac{e^{*^{'}} We^* - (G-K)\hat{\sigma}_\varepsilon^2}{N - tr[(X'WX)^{-1}X'ZZ'X]}. \tag{3}$$

The terms calculated in eqs. 2 and 3 are then used to form $\hat{\Sigma}_v$ to estimate the coefficient standard errors,

$$\hat{\Sigma}_b = S_b = (X^{'}X)^{-1} \left[\sum_{g=1}^{G}(X_g^{'}\hat{\Sigma}_g X_g)\right] (X^{'}X)^{-1}. \tag{4}$$

C.2 Monte Carlo Experiments Comparing BCSE and CESE

Fig. C1 and Table C1 compare the BCSE and CESE performance for the Monte Carlo simulations shown in Fig. 2 for homogeneously normally and heteroskedastically exponentially distributed errors. These plots show that in most of the experiments CESE perform better than do BCSE. The differences favoring CESE increase strongly as the number of clusters and the amount of covariate clustering decrease. For example with twelve clusters and no to moderate covariate clustering the BCSE *amse* is -0.02 to -0.03 more negative than is the CESE *amse* and the BCSE rejection rates are two to three percent higher. With forty-eight clusters and moderate to high covariate clustering the *amse* differences are -0.015 or less and the differences in rejection rates range from -0.4 to 0.7 percent. With seventy two clusters CESE is slightly better with no covariate clustering. With moderate to high covariate clustering there is virtually no difference in performance. With a larger number of clusters and a high degree of covariate clustering the choice is largely one of computational convenience.

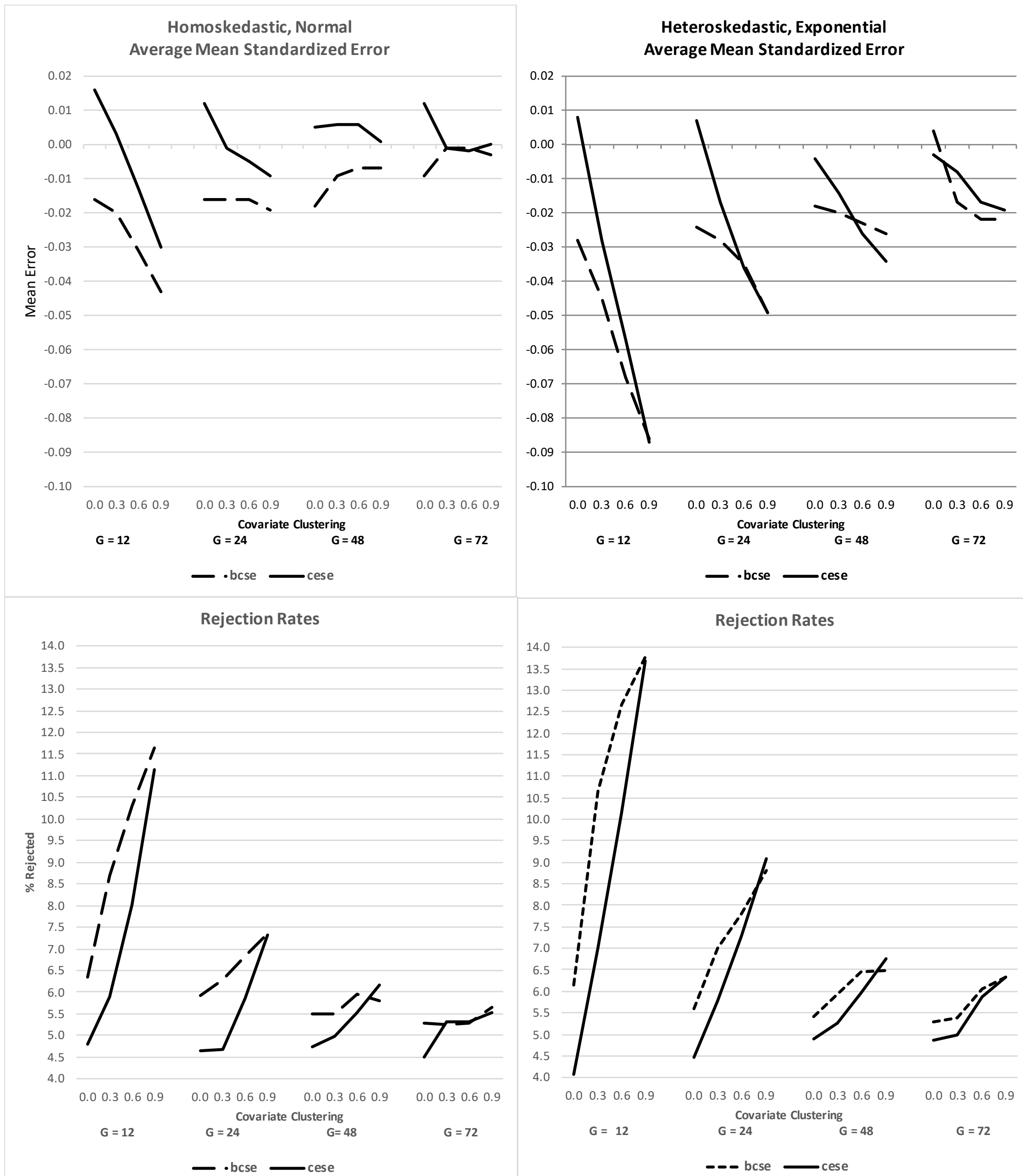

Fig. C1: BCSE and CESE Estimator Performance

G	covariance clustering[a]	AMSE				Rejection Rates, $\alpha = 0.05$			
		Homo, BCSE	Normal $CESE_2$	Hetero, BCSE	Expon $CESE_3$	Homo, BCSE	Normal $CESE_2$	Hetero, BCSE	Expon $CESE_3$
12	0.0	-0.016	0.016	-0.028	0.008	6.34	4.80	6.04	4.14
	0.3	-0.020	0.003	-0.045	-0.028	8.68	5.88	10.42	6.88
	0.6	-0.031	-0.013	-0.068	-0.057	10.30	8.01	12.63	9.49
	0.9	-0.043	-0.030	-0.086	-0.087	11.64	11.14	14.07	12.88
24	0.0	-0.016	0.012	-0.024	0.007	5.92	4.64	5.99	4.46
	0.3	-0.016	-0.001	-0.028	-0.017	6.30	4.69	6.99	4.98
	0.6	-0.016	-0.005	-0.035	-0.036	6.84	5.85	7.85	6.71
	0.9	-0.019	-0.009	-0.049	-0.049	7.35	7.31	8.23	8.72
48	0.0	-0.018	0.005	-0.018	-0.004	5.49	4.73	5.17	5.07
	0.3	-0.009	0.006	-0.020	-0.014	5.49	4.98	6.03	5.43
	0.6	-0.007	0.006	-0.023	-0.026	5.95	5.52	6.49	6.07
	0.9	-0.007	0.001	-0.026	-0.034	5.79	6.17	6.35	6.90
72	0.0	-0.009	0.012	0.004	-0.003	5.28	4.50	5.41	4.68
	0.3	-0.001	-0.001	-0.017	-0.008	5.25	5.31	5.58	5.34
	0.6	-0.001	-0.002	-0.022	-0.017	5.29	5.32	5.89	5.76
	0.9	-0.003	0.000	-0.022	-0.019	5.64	5.53	5.74	6.25

a. Average between variance/total variance.

Table C1: BCSE and CESE Comparisons

D Bootstrapping Variations with # of Replications and Random Seed

Replicating the Brown, et al. and Harden equations illustrates that bootstrapping results are sensitive to the random number seed and the number of replications. The initial replication used the seed 441022 used to generate the explanatory variables in the simulations and 1000 replications as used by Harden. Three additional alternative seeds are used and the number of replications extended to 50,000.[4] Table D.1 reports the p-values testing the null hypothesis that the Liberal Control coefficient is zero. These values range from 0.037 to 0.056 with 1,000 replications. Even with 10,000 replications the p-values range from 0.048 to 0.053. Only with 50,000 replications do the p-values become consistent, ranging from 0.050 to 0.052. There are two important lessons here. The first is that in practice bootstrap users should be careful to use a large number of replications, or at least explore the sensitivity of the results to the number of replications and starting seed. These results also point out the fragility of relying on preset confidence intervals and p-values, as done using coefficient and marginal effects plots, to report results. This procedure is both sensitive to arbitrary choices made during the analysis and susceptible to manipulation to get desired results. As Wasserstein, et al. (2019) argue, better to present estimated standard errors and if necessary the continuous p-values derived from those standard errors and distribution assumptions.

	Random Number Seed			
Reps	441022	8623	813	13579
1,000	0.037	0.056	0.046	0.040
2,000	0.043	0.054	0.053	0.042
4,000	0.046	0.048	0.053	0.049
5,000	0.049	0.050	0.054	0.048
10,000	0.053	0.049	0.053	0.048
50,000	0.052	0.052	0.052	0.050

Table D.1: Liberal Control Coefficient p-Values

[4] The original seed, 441022, is the last six digits of the number of shares traded on the NYSE on August 9, 2016; the seed 8623 is the last four digits of the U. S. national debt on Feb. 7, 2006; 813 are my favorite numbers; and 13579 are obviously the series of odd digits.

References

Baltagi, B. H. and Y-J. Chang(1994). "Incomplete Panels: A Comparative Study of Alternative Estimators for the Unbalanced One-Way Error Component Regression Model." *Journal of Econometrics*. 62: 67 - 89.

Brown, Robert, D., Robert A. Jackson, & Gerald C. Wright (1999). "Registration, Turnout, and State Party Systems." *Political Research Quarterly*. 52(3): 463 - 479.

Esarey, Justin & Andrew Menger (2018). "Practical and Effective Approaches to Dealing with Clustered Data." *Political Science Research & Methods*. Jan., 2018. https://doi.org/10.1017/psrm.2017.42.

Harden, Jeffrey, J. (2011). "A Bootstrap Method for Conducting Statistical Inference with Clustered Data." *State Politics & Policy Quarterly*. 11(2): 223 - 246.

MacKinnon, J. G. & M. D. Webb (2017) "Wild Bootstrap Inference for Wildly Different Cluster Sizes." *Journal of Applied Econometrics*. 32(2): 233 - 254.

Swamy, P. A. V. B. and S. S. Arora (1972). The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models. *Econometrica*. 40: 261 - 275.

Wasserstein, Ronald, L., Allen L. Schirm and Nicole A Lazar (2019) "Moving to a World Beyond '$p < 0.05$' " *The American Statistician*, 73: sup1, 1 - 19, DOI: 10.1080/00031305.2019.1583913