Cluster Robust Inference in Linear Models with Many Covariates Supplemental Appendix: Simulation and Empirical Application

Matias D. Cattaneo* Aibo Gong† Michael Jansson‡ Whitney K. Newey§

September 20, 2022

Abstract

This appendix summarizes the simulation designs and empirical applications used for the project regarding cluster robust inference in linear models with many covariates. It provides the data generating process and results for different setups and applications. It also provides an empirical application to compare the performances of multiple standard errors.

*Department of Operations Research and Financial Engineering, Princeton University.
†School of Economics, Peking University
‡Department of Economics, UC Berkeley and *CREATES*.
§Department of Economics, MIT

1 Simulation

1.1 Setup

We conducted a simulation study to assess the finite sample properties of the standard errors we proposed here and compared them with other cluster robust standard errors available in the literature. We considered a linear regression model with growing dimension. In this model, we examine the performance of a series of standard errors including the well known Liang-Zeger estimator for clusters with and without the degree of freedom adjustment (LZ and LZ-df). These two estimators degenerate to HC0 and HC1 standard errors when each cluster only contains one unit. For comparison purposes, we also included the 'bias-reduction' standard error (BR) and the 'jackknife' standard error (JK) that are clustered versions of the HC2 and HC3 standard errors. We also contained the cluster robust standard error (CR) we proposed here and the leave-out standard error (LO) we discussed. For all standard errors that are not guaranteed to be positive semi-definite in the finite sample, we use the transformation $(A'B + B'A)/2$ to adjust.

Our paper presents theory for Gaussian-based inference methods and report both empirical coverage and the average interval length. The latter provides a summary of efficiency for each inference method.

The Gaussian-based confidence interval takes the form

$$I_l = \left[\hat{\beta} - \Phi^{-1}(1-\alpha/2)\sqrt{\frac{\hat{\Omega}_l}{n}}, \hat{\beta} - \Phi^{-1}(\alpha/2)\sqrt{\frac{\hat{\Omega}_l}{n}}\right] \qquad \hat{\Omega}_l = \hat{\Gamma}^{-1}\hat{\Sigma}_l\hat{\Gamma}^{-1}$$

where Φ^{-1} denotes inverse of the standard normal cumulative distribution function and $\hat{\Sigma}_l$ with $l \in \{\text{LZ,LZ-df,BR,JK,LO,CR}\}$ represents the corresponding variance estimator discussed in the paper.

The data generating process for the linear regression model with many covariates follows

Cattaneo, Jansson, and Newey (2018). For simplicity, we dropped the subindex n here.

$$\begin{aligned} y_i &\sim \beta x_i + \boldsymbol{\gamma}_{\boldsymbol{n}}' \mathbf{w}_i + u_i, \qquad i = 1, \ldots, n, \\ x_i | \mathbf{w}_i &\sim \mathcal{N}(0, \sigma^2_{x,i}) \qquad \sigma^2_{x,i} = \varkappa_x (1 + (\boldsymbol{\iota}' \mathbf{w}_i)^2) \end{aligned}$$

The dimension of the parameter of interest β is d and the dimension of the covariates $\mathbf{w}_i$ is K. The data generating process for the error term u_i follows MacKinnon, Nielsen, and Webb (2020). For each cluster of the error term, we define

$$\mathbf{u}_n(g) = (u_{t_{g,n}(1)}, \ldots, u_{t_{g,n}(\#\mathcal{T}_{g,n})})', \qquad g = 1, 2, \ldots, N_{\mathcal{T},n}$$

which follows

$$\mathbf{u}_n(g) = \mathbf{P}_\xi \boldsymbol{\xi}_n(g) + p_\varepsilon \boldsymbol{\varepsilon}_n(g), \qquad \boldsymbol{\varepsilon}_n(g) \sim \mathcal{N}\left(\mathbf{0}, \mathrm{diag}\{\sigma^2_{u,t_{g,n}(1)}, \ldots, \sigma^2_{u,t_{g,n}(\#\mathcal{T}_{g,n})}\}\right)$$

$\boldsymbol{\xi}_n(g) = [\xi_{g1,n}, \ldots, \xi_{gJ,n}]$ is a J dimension vector with unobserved random factors. The $\#\mathcal{T}_{g,n} \times J$ loading matrix $\mathbf{P}_\xi$'s (i,j)-th entry is $p_\xi \mathbb{I}\left(j = \lfloor (i-1)J/\#\mathcal{T}_{g,n} \rfloor + 1\right)$ where $\lfloor . \rfloor$ denotes the integer part of the argument. When $J = 1$, entries of the loading matrix $\mathbf{P}_\xi$ are ones and the error term degenerates to the commonly known random effect model. In practice, if one would like to pick $\#\mathcal{T}_{g,n}$ as a multiple of J then $\#\mathcal{T}_{g,n}/J$ observations in each cluster depend on each of the J unobserved vectors. In order to avoid the situation where cluster fixed effect fully captured the correlation within each cluster, we required that the J unobserved random factors are also correlated with each other following

$$\xi_{g1,n} \sim \mathcal{N}(0,1), \qquad \xi_{gj,n} = \rho \xi_{gj-1,n} + e_{gj,n} \qquad e_{gj,n} \sim \mathcal{N}(0, 1-\rho^2), \qquad j = 2, \ldots, J$$

$\boldsymbol{\varepsilon}_n(g)$ is a noise term with independent normal distribution and conditional heterogeneous variance where

$$\sigma^2_{u,i} = \varkappa_u (t(x_i) + \boldsymbol{\iota}' \mathbf{w}_i)^2$$

$\varkappa_x$ and $\varkappa_u$ are constants that normalize the variance of covariates and error terms.

We also allowed the covariates to be correlated with each other following a cluster design. Further, the cluster assignment is allowed to be different from that of the error term. If the regressors are designed to be independent, we picked covariates following $\mathbf{w_i} \overset{\text{i.i.d}}{\sim} \mathbf{U}[-1,1]^K$. If the covariates are designed to be correlated, we picked a data generating process similar to the one we used to get the error terms. The $1 \le k \le K$-th regressor in cluster G,

$$\mathbf{w}_n^k(G) = (w^k_{t_{G,n}(1)}, \dots, w^k_{t_{G,n}(\#\mathcal{S}_{G,n})})', \qquad G = 1, 2, \dots, N_{\mathcal{S},n}$$

has the form

$$\mathbf{w}_n^k(G) = \mathbf{L}_\lambda \boldsymbol{\lambda}_n^k(G) + l_\epsilon \boldsymbol{\epsilon}_n^k(G), \qquad \boldsymbol{\epsilon}_i^k \overset{\text{i.i.d}}{\sim} \mathbf{U}[-1,1]$$

$\boldsymbol{\lambda}_n^k(G) = [\lambda^k_{G1,n}, \dots, \lambda^k_{GJ,n}]$ is a J dimension vector with unobserved random factors. The $\#\mathcal{S}_{G,n} \times J$ loading matrix $\mathbf{L}_\lambda$'s (i,j)-th entry is $l_\lambda \mathbb{I}(j = \lfloor (i-1)J/\#\mathcal{S}_{G,n} \rfloor + 1)$. The J unobserved random factors are also correlated with each other following

$$\lambda_{G1,n} \sim \mathbf{U}[-1,1], \qquad \lambda_{Gj,n} = \rho\lambda_{Gj-1,n} + \tilde{e}_{Gj,n} \qquad \tilde{e}_{Gj,n} \sim \sqrt{1-\rho^2}\mathbf{U}[-1,1], \qquad j = 2, \dots, J$$

Further, for other parameters we have that $\boldsymbol{\iota} = (1, 1, \dots, 1)'$, $d = 1$, $\beta = 1$, $\boldsymbol{\gamma_n} = \mathbf{0}$, $t(a) = a\mathbb{I}(-2 \le a \le 2) + 2sgn(a)(1 - \mathbb{I}(-2 \le a \le 2))$. $\varkappa_x$ and $\varkappa_u$ are normalization constants that make $\mathbb{V}[x_{t_{G,n(1)}}] = \mathbb{V}[u_{t_{g,n(1)}}] = 1$. We also pick $\rho = 0.5$, $p_\xi = l_\lambda = 0.7$ and $p_\varepsilon = l_\epsilon = \sqrt{1 - p_\xi^2}$. Regarding the dimension of the random factors, we pick $J = 2$ for error terms and $J = 8$ for covariates.

Regarding the cluster designs, we would like to consider both homogeneous and heterogeneous cluster sizes cases. Under homogeneous design, all the clusters have the same cluster size $\#\mathcal{T}_{g,n} = 4$ while for the heterogeneous case half of the clusters have a cluster size of 3 and half of them have a cluster size of 5. The regressors are allowed to be either independent or follow the correlated data generation process mentioned above with $\#\mathcal{S}_{G,n} = 15$. The sample size $n = 600$ while the number of simulation $s = 1000$. These designs would provide us with four different models by choosing between homogeneous and heterogeneous cluster size as well as independent and dependent regressors. We compare the performance of these standard errors as we change the dimension of the covariates K. We pick K as $\{1, 26, 51, 76, 101, 151, 201, 251\}$ where in all cases the first covariate is an intercept. From the theory above we know that when $\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n = O(1)$, the classical standard errors may fall into problems and this value is related to the number of regressors so we also report the average value of this parameter as well as the covariates dimension K for each design.

1.2 Results

In the first column, we report the average value of $\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$, which is the key parameter that represents the extent how we deviate from the case of low dimension regressors. In practice, one thing worth mentioning is that the leave one out estimator has a low but not zero probability of being negative so we implement an adjustment. Recall that

$$\hat{\Sigma}_n^{LO} = \frac{1}{n}\sum_{g=1}^{\mathcal{N}_{\mathcal{T},n}} \hat{\mathbf{V}}_n(g)'\mathbf{y}_n(g)\mathbf{M}_n(g,g)^{-1}\hat{\mathbf{u}}(g)\hat{\mathbf{V}}_n(g)$$

For those have $\hat{\Sigma}_n < 0$, we build $\tilde{\Sigma}_n^{LO}(g) = \max\{\hat{\mathbf{V}}_n(g)'\mathbf{y}_n(g)\mathbf{M}_n(g,g)^{-1}\hat{\mathbf{u}}(g)\hat{\mathbf{V}}_n(g), 0\}$ and then calculate $\hat{\Sigma}_n = \sum_g \tilde{\Sigma}_n(g)$. In the table, we also report this negative ratio and we can see the ratio is really low and won't have a great effect on the results. For each standard error, we report corresponding empirical coverage and interval lengths in four different models with dependent or independent regressors as well as homogeneous and heterogeneous cluster sizes. The results are shown below and we can see that as the dimension of regressors is increasing, the commonly used standard errors perform worse. Our proposed cluster robust standard error and the leave-out standard error have relatively consistent performance while the HC-3 standard error is becoming more and more conservative during this process.

Further, I would like to examine the performance of different standard errors when we include dummy variables. I use the setup that has heterogeneous cluster sizes and dependent regressors. Further I include dummy variables besides the regressors and end

Table 1: Simulation Results, Independent Regressors, n=600, S=1000
Homogeneous Cluster Size

	(a): Empirical Coverage							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.027	0.940	0.941	0.941	0.941	0.938	0.941	0.000
K= 26	1.094	0.933	0.937	0.936	0.942	0.924	0.938	0.000
K= 51	1.924	0.908	0.917	0.917	0.933	0.911	0.928	0.007
K= 76	2.720	0.910	0.929	0.929	0.944	0.926	0.938	0.002
K=101	3.488	0.904	0.936	0.936	0.957	0.941	0.950	0.001
K=151	4.990	0.885	0.938	0.937	0.968	0.921	0.947	0.005
K=201	6.444	0.855	0.921	0.919	0.968	0.932	0.936	0.005
K=251	7.854	0.825	0.922	0.920	0.972	0.937	0.945	0.001
	(b): Interval Length							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.027	0.198	0.199	0.199	0.199	0.199	0.199	0.000
K= 26	1.094	0.263	0.270	0.269	0.276	0.271	0.272	0.000
K= 51	1.924	0.246	0.258	0.257	0.269	0.259	0.263	0.007
K= 76	2.720	0.266	0.285	0.285	0.305	0.292	0.294	0.002
K=101	3.488	0.259	0.284	0.284	0.311	0.293	0.296	0.001
K=151	4.990	0.230	0.267	0.266	0.308	0.279	0.282	0.005
K=201	6.444	0.240	0.295	0.294	0.361	0.316	0.317	0.005
K=251	7.854	0.235	0.309	0.308	0.404	0.338	0.335	0.001

up with K as $\{1, 40, 80, 120, 160\}$. To examine the performance of standard errors when error terms have different distributions, we change the distribution of the error term to Asymmetric Distribution and Bimodal Distribution as defined in Cattaneo et al. (2018). The performance are presented in 5 and 6.

2 Empirical Application

In this section we consider two different applied papers that encountered with the problems of having nonignorable $\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$ with clustering structure. One has an experiment design while the other one considers a policy evaluation setup. For each empirical paper, we present different standard errors under the cluster setup.

The first paper is Ambuehl, Bernheim, and Ockenfels (2021) that studied experimentally when, why, and how people intervene in others' choices. They conduct an experiment where

Table 2: Simulation Results, Independent Regressors, n=600, S=1000
Heterogeneous Cluster Size

	(a): Empirical Coverage							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.042	0.957	0.957	0.957	0.957	0.954	0.957	0.000
K= 26	1.711	0.933	0.939	0.938	0.945	0.930	0.941	0.003
K= 51	3.015	0.924	0.939	0.938	0.952	0.926	0.944	0.001
K= 76	4.256	0.910	0.929	0.930	0.941	0.934	0.936	0.001
K=101	5.458	0.883	0.925	0.925	0.948	0.924	0.941	0.002
K=151	7.793	0.859	0.908	0.908	0.947	0.914	0.924	0.007
K=201	10.055	0.866	0.928	0.929	0.968	0.929	0.942	0.003
K=251	12.289	0.824	0.922	0.921	0.980	0.935	0.945	0.002
	(b): Interval Length							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.042	0.197	0.198	0.197	0.198	0.197	0.197	0.000
K= 26	1.711	0.260	0.266	0.266	0.272	0.268	0.269	0.003
K= 51	3.015	0.263	0.276	0.276	0.288	0.279	0.281	0.001
K= 76	4.256	0.246	0.264	0.263	0.282	0.270	0.271	0.001
K=101	5.458	0.248	0.273	0.272	0.299	0.283	0.284	0.002
K=151	7.793	0.231	0.268	0.267	0.309	0.282	0.284	0.007
K=201	10.055	0.231	0.284	0.283	0.347	0.303	0.305	0.003
K=251	12.289	0.224	0.295	0.293	0.385	0.321	0.320	0.002

subjects are divided into Choice Architects and Choosers. Choice Architects determine the set of options available to Choosers. Each choice option consists of an immediate payment and one in half a year. Impatience is costly in the sense that larger earlier payments are associated with smaller total payments. The main interest is in the Choice Architects as the authors would like to understand more about paternalism. They conducted 16 sessions with a total of 303 Choice Architects and 124 subjects participated as Choosers.

We would like to use their data to replicate column 1 in Table 3 and column 3 in Table 4 that use OLS methods. In the original paper, Table 3 tries to examine whether the Choice Architects believe the option set they constructed for Choosers are beneficial as a restriction on free choice is considered paternalistic only if people responsible for it believes that it is better for those affected. That is one important step in understanding the paternalism behavior in the authors' paper and we would like to see if introducing our methods here make any difference in the conclusion. The column 3 in Table 4 in the original paper answers

Table 3: Simulation Results, Dependent Regressors, n=600, S=1000
Homogeneous Cluster Size

	(a): Empirical Coverage							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.027	0.939	0.939	0.939	0.939	0.937	0.939	0.000
K= 26	1.257	0.926	0.936	0.937	0.942	0.926	0.940	0.004
K= 51	2.114	0.933	0.946	0.948	0.963	0.926	0.952	0.003
K= 76	2.907	0.910	0.928	0.929	0.951	0.920	0.937	0.003
K=101	3.689	0.892	0.932	0.933	0.956	0.915	0.940	0.002
K=151	5.199	0.880	0.928	0.929	0.966	0.944	0.944	0.001
K=201	6.684	0.837	0.915	0.913	0.969	0.922	0.929	0.001
K=251	8.131	0.805	0.912	0.908	0.979	0.927	0.931	0.002
	(b): Interval Length							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.027	0.192	0.193	0.192	0.193	0.192	0.192	0.000
K= 26	1.257	0.235	0.241	0.241	0.247	0.243	0.244	0.004
K= 51	2.114	0.258	0.270	0.270	0.284	0.274	0.276	0.003
K= 76	2.907	0.258	0.276	0.277	0.298	0.283	0.285	0.003
K=101	3.689	0.255	0.280	0.280	0.309	0.289	0.291	0.002
K=151	5.199	0.252	0.292	0.291	0.339	0.309	0.308	0.001
K=201	6.684	0.258	0.318	0.315	0.389	0.337	0.338	0.001
K=251	8.131	0.214	0.282	0.278	0.367	0.300	0.300	0.002

the question how introducing front-end delay, one method that reduces perceived benefits of paternalistic intervention, affects mandates and this helps us understand better about the motivation of paternalism. We are using the same datasetand only adjust standard errors used to conduct inference. The results are presented below in Table 7.

The second paper is Gratton, Guiso, Michelacci, and Morelli (2021) that found the political instability can cause the introduction of excessive and low-quality legislation, and thereby triggering a chain reaction eventually leading to a Kafkaesque economy. They used microdata on Italian MPs during the Second Republic to provide micro evidence for their mechanisms. Among them, we focus on the results got from Equation (9) and exhibited in Table 5 of that paper. Those results test whether the effects of a shorter political horizon on the pass of laws differ for competent and incompetent politicians, one main prediction from the authors' mechanism using an OLS regression with clusters that fall into our setup. We focus on the standard errors of the interaction coefficient. The results are also presented

Table 4: Simulation Results, Dependent Regressors, n=600, S=1000
Heterogeneous Cluster Size

	(a): Empirical Coverage							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.042	0.940	0.940	0.940	0.940	0.939	0.940	0.000
K= 26	1.970	0.935	0.938	0.939	0.944	0.935	0.941	0.002
K= 51	3.291	0.924	0.939	0.939	0.946	0.940	0.942	0.005
K= 76	4.546	0.901	0.917	0.917	0.939	0.918	0.925	0.002
K=101	5.762	0.900	0.924	0.926	0.956	0.932	0.938	0.003
K=151	8.121	0.877	0.925	0.926	0.968	0.935	0.938	0.005
K=201	10.434	0.855	0.926	0.924	0.969	0.934	0.941	0.003
K=251	12.697	0.810	0.908	0.904	0.977	0.922	0.932	0.002
	(b): Interval Length							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.042	0.193	0.194	0.193	0.194	0.194	0.194	0.000
K= 26	1.970	0.263	0.269	0.270	0.277	0.270	0.273	0.002
K= 51	3.291	0.259	0.272	0.272	0.286	0.280	0.278	0.005
K= 76	4.546	0.272	0.292	0.292	0.315	0.300	0.301	0.002
K=101	5.762	0.263	0.289	0.290	0.320	0.300	0.302	0.003
K=151	8.121	0.241	0.279	0.278	0.324	0.293	0.294	0.005
K=201	10.434	0.235	0.289	0.287	0.354	0.305	0.308	0.003
K=251	12.697	0.208	0.274	0.270	0.356	0.292	0.292	0.002

in Table 8.

Table 5: Simulation Results, Dependent Regressors, n=600, S=1000
Heterogeneous Cluster Size, Asymmetric Distribution

	(a): Empirical Coverage							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.042	0.940	0.941	0.940	0.940	0.935	0.940	0.000
K= 40	2.564	0.930	0.942	0.944	0.951	0.930	0.946	0.007
K= 80	4.609	0.901	0.928	0.928	0.945	0.923	0.929	0.005
K=120	6.803	0.881	0.918	0.913	0.949	0.917	0.924	0.003
K=160	8.603	0.880	0.927	0.920	0.952	0.911	0.929	0.001
	(b): Interval Length							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.042	0.190	0.191	0.191	0.191	0.191	0.191	0.000
K= 40	2.564	0.246	0.255	0.256	0.267	0.257	0.260	0.007
K= 80	4.609	0.254	0.273	0.273	0.295	0.280	0.280	0.005
K=120	6.803	0.247	0.277	0.275	0.310	0.285	0.286	0.003
K=160	8.603	0.252	0.295	0.289	0.340	0.302	0.304	0.001

References

Ambuehl, S., B. D. Bernheim, and A. Ockenfels (2021). What motivates paternalism? an experimental study. *American economic review 111*(3), 787–830.

Cattaneo, M. D., M. Jansson, and W. K. Newey (2018). Inference in linear regression models with many covariates and heteroskedasticity. *Journal of the American Statistical Association 113*(523), 1350–1361.

Gratton, G., L. Guiso, C. Michelacci, and M. Morelli (2021). From weber to kafka: Political instability and the overproduction of laws. *American Economic Review 111*(9), 2964–3003.

MacKinnon, J. G., M. Ø. Nielsen, and M. Webb (2020). *Testing for the appropriate level of clustering in linear regression models.* Department of Economics, Queen's University.

Table 6: Simulation Results, Dependent Regressors, n=600, S=1000
Heterogeneous Cluster Size, Bimodal Distribution

	(a): Empirical Coverage							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.042	0.943	0.945	0.944	0.946	0.942	0.944	0.000
K= 40	2.563	0.926	0.938	0.941	0.954	0.924	0.943	0.000
K= 80	4.605	0.910	0.926	0.932	0.954	0.927	0.941	0.006
K=120	6.793	0.893	0.925	0.929	0.962	0.921	0.933	0.001
K=160	8.595	0.870	0.920	0.921	0.972	0.933	0.932	0.003
	(b): Interval Length							
	$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	LZ	HC1	HC2	HC3	LO	CR	Neg
K= 1	0.042	0.232	0.233	0.233	0.233	0.233	0.233	0.000
K= 40	2.563	0.273	0.283	0.287	0.302	0.289	0.289	0.000
K= 80	4.605	0.285	0.307	0.311	0.344	0.318	0.318	0.006
K=120	6.793	0.277	0.310	0.314	0.363	0.319	0.323	0.001
K=160.000	8.595	0.277	0.324	0.323	0.394	0.337	0.336	0.003

Table 7: Empirical Application for Ambuehl, Bernheim, and Ockenfels (2021)

Dependent Variables	Belief smaller opportunity set better for Chooser		Max € Chooser takes early
Coefficient			
1 option withheld	0.320		
2 options withheld	0.496		
Front End Delay			-0.104
Standard Errors	1 option withheld	2 options withheld	
LZ	0.0662	0.1016	0.0906
HC1(Stata Proposed)	0.0675	0.1037	0.0931
HC2	0.0682	0.1054	0.0919
HC3	0.0704	0.1096	0.0932
CR	0.0691	0.1068	0.0908
LO	0.0685	0.1185	0.0894
Regression Parameters			
Observations n	909		606
Number of Clusters G		303	
Number of Regressors K	37		32
Largest cluster size $\mathcal{C}_{\mathcal{T},n}$	3		2
$\mathcal{M}_n$	0.078		0.068
$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	0.702		0.272
Cutoff for $\mathcal{M}_n$ with given $\mathcal{C}_{\mathcal{T},n}$	0.0762		0.1524

Table 8: Empirical Application for Gratton, Guiso, Michelacci, and Morelli (2021)

	Fixed Effect		Mean Residuals	
Coefficient				
Incompetent politician	-0.45		0.07	
Incompetent politician × completed legislature	-1.33		-1.19	
Standard Errors	Incompetent	Interaction	Incompetent	Interaction
LZ	0.581	0.579	0.419	0.542
HC1(Stata Proposed)	0.585	0.583	0.422	0.545
HC2	0.588	0.583	0.424	0.544
HC3	0.596	0.586	0.430	0.547
CR	0.589	0.581	0.422	0.535
LO	0.593	0.574	0.424	0.543
Regression Parameters				
Observations n	4904			
Number of Clusters G	2786			
Number of Regressors K	70		67	
Largest cluster size $\mathcal{C}_{\mathcal{T},n}$	6			
$\mathcal{M}_n$	0.502			
$\mathcal{C}^2_{\mathcal{T},n}\mathcal{M}_n$	0.502			
Cutoff of $\mathcal{M}_n$ for given $\mathcal{C}_{\mathcal{T},n}$	0.021			