

**ONLINE APPENDIX**  
**Beyond “Fixed Versus Random Effects”: A Framework for Improving Substantive and Statistical Analysis of Panel, Time-Series Cross-Sectional, and Multilevel Data**

**Appendix A: Data Transformation and Estimation**

**Stata Program for Calculating Within- and Between-Cluster Transformations**

I have created a Stata program called **clustergen** that creates within- and between-cluster transformations of all level-1 variables by issuing a single command. To do this the long way (without **clustergen**), one would first have to create the between-cluster operationalization of each variable one command at a time:

```
egen x1_bw=mean(x1) , by(id)
egen x2_bw=mean(x2) , by(id)
egen x3_bw=mean(x3) , by(id)
etc...
```

Generating the within-cluster transformations would entail the following commands:

```
gen x1_wi=x1-x1_bw
gen x2_wi=x2-x2_bw
gen x3_wi=x3-x3_bw
etc...
```

The **clustergen** program greatly simplifies this task by transforming all variables in one fell swoop. To install the program and the help file, first enter the following command in Stata (assumes an active internet connection):

```
net from https://home.gwu.edu/~bartels/stata
```

Then, click on the blue link called **clustergen**. A new window will pop up. Simply click on the link on the right side of the page that says “click here to install.”

This program assumes that your cluster identification variable is called “id.” If your cluster variable is called, e.g., “countryid,” simply issue the following command before using **clustergen**:

```
gen id=countryid
```

See the help file associated with the program for more details. To use the program, simply type “clustergen” followed by the list of level-1 variables that you want to transform. For each variable in the variable list, two new variables will be created: (1) a between-cluster version, which is the original variable name followed by “\_bw”; and (2) a within-cluster version, which is the original variable name followed by “\_wi.” For example, for the Poe, Tate, and Keith analysis, I transformed the level-1 variables (i.e., time-varying covariates) by issuing this command:

```
cluster gen ailag polrt lpop perchpop pcgthou perchpcg left mil2
brit iwarcow2 cwarcow
```

This created 22 new variables, e.g., ailag\_bw, ailag\_wi, polrt\_bw, polrt\_wi, etc.

## Estimation and Software

All models in the paper were estimated using Stata 10.0 and the “xt” commands. For the Poe et al. and Blaydes/Goodrich analyses, I estimated linear random intercept models via maximum likelihood estimation. Continuing with the Poe et al. example, the model from equation 4 in the paper (reported in Table 1) was estimated using the following command:

```
xtreg ainew ailag_wi polrt_wi lpop_wi perchpop_wi pcgthou_wi
perchpcg_wi left_wi mil2_wi iwarcow2_wi cwarcow_wi polrt_bw
lpop_bw perchpop_bw pcgthou_bw perchpcg_bw left_bw mil2_bw brit
iwarcow2_bw cwarcow_bw, i(id) mle
```

To estimate the model using FGLS, one would omit the “mle” after the comma. To generate estimates of the differences between the within- and between-cluster effects, which is the specification from equation 5 in the paper, I used the following command:

```
xtreg ainew ailag polrt lpop perchpop pcgthou perchpcg left mil2
iwarcow2 cwarcow polrt_bw lpop_bw perchpop_bw pcgthou_bw
perchpcg_bw left_bw mil2_bw brit iwarcow2_bw cwarcow_bw, i(id)
mle
```

Recall that in this model, the coefficients for all of the “\_bw” variables represent differences between the within- and between-cluster effects (see text of the paper for details). To estimate the random intercept probit model for the analysis of Senate voting on Supreme Court nominations (Epstein et al. 2006), I used the **xtprobit** command. The default integration method for this command is now adaptive quadrature with 12 integration points. One can also use regular Gauss-Hermite quadrature.

Stata 10.0 has implemented routines for estimating random coefficient models for continuous responses (**xtmixed**), binary responses (**xtmelogit**), and counts (**xtmepoisson**). These commands can also be used to estimate random intercept models. In addition, Rabe-Hesketh et al.’s (2004) **gllamm** package estimates both random intercept and random coefficient models for continuous, binary, ordinal, count, and nominal responses. For non-linear models, **gllamm** offers the option of using Gauss-Hermite quadrature or adaptive quadrature. Skrondal and Rabe-Hesketh (2004; Rabe-Hesketh and Skrondal 2005) show that adaptive quadrature provides more accurate results than Gauss-Hermite quadrature.

R and S-Plus can also estimate random intercept and random coefficient models using the **lme** and **nlme** packages (see Venables and Ripley 2002). Finally, WinBugs is capable of estimating these models via MCMC.

## **REFERENCES**

(Not cited in paper)

Rabe-Hesketh, Sophia, Anders Skrondal, and Andrew Pickles. 2004. *GLLAMM Manual*. UC Berkeley Division of Biostats.

Venables, W.N., and B.D. Ripley. 2002. *Modern Applied Statistics with S* (Fourth Edition). Springer.

## Appendix B: Standard Errors

As mentioned in the paper, TSCS analysts have given a great deal of attention to issues associated with standard errors (e.g., Beck and Katz 1995; Kristensen and Wawro 2003). Panel-corrected standard errors (PCSEs) (Beck and Katz 1995) are analogous to robust standard errors in that they correct for types of non-constant error variance induced by the structure of the data. In particular, PCSEs correct for panel heteroskedasticity (unequal error variance across countries) and contemporaneous error correlation (induced by shocks at a particular time point that affect countries contemporaneously). The random intercept model proposed in the paper does not make any after-the-fact corrections for standard errors (though one could perform such corrections, which I discuss below). However, the random intercept model's explicit accounting of unobserved cluster heterogeneity and the estimation of within- and between-cluster effects should reduce any threats to the accuracy of the standard errors. Thus, modeling the process the way I do—which I believe is superior to alternative approaches discussed in the paper—should lead to accurate standard errors.

To investigate the issue of standard errors in the proposed modeling approach, I compare the results from the random intercept model for the Poe, Tate, and Keith analysis of human rights abuse (in Table 1) with an alternative estimation procedure for estimating within- and between-country effects. In the alternative procedure, which is the basic approach used by Goodrich (2006), separate models are estimated for calculating within- and between-country effects. To calculate the within-country effects, I estimate a fixed-effects (FE) model with PCSEs. For the between-country effects, I estimate a between regression (a regression using the cluster averages of each variable) with robust standard errors. The point of this exercise is to compare the standard errors and statistical inferences from the random intercept model that simultaneously estimates within- and between-cluster effect with those from the separate models that explicitly adjust the standard errors. I contend that estimating within- and between-cluster effects simultaneously via the random intercept model is a superior strategy to estimating them separately, due to the more complete model specification and the ability to test for cluster confounding.

Results of this exercise are presented in Table B1. The first half of the table presents within-country effects using the two methods, and the second half presents between-country effects. As mentioned, the within and between results under the “random intercept model”

heading are identical to those from Table 1 in the paper. Looking first to the within-country effects, the results indicate that the statistical inferences from the FE model with PCSEs and the random intercept model are generally quite similar. While the coefficients are the same, one should expect differences in the standard errors since the models are estimated with different sets of variables and the random intercept model is estimated via maximum likelihood. In general, though, standard errors and significance tests are in the same ballpark. The one major difference concerns the effect of economic standing. In the FE model with PCSEs, economic standing exhibits a marginally significant impact ( $p=.08$ ), while in the random intercept model, the effect is far from statistically significant ( $p=.28$ ). Note how the standard error is larger in the random intercept model, meaning the random intercept model produces a more conservative hypothesis test for economic standing than the FE model with PCSEs.

Moving to the second half of Table B1, which compares estimates of between-country effects, the inferences are again generally similar between models. Two key differences stand out. In the between regression with robust standard errors, the impact of leftist government is statistically insignificant, while in the random intercept model, the effect is statistically significant. Also, the effect of British cultural influence is marginally significant ( $p<.10$ ) in the between regression and statistically significant ( $p<.05$ ) in the random intercept model.

In sum, the comparison in Table B1 shows that the standard errors from the random intercept model appear to produce quite similar types of inferences to alternative methods that explicitly make corrections to standard errors. I believe the more fully specified random intercept model is superior to these alternative procedures. It simultaneously estimates within- and between-cluster effects, allows for explicit tests of cluster confounding, and accounts for cluster-level heterogeneity, which parses the total error into separate within- and between-country error components. Thus, the standard errors from this model should be highly accurate. As I state in the paper, analysts can still investigate various forms of heteroskedasticity induced by the structure in the data. For instance, one might conclude that the random intercept model should be estimated with robust standard errors.

**Table B1: Comparing Inferences from Proposed Framework to Alternative Models (Using the Human Rights Abuse Substantive Application)**

<i>Within-Country Effects</i>							<i>Between-Country Effects</i>					
<b>Variable</b>	<b>Fixed Effects Model with PCSEs (Separate)</b>			<b>Random Intercept Model</b>			<b>Between Regression with Robust SEs (Separate)</b>			<b>Random Intercept Model</b>		
	<b>Coef.</b>	<b>(PCSE)</b>	<b>p</b>	<b>Coef.</b>	<b>(SE)</b>	<b>p</b>	<b>Coef.</b>	<b>(RSE)</b>	<b>p</b>	<b>Coef.</b>	<b>(SE)</b>	<b>p</b>
Rights Abuse <sub>t-1</sub>	0.380	(0.022)	0.000	0.380	(0.019)	0.000	-	-	-	-	-	-
Democracy	-0.105	(0.015)	0.000	-0.105	(0.014)	0.000	-0.090	(0.036)	0.014	-0.105	(0.033)	0.002
Population Size	0.224	(0.097)	0.021	0.224	(0.095)	0.018	0.183	(0.027)	0.000	0.181	(0.024)	0.000
Population Change	0.001	(0.004)	0.733	0.001	(0.004)	0.702	0.093	(0.035)	0.009	0.074	(0.033)	0.025
Economic Standing	-0.005	(0.003)	0.079	-0.005	(0.005)	0.278	-0.043	(0.010)	0.000	-0.047	(0.009)	0.000
% Economic Change	-0.001	(0.001)	0.213	-0.001	(0.001)	0.217	-0.004	(0.009)	0.662	-0.002	(0.009)	0.790
Leftist Government	-0.041	(0.074)	0.578	-0.041	(0.077)	0.594	-0.202	(0.171)	0.238	-0.323	(0.153)	0.035
Military Control	0.002	(0.060)	0.972	0.002	(0.054)	0.969	0.383	(0.139)	0.007	0.307	(0.123)	0.012
British Cultural Influence	-	-	-	-	-	-	-0.178	(0.106)	0.094	-0.231	(0.093)	0.013
International War	0.046	(0.054)	0.396	0.046	(0.056)	0.415	0.547	(0.229)	0.018	0.633	(0.209)	0.003
Civil War	0.465	(0.076)	0.000	0.465	(0.065)	0.000	1.556	(0.213)	0.000	1.525	(0.190)	0.000
Constant	-2.776	(1.863)	0.136	-0.134	(0.423)	0.751	-0.353	(0.471)	0.455	-	-	-