Improving the Interpretation of Random Effects Regression Results

Soren Jordan and Andrew Q. Philips

2023

Abstract

Mummolo and Peterson improve the use and interpretation of fixed effects models by pointing out that unit intercepts fundamentally reduce the amount of variation of variables in fixed effects models. Along a similar vein, we make two claims in the context of random effects models. First, we show that potentially large reductions in variation, in this case caused by quasi-demeaning, also occur in models using random effects. Second, in many instances, what authors claim to be a random effects model is actually a pooled model after the quasi-demeaning process, affecting how we should interpret the model. A literature review of random effects models in top journals suggests both points are currently not well understood. To better help users interested in improving their interpretation of random effects models, we provide Stata and R programs to easily obtain post-estimation quasi-demeaned variables.

Keywords: methodology, random effects, demeaning, interpretation

This is the final version as it appears for the citation:

Jordan, Soren and Andrew Q. Philips. 2023. "Improving the Interpretation of Random Effects Regression Results." *Political Studies Review* 21 (1): 210-220.
DOI: 10.1177/14789299211068418

The page numbers here are consistent with the published version. Corresponding author: Soren Jordan (sorenjordanpols@gmail.com).

Mummolo and Peterson (2018) discuss how fixed effects (FE) models in panel data—including unit-specific dummy variables to allow the intercept to shift up or down for each unit—estimate only within-unit variance. Thus, coefficients describe only the effect of a one-unit increase occurring *within* a unit, not between units.FE models have several advantages, namely that they exclude the possibility of endogeneity caused by a correlation between the included covariates and any omitted time-invariant unit effects. However, FE complicate interpretation, since variation over time is generally appreciably smaller than overall variation (i.e., what we would interpret with a global summary). To avoid problematic interpretation, especially overstating the substantive effects of the model, Mummolo and Peterson suggest accounting for this within-transformation, since it will almost always result in smaller levels of variance for the variable of interest to consider when constructing counterfactuals.

In the spirit of Mummolo and Peterson's contribution, we note that a related panel data model also has the potential to complicate substantive conclusions about the magnitude of effects due to the nature of the estimation process: models incorporating random unit intercepts, commonly referred to as "random" or "mixed" effects (RE). RE offer a data-driven compromise between a fully pooled model and FE model by partially ("quasi") demeaning observations by their unit averages. Moreover, this estimator is popular; recent applications range from economic voting (Valdini and Lewis-Beck, 2018) to the spread of economic, social, and cultural rights (Bjornskov and Mchangama, 2019), to public opinion about government expenditures (Busemeyer et al., 2019).

In this note, we make two novel claims for scholars interested in using RE models. First, like FE, RE models have the potential to appreciably reduce variation in a variable. Thus, if users are interested in exploring the substantive significance of their results they run the risk of making "extreme counterfactuals" far away from the transformed data space the model was estimated on (King and Zeng, 2006). Since no popular software currently informs users of the range of these quasi-demeaned variables, we show that it is often the case that users are making interpretations—and thus theoretical conclusions—based on unrealistic counterfactual shifts in their variables of interest. Second, due to characteristics in the data, sometimes what authors claim is a RE model is in fact no different from a fully pooled model. While this does not alter the reported estimates, it *does* mean that any reported advantages of using RE over a pooled model are not applicable, and might leave users ill-informed about the nature of their data (for instance, by thinking that unit heterogeneity is resulting in demeaning, while in reality units are so similar that they can be fully pooled).

These are claims about RE models, *not* criticisms. Instead, like Mummolo and Peterson, we seek to help applied users better understand what is going on "under the hood" and interpret their RE results: specifically by (1) adjusting counterfactual interpretations for the range of quasi-demeaned data, and (2) always reporting the variance components of RE models (and noting when they result in fully pooled models). We proceed as follows. First, since the notion that RE models lie somewhere between a fully pooled and FE model might be unfamiliar to most practitioners—for instance, Clark and Linzer (2015) devote only a few sentences to this—we discuss how quasi-demeaning occurs. Second, we survey articles using RE in the literature to demonstrate that our claims are currently not well understood. Third, we use two examples to illustrate why our claims matter substantively. We conclude by offering suggestions on reporting these quantities of interest with software we provide as well as describing alternate modeling approaches.

The Link Between Several Panel Data Models

Consider a model with one regressor that does not take into account any potential heterogeneity across units i observed across time t:

$$y_{it} = \alpha + \beta x_{it} + \epsilon_{it} \tag{1}$$

This is often called a fully pooled model since a single global intercept, α, is estimated. No unit-specific differences are captured in this intercept; thus, any unit-specific (i.e., time-invariant) differences in y_{it} will end up in the error term. Moreover, standard errors will likely be incorrect because the estimator assumes observations are independent and identically distributed.

A popular alternative to this adds intercepts for each unit:

$$y_{it} = \alpha_i + \beta x_{it} + \epsilon_{it} \tag{2}$$

and is known as a FE model.[1] Here, any time-invariant unit-specific differences—in other words, any variation between units—are soaked up by the α_i terms. The remaining variation explained by β is only within-unit.[2] A third modeling choice employs a random intercept:

$$y_{it} = \alpha + \beta x_{it} + \nu_{it} \tag{3}$$

where the stochastic term is now a composite of fundamental error, ϵ_{it} (which we assume is well-behaved, as in any model), as well as error specific to a particular unit (after partialing out the "global" intercept α); thus, $\nu_{it} = u_i + \epsilon_{it}$. RE models account for unit-specific differences by allowing for variation around a global intercept, α. To do this, both the variance around the stochastic term ϵ_{it}, σ^2_ϵ, and the variance around α, which is given by σ^2_α, are estimated.[3]

The tradeoffs between these three approaches are not strictly of interest here, although there are well-known differences between pooled, FE and RE (Clark and Linzer, 2015). Generally speaking, a fully pooled model (equation 1) is never advisable in the panel data context (Clark and Linzer, 2015), since any correlation between the observables, x_{it}, and any time invariant unit effects (α_i) will lead to biased estimates of β; at best, it is still inefficient, since it does not take into account the structure of the data (i.e., observations across time are nested within units), which will likely to lead to unit-specific heteroskedasticity. Like the fully pooled model, RE (equation 3) also suffers from potential endogeneity caused by any correlation between the included covariates in the model and the (unobserved) time-invariant unit-specific effects in the error term, although efficiency gains can sometimes outweigh such bias (Clark and Linzer, 2015). FE models in contrast, are unable to include time-invariant regressors (since they will be perfectly correlated with α_i), and are biased in dynamic models with short time points, or under certain forms of dynamic misspecification (Plümper and Troeger, 2019).

This choices might appear somewhat limited; FE models estimate only within effects, while pooled and RE models combine both between and within effects. An emerging alternative is to estimate these effects separately—variants of which are the Mundlak, "hybrid", or random effects within-between (REWB) models (Bell and Jones, 2015; Bell, Fairbrother, and Jones, 2019):

$$y_{it} = \alpha + \beta_W (x_{it} - \bar{x}_i) + \beta_B \bar{x}_i + \nu_{it} \tag{4}$$

Including both the unit-demeaned and unit-means of x_{it} separately provides β_W, the within effect (i.e., same as β in the FE model), and β_B, the between effect. Equation 4 is estimated using RE, as the error components are assumed to be drawn from a normal distribution. REWB has clear advantages. In the context of the "standard" RE model of equation 3, REWB can "test whether the assumption of equal within and between effects is true" by seeing if $\beta_W = \beta_B$ (Bell, Fairbrother, and Jones, 2019, p. 1057). More generally, REWB helps avoid the RE vs. FE false dichotomy often discussed by scholars. Practitioners should consider estimating a REWB model to determine if there are separate within and between effects to uncover. Despite REWB's advantages, we believe there is still a need to offer interpretative advice to users of the "standard" RE shown in equation 3. For one, when certain conditions are met, the RE model remains an appropriate choice as it is a more efficient estimator than REWB.[4] Moreover, since FE and RE are still the predominant models in the field,[5] we stress that there is a need for interpreting these models correctly when reviewers or readers expect to see them.

The pooled, FE, RE and REWB models discussed above offer different approaches to handling unit heterogeneity. In fact, we can explicitly formulate the RE model as a compromise between the completely pooled and completely within-variance (FE) approaches. This formulation better allows us to understand how we might arrive at interpretation challenges due to quasi-demeaning when using RE. To better see this, first write the unit-specific average value of each variable (i.e., averaging across time points t) for each unit i as $\bar{y}_i = \bar{\alpha}_i + \beta\bar{x}_i + \bar{\epsilon}_i$. The RE approach operates by conceptualizing and estimating a demeaning parameter, θ, that accounts for unit-specific averages (Wooldridge, 2010, 326-328):

$$(y_{it} - \theta\bar{y}_i) = (\alpha - \theta\bar{\alpha}_i) + \beta(x_{it} - \theta\bar{x}_i) + (\epsilon_{it} - \theta\bar{\epsilon}_i) \tag{5}$$

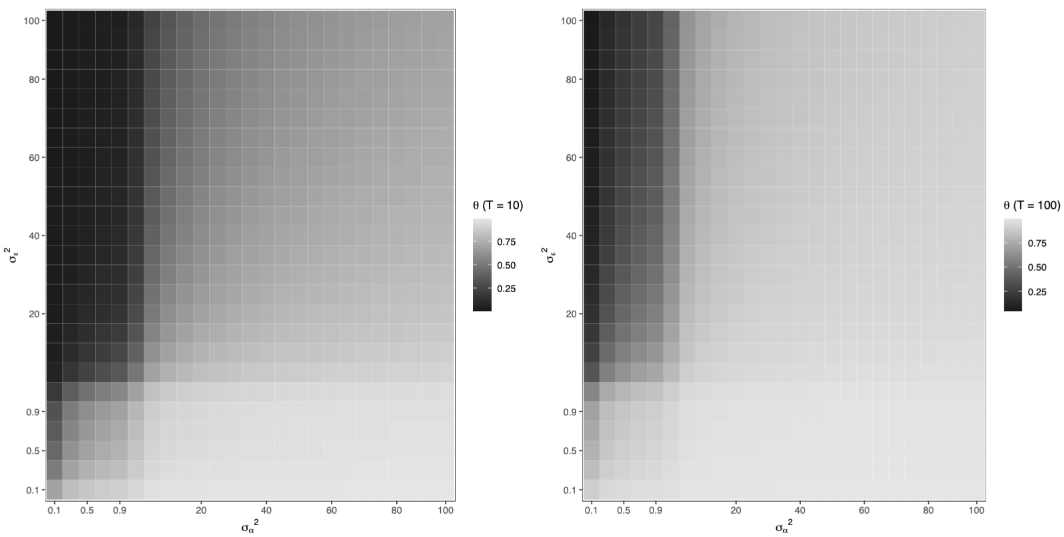
If we let $\theta = 0$, then equation 5 becomes the pooled model; that is to say, there is no demeaning by the unit means for each variable. Another way of picturing this is that the pooled model fully—and equally—incorporates variation within units as well as between units. In contrast, if we let $\theta = 1$, then equation 5 becomes the FE model, since unit means are subtracted out, leaving only within variance.

In contrast to the fully pooled and FE model, the RE model allows θ, the proportion of demeaning from the unit averages, to vary, such that $0 \leq \theta \leq 1$. Therefore, RE can be thought of as a compromise between the fully pooled model (where within and between variance are equally weighted) and the fixed effects model (where within variance is completely removed since $\theta = 1$). As θ gets closer to one, estimation results are based on more within variation relative to between, and thus results should approximate a fixed effects model. In practice, we do not need to make a theoretical determination about the value of θ, as software programs automatically perform this quasi-demeaning. This requires the total number of T time points in the estimated sample for each unit, as well as the two variance components from equation 3; "fundamental" error variance, σ^2_ϵ, and error variance attributable to unit-specific differences, σ^2_α (Clark and Linzer, 2015). With these three values, θ is estimated as[6]

$$\hat{\theta} = 1 - \sqrt{\frac{\hat{\sigma}^2_\epsilon}{T\hat{\sigma}^2_\alpha + \hat{\sigma}^2_\epsilon}} \tag{6}$$

Estimation of $\hat{\sigma}^2_\epsilon$ and $\hat{\sigma}^2_\alpha$ is typically performed via feasible generalized least squares; for instance, by using the `xtreg, re` command in Stata or in the `plm` package in `R` (Croissant and Millo, 2008). The estimation of the unit variance and error variance differs across various algorithms, typically, by how they handle unbalanced panels and estimation in small samples (c.f., Amemiya, 1971; Swamy and Arora, 1972). RE estimation can also be performed using maximum likelihood, which involves simultaneously estimating σ^2_ϵ and σ^2_α and the coefficients in a single likelihood function. However, the same demeaning occurs.

To illustrate how θ depends on T, between-unit ($\hat{\sigma}^2_\alpha$) and fundamental ($\hat{\sigma}^2_\epsilon$) error variance, Figure 1(a) plots values of θ from equation 6 across various combinations of σ^2_α and σ^2_ϵ when $T = 10$. Regardless of σ^2_ϵ—the variance in the stochastic



(a) $T = 10$ (b) $T = 100$

Figure 1: The Proportion of Quasi-Demeaning, θ, Varies by σ_α^2, σ_ϵ^2 and T. (a) $T = 10$ and (b) $= 100$.

term—greater variance between units (higher values of σ_α^2) leads to larger values of θ. This means that more quasi-demeaning is occurring, and thus the RE model more closely approximates FE, which only uses within-variation. θ tends to be highest when the between-unit variance is large relative to the stochastic error variance. As the stochastic error variance increases, the proportion of quasi-demeaning decreases; quasi-demeaning tends to be lowest when the between-unit variance is small relative to the stochastic error variance. Figure 1(b) shows values of θ when $T = 100$. Now, θ very quickly increases as the between-unit variance increases, meaning that more within-unit variation is being analyzed, relative to between. This relationship between θ, $\hat{\sigma}_\epsilon^2$, and $\hat{\sigma}_\alpha^2$ echoes other measures of the division of variance between levels, like the Variance Partition Coefficient (VPC) or the Intra-Class Correlation (ICC).[7] When choosing a measure to report, though, θ explicitly accounts for unbalanced panels, has a consistent direction (as it is always expressed with $\hat{\sigma}_\epsilon^2$ in the numerator), and plays a direct role in the underlying quasi-demeaning happening in RE models.

There are a few other important features surrounding quasi-demeaning in RE models. For one, the quasi-demeaning transformation applies to *all* variables in a model. In addition, the quasi-demeaning process sometimes estimates a higher level variance, σ_α^2, as zero. While this may not be an econometric problem,[8] it presents a poorly understood issue for interpretation: namely that the RE results will be identical to the pooled model. In this world, special care must be taken to *interpret* it as a fully pooled model, lacking any particular benefits of the RE approach. Taken together, quasi-demeaning and the RE model suggest special attention needs to be paid to the counterfactuals used when generating inferences (as the quasi-demeaning process explicitly changes the estimated variance of the variables), as well as whether RE results in pooled modeling.

New Insights or Old Hat?

A skeptic may claim that the above discussion and its implications are already well understood by authors using RE. To evaluate this, we reviewed the use of RE modeling in top

Table 1: RE Models and Interpretation in Top Journals

Category	Number of Articles
Total that use RE	160
Report $\hat{\sigma}_{\epsilon}^2$ and $\hat{\sigma}_{\alpha}^2$	5
Report numeric "variance of ..."	57
Report Intra-Class Correlation	3
Report Variance Partition Coefficient	1
Report $\hat{\theta}$	0
Report a variance estimate of 0	20 of 57 (35%)
Interpret substantively beyond a one-unit effect	102 of 130 (78%)
Considers within-unit range through interpretation	1 of 130 (1%)

Includes all tables of "random effects" or "mixed" models in the *American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, and *British Journal of Political Science* from 2012 to 2020.

political science journals (*American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, and *British Journal of Political Science*) from 2012 to 2020. A discipline that understands the link between FE, RE, and the fully pooled model should consistently be reporting the variance at each level ($\hat{\sigma}_{\epsilon}^2$ and $\hat{\sigma}_{\alpha}^2$), the quasi-demeaning parameter $\hat{\theta}$ (or similar quantities like ICC or VPC), and should be accounting for this quasi-demeaning through the counterfactuals provided in the text. The findings are summarized in Table 1.

We found 160 papers use RE models in the main text (instead of a robustness check in an appendix). Of these, well under half report $\hat{\sigma}_{\epsilon}^2$ or $\hat{\sigma}_{\alpha}^2$, or any measure of variance, precluding an inference about the extent of quasi-demeaning in the data. Only four report a quantity describing the relative variance, and none of the articles explicitly reported θ, the parameter that captures the quasi-demeaning. Moreover, 20 of the articles reported an estimated variance of 0, suggesting that the authors were actually estimating a fully pooled model *despite* no discussion as such in the text. In fact, it was the opposite; authors would often vaunt the appropriateness of RE because of its between-unit variance properties, even though these were not being reflected in the model.

With regards to substantive interpretation, 130 articles drew inferences beyond basic sign and significance. Only 28 reported the effect of a one-unit change. The remaining 102 reported counterfactuals relative to the original (not quasi-demeaned) range of the variables, sometimes as large as the full minimum to maximum: a comparison which is *never* observed within any unit after quasi-demeaning. Therefore, we think our claims, despite perhaps being transparent to experts, are still mishandled among practitioners, even in top journals. We further elaborate through applied examples below.

Substantive Examples From the Literature

Example I: Quasi-Demeaning and Avoiding Extreme Counterfactuals

To see the extent to which variation may be reduced beyond a counterfactual given in a paper, we replicate Williams (2017), who examines why some development projects like schools are started but never finished. Williams reports using the RE GLS estimator (Table 2, Model 1 in his article) to analyze Ghanaian local development projects in 327 districts from 2011 to 2013. The probability a project is unfinished is a function of

Table 2: Replication of Williams (2017), Table 2, Model 1

	RE Model
Government-funded	-0.109*** (0.015)
NDC Vote Share	0.109 (0.082)
Constant	1.017*** (0.157)
σ^2_ϵ	0.412
σ^2_α	0.226

RE: random effects; NDC: National Democratic Congress.
Note: Dependent variable is probability of project completion.
Not shown: FE for 22 project types, FE for construction type, FE for years, and number of years since project start. Model appears exactly as in original article. $N = 4563$. $^*(p < 0.10)$,$^{**}(p < 0.05)$,$^{***}(p < 0.01)$.

whether the project is government funded, the vote share for the National Democratic Congress (NDC) party, as well as controlling for project type, construction type, project duration, and year. RE are included for districts, which are the subject of our investigation here. The model is strictly replicated in Table 2.

Our principal focus is the range of the transformed data once quasi-demeaning has occurred. Williams' key predictor is NDC vote share: the proportion of the vote for the National Democratic Congress, which can range from 0 (no votes) to 1 (all of the votes). Quasi-demeaning implies, however, that the proportion in any given district can *never* range from 0 to 1, since it will be adjusted downward by the unit-specific mean, according to equation 3. Following the formula in equation 6, we calculate the quasi-demeaned data relative to the original data for the key predictor, NDC vote share. Compared to the original data, the range of quasi-demeaned data is dramatically reduced. Figure 2 illustrates exactly this; Williams interprets NDC across the range of 0.1 to 0.9, implying an effect of $[0.9 - 0.1] \times 0.109 = 0.8 \times 0.109 \approx 0.09$ unit increase in the probability a project is completed. However, once the data are quasi-demeaned, instead of moving NDC vote share from 0.1 to 0.9, the new $\pm$ one standard deviation moves NDC vote share from 0.1 to just 0.29, implying an effect of only a $[0.29 - 0.10] \times 0.109 = 0.19 \times 0.109 \approx 0.02$ unit increase in the probability of a project being completed. Interpreting the original counterfactual leads to an estimated effect that is quadruple the size of a reasonable effect calculated using the demeaned data.

Williams goes on to estimate an interactive model between whether a project is government-funded and the NDC vote share, then shows a marginal effects plot displaying the effect of government funding across a range of NDC vote share from 0.1 to 0.9. Accounting for the quasi-demeaning, as shown in Figure 2, the 99.5 percentile of cases on the demeaned NDC vote share variable is 0.499. In other words, over half of the range of the interpreted interaction (all values over 0.50) *has no cases*, since less than 0.5% of units have a quasi-demeaned value of NDC vote share of over 0.5. Without special attention to this quasi-demeaning post-estimation, we are at significant risk of drawing implausible (at best) or impossible (at worst) counterfactuals from our RE models.

Example II: Quasi-Demeaning Estimates No Variance; RE *Are* a Pooled Model

For the second applied example, we show how in certain instances, estimation of the quasi-demeaning parameter, $\hat{\theta}$, results in a model that is no different from the fully

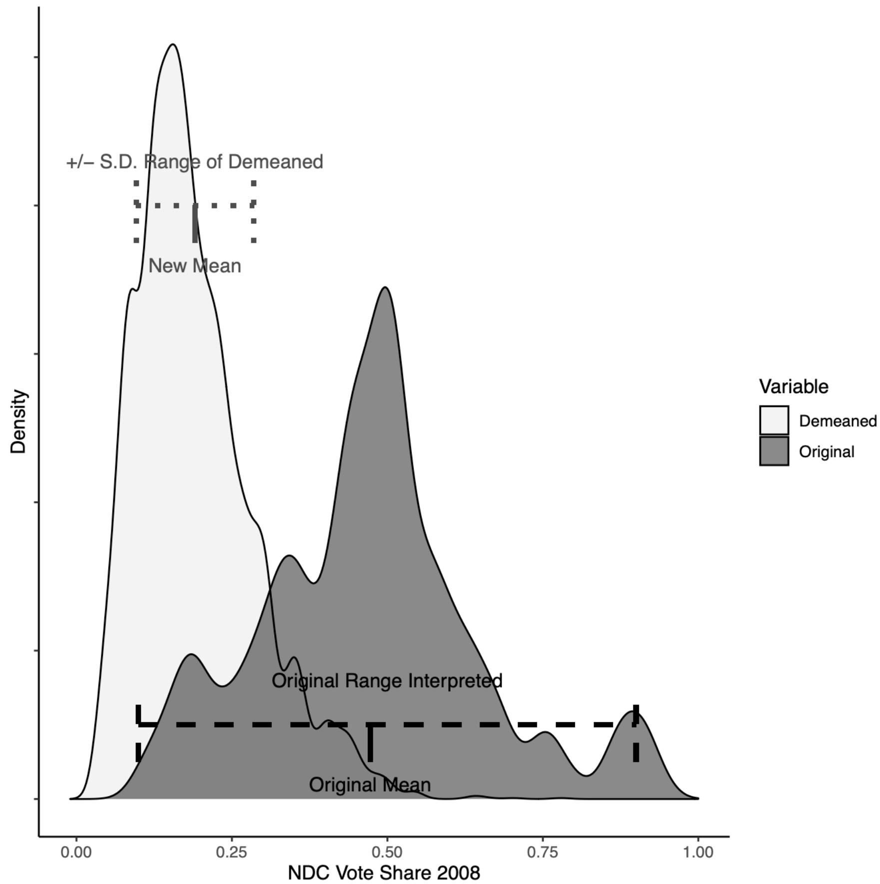


Figure 2: Quasi-Demeaned Versus Untransformed Vote Share

pooled model. We replicate Valdini and Lewis-Beck (2018), who examine how institutional rules affect the vote share of incumbent parties. Table 3, Model 1 shows the same results as the authors, using the RE GLS estimator as they do: except we report the estimated variances σ_{ϵ}^2 and σ_{α}^2.[9] Note that σ_{α}^2, the partitioned error variance attributable to unit-specific differences, is estimated at 0. This means that $\hat{\theta} = 0$, thus no quasi-demeaning is taking place, and the estimator is equivalent to the fully-pooled model. As evidence of this, Model 2 in Table 3 shows the results using the fully pooled model estimator, which has exactly the same results at in Model 1. In their text, though, the authors suggest that a "full set of corrections for statistical efficiency" are employed, even though they are "limited by sample size" and the amount of variation in their data (Valdini and Lewis-Beck, 2018, 418-419). However, the statistical "correction" did not include any quasi-demeaning: negating any supposed advantage of the RE model.

By no means are we criticizing the modeling strategy of Valdini and Lewis-Beck (2018). Moreover, they are in good company; recall from Table 1 that 35% of the articles that reported variance estimates actually estimated 0 variance: a RE model where quasi-demeaning resulted in a fully-pooled model without the researchers noting it as such. As discussed above, while econometricians differ about whether this implies fundamental model misspecification or something more benign (Baltagi, 2008; Greene, 2018), we

Table 3: RE *Are* a Fully Pooled Model When Quasi-Demeaning Estimates No Variance

	(1) RE-GLS		(2) Pooled Model	
Previous Incumbent Vote	0.530*	(0.284)	0.530*	(0.284)
GDP Growth Rate$_{t-1}$	2.076***	(0.638)	2.076***	(0.638)
Electoral Stability	0.309***	(0.068)	0.309***	(0.068)
Trade Openness	0.134**	(0.058)	0.134**	(0.058)
Concurrent Elections	2.756	(2.539)	2.756	(2.539)
GDP Growth Rate$_{t-1}\times$ Concurrent Elections	-0.894**	(0.447)	-0.894*	(0.447)
Constant	-19.383	(18.795)	-19.383	(18.795)
σ^2_ϵ	261.79		230.81	
σ^2_α	0		NA	

Note: Dependent variable is incumbent party vote share.
Coefficients shown with robust standard errors clustered on 18 countries in parentheses for both models. $N = 92$.
$^{*}(p < 0.10)$,$^{**}(p < 0.05)$,$^{***}(p < 0.01)$.

observe that these instances of $\sigma^2_\alpha = 0$ are never discussed in the main results as a fully pooled model. In other words: scholars are introducing, reporting, and interpreting fully-pooled models as if they are RE models. This misrepresentation indicates the need for a broad prescription on what to report when estimating RE models.

Conclusions and Recommendations

We emphasize two claims. First—similar to FE models (Mummolo and Peterson, 2018)—scholars need to account for the quasi-demeaned nature of their independent variables when interpreting RE models. Second, it is possible, maybe even common, for a model that a scholar believes to be a RE model to actually just be a fully pooled model due to a lack of higher-level variance to be estimated. This may not be a statistical problem, but when it happens, we should interpret the model more faithfully to what's actually being estimated.
As such, we offer the following recommendations:

1. Scholars should discuss the substantive and statistical significance of their effects with plausible counterfactual changes that actually occur in the (transformed) sample space of data the model is estimated on. In order to assist users, we have written `qdmean`, functions for both Stata and `R` that automatically calculate the proportion of demeaning that is occurring in their variables when a RE model is estimated.[10] These programs generate these quasi-demeaned variables, from which users can create more realistic counterfactual scenarios (i.e., avoiding the extreme counterfactuals shown in Figure 2).

2. Scholars should report the estimated variance components of the RE model (σ^2_ϵ and σ^2_α): currently reported by under half of RE models (Table 1). Additionally, it's useful to report $\hat{\theta}$ (the estimated proportion of quasi-demeaning), which summarizes the partitioning of the variance at different levels in a single statistic.

Values of $\hat{\theta}$ close to 1 indicate that the RE model is utilizing mostly within-variation, while small values of $\hat{\theta}$ close to 0 indicate that a substantial proportion of between variation is being used in estimation, which may lead to differences between the random and FE models. Using $\hat{\theta}$ as a diagnostic would help better identify cases where RE devolves to the fully pooled model. In these cases, scholars should simply use (and report) the results from a pooled model, or consider using a FE model, since none of the claimed advantages of RE apply.

3. Scholars might consider another approach entirely. While the points above are crucial given the popularity of RE in political science, alternatives exist. A fully-pooled model is one option; while it retains overall variation, it makes heroic assumptions about observation independence that are probably not met in most panel data applications. Despite this, under certain conditions it may still perform better than other models in the presence of dynamic misspecification (Plümper and Troeger, 2019). As discussed above, an alternative approach is to parse the effect of covariates into specific within- and between-variation components (Bell and Jones, 2015; Bell, Fairbrother, and Jones, 2019). This would allow users to examine counterfactual changes within a unit separately from those between units, as well as examine whether combining within and between effects (as the pooled does and the RE does to some extent) is valid.

Regardless of the modeling approach used, when working with panel data, users should always be aware of the assumptions behind each strategy, as well as understand any data transformations (or lack thereof) that have occurred.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Soren Jordan `https://orcid.org/0000-0003-4201-1085`
Andrew Q. Philips `https://orcid.org/0000-0002-6551-1296`

Notes

[1]An equivalent model is often estimated by subtracting unit means from each variable (and the constant), and is known as the within-transformation: $(y_{it} - \bar{y}_i) = \beta(x_{it} - \bar{x}_i) + (\epsilon_{it} - \bar{\epsilon}_i)$.

[2]Within variance for, say, x_{it}, is given as $\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2}{N \times T - 1}$, while between variance is $\frac{\sum_{i=1}^{N}(\bar{x}_i-\bar{\bar{x}})^2}{N-1}$, where N and T are the total number of units and time points (respectively), $\bar{x}_i$ are unit-means, and $\bar{\bar{x}}$ is the grand mean.

[3]That is, we assume $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ and $\alpha_i \sim N(\alpha, \sigma_\alpha^2)$. Although we use panel data notation with time and unit notation, our claims extend to any model that estimates random intercepts for a hierarchical, upper, "level 2" structure in which observations are nested (level i in our notation (c.f., Bell and Jones, 2015; Bell, Fairbrother, and Jones, 2019)).

[4]When the within and between effects of equation 4 are equivalent (Bell, Fairbrother, and Jones, 2019, p. 1057), which practitioners can test by estimating the REWB model.

[5]In Bell and Jones (2015, p. 134), for instance, less than 10% of random effects models in articles also mentioned "Mundlak." Even leading proponents of REWB like Bell, Fairbrother, and Jones (2019, p. 1056), refer to contemporary examples of this approach as "rare."

[6]Equation 6 assumes a balanced dataset; in other words, the total number of time points is the same for all units. If this is not the case, we end up indexing T_i, meaning θ_i can differ by unit: $\hat{\theta}_i = 1 - \sqrt{\frac{\hat{\sigma}_\epsilon^2}{T_i\hat{\sigma}_\alpha^2 + \hat{\sigma}_\epsilon^2}}$.

[7]In the case of a model with a continuous dependent variable and random intercepts, the VPC and ICC are both calculated as the same proportion: $\frac{\hat{\sigma}_x^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}_\epsilon^2}$, where x can represent either ϵ or α (Leckie et al., 2020). Hence, $\hat{\theta}$ resembles the VPC weighted by the length of the panel.

[8]While Greene (2018, p. 409) suggests this is a model specification issue, Baltagi (2008, p. 24) argues it is "not a bad sin" since σ_α^2 is likely very small.

[9]Table 4, Model 1 in their article.

[10]Available at https://github.com/andyphilips/qdmean.

Author Biographies

Soren Jordan is an Associate Professor of Political Science at Auburn University. He studies American politics and political methodology.

Andrew Q Philips is an assistant professor in the Department of Political Science, University of Colorado Boulder, Boulder, CO, USA. His research interests are in political economy and methodology.

References

Amemiya, Takeshi. 1971. "The estimation of the variances in a variance-components model." *International Economic Review* pp. 1–13.

Baltagi, Badi. 2008. *Econometric analysis of panel data*. John Wiley & Sons.

Bell, Andrew, and Kelvyn Jones. 2015. "Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data." *Political Science Research and Methods* 3 (1): 133–153.

Bell, Andrew, Malcolm Fairbrother, and Kelvyn Jones. 2019. "Fixed and Random effects models: making an informed choice." *Quality & Quantity* 53 (2): 1051–1074.

Bjornskov, Christian, and Jacob Mchangama. 2019. "Do social rights affect social outcomes?" *American Journal of Political Science* 63 (2): 452–466.

Busemeyer, Marius R, Aurélien Abrassart, Spyridoula Nezi, and Roula Nezi. 2019. "Beyond Positive and Negative: New Perspectives on Feedback Effects in Public Opinion on the Welfare State." *British Journal of Political Science* pp. 1–26.

Clark, Tom S, and Drew A Linzer. 2015. "Should I use fixed or random effects?" *Political Science Research and Methods* 3 (2): 399–408.

Croissant, Yves, and Giovanni Millo. 2008. "Panel data econometrics in R: The plm package." *Journal of Statistical Software* 27 (2): 1–43.

Greene, William H. 2018. *Econometric analysis*. 8 ed. Pearson.

King, Gary, and Langche Zeng. 2006. "The dangers of extreme counterfactuals." *Political Analysis* 14 (2): 131–159.

Leckie, George, William J. Browne, Harvey Goldstein, Juan Merlo, and Peter C. Austin. 2020. "Partitioning variation in multilevel models for count data." *Psychological Methods* 25 (6): 787–801.

Mummolo, Jonathan, and Erik Peterson. 2018. "Improving the interpretation of fixed effects regression results." *Political Science Research and Methods* 6 (4): 829–835.

Plümper, Thomas, and Vera E Troeger. 2019. "Not so Harmless After All: The Fixed-Effects Model." *Political Analysis* 27 (1): 21–45.

Swamy, PAVB, and Swarnjit S Arora. 1972. "The exact finite sample properties of the estimators of coefficients in the error components regression models." *Econometrica: Journal of the Econometric Society* pp. 261–275.

Valdini, Melody E, and Michael S Lewis-Beck. 2018. "Economic Voting in Latin America: Rules and Responsibility." *American Journal of Political Science* 62 (2): 410–423.

Williams, Martin J. 2017. "The political economy of unfinished development projects: Corruption, clientelism, or collective choice?" *American Political Science Review* 111 (4): 705–723.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2 ed. MIT Press.