Political Science Research and Methods Vol 3, No. 1, 133–153 January 2015

© The European Political Science Association, 2014. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited. doi:10.1017/psrm.2014.7

Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data*

ANDREW BELL AND KELVYN JONES

This article challenges Fixed Effects (FE) modeling as the 'default' for time-series-cross-sectional and panel data. Understanding different within and between effects is crucial when choosing modeling strategies. The downside of Random Effects (RE) modeling—correlated lower-level covariates and higher-level residuals—is omitted-variable bias, solvable with Mundlak's (1978a) formulation. Consequently, RE can provide everything that FE promises and more, as confirmed by Monte-Carlo simulations, which additionally show problems with Plümper and Troeger's FE Vector Decomposition method when data are unbalanced. As well as incorporating time-invariant variables, RE models are readily extendable, with random coefficients, cross-level interactions and complex variance functions. We argue not simply for technical solutions to endogeneity, but for the substantive importance of context/heterogeneity, modeled using RE. The implications extend beyond political science to all multilevel datasets. However, omitted variables could still bias estimated higher-level variable effects; as with any model, care is required in interpretation.

Two solutions to the problem of hierarchical data, with variables and processes at both higher and lower levels, vie for prominence in the social sciences. Fixed effects (FE) modeling is used more frequently in economics and political science, reflecting its status as the "gold standard" default (Schurer and Yong 2012, 1). However, random effects (RE) models—also called multilevel models, hierarchical linear models and mixed models—have gained increasing prominence in political science (Beck and Katz 2007) and are used regularly in education (O'Connell and McCoach 2008), epidemiology (Duncan, Jones and Moon 1998), geography (Jones 1991) and biomedical sciences (Verbeke and Molenberghs 2000, 2005). Both methods are applicable to research questions with complex structure, including place-based hierarchies (such as individuals nested within neighborhoods, for example Jones, Johnston and Pattie 1992), and temporal hierarchies (such as panel data and time-series cross-sectional (TSCS) data,[1] in

* Andrew Bell is a lecturer and PhD candidate, School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS (andrew.bell@bristol.ac.uk). Kelvyn Jones is co-director of the Centre for Multilevel Modelling, and Professor of Quantitative Human Geography, School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS (kelvyn.jones@bristol.ac.uk). Thanks to Fiona Steele, Paul Clarke, Malcolm Fairbrother, Alastair Leyland, Mark Bell, Ron Johnston, George Leckie, Dewi Owen, Nathaniel Beck, Chris Adolph, and Thomas Plümper for their help and advice. Also, thanks to the two anonymous reviewers for their suggestions. All mistakes are our own.

[1] The difference between TSCS and panel data lies partly in its sample structure: TSCS data has comparatively few higher-level entities (usually groups of individuals such as countries, rather than individuals) and comparatively many measurement occasions (Beck and Katz 1995). In addition, TSCS data, used mainly in political science, often contains more slowly changing, historically determined variables (such as GDP per capita), and researchers using it are often more interested in specific effects in specific higher-level entities. This makes the issues we discuss here particularly important to researchers using TSCS data.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

which measurement occasions are nested within entities such as individuals or countries (see Beck 2007)). While this article is particularly concerned with the latter, its arguments apply equally to all forms of hierarchical data.[2]

One problem with the disciplinary divides outlined above is that much of the debate between the two methods has remained separated by subject boundaries; the two sides of the debate often seem to talk past each other. This is a problem, because we believe that both sides are making important points that are currently not taken seriously by their counterparts. Since this article draws on a wide, multidisciplinary literature, we hope it will help inform each side of the relative merits of both sides of the argument.

Yet we take the strong, and rather heterodox, view that there are few, if any, occasions in which FE modeling is preferable to RE modeling. If the assumptions made by RE models are correct, RE would be the preferred choice because of its greater flexibility and generalizability, and its ability to model context, including variables that are only measured at the higher level. We show in this article that the assumptions made by RE models, including the exogeneity of covariates and the Normality of residuals, are at least as reasonable as those made by FE models when the model is correctly specified. Unfortunately, this correct formulation is used all too rarely (Fairbrother 2013), despite being fairly well known (it is discussed in numerous econometrics textbooks (Greene 2012; Wooldridge 2002), if rather too briefly). Furthermore, we argue that, in controlling out context, FE models effectively cut out much of what is going on—goings-on that are usually of interest to the researcher, the reader and the policy maker. We contend that models that control out, rather than explicitly model, context and heterogeneity offer overly simplistic and impoverished results that can lead to misleading interpretations.

This article's title has two meanings. First, we hope to explain the technique of FE estimation to those who use it too readily as a default option without fully understanding what they are estimating and what they are losing by doing so. Second, we show that while the fixed dummy coefficients in the FE model are measured unreliably, RE models can explain (and thus reveal) specific differences between higher-level entities.

This article has three distinctive contributions. First, it is a central attack on the dominant method (FE) in much of the quantitative social sciences. Second, it advocates an alternative approach to endogeneity, in which its causes (separate 'within' and 'between' effects) are modeled explicitly. Third, it emphasizes the importance of modeling heterogeneity (not just overall mean effects) using random coefficients and cross-level interactions.

It is important to reiterate that our recommendations are not entirely one-sided: our proposed formulation is currently not used enough,[3] and in many disciplines endogeneity

[2] Indeed, this includes non-hierarchical data with cross-classified or multiple membership structures (see Snijders and Bosker 2012, 205).

[3] Endogeneity is notable in its absence from multilevel modeling quality checklists (such as Ferron *et al.* 2008). Indeed, the following Google scholar 'hits' of combinations of terms (24 April 2012) tells its own story:

Terms	With ''Hausman''	With ''Mundlak''
''Fixed effects''	25,000	1,960
''Random effects''	18,900	1,610
''Multilevel''	2,400	170

The multilevel modeling literature has not significantly engaged with the Mundlak formulation or the issue of endogeneity.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

is often ignored. Furthermore, we want to be clear that the model is no panacea; there will remain biases in the estimates of higher-level effects if potential omitted variables are not identified, which needs to be considered carefully when the model is interpreted. However, our central point remains: a well-specified RE model can be used to achieve everything that FE models achieve, and much more besides.

THE PROBLEM OF HIERARCHIES IN DATA, AND THE RE SOLUTION

Many research problems in the social sciences have a hierarchical structure; indeed, "once you know hierarchies exist, you see them everywhere" (Kreft and De Leeuw 1998, 1). Such hierarchies are produced because the population is hierarchically structured—voters at Level 1 are nested in constituencies at Level 2—and/or a hierarchical structure is imposed during data collection so that, for example in a longitudinal panel, there are repeated measures at Level 1 nested in individuals at Level 2. In the discussion that follows, and to make things concrete, we use 'higher-level entities' to refer to Level 2, and occasions to refer to Level 1. Consequently, time-varying observations are measured at Level 1 and time-invariant observations at Level 2; the latter are unchanging attributes. Thus in a panel study, higher-level entities are individuals, and time-invariant variables may include characteristics such as gender. In a TSCS analysis, the higher-level entities may be countries, and time-invariant variables could be whether they are located in the global south.[4]

The technical problems of analyzing hierarchies in data are well known. Put briefly, standard 'pooled' linear regression models assume that residuals are independently and identically distributed. That is, once all covariates are considered, there are no further correlations (that is, dependence) between measures. Substantively, this means that the model assumes that any two higher-level entities are identical, and thus can be completely 'pooled' into a single population. With hierarchical data, particularly with temporal hierarchies that are often characterized by marked dependence over time, this is a patently unreasonable assumption. Responses for measurement occasions within a given higher-level entity are often related to each other. As a result, the effective sample size of such datasets is much smaller than a simple regression would assume: closer to the number of higher-level entities (individuals or countries) than the number of lower-level units (measurement occasions). As such, standard errors will be incorrect[5] if this dependence is not taken into account (Moulton 1986).

The RE solution to this dependency is to partition the unexplained residual variance into two: higher-level variance between higher-level entities and lower-level variance within these entities, between occasions. This is achieved by having a residual term at each level; the higher-level residual is the so-called random effect. As such, a simple standard RE model would be:

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + e_{ij},$$

where

$$\beta_{0j} = \beta_0 + \beta_2 z_j + u_j.$$

[4] Duncan, Jones and Moon (1998) develop this perspective, whereby a range of research questions and different research designs is seen as having a hierarchical or more complex structure.

[5] Standard errors will usually be underestimated in pooled ordinary least squares (OLS), which ignores the hierarchical structure but can also be biased up (see Arceneaux and Nickerson 2009, 185).

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

These are the 'micro' and 'macro' parts of the model, respectively, and they are estimated together in a combined model that is formed by substituting the latter into the former:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 z_j + (u_j + e_{ij}), \tag{1}$$

where y_{ij} is the dependent variable. In the 'fixed' part of the model, β_0 is the intercept term, x_{1ij} is a (series of) covariate(s) that are measured at the lower (occasion) level with coefficient β_1, and z_j is a (series of) covariate(s) measured at the higher level with coefficient β_2. The 'random' part of the model (in brackets) consists of u_j, the higher-level residual for higher-level entity j, allowing for differential intercepts for higher-level entities, and e_{ij}, the occasion-level residual for occasion i of higher-level entity j. The u_j term is in effect a measure of 'similarity' that allows for dependence, as it applies to all the repeated measures of a higher-level entity. The variation that occurs at the higher level (including u_j and any time-invariant variables) is considered in terms of the (smaller) higher-level entity sample size, meaning that the standard errors are correct. By assuming that u_j and e_{ij} are Normally distributed, an overall measure of their respective variances can be estimated:

$$\begin{aligned} u_j &\sim N(0, \sigma_u^2) \\ e_{ij} &\sim N(0, \sigma_e^2). \end{aligned} \tag{2}$$

As such, we can say that we are 'partially pooling' our data by assuming that our higher-level entities, though not identical, come from a single distribution σ_u^2—which is estimated from the data, much like the occasion-level variance σ_e^2—and can itself be interpreted substantively.

These models must not only be specified, but also estimated on the basis of assumptions. Beck and Katz (2007) show that, with respect to TSCS data, RE models perform well, even when the Normality assumptions are violated.[6] Therefore they are preferred to both 'complete pooling' methods, which assume no differences between higher-level entities and FE, which do not allow for the estimation of higher-level, time-invariant parameters or residuals (see below). Shor *et al.* (2007) use similar methods, but estimated using Bayesian Markov Chain Monte Carlo (MCMC) (rather than maximum likelihood) estimation, which they find produces estimates that are as good, or better than,[7] maximum likelihood RE and other methods.

THE PROBLEM OF OMITTED VARIABLE BIAS AND ENDOGENEITY IN RE MODELS

Considering this evidence, one must consider why RE is not employed more widely, and remains rarely used in disciplines such as economics and political science. The answer lies in the exogeneity assumption of RE models: that the residuals are independent of the covariates; in particular, the assumptions concerning the occasion-level covariates and the

[6] Outliers, however, are a different matter, but these can be dealt with using dummy variables for those outliers in an RE framework.

[7] The reason for this is that there is 'full error propagation' in Bayesian estimation, as the uncertainty in both constituent parts of the model is taken into account, so that the variances of the random part are estimated on the basis that the fixed part are estimates and not known values, and vice versa. Simulations have shown that the improvement of MCMC-estimated models over likelihood methods is greatest when there are there a small number of higher-level units, for example few countries (Browne and Draper 2006; Stegmueller 2013).

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

two variance terms, such that:

$$E(u_j|x_{ij}, z_j) = E(e_j|x_{ij}, z_j) = 0.$$

In most practical applications, this is synonymous with:

$$\begin{aligned} Cov(x_{ij}, u_j) &= 0 \\ Cov(x_{ij}, e_{ij}) &= 0. \end{aligned} \tag{3}$$

The fact is that the above assumptions[8] often do not hold in many standard RE models as formulated in Equation 1. Unfortunately, little attention has been paid to the substantive reasons why not. Indeed, the discovery of such endogeneity has regularly led to the abandonment of RE in favor of FE estimation, which models out higher-level variance and makes any correlations between that higher-level variance and covariates irrelevant, without considering the source of the endogeneity. This is unfortunate, because the source of the endogeneity is often itself interesting and worthy of modeling explicitly.

This endogeneity most commonly arises as a result of multiple processes related to a given time-varying covariate.[9] In reality, such covariates contain two parts: one that is specific to the higher-level entity that does not vary between occasions, and one that represents the difference between occasions, within higher-level entities:

$$x_{ij} = x_j^B + x_{ij}^W. \tag{4}$$

These two parts of the variable can have their own different effects: called 'between' and 'within' effects, respectively, which together comprise the total effect of a given Level 1, time-varying variable. This division is inherent to the hierarchical structure present in both FE and RE models.

In Equation 1 above, it is assumed that the within and between effects are equal (Bartels 2008). That is, a one-unit change in x_{ij} for a given higher-level entity has the same statistical effect (β_1) as being a higher-level entity with an inherent time-invariant value of x_{ij} that is 1 unit greater. While this might well be the case, there are clearly many examples in which this is unlikely. Considering an example of TSCS country data, an increase in equality may have a different effect to generally being an historically more equal country, for example due to some historical attribute(s) (such as colonialism) of that country. Indeed, as Snijders and Bosker (2012, 60) argue, "it is the rule rather than the exception that within-group regression coefficients differ from between-group coefficients."

Where the within and between effects are different, β_1 in Equation 1 will be an uninterpretable weighted average of the two processes (Krishnakumar 2006; Neuhaus and Kalbfleisch 1998; Raudenbush and Bryk 2002, 137), while variance estimates are also affected (Grilli and Rampichini 2011). This can be thought of as omitted variable bias (Bafumi and Gelman 2006; Palta and Seplaki 2003); because the between effect is omitted, β_1 attempts to account for both the within and the between effect of the covariate on the response, and if the two effects are different, it will fail to account fully for either. The variance that is left unaccounted for will be absorbed into the error terms u_{0j} and e_{0ij}, which will consequently both be correlated with the covariate, violating the assumptions

[8] An additional assumption implied here is that $Cov(z_j, u_{0j})=0$. While this is an important assumption, it is not a good reason to choose FE, as the latter cannot estimate the effect of z_j at all.

[9] While there may be other additional causes for correlation between x_{ij} and e_{ij}, this is the only cause of correlation between x_{ij} and u_j.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

of the RE model. When viewed in these terms, it is clear that this is a substantive inadequacy in the theory behind the RE model, rather than simply a statistical misspecification (Spanos 2006) requiring a technical fix.

The word 'endogenous' has multiple forms, causes and meanings. It can be used to refer to bias caused by omitted variables, simultaneity, sample selection or measurement error (Kennedy 2008, 139). These are all different problems that should be dealt with in different ways. Therefore we consider the term to be misleading and, having explained it, do not use it in the rest of the article. The form of the problem that this article deals with is described rather more clearly by Li (2011) as 'heterogeneity bias,' and we use that terminology from now on. Our focus on this does not deny the existence of other forms of bias that cause and/or result from correlated covariates and residuals.[10]

FE ESTIMATION

The rationale behind FE estimation is simple and persuasive, which explains why it is so regularly used in many disciplines. To avoid the problem of heterogeneity bias, all higher-level variance, and with it any between effects, are controlled out using the higher-level entities themselves (Allison 2009), included in the model as dummy variables D_j:

$$y_{ij} = \sum_{j=1}^{j} \beta_{0j} D_j + \beta_1 x_{ij} + e_{ij}. \tag{5}$$

To avoid having to estimate a parameter for each higher-level unit, the mean for higher-level entity is taken away from both sides of Equation 5, such that:

$$(y_{ij} - \bar{y}_j) = \beta_1 (x_{ij} - \bar{x}_j) + (e_{ij} - \bar{e}_j). \tag{6}$$

Because FE models only estimate within effects, they cannot suffer from heterogeneity bias. However, this comes at the cost of being unable to estimate the effects of higher-level processes, so RE is often preferred where the bias does not exist. In order to test for the existence of this form of bias in the standard RE model as specified in Equation 1, the Hausman specification test (Hausman 1978) is often used. This takes the form of comparing the parameter estimates of the FE and RE models (Greene 2012; Wooldridge 2002) via a Wald test of the difference between the vector of coefficient estimates of each.

The Hausman test is regularly used to test whether RE can be used, or whether FE estimation should be used instead (for example Greene 2012, 421). However, it is problematic when the test is viewed in terms of fixed and random effects, and not in terms of what is actually going on in the data. A negative result in a Hausman test tells us only that the between effect is not significantly biasing an estimate of the within effect in Equation 1. It "is simply a diagnostic of one particular assumption behind the estimation procedure usually associated with the random effects model… it does not address the decision framework for a wider class of problems" (Fielding 2004, 6). As we show later, the RE model that we propose in this article solves the problem of heterogeneity bias described above, and so makes the Hausman test, as a test of FE against RE, redundant.

[10] Although we do deny that FE models are any better able than RE models to deal with these other forms of bias.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

It is "neither necessary nor sufficient" (Clark and Linzer 2012, 2) to use the Hausman test as the sole basis of a researcher's ultimate methodological decision.

PROBLEMS WITH FE MODELS

Clearly, there are advantages to the FE model of Equations 5-6 over the RE models in Equation 1. By clearing out any higher-level processes, the model deals only with occasion-level processes. In the context of longitudinal data, this means considering differences over time, controlling out higher-level differences and processes absolutely, and supposedly "getting rid of proper nouns" (King 2001, 504)—that is, distinctive, specific characteristics of higher-level units. This is why it has become the "gold standard" method (Schurer and Yong 2012, 1) in many disciplines. There is no need to worry about heterogeneity bias, and β_1 can be thought to represent the 'causal effect.'

However, by removing the higher-level variance, FE models lose a large amount of important information. No inferences can be made about that higher-level variance, including whether or not it is significant (Schurer and Yong 2012, 14). Thus it is impossible to measure the effects of time-invariant variables at all, because all degrees of freedom at the higher level have been consumed. Where time-invariant variables are of particular interest, this is obviously critical. And yet even in these situations, researchers have suggested the use of FE, on the basis of a Hausman test. For example, Greene's (2012, 420) textbook gives an example of a study of the effect of schooling on future wages:

> The value of the [Hausman] test statistic is 2,636.08. The critical value from the chi-squared table is 16.919 so the null hypothesis of a random effects model is rejected. We conclude that the fixed effects model is the preferred specification for these data. This is an unfortunate turn of events, as the main object of the study is the impact of education, which is a time invariant variable in this sample.

Unfortunate indeed! To us, explicating a method that fails to answer your research question is nonsensical. Furthermore, because the higher-level variance has been controlled out, any parameter estimates for time-varying variables deal with only a small subsection of the variance in that variable. Only within effects can be estimated (that is, the lower-level relationship net of any higher-level attributes), and so nothing can be said about a variable's between effects or a general effect (if one exists); studies that make statements about such effects on the basis of FE models are over-interpreting their results.

Beck and Katz (Beck 2001; Beck and Katz 2001) consider the example of the effect of a rarely changing variable, democracy, on a binary variable representing whether a pair of countries is at peace or at war (Green, Kim and Yoon 2001; see also King 2001; Oneal and Russett 2001). They show that estimates obtained under FE fail to show any relation between democracy and peace because it filters out all the effects of unchanging, time-invariant peace, which has an effect on time-variant democracy. In other words, time-invariant processes can have effects on time-varying variables, which are lost in the FE model. Countries that do not change their political regime, or do not change their state of peace (that is, most countries) are effectively removed from the sample. While this problem particularly applies to rarely changing, almost time-invariant variables (Plümper and Troeger 2007), any time-varying covariate can have such time-invariant 'between' effects, which can be different from time-varying effects of the same variable, and these processes cannot be assessed in an FE model. Only an RE model can allow these processes to be modeled simultaneously.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

PLÜMPER AND TROEGER'S FE VECTOR DECOMPOSITION

A method proposed by Plümper and Troeger (2007) allows time-invariant variables to be modeled within the framework of the FE model. They use an FE model before 'decomposing' the vector of fixed-effects dummies into that explained by a given time-invariant (or rarely changing) variable, and that which is not. They begin by estimating a standard dummy variable FE model as in Equation 5:

$$y_{ij} = \sum_{j=1}^{j} \beta_{0j} D_j + \beta_1 x_{ij} + e_{ij}. \tag{7}$$

Here, D_j is a series of higher-level entity dummy variables, each with an associated intercept coefficient β_{0j}. Plümper and Troeger then regress in a separate higher-level model the vector of these estimated FE coefficients on time-invariant variables, such that

$$\beta_{0j} = \beta_0 + \beta_2 z_j + R_j, \tag{8}$$

where z_j is a (series of) higher-level variable(s) and R_j is the residual. This equation can be rearranged so that, once estimated, the values of R_j can be estimated as

$$R_j = \beta_{0j} - \beta_2 x_{2j} - \beta_0. \tag{9}$$

Finally, Equation 8 is substituted into Equation 7 such that

$$\begin{aligned} y_{ij} &= \sum_{j=1}^{j} (\beta_0 + \beta_2 z_j + R_j) D_j + \beta_1 x_{1ij} + e_{ij} \\ y_{ij} &= \beta_0 + \beta_1 x_{1ij} + \beta_2 z_j + \beta_3 R_j + e_{ij}, \end{aligned} \tag{10}$$

where β_3 will equal exactly one (Greene 2012, 405). The residual higher-level variance not explained by the higher-level variable(s) is modeled as a fixed effect, leaving no higher-level variance unaccounted for. As such, the model is very similar to an RE model (Equation 1), which does a similar thing but in a single overall model.[11] Stage 1 (Equation 7) is equivalent to the RE micro model, Stage 2 (Equation 8) to the macro model and Stage 3 (Equation 10) to the combined model. Just as with RE, the higher-level residual is assumed to be Normal (from the regression in Equation 8). What it does do differently is also control out any between effect of x_{1ij} in the estimation of β_1, meaning that these estimates will only include the within effect, as in standard FE models.

The FE vector decomposition (FEVD) estimator has been criticized by many in econometrics, who argue that the standard errors are likely to be incorrectly estimated (Breusch *et al.* 2011a, 2011b; Greene 2011a, 2011b, 2012). Plümper and Troeger (2011) provide a method for calculating more appropriate standard errors, and so the FEVD model does work (at least with balanced data) when this method is utilized. However, our concern is that it retains many of the other flaws of FE models, which we have outlined above. It remains much less generalizable than an RE model—it cannot be

[11] In the early stages of the development of the multilevel model, a very similar process to the two-stage FEVD model was used to estimate processes at multiple levels (Burstein *et al.* 1978; Burstein and Miller 1980) before being superseded by the modern multilevel RE model in which an overall model is estimated (Raudenbush and Bryk 1986). As Beck (2005, 458) argues: "perhaps at one time it could have been argued that one-step methods were conceptually more difficult, but, given current training, this can no longer be an excuse worth taking seriously."

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

extended to three (or more) levels, nor can coefficients be allowed to vary (as in a random coefficients model). It does not provide a nice measure of variance at the higher level, which is often interesting in its own right. Finally, it is heavily parameterized, with a dummy variable for each higher-level entity in the first stage, and thus can be relatively slow to run when there is a large number of higher-level units.

Plümper and Troeger also attempt to estimate the effects of 'rarely changing' variables, and their desire to do so using FE modeling suggests that they do not fully appreciate the difference between within and between effects. While they do not quantify what 'rarely changing' means, their motivation is to get significant results where FE produces insignificant results. FE models only estimate within effects, and so an insignificant effect of a rarely changing variable should be taken as saying that there is no evidence for a within effect of that variable. When Plümper and Troeger use FEVD to estimate the effects of rarely changing variables, they are in fact estimating between effects. Using FEVD to estimate the effects of rarely changing variables is not a technical fix for the high variance of within effects in FE models—it is shifting the goalposts and measuring something different. Furthermore, if between effects of rarely changing variables are of interest, then there is no reason why the between effects of other time-varying variables would not be, and so these should potentially be modeled as time-invariant variables as well.

AN RE SOLUTION TO HETEROGENEITY BIAS

What is needed is a solution, within the parsimonious, flexible RE framework, which allows for heterogeneity bias not to simply be corrected, but for it to be explicitly modeled. As it turns out, the solution is well documented, starting from an article by Mundlak (1978a). By understanding that heterogeneity bias is the result of attempting to model two processes in one term (rather than simply a cause of bias to be corrected), Mundlak's formulation simply adds one additional term in the model for each time-varying covariate that accounts for the between effect: that is, the higher-level mean. This is treated in the same way as any higher-level variable. Therefore in the simple case, the micro and macro models, respectively, are:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$

and

$$\beta_{0j} = \beta_0 + \beta_2 z_j + \beta_3 \bar{x}_j + u_j.$$

This combines to form

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_3 \bar{x}_j + \beta_2 z_j + (u_j + e_{ij}), \tag{11}$$

where x_{ij} is a (series of) time-variant variables, while $\bar{x}_j$ is the higher-level entity j's mean; as such, the time-invariant component of those variables (Snijders and Bosker 2012, 56). Here β_1 is an estimate of the within effect (as the between effect is controlled by $\bar{x}_j$); β_3 is the 'contextual' effect that explicitly models the difference between the within and between effects. Alternatively, this can be rearranged by writing β_3 explicitly as this difference (Berlin *et al.*, 1999):

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + (\beta_4 - \beta_1)\bar{x}_j + \beta_2 z_j + (u_j + e_{ij}).$$

This rearranges to:

$$y_{ij} = \beta_0 + \beta_1(x_{ij} - \bar{x}_j) + \beta_4 \bar{x}_j + \beta_2 z_j + (u_j + e_{ij}). \tag{12}$$

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

Now β_1 is the within effect and β_4 is the between effect of x_{ij} (Bartels 2008; Leyland 2010). This 'within-between' formulation (see Table 1) has three main advantages over Mundlak's original formulation. First, with temporal data it is more interpretable, as the within and between effects are clearly separated (Snijders and Bosker 2012, 58). Second, in the first formulation, there is correlation between x_{ij} and $\bar{x}_j$; by group mean centering x_{ij}, this collinearity is lost, leading to more stable, precise estimates (Raudenbush 1989). Finally, if multicollinearity exists between multiple $\bar{x}_j$s and other time-invariant variables, $\bar{x}_j$s can be removed without the risk of heterogeneity bias returning to the occasion-level variables (as in the within model in Table 1).[12]

Just as before, the residuals at both levels are assumed to be Normally distributed:

$$u_j \sim N(0, \sigma_u^2)$$
$$e_{ij} \sim N(0, \sigma_e^2).$$

As can be seen, this approach is algebraically similar to the FEVD estimator (Equation 10)—the mean term(s) are themselves interpretable time-invariant variables[13] (Begg and Parides 2003), measuring the propensity of a higher-level entity to be x_{ij} (in the binary case) or the average level of x_{ij} (in the continuous case) across the sample time period.[14] There are a few differences. First, estimates for the effects of time-invariant variables are controlled for by the means of the time-varying variables. While this could be done in Stage 2 of FEVD, it rarely is; nor is it suggested by Plümper and Troeger (2007) except for in the case of 'rarely changing' variables. Second, correct standard errors are automatically calculated, accounting for "multiple sources of clustering" (Raudenbush 2009, 473). Crucially, there can be no correlation between the group mean centered covariate and the higher-level variance, because each group mean centered covariate has a mean of zero for each higher-level entity j. Equally, at the higher level, the mean term is no longer constrained by Level 1 effects, so it is free to account for all the higher-level variance associated with that variable. As such, the estimate of β_1 in

[12] Instead of using the higher-level unit mean (an aggregate variable), Clarke *et al.* (2010) suggest using global (Diez-Roux 1998) unit characteristics that are correlated with that mean. These global variables express the causal mechanism underlying the association expressed by β_4, which may not be linear, as is assumed by Models 10 and 11. Including $\bar{x}_j$ would be over-controlling in this case, and such a model has a different interpretation of the higher-level residual, but it is harder to reliably control out all (or even most) of the between effect from the within effect without using $\bar{x}_j$ (Clarke *et al.* 2010) in Equation 11. However, this is not a problem when using the formulation in Equation 12, as the within variable is already group mean centered, so the inclusion of $\bar{x}_j$ is optional depending on the research question at hand, as in the 'within' model in Table 1.

[13] Because of this, the number of higher-level units in the sample must be considered, and as such caution should be taken regarding how many higher-level variables (including $\bar{x}_j$s) the model can estimate reliably. The MLPowSim software (Browne *et al.* 2009) can be used to judge this in the research design phase.

[14] Note that when interpreting these terms, we are usually interested in an individual's general, latent characteristics that are invariant beyond the sample period. From this perspective, it is not the case that we are conditioning on the future (as argued by Kravdal 2011), any more than with any other time-invariant variable. However, because these means are measured from a finite sample, they are subject to measurement error and their coefficients are subject to bias. This can be corrected for by shrinking them back toward the grand mean, in a similar way to the residuals, through Equation 13 (see Grilli and Rampichini 2011; Shin and Raudenbush 2010). However, more detailed explication of this is beyond the scope of this article.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

TABLE 1 *RE Model Formulations Considered*

Model Name	Fixed Part of Model
Standard RE	$y_{ij} = \beta_0 + \beta_1 x_{1ij}$
Mundlak	$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_3 \bar{x}_j$
Within-Between	$y_{ij} = \beta_0 + \beta_1 (x_{ij} - \bar{x}_j) + \beta_4 \bar{x}_j$
Within	$y_{ij} = \beta_0 + \beta_1 (x_{ij} - \bar{x}_j)$

Note: the within-between and within-RE model involve group mean centering of the covariate. This is different from centering on the grand mean, which has a different purpose: to keep the value of the intercept (β_0) within the range of the data and to aid convergence of the model. Indeed, x_{1ij} and $\bar{x}_j$ can be grand mean centered if required (the group mean-centered variables will already be centered on their grand mean by definition).

Equations 11 and 12 above will be identical to that obtained by FE, as Mundlak (1978a, 70) stated clearly:

When the model is properly specified, the GLSE [that is RE] is identical to the 'within' [that is, the FE] estimator. Thus there is only one estimator. The whole literature which has been based on an imaginary difference between the two estimators … is based on an incorrect specification which ignores the correlation between the effects and the explanatory variables.

While it is still possible that there is correlation between the group mean centered x_{ij} and e_{ij}, and between $\bar{x}_j$ (and other higher-level variables) and u_j (Kravdal 2011), this is no more likely than in FE models for the former and aggregate regression for the latter, because we have accounted for the key source of this correlation by specifying the model correctly (Bartels 2008).[15] After all, "all models are wrong; the practical question is how wrong do they have to be to not be useful" (Box and Draper 1987, 74). How useful the model is depends, as with any model, on how well the researcher has accounted for possible omitted variables, simultaneity or other potential model misspecifications.

We see the FE model as a constrained form of the RE model,[16] meaning that the latter can encompass the former but not vice versa. By using the RE configuration, we keep all the advantages associated with RE modeling.[17] First, the 'problem' of heterogeneity bias across levels is not simply solved; it is explicitly modeled. The effect of x_{ij} is separated into two associations, one at each level, which are interesting, interpretable and relevant to the researcher (Enders and Tofighi 2007, 130). Second, by assuming Normality of the higher-level variance, the model need only estimate a single term for each level (the variance), which are themselves useful measures, allowing calculation of the variance partitioning coefficient (VPC),[18] for

[15] If covariates remain correlated with residuals (for example as a result of simultaneity, or other omitted variables), they can potentially be dealt with within this RE framework through other means, such as instrumental variable methods (Heckman and Vytlacil 1998) using simultaneous equations (Steele *et al*. 2007)—assuming, of course, that appropriate instruments can be found. While all heterogeneity bias of lower-level variables has been dealt with, a variant of the Hausman-Taylor IV estimator (Greene 2012, 434; Hausman and Taylor 1981) can be used to deal with correlated time-invariant variables (Chatelain and Ralf 2010).

[16] Demidenko (2004, 54–5) proves that the FE model is equivalent to an RE model in which the higher-level variance is constrained to be infinite.

[17] Note that it is still necessary to use RE estimation methods (rather than OLS) to calculate correct SEs.

[18] The VPC is the proportion of variance that occurs at Level 2. In the simple two-level RE case, it is calculated as $\frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$, and is a standardized measure of the similarity between higher-level units.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

example. Further, higher-level residuals (conditional on the variables in the fixed part of the model) are precision weighted or shrunken by multiplying by the higher-level entity's reliability λ_j (see Snijders and Bosker 2012, 62),[19] calculated as:

$$\lambda_j = \frac{\sigma_u^2}{\sigma_u^2 + \left(\frac{\sigma_e^2}{n_j}\right)}, \tag{13}$$

where n_j is the sample size of higher-level entity j, σ_u^2 is the between-entity variance and σ_e^2 is the variance within higher-level entities, between occasions. One can thus estimate reliable residuals for each higher-level entity that are less prone to measurement error than FE dummy coefficients. By partially pooling by assuming that u_j comes from a common distribution with a variance that has been estimated from the data, we can obtain much more reliable predictions for individual higher-level units (see Rubin 1980 for an early example of this).[20] While this is rarely of interest in individual panel data, it is likely to be of interest with TSCS data with repeated measures of countries.

The methods we are proposing here are beginning to be taken up by researchers under the guise of a 'hybrid' or 'compromise' approach between FE and RE (Allison 2009, 23; Bartels 2008; Greene 2012, 421). Yet this is to misrepresent the nature of the model. There is nothing FE-like about the model at all—it is an RE model with additional time-invariant predictors. Perhaps as a consequence of this potentially misleading terminology, many of those who use such models fail to recognize its potential as an RE model. Allison (2009, 25), for example, argues that the effects of the mean variables ($\bar{x}_j$) "are not particularly enlightening in themselves," while many have suggested using the formulation as a form of the Hausman test and use the results to choose between fixed and random effects (Allison 2009, 25; Baltagi 2005; Greene 2012, 421; Hsiao 2003, 50; Wooldridge 2002, 290, 2009). Thus β_3 in Equation 11 is thought of simply as a measure of 'correlation' between x_{ij} and u_j and when $\beta_3 \neq 0$ in Equation 11, or $\beta_1 \neq \beta_4$ in Equation 12, the Hausman test fails and it is argued that the FE model should be used. It is clear to us (and to Skrondal and Rabe-Hesketh 2004, 53; Snijders and Berkhof 2007, 145), however, that the use of this model makes that choice utterly unnecessary.

To reiterate, the Hausman test is not a test of FE versus RE; it is a test of the similarity of within and between effects. An RE model that properly specifies the within and between effects will provide identical results to FE, regardless of the result of a Hausman test. Furthermore, between effects, other higher-level variables and higher-level residuals (none of which can be estimated with FE) should not be dismissed lightly; they are often enlightening, especially for meaningful entities such as countries. For these reasons, and the ease with which they can now be fitted in most statistical software packages, RE models are the obvious choice.

SIMULATIONS

We now present simulations results that show that, under a range of situations, the RE solution that we propose performs at least as well as the alternatives on offer—it predicts

[19] A detailed comparison between the fixed and random effects estimates is given algebraically and empirically in Jones and Bullen (1994)

[20] We are assuming here that higher-level units come from a single distribution. This is usually a reasonable assumption, and it can be readily evaluated.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

the same effects as both FE and FEVD for time-varying variables, and the same results as FEVD for time-invariant variables. Furthermore, the simulations show that standard errors are poorly estimated by FEVD when there is imbalance in the data.

The simulations[21] are similar to those conducted by Plümper and Troeger (2007), using the following underlying data generating process:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_{3C} \bar{x}_{3j} + \beta_4 z_{1j} + \beta_5 z_{2j} + \beta_6 z_{3j} + u_j + e_{ij}, \tag{14}$$

where

$\beta_0 = 1$, $\beta_1 = 0.5$, $\beta_2 = 2$, $\beta_3 = -1.5$, $\beta_4 = -2.5$, $\beta_5 = 1.8$, $\beta_6 = 3$.

In order to simulate correlation between x_{3ij} and u_j, the value of β_{3C} varies (-1, 0, 1, 2) between simulations. This parameter is also estimated in its own right—as we have argued, it is often of substantive interest in itself. We also vary the extent of correlation between z_{3j} and u_j (-0.2, 0, 0.2, 0.4, 0.6). All variables were generated to be Normally distributed with a mean of zero—fixed part variables with a standard deviation of 1, Level 1 and 2 residuals with standard deviations of 3 and 4, respectively. In addition, we varied the sample size—both the number of Level 2 units (100, 30) and the number of time points (20, 70). Additionally we tested the effect of imbalance in the data (no missingness, 50 percent missingness in all but five of the higher-level units) on the performance of the various estimators. The simulations were run in Stata using the xtreg and xtfevd commands.

For each simulation scenario, the data were generated and models estimated 1,000 times, and three quantities were calculated: bias, root mean square error (RMSE) and optimism, calculated as in Shor *et al.* (2007) and in line with the simulations presented by Plümper and Troeger (2007, 2011). Bias is the mean of the ratios of the true parameter value to the estimated parameter, and so a value of 1 suggests that the model estimates are, on average, exactly correct. RMSE also assesses bias, as well as efficiency: the lower the value, the more accurate and precise the estimator. Finally, optimism evaluates how the standard errors compare to the true sampling variability of the simulations; values greater than 1 suggest that the estimator is overconfident in its estimates, while values below 1 suggest that they are more conservative than necessary.

Table 2 presents the results from some permutations of the simulations when the data is balanced. As can be seen, and as expected, the standard RE estimator is outperformed by the other estimators, because of bias resulting from the omission of the between effect associated with X3 from the model. It can also be seen that the within-between RE model (REWB) performs at least as well as both FE and FEVD for all three measures. What is more surprising is the effect of data imbalance (see Table 3) on the performance of the estimators—while for RE, FE and REWB the results remain much the same, the standard errors are estimated poorly by the FEVD—too high (type 2 errors) for lower-level variables and too low (type 1 errors) for higher-level variables. The online appendices show that this result is repeated for all the simulation scenarios that we tested, regardless of the size of correlations present in the data and the data sample size. It is clear that it would be unwise to use the FEVD with unbalanced data, and even when data is balanced, Mundlak's (1978a) claim, that the models will produce identical results, is fully justified.

Having shown that the REWB model produces results that are at least as unbiased as alternatives including FE and FEVD, the question remains why one should choose the

[21] Do files for the replication of these simulations can be found in the online appendices.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

TABLE 2 *RMSE, Bias and Optimism from the Simulation Results over Five Permutations (times 1,000 estimations)*

	FE	RE	REWB	FEVD
Bias (perfect =1)				
β_3 (within effect of x_{3ij})	0.998	0.978	0.998	0.998
β_6 (effect of t-invariant z_{3j})		1.262	1.261	1.262
β_{3B} (between effect of x_{3ij})			0.969	
RMSE (perfect=0)				
β_3	0.127	0.130	0.127	0.127
β_6		1.461	1.455	1.463
β_{3B}			0.778	
Optimism (perfect =1)				
β_3	1.007	1.010	1.006	1.007
β_6		1.003	0.975	1.029
β_{3B}			1.003	

Note: units (30), time periods (20) and the contextual effect size (1) are kept constant. Correlation between Z3 and *uj* varies, with values -0.2, 0, 0.2, 0.4 and 0.6. The data are balanced.

TABLE 3 *RMSE, Bias and Optimism from the Simulation Results over Five Permutations (times 1,000 estimations)*

	FE	RE	REWB	FEVD
Bias (perfect =1)				
β_3 (within effect of x_{3ij})	1.000	0.966	1.000	1.000
β_6 (effect of t-invariant z_{3j})		1.267	1.267	1.267
β_{3B} (between effect of x_{3ij})			0.969	
RMSE (perfect=0)				
β_3	0.165	0.170	0.165	0.165
β_6		1.484	1.474	1.518
β_{3B}			0.793	
Optimism (perfect =1)				
β_3	0.978	0.987	0.977	0.780
β_6		1.030	1.003	1.333
β_{3B}			1.010	

Note: units (30), time periods (20) and the contextual effect size (1) are kept constant. Correlation between Z3 and *uj* varies, with values -0.2, 0, 0.2, 0.4 and 0.6. The data are unbalanced.

RE option over these others. If higher-level variables and/or shrunken residuals are not of substantive interest, why not simply estimate an FE regression (or the FEVD estimator if time-invariant variables or other between effects happen to be of interest and the data is balanced)? The answer is twofold. First, with the ability to estimate both effects in a single model (rather than the three steps of the FEVD estimator), the RE model is more general than the other models. We believe it is valuable to be able to model things in a single coherent framework. Second, and more importantly, the RE model can be extended to allow for variation in effects across space and time to be explicitly modeled, as we show in the following section. That is, while FE models assume *a priori* that there is a single effect that affects all higher-level units in the same way, the RE framework allows for that

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

assumption to be explicitly tested. This does not simply provide additional results to those already found—failing to do this can lead to results that are seriously and substantively misleading.

EXTENDING THE BASIC MODEL: RANDOM COEFFICIENT MODELS AND CROSS-LEVEL INTERACTIONS

We have argued that the main advantage of RE models is their generalizability and extendibility, and this section outlines one[22] such extension: the random coefficient model (RCM). This allows the effects of β coefficients to vary by the higher-level entities (Bartels 2008; Mundlak 1978b; Schurer and Yong 2012). Heteroscedasticity at the occasion level can also be explicitly modeled by including additional random effects at Level 1. Thus our model could become:

$$y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - \bar{x}_j) + e_{oij} + e_{1ij}(x_{ij} - \bar{x}_j),$$

where

$$\beta_{0j} = \beta_0 + \beta_4\bar{x}_j + \beta_2 z_j + u_{0j}$$
$$\beta_{1j} = \beta_1 + u_{1j}.$$

These equations (one micro and two macro) combine to form:

$$y_{ij} = \beta_0 + \beta_1(x_{ij} - \bar{x}_j) + \beta_4\bar{x}_j + \beta_2 z_j + [u_{0j} + u_{1j}(x_{ij} - \bar{x}_j) + e_{0ij} + e_{1ij} + (x_{ij} - \bar{x}_j)] \quad (15)$$

with the following distributional assumptions:

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u0u1} & \sigma^2_{u1} \end{bmatrix}\right)$$
$$\begin{bmatrix} e_{0ij} \\ e_{1ij} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{e0} & \\ \sigma_{e0e1} & \sigma^2_{e1} \end{bmatrix}\right).$$

These variances and covariances can be used to form quadratic 'variance functions' (Goldstein 2010, 73) to see how the variance varies with $(x_{ij} - \bar{x}_j)$. At the higher level, the total variance is calculated by

$$var[u_{0ij} + u_{1ij}(x_{ij} - \bar{x}_j)] = \sigma^2_{u0} + 2\sigma_{u0u1}(x_{ij} - \bar{x}_j) + \sigma^2_{u1}(x_{ij} - \bar{x}_j)^2 \quad (16)$$

and at Level 1, it is

$$var[e_{0ij} + e_{1ij}(x_{ij} - \bar{x}_j)] = \sigma^2_{e0} + 2\sigma_{e0e1}(x_{ij} - \bar{x}_j) + \sigma^2_{e1}(x_{ij} - \bar{x}_j)^2. \quad (17)$$

These can often be substantively interesting, as well as correct misspecification of a model that would otherwise assume homogeneity at each level (Rasbash *et al.* 2009, 106). As such, even when time-invariant variables are not of interest, the RE model is preferable because it means that "a richer class of models can be estimated" (Raudenbush 2009, 481), and rigid assumptions of FE and FEVD can be relaxed.

[22] Other potential model extensions could include three-level models, or multiple-membership or cross-classified (Raudenbush 2009) data structures.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

RCMs additionally allow cross-level interactions between higher- and lower-level variables. In the TSCS case, that is an interaction between a variable measured at the country level and one measured at the occasion level. This is achieved by extending Equation 15 to, for example:

$$y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - \bar{x}_j) + e_{0ij} + e_{1ij}(x_{ij} - \bar{x}_j)$$
$$\beta_{0j} = \beta_0 + \beta_4\bar{x}_j + \beta_2 z_j + u_{0j}$$
$$\beta_{1j} = \beta_1 + \beta_5\bar{x}_j + u_{1j}$$

which combine to form:

$$\begin{aligned} y_{ij} = \beta_0 + \beta_1(x_{ij} - \bar{x}_j) + \beta_4\bar{x}_j + \beta_5(x_{ij} - \bar{x}_j)\bar{x}_j + \beta_2 z_j \\ + [u_{0j} + u_{1j}(x_{ij} - \bar{x}_j) + e_{0ij} + e_{1ij}(x_{ij} - \bar{x}_j)]. \end{aligned} \tag{18}$$

The models can thus give an indication of whether the effect of a time-varying predictor varies by time-invariant predictors (or vice versa), and this is quantified by the coefficient β_5. Note that these could include interactions between the time-variant and time-invariant parts of the same variable, as is the case above, or could involve other time-invariant variables. The possibility of such interactions is not new (Davis, Spaeth and Huson 1961); they have been an established part of the multilevel modeling literature for many years (Jones and Duncan 1995, 33). While the interaction terms themselves can be included in an FE model (for example see Boyce and Wood 2011; Wooldridge 2009), it is only when they are considered together with the additive effects of the higher-level variable (β_4) that their full meaning can be properly established. This can only be done in a random coefficient model. Such relationships ought to be of interest to any researcher studying time-varying variables. If the effect of a time-varying education policy is different for boys and girls, the researcher needs to know this. It is even conceivable that such relationships could be in opposite directions for different types of higher-level entities. Therefore, an FE study that suggests that a policy generally helps everyone could be hiding the fact that it actually hinders certain types of people. Resources could be wasted applying a policy to individuals who are harmed by it. Following Pawson (2006), we believe that context should be central to any evidence-based policy.

To reiterate this point, even when time-invariant variables are not directly relevant to the research question itself, it is important to think about what is happening at the higher level, in a multilevel RE framework. Simpler models that control out context assume that occasion-level covariates have only 'stylized' (see Clark 1998; Kaldor 1961; Solow 1988, 2) mean effects that affect all higher-level entities in exactly the same way. This leads to nice simple conclusions (a policy either works or does not), but it misses out important information about what is going on:

> Continuing to do individual-level analyses stripped out of its context will never inform us about how context may or may not shape individual and ecological outcomes (Subramanian *et al.* 2009a, 355).

An example illustrating the ideas presented in this article can be found online,[23] where we reanalyze the data used by Milner and Kubota (2005) in their FE study of democracy and free trade. A Hausman test would suggest that for this dataset, an FE model should be used; we show that doing so leads to considerably impoverished results.

[23] Note that we are currently preparing a more comprehensive critique of Milner and Kubota's article (Bell *et al.* forthcoming).

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

CONCLUSIONS

In the introduction to his book on FE models, Allison (2009, 2) criticizes an early proponent of RE:

> Such characterisations are very unhelpful in a nonexperimental setting, however, because they suggest that a random effects approach is nearly always preferable. Nothing could be further from the truth.

We have argued in this article that the RE approach is, in fact, nearly always preferable. We have shown that the main criticism of RE, the correlation between covariates and residuals, is readily solvable using the within-between formulation espoused here, although the solution is used all too rarely in RE modeling. This is why, in fact, Allison argues in favor of the same RE formulation that we have used, even though he calls it a 'hybrid' solution. Our strong position is not simply based on finding a technical fix, however. We believe that understanding the role of context (households, individuals, neighborhoods, countries, etc.) that defines the higher level, is usually of profound importance to a given research question—one must model it explicitly—and requires the use of an RE model that analyzes and separates both the within and between components of an effect explicitly, and assesses how those effects vary over time and space rather than assuming heterogeneity away with FE:

> Heterogeneity is not a technical problem calling for an econometric solution but a reflection of the fact that we have not started on our proper business, which is trying to understand what is going on (Deaton 2010, 430).

This point is as much philosophical as it is statistical (Jones 2010). We as researchers are aiming to understand the world. FE models attempt to do this by cutting out much of "what is going on," leaving only a supposedly universal effect and controlling out differences at the higher level. In contrast, an RE approach explicitly models this difference, leading "to a richer description of the relationship under scrutiny" (Subramanian *et al.* 2009b, 373). To be absolutely clear, this is not to say that within-between RE models are perfect—no model is. If there are only a very small number of higher-level units, RE may not be appropriate. As with any model, it is important to consider whether important variables have been omitted and whether causal interpretations are justified, using theory—particularly regarding time-invariant variables. No statistical model can act as a substitute for intelligent research design and forethought regarding the substantive meaning of parameters. However, the advantages of within-between RE over the more restrictive FE are at odds with the dominance of FE as the 'default' option in a number of social science disciplines. We hope this article will go some way toward ending that dominance and stimulating much-needed debate on this issue.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/psrm.2014.7

REFERENCES

Allison, Paul D. 2009. *Fixed Effects Regression Models*. London: Sage.

Arceneaux, Kevin, and David W. Nickerson. 2009. 'Modeling Certainty with Clustered Data: A Comparison of Methods'. *Political Analysis* 17(2):177–90.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

Bafumi, Joseph, and Andrew Gelman. 2006. 'Fitting Multilevel Models when Predictors and Group Effects Correlate'. Paper presented at the Annual Meeting of the Midwest Political Science Association. Chicago, IL. 20-23 April. Available at http://www.stat.columbia.edu/ ~ gelman/research/unpublished/Bafumi_Gelman_Midwest06.pdf, accessed 21 March 2012.

Baltagi, Badi H. 2005. *Econometric Analysis of Panel Data*, 3rd edition. Chichester: Wiley.

Bartels, Brandon L. 2008. 'Beyond "Fixed versus Random Effects": A Framework for Improving Substantive and Statistical Analysis of Panel, Time-Series Cross-Sectional, and Multilevel Data'. Paper presented at the Political Methodology Conference. Ann Arbor, MI. 9-12 July. Available at http://home.gwu.edu/ ~ bartels/cluster.pdf, accessed 1 March 2012.

Beck, Nathaniel. 2001. 'Time-series-cross-section-data: What Have We Learned in the Past Few Years?' *Annual Review of Political Science* 4:271–93.

——. 2005. 'Multilevel Analyses of Comparative Data: A Comment'. *Political Analysis* 13(4):457–8.

——. 2007. 'From Statistical Nuisances to Serious Modeling: Changing How We Think about the Analysis of Time-Series-Cross-Section Data'. *Political Analysis* 15(2):97–100.

Beck, Nathaniel, and Jonathan N. Katz. 1995. 'What to Do (and not to Do) with Time-Series Cross-Section Data. *American Political Science Review* 89(3):634–47.

——. 2001. 'Throwing out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon'. *International Organization* 55(2):487–95.

——. 2007. 'Random Coefficient Models for Time-Series-Cross-Section Data: Monte Carlo Experiments'. *Political Analysis* 15(2):182–95.

Begg, Melissa D., and Michael K. Parides. 2003. 'Separation of Individual-level and Cluster-level Covariate Effects in Regression Analysis of Correlated Data'. *Statistics in Medicine* 22:2591–602.

Bell, Andrew, Kelvyn Jones, and Malcolm Fairbrother. (forthcoming). 'Does Democracy Encourage Globalisation? A Random Effects Reanalysis of Milner and Kubota's (2005) Study of Free Trade'.

Berlin, Jesse A., Stephen E. Kimmel, Thomas R. Ten Have, and Mary D. Sammel. 1999. 'An Empirical Comparison of Several Clustered Data Approaches under Confounding due to Cluster Effects in the Analysis of Complications of Coronary Angioplasty'. *Biometrics* 55(2):470–6.

Box, George E., and Norman R. Draper. 1987. *Empirical Model-building and Response Surfaces*. New York: Wiley.

Boyce, Christopher J., and Alex M. Wood. 2011. 'Personality and the Marginal Utility of Income: Personality Interacts with Increases in Household Income to Determine Life Satisfaction'. *Journal of Economic Behavior & Organization* 78(1-2):183–91.

Breusch, Trevor, Mickael B. Ward, Hoa T.M. Nguyen, and Tom Kompas. 2011a. 'On the Fixed-Effects Vector Decomposition'. *Political Analysis* 19(2):123–34.

——. 2011b. 'FEVD: Just IV or Just Mistaken?'. *Political Analysis* 19(2):165–9.

Browne, William J., Mousa G. Lahi, and Richard M.A. Parker. 2009. *A Guide to Sample Size Calculations for Random Effect Models via Simulation and the MLPowSim Software Package*. University of Bristol. Available at http://www.bristol.ac.uk/cmm/software/mlpowsim/, accessed 21 June 2012.

Browne, William J., and David Draper. 2006. 'A Comparison of Bayesian and Likelihood-based Methods for Fitting Multilevel Models'. *Bayesian Analysis* 1(3):473–513.

Burstein, Leigh, Robert L. Linn, and Frank J. Capell. 1978. 'Analyzing Multilevel Data in the Presence of Heterogeneous Within-class Regressions'. *Journal of Educational Statistics* 3(4):347–83.

Burstein, Leigh, and Michael David Miller. 1980. 'Regression-based Analyses of Multilevel Education Data'. *New Directions for Methodology of Social and Behavioral Sciences* 6:194–211.

Chatelain, Jean-Bernard, and Kirsten Ralf. 2010. 'Inference on Time-invariant Variables using Panel Data: A Pre-test Estimator with an Application to the Returns to Schooling'. PSE Working Paper. Available at http://hal-paris1.archives-ouvertes.fr/docs/00/49/20/39/PDF/Chatelain_Ralf_Time_Invariant_Panel.pdf, accessed 16 April 2012.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

Clark, Gordon L. 1998. 'Stylized Facts and Close Dialogue: Methodology in Economic Geography'. *Annals of the Association of American Geographers* 88(1):73–87.

Clark, Tom S., and Drew A. Linzer 2012. 'Should I Use Fixed or Random Effects?', Emory University. Available at http://polmeth.wustl.edu/mediaDetail.php?docId=1315, accessed 3 May 2012.

Clarke, Paul, Claire Crawford, Fiona Steele, and Anna Vignoles. 2010. 'The Choice between Fixed and Random Effects Models: Some Considerations for Educational Research'. CMPO working paper, University of Bristol. Available at http://www.bristol.ac.uk/cmpo/publications/papers/2010/wp240.pdf, accessed 1 March 2012.

Davis, James A., Joe L. Spaeth, and Carolyn Huson. 1961. 'A Technique for Analyzing the Effects of Group Composition'. *American Sociological Review* 26(2):215–25.

Deaton, Angus. 2010. 'Instruments, Randomization, and Learning about Development'. *Journal of Economic Literature* 48(2):424–55.

Demidenko, Eugene. 2004. *Mixed Models: Theory and Applications*. Hoboken, NJ: Wiley.

Diez-Roux, Ana V. 1998. 'Bringing Context back into Epidemiology: Variables and Fallacies in Multilevel Analysis'. *American Journal of Public Health* 88(2):216–22.

Duncan, Craig, Kelvyn Jones, and Graham Moon. 1998. 'Context, Composition and Heterogeneity: Using Multilevel Models in Health Research'. *Social Science & Medicine* 46(1):97–117.

Enders, Craig K., and Davood Tofighi. 2007. 'Centering Predictor Variables in Cross-sectional Multilevel Models: A New Look at an Old Issue'. *Psychological Methods* 12(2):121–38.

Fairbrother, Malcolm. 2013. 'Two Multilevel Modeling Techniques for Analyzing Comparative Longitudinal Survey Datasets'. *Political Science Research and Methods* 2(1):119–40.

Ferron, John M., Kristin Y. Hogarty, Robert F. Dedrick, Melinda R. Hess, Jonhn D. Niles, and Jeffrey D. Kromrey. 2008. 'Reporting results from multilevel analysis'. In *Multilevel Modeling of Educational Data*, edited by Ann A. O'Connell and Betsy McCoach, 391–426. Charlotte, NC: Information Age.

Fielding, Antony. 2004. The Role of the Hausman Test and Whether Higher Level Effects Should be Treated as Random or Fixed. *Multilevel Modelling Newsletter* 16(2):3–9.

Goldstein, Harvey. 2010. *Multilevel statistical models*, 4th edition. Chichester: Wiley.

Green, Donald P., Soo Y. Kim, and David H. Yoon. 2001. 'Dirty Pool'. *International Organization* 55(2):441–68.

Greene, William H. 2011a. 'Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time-invariant Variables in Fixed Effects Models?' *Political Analysis* 19(2):135–46.

——. 2011b. 'Fixed-Effects Vector Decomposition: Properties, Reliability, and Instruments—Reply'. *Political Analysis* 19(2):170–2.

——. 2012. *Econometric Analysis*, 7th ed. Harlow: Pearson.

Grilli, Leonardo, and Carla Rampichini. 2011. 'The Role of Sample Cluster Means in Multilevel Models: A View on Endogeneity and Measurement Error Issues'. *Methodology-European Journal of Research Methods for the Behavioral and Social Sciences* 7(4):121–33.

Hausman, Jerry A. 1978. 'Specification Tests in Econometrics'. *Econometrica* 46(6):1251–71.

Hausman, Jerry A., and William E. Taylor. 1981. 'Panel Data and Unobservable Individual Effects'. *Econometrica* 49(6):1377–98.

Heckman, James J., and Edward Vytlacil. 1998. 'Instrumental Variables Methods for the Correlated Random Coefficient Model—Estimating the Average Rate of Return to Schooling when the Return is Correlated with Schooling'. *Journal of Human Resources* 33(4):974–87.

Hsiao, Cheng. 2003. *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Jones, Kelvyn. 1991. 'Specifying and Estimating Multi-level Models for Geographical Research'. *Transactions of the Institute of British Geographers* 16(2):148–59.

——. 2010. 'The Practice of Quantitative Methods'. In *Research Methods in the Social Sciences*, edited by Bridget Somekh and Cathy Lewin. 2nd ed. London: Sage.

Jones, Kelvyn, and Nina Bullen. 1994. 'Contextual Models of Urban House Prices—a Comparison of Fixed-coefficient and Random-coefficient Models Developed by Expansion'. *Economic Geography* 70(3):252–72.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

Jones, Kelvyn, and Craig Duncan. 1995. 'Individuals and their Ecologies: Analysing the Geography of Chronic Illness within a Multilevel Modelling Framework'. *Health and Place* 1(1):27–40.

Jones, Kelvyn, Ron J. Johnston, and Charles J. Pattie. 1992. 'People, Places and Regions—Exploring the Use of Multilevel Modeling in the Analysis of Electoral Data'. *British Journal of Political Science* 22:343–80.

Kaldor, Nicholas. 1961. 'Capital Accumulation and Economic Growth'. In *The Theory of Capital: Proceedings of a Conference Held by the International Economic Association*, edited by Friedrich A Lutz and Douglas Hague, 177–222. London: Macmillan.

Kennedy, Peter. 2008. *A Guide to Econometrics*, 6th ed. Malden, MA: Blackwell.

King, Gary. 2001. 'Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data'. *International Organization* 55(2):497–507.

Kravdal, Øystein. 2011. 'The Fixed-effects Model Admittedly no Quick Fix, but still a Step in the Right Direction and Better than the Suggested Alternative'. *Journal of Epidemiology and Community Health* 65(4):291–2.

Kreft, Ita, and Jan De Leeuw. 1998. *Introducing Multilevel Modeling*. London: Sage.

Krishnakumar, Jaya. 2006. 'Time Invariant Variables and Panel Data Models: A Generalised Frisch-Waugh Theorem and its Implications'. In:*Panel Data Econometrics: Theoretical Contributions and Empirical Applications*, edited by Badi H. Baltagi, 119–32. Amsterdam: Elsevier.

Leyland, Alastair H. 2010. 'No Quick Fix: Understanding the Difference between Fixed and Random Effect Models'. *Journal of Epidemiology and Community Health* 64(12):1027–8.

Li, Xiaomei. 2011. 'Approaches to Modelling Heterogeneity in Longitudinal Studies'. Victoria University. Available at http://researcharchive.vuw.ac.nz/bitstream/handle/10063/1695/thesis.pdf?sequence=1, accessed 26 April 2012.

Milner, Helen V., and Keito Kubota. 2005. 'Why the Move to Free Trade? Democracy and Trade Policy in the Developing Countries'. *International Organization* 59(1):107–43.

Moulton, Brent R. 1986. 'Random Group Effects and the Precision of Regression Estimates'. *Journal of Econometrics* 32(3):385–97.

Mundlak, Yair. 1978a. 'Pooling of Time-series and Cross-section Data'. *Econometrica* 46(1):69–85.

——. 1978b. 'Models with Variable Coefficients: Integration and Extension'. *Annales de l'inséé* 30/31:48–509.

Neuhaus, John M., and Jack D. Kalbfleisch. 1998. 'Between- and Within-cluster Covariate Effects in the Analysis of Clustered Data'. *Biometrics* 54(2):638–45.

O'Connell, Ann A., and D. Betsy McCoach. 2008. *Multilevel Modelling of Educational Data*. Charlotte, NC: Information Age.

Oneal, John R., and Bruce Russett. 2001. 'Clear and Clean: The Fixed Effects of the Liberal Peace'. *International Organization* 55(2):469–85.

Palta, Mari, and Chris Seplaki. 2003. 'Causes, Problems and Benefits of Different between and within Effects in the Analysis of Clustered Data'. *Health Services and Outcomes Research Methodology* 3:177–3.

Pawson, Ray. 2006. *Evidence-based Policy: A Realist Perspective*. London: Sage.

Plümper, Thomas, and Vera E. Troeger. 2007. 'Efficient Estimation of Time-invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects'. *Political Analysis* 15(2):124–39.

——. 2011. 'Fixed-Effects Vector Decomposition: Properties, Reliability, and Instruments'. *Political Analysis* 19(2):147–64.

Rasbash, Jon, Fiona Steele, William J. Browne, and Harvey Goldstein. 2009. *A User's Guide to MLwiN, version 2.10*. Centre for Multilevel Modelling, University of Bristol.

Raudenbush, Stephen W. 1989. 'Centering Predictors in Multilevel Analysis: Choices and Consequences'. *Multilevel Modelling Newsletter* 2(1):10–2.

——. 2009. 'Adaptive Centering with Random Effects: An Alternative to the Fixed Effects Model for Studying Time-varying Treatments in School Settings'. *Education, Finance and Policy* 4(4):468–91.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7

Raudenbush, Stephen W., and Anthony Bryk. 1986. 'A Hierarchical Model for Studying School Effects'. *Sociology of Education* 59(1):1–17.

——. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed. London: Sage.

Rubin, Donald B. 1980. 'Using Empirical Bayes Techniques in the Law-school Validity Studies'. *Journal of the American Statistical Association* 75(372):801–16.

Schurer, Stefanie, and Jongsay Yong. 2012. 'Personality, Well-being and the Marginal Utility of Income: What Can We Learn from Random Coefficient Models?' *Health, Economics and Data Group*. Working Paper, University of York. Available at http://www.york.ac.uk/res/herc/documents/wp/12_01.pdf, accessed 16 March 2012.

Shin, Yongyun, and Stephen W. Raudenbush. 2010. 'A Latent Cluster-mean Approach to the Contextual Effects Model with Missing Data'. *Journal of Educational and Behavioral Statistics* 35(1):26–53.

Shor, Boris, Joseph Bafumi, Luke Keele, and David Park. 2007. 'A Bayesian Multilevel Modeling Approach to Time-series Cross-sectional Data'. *Political Analysis* 15(2):165–81.

Skrondal, Anders, and Sophia Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton: Chapman and Hall.

Snijders, Tom A. B., and Johannes Berkhof. 2007. 'Diagnostic Checks for Multilevel Models'. In *Handbook of Multilevel Analysis*, edited by Jan de Leeuw and Erik Meijer, 139–73. New York: Springer.

Snijders, Tom A.B., and Roel J. Bosker. 2012. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*, 2nd ed. London: Sage.

Solow, Robert M. 1988. *Growth Theory: An Exposition*. New York: Oxford University Press.

Spanos, Aris. 2006. 'Revisiting the Omitted Variables Argument: Substantive vs Statistical Adequacy'. *Journal of Economic Methodology* 13(2):179–218.

Steele, Fiona, Anna Vignoles, and Andrew Jenkins. 2007. 'The Effect of School Resources on Pupil Attainment: A Multilevel Simultaneous Equation Modelling Approach'. *Journal of the Royal Statistical Society Series A-Statistics in Society* 170:801–24.

Stegmueller, Daniel. 2013. How Many Countries Do You Need for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches. *American Journal of Political Science* 57(3):748–61.

Subramanian, S.V., Kelvyn Jones, Afamia Kaddour, and Nancy Krieger. 2009a. Revisiting Robinson: The Perils of Individualistic and Ecologic Fallacy. *International Journal of Epidemiology* 38(2):342–60.

——. 2009b. Response: The Value of a Historically Informed Multilevel Analysis of Robinson's Data. *International Journal of Epidemiology* 38(2):370–3.

Verbeke, Geert, and Geert Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data*. New York: Springer.

——. 2005. *Models for Discrete Longitudinal Data*. New York: Springer.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

——. 2009. Correlated Random Effects Models with Unbalanced Panels. Michigan State University. Available at http://www.bancaditalia.it/studiricerche/seminari/2011/Wooldridge/paperwooldridge.pdf, accessed 16 April 2012.

Downloaded from https:/www.cambridge.org/core. IP address: 50.16.122.200, on 30 Mar 2017 at 22:40:27, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/psrm.2014.7