The Choice Between Fixed and Random Effects

Zac Townsend
Mayor's Office, City of Newark,
NJ and New York University, USA

Jack Buckley
National Center for Educational
Statistics and New York University, USA

Masataka Harada
New York University, USA

Marc A. Scott
New York University, USA

5.1 INTRODUCTION

In prior chapters, group-specific effects were assumed to be drawn from a distribution, typically Gaussian. In applied research in the social and behavioral sciences, economics, public health, public policy, and many other fields, alternatives to this choice are often made, with the most common being the "fixed effects" approach. In its most basic formulation, group-specific intercepts are modeled using indicator variables, effectively making them open parameters in the model and not assigning them a distributional form. The choice between random and fixed effects has implications for the interpretation of parameter estimates and for estimation efficiency.[1] Specifically,

- the effect estimates for each group will be different depending on the modeling choice;
- the β estimates for the predictors in the model may be different depending on the modeling choice;
- the reliability with which one can make predictions for new groups differs; and
- there is a tradeoff between ef ciency and bias, with random effects being more ef cient than xed effects, but potentially biased (we defer formalizing our de nition of bias to a later section).

The remainder of this chapter will discuss these options as well as a hybrid version of the two. It provides guidance as to when an applied researcher should use each model, and describes situations when neither is appropriate and other models should be considered.

We proceed by describing the two models in Section 5.2 before discussing the different assumptions, describing estimation, and giving advice on when to use each in Section 5.3. Section 5.4 outlines an example of using fixed effects and random effects with data from the National Assessment of Educational Progress, a series of large-scale assessments of schoolchildren in the United States.

5.2 THE FIXED EFFECTS AND RANDOM EFFECTS MODELS

In this chapter, we outline both random and fixed effects models. We will refer to models as fixed if they model unit-specific components in longitudinal data or group-specific components in clustered data as separate parameters, and random effects if they are drawn from a (often Gaussian) probability distribution.

We will delve into model specifics shortly, but here we mention that with grouped data there are two types of relationship that are in play: between-group and within-group. Failure to understand this interplay can result in an ecological fallacy, in which one makes inferences about the nature of units based on information about the groups containing them. In the context of multilevel models, we will show that variation naturally divides into within- and between-group components, so that potential for misinterpretation exists whenever the two types of relationship differ for a predictor—for example, when neighborhood income matters for some outcome, but relative income within that group does not. Diez Roux (2004) and Wakefield (2003) give extensive overviews (see also Gelman, 2006). We note that the standard ecological fallacy typically occurs when one does not have both unit- and group-level data available to disentangle the two types of effects.

5.2.1 The Random Effects Model

Consider the multilevel model

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_{0j} + \varepsilon_{ij}, \qquad (5.1)$$

where y_{ij} is the outcome for the ith subject in the jth group. It is also represented more explicitly as a multilevel model as

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \varepsilon_{ij}, \qquad (5.2)$$

where

$$\beta_{0j} = \beta_0 + u_{0j}. \qquad (5.3)$$

Equation (5.2) is often referred to as the level one equation and (5.3) is the level two equation. In both formulations, it is usually assumed that $u_{0j} \sim N(0, \sigma^2_{u_0})$ is independent of $\varepsilon_{ij} \sim N(0, \sigma^2_{\varepsilon})$. This model proposes that the values of the outcome, y_{ij}, are based on the population-level constant, β_0, the single explanatory variable, x_{1ij} and the associated parameter β_1, and a group-specific component, u_{0j}. Although the random effects model can include more complex random components such as two-way (e.g., time- and unit-specific) intercepts or multiple group-specific intercepts that have a nested structure (e.g., classrooms and schools), in this chapter we consider a single group-specific intercept to focus on the choice between random and fixed effects.

Researchers using random effects models often focus on the variance components, such as $\sigma^2_{u_0}$. It is instructive, however, to examine the group-specific effects, as these serve as controls for all other effects, and thus have the potential to alter the other effect estimates. In the random effects model, when $u_{0j} \sim N(0, \sigma^2_{u_0})$, given constant β_0, we have $\beta_{0j} \sim N(\beta_0, \sigma^2_{u_0})$, which are group-specific intercepts. Gelman and Hill (2007, p. 257) call this a “soft constraint” that is applied to the intercepts, which shrinks the estimates of β_{0j} toward the common mean β_0, a phenomenon they call a partial pooling (see Lindley and Smith, 1972; Smith 1973; Efron and Morris, 1975; Morris, 1983; and Kreft and de Leeuw, 1998, for some background). This term derives from the “pooled” and “unpooled” modeling approaches described in Gelman and Hill (2007) and introduced in Chapter 1. A pooled model simply applies OLS or similar estimation techniques to grouped data, *ignoring the* *grouping*. An unpooled model may be defined in at least two ways. First, it could be understood as separate regressions, one for each group. This would entail unique regression parameters for every predictor as well as the intercept, and small sample sizes may yield highly imprecise estimates. A more common approach is to estimate group-specific intercepts, effectively using one parameter per group, and to pool the estimates of all other predictor effects; this definition is suggested by Gelman and Hill (2007). This is equivalent to estimating the so-called fixed effects model and will be what we adopt in our subsequent discussion.

The intercepts in the random effects model are a weighted average of estimates from the two types of pooling. In our simple model (5.1), assuming the data are balanced, the random intercepts and slope coefficients are

$$\hat{\beta}_{0j} \approx \frac{n_j/\sigma_y^2}{n_j/\sigma_y^2 + 1/\sigma_{u_0}^2}(\bar{y}_j - \hat{\beta}_1\bar{x}_{1j}) + \frac{1/\sigma_{u_0}^2}{n_j/\sigma_y^2 + 1/\sigma_{u_0}^2}\beta_0 \tag{5.4}$$

$$\hat{\beta}_1 = \left(\sum_i \sum_j (x_{1ij} - \lambda\bar{x}_{1j})^2\right)^{-1} \cdot \left(\sum_i \sum_j (x_{1ij} - \lambda\bar{x}_{1j})(y_{ij} - \lambda\bar{y}_j)\right), \tag{5.5}$$

where

$$\lambda = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + n_j\sigma_{u_0}^2}},$$

n_j is the number of observations within a group (group size), $\bar{y}_j$ and $\bar{x}_{1j}$ are within-group means of the outcome variable and the explanatory variable, and σ_y^2, $\sigma_{u_0}^2$, and σ_ε^2 are variances of y, u_0, and ε, respectively (Gelman and Hill, 2007, p. 258; Afshartous and de Leeuw, 2005, pp. 112–13; Wooldridge, 2010, p. 287).[2] In order to utilize these formulas to get estimates of the coefficients, one must replace the variances and the intercept with their estimates.

In Equation (5.4), the group-specific residuals $(\bar{y}_j - \beta_1\bar{x}_{1j})$ employ the β_1 from the no-pooling model (or fixed effects model) and the overall intercept (β_0) is obtained from the complete pooling model (or OLS). Likewise, the slope estimate in Equation (5.5), $\hat{\beta}_1$, reduces to that of the fixed effects model when $\lambda = 1$ and that of OLS when $\lambda = 0$. That is, two weight parameters, $(n_j/\sigma_y^2)/(n_j/\sigma_y^2 + 1/\sigma_{u_0}^2)$ and λ, serve as a measure of the shrinkage toward the unpooled estimators. When group-level effects are estimable with great precision, $\sigma_{u_0}^2$ becomes large relative to σ_y^2 and σ_ε^2, and these weight parameters approach one, very little shrinkage occurs, and the random and fixed effects estimates are nearly identical. Conversely, when $\sigma_{u_0}^2$ becomes relatively small, these two weight parameters approach zero, and shrinkage to the population-level effect (ignoring group) becomes nearly complete.

When the assumptions in the random effects model hold, these above estimates are more efficient than their fixed effects counterparts, and this is one of the key reasons for choosing such models. The assumptions, however, may be more or less plausible in different applications, such as randomized experiments versus observational studies. See Chapter 12 for further discussion.

5.2.2 Fixed Effects Model

If one wants to estimate group-specific intercepts without imposing a distribution or utilizing information from the other groups, one may choose to estimate a fixed effects model. The model can be represented in a similar way to equation (5.1), but with a modification to the group effects:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \sum_{j'=2}^{J} u_{0j'} I\left\{j = j'\right\} + \varepsilon_{ij}, \tag{5.6}$$

with the indicator function $I\left\{j = j'\right\}$ tracking the $J - 1$ group-level effects (group 1 taken as the reference category). The implication of this reformulation is that the group effects u_{0j} may be correlated with the other predictor (or predictors, more generally). In situations in which the orthogonality of the random effects and other predictors is implausible, fixed effects models are one alternative.

Another way to characterize this model in the context of multilevel models is to describe the random effects distribution as having infinite variance, $\sigma^2_{u_0} = \infty$, so the model does not borrow any information across groups.[3] Although, in this sense, the fixed effects model is a special case of random effects, our simplest fixed effects model is not really a multilevel model in the traditional use of the term, as the group level has been reduced to a set of indicators and their corresponding effects.

A tradeoff that may not be immediately apparent is that one cannot include any group-level explanatory variables, such as each group's average income, due to collinearity with the group fixed effect. If one is interested in variance components and how they change with the addition of predictors at group and subject level, then one is severely hampered by the fixed effects approach. However, one may not have this type of interest.

Fixed effects models allow group-specific components u_{0j} to be correlated with other covariates x_{1ij}, which is certainly a less restrictive approach. This has led some researchers to interpret these components as controls that proxy for any omitted variable that is "fixed" or constant for the entire group. However, the form of "omitted variable" that may be captured this way is quite restrictive (see Baltagi, 2005; Wooldridge, 2010, for some discussion of omitted variable bias and how it relates to this model; see also Chapter 12). At a minimum, it is clear that the omitted confounder must be time- (or group-) constant.

Of course, one may estimate (5.6) using OLS, but in practice, a demeaning approach, to be described in the next section, is usually employed. Given the estimators of random effects model, the derivation of fixed effects estimators is straightforward. By setting $\sigma^2_{u_0} = \infty$ in (5.4) and (5.5),

$$\hat{\beta}_{0j} = \bar{y}_j - \hat{\beta}_1 \bar{x}_{1j}, \text{ where}$$

$$\hat{\beta}_1 = \left(\sum_i \sum_j (x_{1ij} - \bar{x}_{1j})^2 \right)^{-1} \cdot \left(\sum_i \sum_j (x_{1ij} - \bar{x}_{1j})(y_{ij} - \bar{y}_j) \right).$$

These fixed effects estimators are often called the within estimators because they do not attempt to explain between-group differences, and are solely based on *deviations* of x_{1ij} and y_{ij} from the respective estimated group means $\bar{x}_{1\cdot j}$ and $\bar{y}_{\cdot j}$. Fixed effect methods control for all stable characteristics of the groups in the study. In a longitudinal study, this would mean that fixed effects control for time-invariant characteristics of each individual, such as personal history before the observation period.

While we have briefly noted some of the differences between the two estimators, our next section will provide more specific details of those estimation procedures and then discuss the relative merits or conditions favoring each of the techniques.

5.3 A COMPARISON OF THE TWO ESTIMATORS

This section describes the assumptions behind the fixed and random effects models, outlines the estimation techniques of both models and provides general advice regarding when to use and not use each technique. We then describe some hybrid models developed that satisfy the "random effects assumption" while still offering some of the benefits of fixed effects models.

5.3.1 The Assumptions of Each Model

Wooldridge (2010) states that the fixed effects model must satisfy the strict exogeneity assumption, $E(\varepsilon_{ij}|x_{1i'j}, u_{0j}) = 0$ for all values of i' given any i, to obtain an unbiased fixed effects estimator in the presence of any unobserved group-specific confounders. That is, ε_{ij} is not only independent of its contemporaneous covariates but also should be independent of the covariates that belong to the same group ($x_{1i'j}$ such that $i' \neq i$). If the condition on ε_{ij} is not met in the fixed effects model, then potential unobserved omitted variables could bias the estimator for x_{1ij}'s effect.

In addition to this strong exogeneity assumption, the random effects model assumes conditional mean independence of group-specific effects u_{0j} given all covariates x_{1ij} (Wooldridge, 2010). Formally,

$$\mathrm{E}(u_{0j}|x_{1ij}) = \mathrm{E}(u_{0j}) = 0.$$

This assumption is also referred to as orthogonality or, somewhat less formally, no correlation between u_{0j} and x_{1ij} ($\mathrm{Cov}(u_{0j}, x_{1ij}) = 0$). Usually it is further assumed that $u_{0j} \sim N(0, \sigma^2_{u_0})$. This is perhaps the fundamental difference between the fixed and random effects approaches: the latter assumes that u_{0j} are orthogonal to ε_{ij} and to any predictors x_{1ij} in the model, while the former implicitly allows correlation with predictors.

This same concern of bias under violation of assumptions exists in the random effects model, but the additional assumption that the random effects are orthogonal to explanatory variables may be harder to justify, particularly where causal inference is the goal. However, in practice the researcher may wish to trade a small amount of bias for efficiency, and we reiterate that the hybrid models we present in Section 5.3.5 provide a useful compromise approach. Causal modeling for multilevel models requires a much deeper framework than we present in this chapter—see Chapter 12 of this volume for further discussion of assumptions in multilevel models for causal inference.

The differences in the above assumptions—and their implications—has led some researchers to frame the choice between random and fixed effects models as a bias/variance trade-off. Random effects models may be biased (for the within-group effect) when the model assumptions fail to hold, but they are more efficient than fixed effects models when they do. The bias is often framed as a form of omitted variable bias (see Bafumi and Gelman, 2006). Specifically, if there is a group-constant confounder c omitted from the model, and it is correlated with another predictor, such as x_{1ij}, (and the outcome), then the corresponding parameter β_1 will be biased when c is omitted. One of motivations for using grouped data is that we can recover the (within-group) β_1 that would have been estimated had we included the confounder c. With respect to recovering this parameter, fixed effects models are unbiased and random effects models are efficient (this is the setup for the Hausman Test). However, this statement applies only to the basic forms of these models; in Section 5.3.5 we show that centering can be used to achieve the same goal for random effects models.

5.3.2 Estimating Each Model

Fixed effects estimation can be conducted either by group demeaning both the dependent and explanatory variables using the *within* transformation (sometimes called the fixed effects transformation), or by adding a set of dummy variables for each group j and using OLS, as previously described. To perform the *within* transformation, we can average equation (5.6) for each group, and then subtract the new group means from the original, obtaining:

$$\begin{aligned} y_{ij} - \bar{y}_j &= \left(\beta_0 + \beta_1 x_{1ij} + u_{0j} + \varepsilon_{ij}\right) \\ &\quad - \left(\beta_0 + \beta_1 \bar{x}_{1j} + u_{0j} + \bar{\varepsilon}_j\right) \qquad (5.7) \\ &= \beta_1(x_{1ij} - \bar{x}_{1j}) + (\varepsilon_{ij} - \bar{\varepsilon}_j). \end{aligned}$$

Thus, the demeaning of the original equation ensures that demeaned covariates are orthogonal to any group effects, u_{0j}. It also implicitly controls for group effects, or partials them out of the model. The *within* nature of the effects is evident from this formulation, for now a one-unit change in the demeaned covariate reflects change relative to one's group. Chapter 6 discusses this concept in some detail. The estimation of equation (5.7) with OLS yields fixed effects estimators.[4]

Alternatively, fixed effects estimators can be obtained by adding a set of group indicators as explanatory variables and applying OLS to equation (5.6). These two methods are equivalent; the interested reader can consult Davidson and MacKinnon (1993) and Wooldridge (2010).

Random effects models may be estimated in a variety of ways, including feasible generalized least squares (FGLS), maximum-likelihood estimation (MLE), or Bayesian approaches with MCMC. See Chapters 3 and 4 of this book, Hsiao (2003), and Gelman and Hill (2007) for an overview of potential approaches to estimation. Here, we introduce a step-by-step approach to calculate random effects estimators given by Johnston and DiNardo (1997).[5] Each step illuminates important properties of random effects models, but the main idea is to find the weight parameter, λ. Random effects estimators are partial-pooling estimators situated between no pooling (or fixed effects) and complete pooling (or OLS) estimators, and λ determines the proportion of these two quantities.

The first step is to find no-pooling and complete-pooling estimators. This can be easily done by estimating (5.6) or (5.7) and by estimating (5.1) without a group-specific component (u_{0j}) using OLS. Second, using the residuals obtained in the first step, we estimate σ_ε^2 and $\sigma_{u_0}^2$ using the following formulae:

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{\varepsilon}'_{FE}\hat{\varepsilon}_{FE}}{n(n_j - k - 1)}$$
$$\hat{\sigma}_{u_0}^2 = \frac{\hat{\varepsilon}'_{OLS}\hat{\varepsilon}_{OLS}}{n - k} - \frac{\hat{\sigma}_\varepsilon^2}{n_j},$$

where $\hat{\varepsilon}_{FE}$ is a residual vector from the no-pooling model, $\hat{\varepsilon}_{OLS}$ is that from the omplete-pooling model, and k is the number of explanatory variables. Now, we have all the information required to calculate λ. Remember, again, that the fixed effects model is defined to be the same as the unpooled model mentioned above, while the final step for this approach to estimating a random effects model is simply to regress $y_{ij} - \hat{\lambda}\bar{y}_j$ on $x_{1ij} - \hat{\lambda}\bar{x}_{1j}$. To repeat, this regression is identical to that of the fixed effects model when $\lambda = 1$, and that of OLS when $\lambda = 0$. Because $0 \leq \lambda \leq 1$, one way to conceptualize the random effects estimator is as the regression of *partially* demeaned variables.

Another way of conceptualizing and then estimating random effects models is through a matrix representation of predictors and error structure and vector notation for outcomes. This requires specification of a complex covariance structure, in which some between-element correlation must be modeled. Typically, observations are assumed independent across groups, but not within groups, resulting in a block-diagonal structure. An example of a classic random intercepts model is given in Chapter 2, row 4 of Table 2.2; in matrix notation, using the terms X, Z, β, G, and R to capture the fixed and random effects designs, the fixed effects parameters, the covariance of the random effects, and the covariance of the errors, respectively, the outcome vector $\mathbf{y} \sim N(X\beta, ZGZ' + R)$. Estimation is usually made using the method of maximum likelihood, but Bayesian approaches are also common. As a by-product of this notation, we can express the best linear unbiased predictor (BLUP; see Robinson, 1991) of the random effect vector for group j as:

$$\begin{aligned} \hat{\mathbf{u}}_j &= E(\mathbf{u}_j | Y_{\cdot j}, X_{\cdot j}, Z_{\cdot j}, \hat{\beta}) \\ &= GZ'_{\cdot j}(Z'_{\cdot j}GZ_{\cdot j} + R)^{-1}(Y_{\cdot j} - X_{\cdot j}\hat{\beta}), \end{aligned} \tag{5.8}$$

where "$\cdot j$" refers to that portion of the vector or matrix associated with group j (over all associated i). In the case of a random intercept model, Z, G, and R have simpler structure, so this becomes:

$$\begin{aligned} \hat{u}_{0j} &= E(u_{0j} | Y_{\cdot j}, X_{\cdot j}, Z_{\cdot j}, \hat{\beta}, \hat{\sigma}_u^2, \hat{\sigma}_\varepsilon^2) \\ &= \hat{\sigma}_u^2(\hat{\sigma}_u^2 Z'_{\cdot j}Z_{\cdot j} + \hat{\sigma}_\varepsilon^2 I)^{-1}(Y_{\cdot j} - X_{\cdot j}\hat{\beta}). \end{aligned}$$

While most of our discussion assumes some interest in estimating these effects, we reiterate that random effects models estimate the parameters *governing* the effects, while fixed effects models incorporate them through group-specific parameters. The BLUPs are post hoc "best guesses" of the group effects, which, under likelihood theory, have asymptotic distributions, and thus known precision. A new group's effect along with its precision could thus be estimated using (5.8). In the case of fixed effects, it is possible to define an analogous prediction for a group effect from residualized outcomes, but the lack of model assumptions offers no shrinkage of that prediction, nor an estimate of its precision that utilizes information from all groups.

5.3.3 The Debate Between Fixed and Random Effects

Whenever a textbook introduces the random and fixed effects models, it discusses which method is to be used when. However, readers are sometimes perplexed by the wide range of prescriptions proposed by different textbooks. The goal of this subsection is to provide some guidance on the opinions and prescriptions encountered throughout the literature. Here we provide three criteria, namely satisfaction of the assumptions, compatibility with the quantities of interest, and trade-off between bias reduction and efficiency, on which researchers can base their judgment in selecting a model.

The first criterion is whether the selected method satisfies the required assumptions. We have pointed out that the random effects assumption, or conditional independence between the covariates and group-specific components, may be unrealistic. An advantage to the fixed effects model is that it allows for group-specific components to be correlated with the covariates. However, when the random effects assumption is satisfied, random effects models are more efficient. As we see in the next section, the Hausman test takes advantage of this property to adjudicate between the choice of fixed or random effects.

Although the fixed effects model requires fewer assumptions than the random effects model, this does not necessarily mean the fixed effects model is always more robust against violation of its assumptions. In particular, the fixed effects model is known to be more susceptible to several violations of the regression assumptions and other problems typical of regression modeling. For example, the impact of measurement errors in covariates is amplified in fixed effects models (Johnston and DiNardo, 1997). A drawback of the fixed effects model is its susceptibility to the violation of strict exogeneity. As discussed in Section 5.3.1, under the assumption of strict exogeneity, the error terms must be orthogonal to any values of the covariates that belong to the same group. Typically, this assumption is violated when there is an endogenous covariate in panel data or when a subject's outcome may affect another subject's outcome *within the same group*. To be fair, the random effects model must also satisfy the strict exogeneity assumption. However, because random effects estimators are derived from both within-group and between-group variations, the latter of which is unaffected by the strict exogeneity assumption, the violation has the potential to bias fixed effects estimators more (Johnston and DiNardo, 1997).

The second criterion is whether the quantities of interest are estimable with the fixed effects model. Specifically, the quantities of interest obtained from a fixed effects model are limited in two ways. First, fixed effects models cannot estimate the effect of a variable that has no within-group variation because fixed effects subsume all observed and unobserved group-specific variation. For example, one cannot estimate the effect of race or gender with the panel data of repeated individual observations because race and gender do not change over time. Also, with grouped data that contain students' academic performance, one cannot use classroom fixed effects if the students are treated on a classroom basis. Further, even if explanatory variables have some within-group variance, they are often imprecisely estimated in fixed effects models due to the additional degrees of freedom used up (Ashenfelter, et al., 2003).

As alluded to previously, fixed effects models have a limited capability to make predictions about the distribution of an outcome variable at different levels. For example, with the random effects model, a researcher whose interest is the effect of extended class hours on students' test scores can predict the distribution of the test scores at classroom, school, or population level. With the random effects model, a researcher can incorporate the information from random intercepts at different levels into the predictive distribution of the outcome variable and this extends to new observations and groups. Although it is possible to estimate fixed effects for new groups in an ad hoc manner in fixed effects models, the precision of those estimates is not readily available, nor does any parameter directly reflect the overall between-group (and, in more nested models, between-level) variation. This observation has led some researchers to frame the tradeoff between fixed and random effects as: random effects are samples from a population while fixed effects represent the entire population of interest or conditional inference on the available groups in the sample (see, for example, Mundlak, 1978). The basis for this characterization is classical presentations of ANOVA; we have shown that the implications of the choice on parameter estimates requires the practitioner to consider additional implications of the choice.

The final criterion is the tradeoff between how much bias can be reduced and how much efficiency is lost with the fixed effects model. Fixed effects models add group-specific indicators as explanatory variables, controlling for time-constant, homogeneous group-specific confounding, including that due to unobservables, provided the underlying assumptions hold. This means that, to return to the example of panel data on individuals, confounding variables such as personal or family background before the observation period and any personal attributes such as race, gender, or genetics are controlled for, up to the validity of the model assumptions, primary among which is the assumption that group effects have no heterogeneity within that group (e.g., no change over time in the panel data setting). In grouped data, e.g., observations on students in classrooms, a fixed effects model would control for quality of teachers, family background, and any experiences that classmates might share. It is the promise of controlling for such confounders, and the expectation that they are always lurking, that has led some researchers to endorse primarily fixed effects modeling with multilevel data (see, for example, Allison, 2005).

However, the contribution to bias reduction with the fixed effects model must also be evaluated in terms of its loss of efficiency. Fixed effects models lose their efficiency, particularly in the following two instances. First, for a large number of groups and a small number of within-group observations, the number of degrees of freedom consumed by the fixed effects will be large. For example, if a researcher has panel data that consist of 100 group observations ($J = 100$), the introduction of a fixed effects model reduces the degree of freedom by 99. Or, in the panel case, if the number of time periods is two, the fixed effects model halves the degrees of freedom. The tails of the t-distribution become thicker as the degrees of freedom decreases, which implies that the coefficients are estimated with greater uncertainty. Goldstein (2003) advises using the fixed effects model when there are a few groups and moderately large numbers of units in each. Snijders and Bosker (1999) also advise using a random effects model if the number of groups is more than ten. Second, when explanatory variables have little within-group variation, they are highly correlated with the fixed effects, and fixed effects estimators will be inefficient (Bartels, 2008). In a simple example, suppose that a researcher is interested in estimating a policy effect using the panel data of 50 states in the U.S. If only 10 states changed the policy during the observed period, the fixed effects estimators are calculated based on these 10 states, ignoring 40 states in the data because fixed effects subsume all time-invariant covariates.

To make the choice between fixed and random effects, a researcher needs to consider all of what we have discussed, such as the plausibility of the random effects assumption and strict exogeneity, the need to include group-level predictors, efficiency, degrees of freedom, the presence and degree of measurement error in the data, and the amount of within-group variance.

5.3.4 Comparing Fixed Effects and Random Effects: The Hausman Test

Although most of the criteria for the choice between fixed and random effects models require subjective judgments, test statistics have been developed for some of them. The Hausman test (Hausman, 1978) is most frequently used to check the validity of the random effects assumption, namely the conditional independence between group-specific intercepts (u_{0j}) and covariates ($\mathbf{x}_{ij}$). Briefly speaking, the Hausman test is a comparison between the parameter estimates of the fixed effects and random effects models, and supports the random effects assumption if the difference between two parameters is sufficiently small.

To see the logic, let us show the formal definition of the test statistic, H, below:

$$\begin{aligned} H &= (\hat{\beta}_{RE} - \hat{\beta}_{FE})'(\hat{V}_{FE} - \hat{V}_{RE})^{-1} \\ &\quad (\hat{\beta}_{RE} - \hat{\beta}_{FE}), \end{aligned}$$

where $H \sim \chi^2$ with k degrees of freedom (the number of time-varying coefficients in both models), $\hat{\beta}$ denotes a vector of estimated coefficients, $\hat{V}$ denotes a variance-covariance matrix of estimators, and the subscripts RE and FE denote the random or fixed effect estimator as the source of the estimates. We test the null hypothesis that the results of the two estimators are not different. When the random effects assumption holds, both the fixed and random effects estimators are consistent. However, when the assumption does not hold, only the fixed effects estimator is consistent, and the random effects estimator is biased. Moreover, when the assumption holds, the fixed effects estimator is inefficient compared to the random effects model. As the violation of the assumption becomes less serious, the difference in point estimates between two parameters decreases in the numerator, while the difference in variances increases in the denominator. Thus, the size of H is interpreted as the extent of the deviation from the assumption.

It must be noted, however, that the Hausman test does not always provide a definitive answer. One problem is that although statistical significance of the test statistic, H, is fairly reliable evidence for the bias in the random effects estimates, statistical insignificance is not necessarily evidence for unbiasedness. This is because the Hausman test relies on the consistency of the two estimators. As (5.4) and (5.5) indicate, fixed and random effects estimators converge only if the group size (n_j) and/or the within-group variance ($\sigma^2_{u_0}$) is large. Thus, in a small finite sample and/or with covariates that have little within-group variance, these two estimators may not significantly differ from each other even when the random effects assumption is violated, and the Hausman test cannot distinguish between the satisfaction of the random effects assumption and the lack of convergence.[6] A number of studies have been conducted to improve the performance of the Hausman test. Ahn and Low (1996) refined the Hausman test to better detect the violation when the composite error term ($u_{0j} + \varepsilon_{ij}$) contains little within-group variance ($\sigma^2_{u_0}$). Bole and Rebec (2004) show that a bootstrap version of the Hausman statistics marginally improved its performance when the violation is minimal. To date, none has been established as a definitive alternative to the Hausman test.

5.3.5 The Benefits of Both Models: Hybrid Models

The classic back and forth of fixed versus random effects is that a researcher should have less concern about bias when using fixed effects, while random effects are more efficient. However, several interesting compromises have been proposed. The two most prominent approaches are correlated random effects (for example, Mundlak, 1978 and Chamberlain, 1982) and adaptive centering (Raudenbush, 2009). Several other recent papers (Gelman, 2006; Bartels, 2008; Leyland, 2010) also propose various solutions to the correlation between predictors and random effects that allow researchers to more comfortably estimate varying parameters.

The literature has accurately redefined the random effects assumption outlined above as one of cluster confounding. When only a single, unadjusted version of a predictor is included in a model, standard random effects approaches implicitly assume that the within- and between-group effects are equal for changes in any component of $\mathbf{x}_{ij}$ (Skrondal and Rabe-Hesketh, 2004; Rabe-Hesketh and Everitt, 2006; Zorn, 2001; Bartels, 2008). We can decompose any particular x_{1ij} into different between- and within-group effects, which we can call x^B_{1j} and x^W_{1ij}, and the result is that β_1 is a weighted average of the two processes. Bafumi and Gelman (2006) label this a type of omitted variable bias, and go on to propose group mean centering as a workaround solution. This is a form of the hybrid model described in Allison (2005).

In the case of one predictor, a hybrid model would be

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_{W1}(x_{1ij} - \bar{x}_{1j}) \\ &\quad + \beta_{G1}(\bar{x}_{1j}) + u_{0j} + \varepsilon_{ij}, \end{aligned}$$

where β_{W1} captures the within-group effect of the predictor x_1, β_{G1} captures the between-group effect, and u_{0j} is a random effect. All other terms, such as the group means $\bar{x}_{1j}$, are as they have previously been defined, as are the distributional assumptions on the stochastic terms. The within-group effect should be the same as that obtained via the fixed effects estimator.[7] Two things are unique about this approach: first, we can include group-constant predictors, such as $\bar{x}_{1j}$, in the model. In fact, any group-constant predictors may be included without affecting β_{W1}, as $(x_{1ij} - \bar{x}_{1j})$ is orthogonal to any group-constant predictor. Second, the random intercept variance σ_{u_0} measures the remaining between-group variation in level, and inference for this parameter is obtainable. In the fixed-effects setting, say, using group-specific indicators, one cannot include additional group-specific predictors, and the inferential framework for assessing between-group variance is somewhat ad hoc (we describe this in the example). In this sense, the hybrid method has all of the advantages of both random- and fixed-effects approaches. Allison (2005), drawing in particular on Neuhaus and Kalbfleisch (1998), describes how this model may be implemented in the SAS Statistical Programming language, and notes that a Wald test of the hypothesis $\beta_{W1} - \beta_{G1} = 0$ provides a Hausman-like test of whether random effects models provide the same coefficients of interest as the fixed effects models.

The above approach is also quite similar to Mundlak's (1978) formulation, and simply adds the mean for each time-varying covariate. This identifies the endogeneity concern as a result of attempting to model two processes in one term (see also Snijders and Bosker, 2012, p. 56; Berlin, Kimmel, Ten Have, and Sammel 1999). Others similarly argue for unit characteristics that are correlated with the mean (Clark, Crawford, Steele, and Vignoles, 2010). Bartels (2008) and Leyland (2010) are more interpretable reformulations of Mundlak (1978). Proponents of fixed effects either call this method a compromise approach or a form of the Hausman test to choose between fixed and random effects (Allison, 2009, pp. 33–5; Hsiao, 2003, pp. 44–50; Wooldridge, 2010; Greene, 2011; and others). In truth, they are essentially the same model as those argued for by Bafumi and Gelman (2006), Gelman and Hill (2007), and throughout the multilevel models literature: random effects models with additional time-invariant predictors, which address the exogeneity assumption.

We also note that in fields such as psychology and education, group mean centering has a long history (e.g., Bryk and Raudenbush, 1992; Raudenbush and Bryk, 2002), but the connections to the econometric fixed-effects approach have only been made more recently (Allison, 2009). Lastly, the practical researcher should always check the plausibility of the assumptions behind the modeling, no matter the name, and realize that not all random effects models place the same demands on the random effects assumption.

5.3.6 Other Extensions

If a researcher is interested in the causal effects of (potentially time-dependent) treatments on time-varying regressors, neither fixed nor random effects may be appropriate (see Sobel, 2012). The methodological challenge arises because the effects of later treatments may themselves be outcomes of earlier instances of the treatment. Structural nested models, proposed in Robins (1993, 1994, 1997), were designed to estimate the effect of time-varying treatments, while marginal structural models (Robins, 1998; Robins, Hernan, and Brumback, 2000; Robins and Hernan, 2009) are a more recently developed alternative approach. This issue is discussed in Chapter 12 as well.

5.4 AN EXAMPLE: PROFICIENCY STANDARDS AND ACADEMIC ACHIEVEMENT

To illustrate the fixed and random effects approaches, we consider an open question in U.S. education policy: are more rigorous or stringent performance standards on state assessments associated with improved academic outcomes? Although states in the U.S. develop their own assessments and set their own proficiency standards, the Federal government assesses student performance for state-representative samples using a series of common tests known as the National Assessment of Educational Progress or NAEP.[8]

The National Center for Education Statistics publishes a series of studies that use the NAEP to place the states' performance or proficiency standards on a common metric—the NAEP scale for a given subject and grade level (NCES, 2011). This enables the creation of a panel dataset at the state level that includes states' performance on NAEP, states' proficiency levels or the "cutscores" that they set on their own state assessments mapped onto the NAEP scale, and an additional set of time-varying covariates measuring various student demographics and other factors shown to be related to academic outcomes. These data can be used to estimate the relationship between the stringency of states' performance standards and the average academic achievement of their students.

Specifically, the models estimated in this section condition the state average score on NAEP grade 4 mathematics (the outcome measure) on the following covariates:

- the state pro ciency or performance standard for its state grade 4 mathematics test mapped to the NAEP scale (the policy variable of interest);
- the percentage of students in the state who are eligible for free or reduced price school lunch as a measure of poverty;
- the percentage of disabled students;
- the percentage of English language learners;
- the percentage of Black students;
- the percentage of Hispanic students;
- the percentage of Asian students; and
- the percentage of Native American or Alaska Native students in each state.

NAEP is given at the state level every other year, but state test mapping data—the policy variable of interest—are currently only available for academic years 2004–5, 2006–7, and 2008–9. Thus, for these examples, $T = 3$. In the following models, dichotomous indicators or fixed effects for time periods are also included to control for any common trends among the states during this time period. Also, not all data are available for all states in each time period, but these data are assumed to be missing completely at random (Little and Rubin, 2002; see Chapter 23 of this volume).

Table 5.1 Estimation Results: Random Effects Model

Variable	Coefficient	(Std Err.)
lagged standard	0.058	(0.024)
% free and reduced lunch	0.086	(0.035)
% disabled	0.000	(0.001)
% ELL	0.001	(0.001)
% Black	−0.212	(0.041)
% Hispanic	−0.115	(0.039)
% Asian	−0.119	(0.051)
% Indian	−0.300	(0.108)
2008	0.061	(0.363)
2010	1.278	(0.451)
Intercept	236.495	(5.577)
N	132	
σ_{u_0}	2.743	
$\chi^2_{(10)}$	85.127	

Finally, because it is theoretically unlikely that a change in state standards would immediately translate into a change in academic achievement, the mapped state proficiency standard is included in each specification below as lagged one time period (two years).

5.4.1 The Random Effects Model

If we are willing to assume that the unobservables are uncorrelated with the other covariates, we can estimate the model:

$$\text{NAEP}_{it} = \mathbf{x}_{it}\beta + \delta\text{Standard}_{i,t-1} + u_{0i} + \varepsilon_{it},$$

with the random effects estimator, assuming that $u_{0i} \sim \text{Normal}(0, \sigma_{u_0})$. The results are presented in Table 5.1. The key coefficient of policy interest, δ, is positive and statistically significant at conventional levels, implying a positive relationship between stringency of academic performance standards and student achievement. The between-states variance component is estimated to be $\hat{\sigma}_{u_0} = 2.743$.

5.4.2 The Fixed Effects Approach

An alternative means of estimating the linear unobserved effects model is by means of the fixed effects transformation (time demeaning) to eliminate the time-invariant unobservables:

$$\begin{aligned}\text{NAEP}_{it} - \overline{\text{NAEP}}_i = \\ &(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\beta \\ &+ \delta(\text{Standard}_{i,t-1} - \overline{\text{Standard}}_i) \\ &+ \varepsilon_{it} - \bar{\varepsilon}_i.\end{aligned}$$

Here, we see a substantively different result (Table 5.2). The estimated coefficient for the policy variable of interest, lagged state proficiency standard, is much smaller in magnitude and no longer statistically significant at conventional levels ($p = .175$, two-tailed). We note that an estimate of the between-states variation captured in this fixed effects specification is much larger, at $\hat{\sigma}_{u_0} = 9.936$. This standard deviation is based directly on the 50 state fixed effects implicitly estimated in this model (values not shown). Inference for this variance component is made indirectly through an F-test on the underlying fixed effects. It is highly significant in this case, but implies something about the states in the sample—at least one has non-zero deviation from the mean outcome level, net of all else—rather than a property of all states in the population (at other historical periods, for example). This value is substantially larger than that estimated in the random effects model, suggesting either that there is additional state-level variation that should be modeled, or that fixed effects models are overestimating the differences we would observe in this population under some sampling mechanism.

5.4.3 The Hausman Test

So which is the correct specification? Although it does not provide a definitive answer, the Hausman test may be useful to the applied researcher to guide the choice between the FE and RE models, assuming unobserved covariates. For the example data,

$$\begin{aligned}H &= (\hat{\beta}_{RE} - \hat{\beta}_{FE})'(\hat{V}_{FE} - \hat{V}_{RE})^{-1} \\ &\quad (\hat{\beta}_{RE} - \hat{\beta}_{FE}) \\ &= 49.78.\end{aligned}$$

Table 5.2 Estimation Results: Fixed Effects Model

Variable	Coeff.	(Std Err.)
lagged standard	0.032	(0.023)
% free and reduced lunch	0.059	(0.037)
% disabled	0.0004	(0.001)
% ELL	0.001	(0.001)
% Black	−0.257	(0.305)
% Hispanic	0.037	(0.234)
% Asian	−0.896	(0.387)
% Indian	0.087	(0.385)
2008	−0.117	(0.349)
2010	0.135	(0.556)
Intercept	237.953	(7.537)
N	132	
σ_{u0}	9.936	
$F_{(59,72)}$	2.676	

Since $H \sim \chi^2$ with $k = 10$ degrees of freedom (recall, the number of time-varying coefficients in both models), the null hypothesis that the difference in estimated coefficients between the two models is not systematic is rejected at $p < .0005$. Thus, the results of the Hausman test suggest that the fixed effects model may be accounting for important unobserved heterogeneity that may otherwise be biasing the observed relationships.

5.4.4 The Hybrid Model

Alternatively, as noted in Section 5.3.5 above, one may instead fit a "hybrid" model of the fixed and random effects approaches:

$$\begin{aligned}\text{NAEP}_{it} &= (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\beta_w \\ &\quad + \bar{\mathbf{x}}_i\beta_g + \delta_w(\text{Standard}_{i,t-1} \\ &\quad - \overline{\text{Standard}}_i) \\ &\quad + \delta_g\overline{\text{Standard}}_i + u_{0i} + \varepsilon_{it}.\end{aligned}$$

The results of this model, presented in Table 5.3, are informative. The hybrid model yields the same coefficients for within-state effects as the fixed effects model, and very similar precision. The variance component σ_{u0} has been reduced to 2.712, a small drop from the random effects model's findings, reflecting the additional explanatory power of group-specific predictors. Group-level effects can be contrasted with within-group effects through the construction of a contextual effect, whose substantive interpretation is described in Raudenbush and Bryk (2002). A Wald test for the difference between any two such between/within effects is analogous to the Hausman test, and in our example, for lagged standard, the difference between 0.032 and 0.164 has an associated $\chi^2_1 = 7.80$, with $p = 0.005$, which results in the same inferential conclusion as the prior Hausman test at conventional significance levels. Note that the estimates from the hybrid model are sensitive to the choice of predictors, just as in any regression, be it using fixed or random effects for groups.

Table 5.3 Estimation Results: Hybrid Model

Variable	Coeff.	(Std Err.)
centered lagged standard	0.032	(0.023)
centered % free/reduced lunch	0.058	(0.037)
centered % disabled	0.000	(0.001)
centered % ELL	0.001	(0.001)
centered % Black	−0.254	(0.305)
centered % Hispanic	0.045	(0.234)
centered % Asian	−0.902	(0.386)
centered % Indian	0.091	(0.384)
centered 2008	−0.120	(0.348)
centered 2010	0.135	(0.555)
mean lagged standard	0.164	(0.041)
mean % free/reduced lunch	−0.412	(0.050)
mean % disabled	−0.009	(0.004)
mean % ELL	0.002	(0.003)
mean % Black	0.017	(0.046)
mean % Hispanic	−0.021	(0.035)
mean % Asian	−0.109	(0.042)
mean % Indian	−0.172	(0.094)
mean 2008	2.869	(5.474)
mean 2010	−2.135	(4.139)
Intercept	222.121	(9.988)
N	132	
σ_{u0}	2.712	
$\chi^2_{(20)}$	216.599	

5.5 CONCLUSION

The literature is full of advice, often conflicting, about when to use fixed or random effects models. Adding to this confusion, random effects models have several extensions which relax the random effects assumption, including the long history of hybrid models outlined above. Standard random effects models can obtain estimators that are potentially biased if the random effects assumption is violated, but fixed effect models come with an inability to account for any time-invariant variables, difficulty in making out-of-sample predictions, and statistical inefficiency. While the Hausman test is often noted as a tool for evaluating bias in the random effects model, Clark and Linzer (2012) have shown through simulations that it may not be reliable. Ultimately, evaluating the balance of bias and variance in the two techniques is not a simple task, but applied researchers should not immediately shy away from using a random effects model because of the bias concerns, as the associated variance reduction may be worthwhile. Our example above serves a useful guide for other potential analyses in which one fits fixed effects, random effects, and hybrid models. By exploring the relationship between all three estimates, researchers can more completely understand the associations and potential effects of some forms of confounding.

NOTES

1 Editors' note: In this chapter, the underlying model is effectively the same throughout, but by changing assumptions about model parameters and, more importantly, by changing the estimator, different goals are achieved. These competing goals are the subject of much debate in the applied statistics literature; this chapter describes the assumptions underpinning various estimators and reframes some of the debate in terms of between- and within-group predictor effects.

2 $\hat{\beta}_{0j}$ is only approximately equal to the right-hand side of equation (5.4) because the actual fitting techniques are slightly more complex. Statistical programs fit the models through either Bayesian inference or an augmented least squares technique (see Gelman and Hill, 2007, sections 18.3–18.4 for detailed descriptions on both techniques and extensions for more complex multilevel models). In the first technique, one uses an iterative algorithm that alternatively estimates the particular intercepts, and the means and variances, called the hyperparameters, which in our case above includes $\sigma^2_{u_0}$, σ^2_y, and β_0 (Lindley and Smith, 1972; Efron and Morris, 1975). In the second technique, one develops a weighted least squares regression (Afshartous and de Leeuw, 2005, pp. 112–13).

3 This is why fixed effects are sometimes called "unmodeled effects" and random effects are called "modeled effects" (Bafumi and Gelman, 2006). Additionally, see Gelman (2005) or Gelman and Hill (2007, p. 245) for a discussion of all the conflicting different meanings of the term "fixed effects". Here, we mean the fifth definition in that list.

4 The degrees of freedom associated with the demeaning of the outcomes require an adjustment to the standard error of the estimated effects. The reader may also be concerned that from a multilevel perspective, these models ignore the within-group correlation between observations. However, the demeaning process accounts for a group-constant effect, which captures a specific form of within-group correlation, analogous to first differencing with pre/post designs.

5 The following account applies to balanced data where the group size is the same for all groups. Unbalanced data require some adjustments.

6 Specifically, if the within variation is small, the fixed effects estimates may not be asymptotically normal and this may invalidate the Hausman test (Hahn, Ham, and Moon, 2011).

7 Little has been written about the relative efficiency of this compared to the fixed effects estimator, but our own simulations suggest that the precision is comparable in singly-nested models.

8 For more information on these assessments, including subjects and grade levels tested, see `http://nces.ed.gov/nationsreportcard/`.

REFERENCES

Afshartous D. and de Leeuw, J. (2005) 'Prediction in Multilevel Models', *Journal of Educational and Behavorial Statistics*, 30: 109–39.

Ahn, S.C. and Low, S. (1996) 'A Reformulation of the Hausman Test for Regression Models with Pooled Cross-Section-Time-Series Data', *Journal of Econometrics*, 71: 309–19.

Allison, P.D. (2005) *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. Cary, NC: SAS Publishing.

Allison, P.D. (2009) *Fixed Effects Regression Models*. London: Sage.

Ashenfelter, O., Levine, P.B., and Zimmerman, D.J. (2003) *Statistics and Econometrics: Methods and Applications*. New York: J. Wiley.

Bafumi, J. and Gelman, A. (2006) 'Fitting Multilevel Models When Predictors and Group Effects Correlate', Paper presented at the annual meeting of the Midwest Political Science Association, Chicago, IL.

Baltagi, B. (2005) *Econometric Analysis of Panel Data*, 3rd edition. New York: Wiley.

Bartels, B.L. (2008) 'Beyond "Fixed Versus Random Effects": A Framework for Improving Substantive and Statistical Analysis of Panel, Time-Series Cross-Sectional, and Multilevel data', Political Methodology Conference. Ann Arbor, MI. Available at `http://home.gwu.edu/~bartels/cluster.pdf` [accessed 26 August 2012].

Berlin, J.A., Kimmel, S.E., Ten Have, T.R., and Sammel, M.D. (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: 470–76.

Bole, V.A. and Rebec, P. (2004) 'Bootstrapping the Hausman Test in Panel Data Models'. Available at SSRN: `http://ssrn.com/abstract=628321` [accessed 26 August 2012].

Bryk, A.S. and Raudenbush, S.W. (1992) *Hierarchical Linear Models in Social and Behavioral Research: Applications and Data Analysis Methods*, 1st edition. Newbury Park, CA: Sage Publications.

Chamberlain, G. (1982) 'Multivariate Regression Models for Panel Data,' *Journal of Econometrics*, 1: 5–46.

Clark, T.S. and Linzer, D.A. (2012) 'Should I Use Fixed Or Random Effects?', Unpublished working paper. Available at `http://userwww.service.emory.edu/~tclark7/randomeffects.pdf` [accessed 26 August 2012].

Clarke, P., Crawford, C., Steele F., and Vignoles, A. (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 26 August 2012].

Davidson, R. and MacKinnon J.G. (1993) *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Diez Roux, A.V. (2004) 'The Study of Group-Level Factors in Epidemiology: Rethinking Variables, Study Designs, and Analytical Approaches', *Epidemiologic Reviews*, 26: 104–11.

Efron, B. and Morris, C. (1975) 'Data Analysis Using Stein's Estimator and its Generalizations', *Journal of the American Statistical Association*, 70: 311–19.

Gelman, A. (2005) 'Analysis of Variance--Why it is More Important than Ever', *Annals of Statistics*, 33: 1–53.

Gelman, A. (2006) 'Multilevel Modeling: What It Can and Can't Do', *Technometrics*, 48: 241–51.

Gelman, A. and Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge.

Goldstein, H. (2003) *Multilevel Statistical Models*, 3rd edition. New York: Wiley.

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

Hahn, J., Ham, J.C., and Moon, H.R. (2011) 'The Hausman Test and Weak instruments', *Journal of Econometrics*, 160: 289–99.

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

Hsiao, C. (2003) *Analysis of Panel Data*, 2nd edition. New York: Cambridge University Press.

Johnston, J. and DiNardo, J. (1997) *Econometric Methods*. New York: McGraw Hill Co.

Kreft, I.G. and de Leeuw, J. (1998) *Introducing Multilevel Modeling*. London: Sage Publications.

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

Lindley D.V. and Smith, A.F.M. (1972) 'Bayes Estimates for the Linear Model', *Journal of the Royal Statistical Society, Series B (Methodological)*, 34: 1–41.

Little, R.J.A. and Rubin, D.B. (2002) *Statistical Analysis With Missing Data*, 2nd edition. New York: Wiley.

Morris, C.N. (1983) 'Parametric Empirical Bayes Confidence intervals', in *Scientific inference, Data Analysis and Robustness*. Box, G.E.P., Leonard, T., and Wu, C.F. (eds) New York: Academic Press, 25–50.

Mundlak, Y. (1978) 'Pooling of Time-Series and Cross-Section Data', *Econometrica*, 46: 69–85.

National Center for Education Statistics (2011) *Mapping State Proficiency Standards On to the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005–2009, (NCES 2011–458)*. Institute of Education Sciences, U.S. Department of Education, Washington, DC.

Neuhaus, J.M. and Kalbfleisch, J.D. (1998) 'Between and Within Cluster Covariate Effects in the Analysis of Clustered Data', *Biometrics*, 54: 726–31.

Rabe-Hesketh, S. and Everitt, B.S. (2006) *A Handbook of Statistical Analyses Using STATA*, 4th edition. Boca Raton, FL: Chapman and Hall/CRC.

Raudenbush, S.W. (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: 468–91.

Raudenbush, S.W. and Bryk, A.S. (2002) *Hierarchical Linear Models*, 2nd edition. Thousand Oaks: Sage Publications.

Robins J.M. (1993) 'Analytic Methods for HIV Treatment and Cofactor Effects', in *AIDS Epidemiology Methodological Issues* (pp. 213–29). Ostrow, D.G. and Kessler, R. (eds). New York: Plenum Publishing.

Robins, J.M. (1994) 'Correcting for Non-Compliance in Randomized Trials Using Structural Nested

Mean Models', *Communications in Statistics*, 23: 2379–412.

Robins, J.M. (1997) 'Causal inference from Complex Longitudinal Data', in *Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120)* (pp. 69–111). Berkane, M. (ed.). New York: Springer Verlag.

Robins, J.M. (1998) 'Marginal Structural Models', *1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science*, 1–10.

Robins, J.M. and Hernan, M. (2009) 'Estimation of the Causal Effects of Time- Varying Exposures', in *Longitudinal Data Analysis*, Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs G. (eds). Boca Raton, FL: Chapman and Hall, 553–97.

Robins J.M., Hernan M.A., and Brumback B. (2000) 'Marginal Structural Models and Causal inference in Epidemiology', *Epidemiology*, 11: 550–60.

Robinson, GK. (1991) 'That BLUP Is A Good Thing: The Estimation of Random Effects', *Statistical Science*, 6: 15–32.

Skrondal, A. and Rabe-Hesketh, S. (2004) *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton: Chapman and Hall.

Smith, A.F.M. (1973) 'A General Bayesian Linear Model', *Journal of the Royal Statistical Society, Series B (Methodological)*, 35: 67–75.

Snijders, T.A.B. and Bosker, R. (1999) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.

Snijders, T.A.B. and Bosker, R. (2012) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd edition. London: Sage.

Sobel, M.E. (2012) 'Does Marriage Boost Men's Wages?: Identification of Treatment Effects in Fixed Effects Regression Models for Panel Data', *Journal of the American Statistical Association*, 107: 521–9.

Wakefield, J.C. (2003) 'Sensitivity Analyses for Ecological Regression', *Biometrics*, 59: 9–17.

Wooldridge, J. (2010) *Economic Analysis of Cross Section and Panel Data*. Boston: MIT Press.

Zorn, C. (2001) 'Estimating Between- and Within-Cluster Covariate Effect, With an Application to Models of international Disputes', *International Interactions*, 27: 433–45.