THE CENTRE FOR MARKET AND PUBLIC ORGANISATION

The Choice Between Fixed and Random Effects Models: Some Considerations For Educational Research

Paul Clark, Claire Crawford,
Fiona Steele and Anna Vignoles

June 2010

Working Paper No. 10/240

Published as DoQSS Working Paper No. 10-10. June 2010

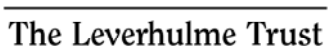
Centre for Market and Public Organisation
Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol BS8 1TX
http://www.bristol.ac.uk/cmpo/

Tel: (0117) 33 10799
Fax: (0117) 33 10705
E-mail: cmpo-office@bristol.ac.uk

The Centre for Market and Public Organisation (CMPO) is a leading research centre, combining expertise in economics, geography and law. Our objective is to study the intersection between the public and private sectors of the economy, and in particular to understand the right way to organise and deliver public services. The Centre aims to develop research, contribute to the public debate and inform policy-making.

CMPO, now an ESRC Research Centre was established in 1998 with two large grants from The Leverhulme Trust. In 2004 we were awarded ESRC Research Centre status, and CMPO now combines core funding from both the ESRC and the Trust.

ISSN 1473-625X



CMPO Working Paper Series No. 10/240

The Choice Between Fixed and Random Effects Models: Some Considerations For Educational Research

Paul Clarke[1]
Claire Crawford[2]
Fiona Steele[3]
and
Anna Vignoles[4]

[1]CMPO, University of Bristol.
[2]Institute of Fiscal Studies and Institute of Education, University of London.
[3] Centre for Multilevel Modelling, Graduate School of Education, University of Bristol.
[4] Institute of Education, University of London.

June 2010

Abstract
We discuss the use of fixed and random effects models in the context of educational research and set out the assumptions behind the two modelling approaches. To illustrate the issues that should be considered when choosing between these approaches, we analyse the determinants of pupil achievement in primary school, using data from the Avon Longitudinal Study of Parents and Children. We conclude that a fixed effects approach will be preferable in scenarios where the primary interest is in policy-relevant inference about the effects of individual characteristics, but the process through which pupils are selected into schools is poorly understood or the data are too limited to adjust for the effects of selection. In this context, the robustness of the fixed effects approach to the random effects assumption is attractive, and educational researchers should consider using it, even if only to assess the robustness of estimates obtained from random effects models. On the other hand, when the selection mechanism is fairly well understood and the researcher has access to rich data, the random effects model should naturally be preferred because it can produce policy-relevant estimates while allowing a wider range of research questions to be addressed. Moreover, random effects estimators of regression coefficients and shrinkage estimators of school effects are more statistically efficient than those for fixed effects.

Keywords: fixed effects, random effects, multilevel modelling, education, pupil achievement.

JEL Classification: C52, I21

Electronic version: www.bristol.ac.uk/cmpo/publications/papers/2010/wp240.pdf

Acknowledgements
The authors gratefully acknowledge funding from the Economic & Social Research Council (grant number RES-060-23-0011), and would like to thank Rebecca Allen, Simon Burgess and seminar participants at the University of Bristol and the Institute of Education for helpful comments and advice. All errors remain the responsibility of the authors.

Address for correspondence
CMPO, Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol BS8 1TX
paul.clarke@bristol.ac.uk
www.bristol.ac.uk/cmpo/

1. Introduction

In this paper, we discuss the use of fixed and random effects models in different research contexts. In particular, we set out the assumptions behind the two modelling approaches, highlight their strengths and weaknesses, and discuss how these factors might relate to any particular question being addressed. To illustrate the issues that should be considered when choosing between fixed and random effects models, we analyse the determinants of pupil achievement in primary school, using data from the Avon Longitudinal Study of Parents and Children.

Regardless of the research question being addressed, any model of pupil achievement needs to reflect the hierarchical nature of the data structure, where pupils are nested within schools. Estimation of hierarchical regression models in this context can be done by treating school effects as either fixed or random. Currently, the choice of approach seems to be based primarily on the types of research question traditionally studied within each discipline. Economists, for example, are more likely to focus on the impact of personal and family characteristics on achievement (Todd and Wolpin, 2003), and hence tend to use fixed effect models.[2] In contrast, an important focus for education researchers is on the role of schools (Townsend, 2007), which is best studied using random effects models because fixed effect approaches do not allow school characteristics to be modelled.

An important aim of this paper is to encourage an inter-disciplinary approach to modelling pupil achievement. In an ideal world, evidence from different disciplines would be brought together to consider the same research question, which in this case is: how can we improve pupil achievement? However, this is only possible if the assumptions underlying the different models used by each discipline are made clear. We hope to contribute to multi-disciplinary understanding and collaboration by highlighting these assumptions and by discussing which approach might be most appropriate in the context of educational research.

[2] Economists also tend to use fixed effects models in the context of panel data, where level two corresponds to the individual and level one corresponds to occasion-specific residual error.

Another important aim of this paper is to highlight that, in the case of economics and education, discipline tradition unnecessarily constrains the types of research question that are addressed. More precisely, we hope to encourage economists that they should not dismiss the random effects approach entirely. In fact, we aim to convince them that, provided they have some understanding of the process through which pupils are selected into schools, and that sufficiently rich data are available to control for the important factors, then the natural approach should be to use random effects because: a) school (level 2) characteristics can be modelled – allowing questions concerning differential school effectiveness for different types of pupils using random coefficients to be addressed – and b) precision-weighted 'shrinkage' estimates of the school effects can be used. Equally, we hope to make education researchers more aware of the issues concerning causality, the potential robustness of fixed effects, and the need for care when using results from random effects models to inform government policy, particularly where such analyses are based on administrative data sources with a limited range of variables.

To illustrate these methodological issues, we consider two topical and important research questions in the field of education. We have chosen these examples to illustrate the selection problems that should influence a researcher's choice of modelling approach. The first example relates to inequality due to the impact of having special educational needs (SEN) on pupil attainment. The second example, continuing with the theme of inequalities in education achievement, is a perennial research question: what impact do children's socio-economic backgrounds, as measured by their eligibility for free school meals (FSM), have on their academic achievement? These examples usefully highlight the importance of choosing the appropriate model for the relevant research question, as we discuss in detail below.

The rest of this paper is set out as follows: in Section 2 we review the key features of fixed and random effects models; in Section 3 we highlight the dangers that selection of pupils into schools can pose for the validity of fixed and random effects models; in Section 4 we discuss in more detail our choice of illustrative examples; in Section 5 we describe these analyses and present the results; and finally in Section 6 we make our concluding remarks.

2. Methods to Allow for School Effects on Pupil Achievement

It is well known that the achievements of pupils in the same school are likely to be clustered due to the influence of unmeasured school characteristics like school leadership. A common way to allow for such school effects on achievement is to fit a hierarchical regression model with a two-level nested structure in which pupils at level 1 are grouped within schools at level 2. If we denote by y_{ij} the achievement of pupil i in school j (i = 1,…, n_j; j = 1,…, J) then a two-level linear model for achievement can be written

$$y_{ij} = \beta_0 + x_{1ij}\beta_1 + \cdots + x_{pij}\beta_p + u_j + e_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}_1 + \mathbf{x}'_j\boldsymbol{\beta}_2 + u_j + e_{ij} \qquad (1)$$

where β_0 is the usual regression intercept; $\mathbf{x}_{ij}$ represents only those covariates that vary between pupils/families; $\mathbf{x}_j$ represents those covariates that vary only between schools; $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the respective coefficients for these vectors; u_j is the 'effect' of school j; and e_{ij} is a pupil-level residual. Typically $\mathbf{x}_{ij}$ will include a measure of prior attainment, so that u_j is interpreted as the effect of a school on academic *progress*. Note that we use the term 'effect' as a synonym for association; in the remainder of the paper we discuss the assumptions required for u_j to be interpretable as a 'causal' effect of school j on pupil achievement.

We will use (1) to express the general model again in Section 3, but for a comparison of the two approaches it is easier to use the more general representation

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_j + e_{ij} \qquad (2)$$

where $\mathbf{x}_{ij}$ is now a vector containing a constant term (the intercept) and all p covariates characterising the pupil, family or school, and $\boldsymbol{\beta}$ is the vector of p + 1 regression coefficients associated with each entry in $\mathbf{x}_{ij}$. It is important to recognise that neither the regression coefficients nor the school effects can be interpreted as *policy-relevant* effects without further assumptions, and it is the nature of these assumptions that is central to this paper.

Two assumptions that are commonly made about the pupil-level residuals are: (*i*) $e_{ij} \sim i.i.d.$ N(0, σ_e^2), and (*ii*) 'exogeneity' of the covariates $\mathbf{x}_{ij}$, i.e., cov(e_{ij}, x_{kij}) = 0 for $k = 1,\ldots, p$.[3] In fact, while the normality assumption (*i*) is desirable for reasons of estimator performance and interpretation, it is not essential for either the random or fixed effects approaches and we need only assume that Var(e_{ij}) = σ_e^2. However, assumption (*ii*) is crucial to ensure that the regression coefficient has a policy-relevant interpretation; we refer to (*ii*) as the *regression assumption* and will discuss it again in Section 3, together with a more precise explanation of what we mean by policy-relevant.

Finally, the researcher must decide whether to treat u_j as fixed or random effects. The purpose of this section is to review both approaches in terms of the underlying assumptions required by each. In Section 3 we consider the validity of these assumptions in the context of education research, and more specifically in Sections 4 and 5, we consider these assumptions' validity for our two example analyses: the impact of special educational needs status and eligibility for free school meals on pupil achievement.

2.1 Random school effects

Random effects models are also known as *multilevel* or *mixed models*. The random effects approach views the clustering of pupils in schools as a feature of interest in its own right, and not just a nuisance to be adjusted for. The school effects (also referred to as school residuals) are usually assumed to be drawn from a normal distribution: $u_j \sim i.i.d.$ N(0, σ_u^2); as with assumption (*i*) above, only Var(u_j) = σ_u^2 (rather than the assumption of normality) is required to make inferences about the regression coefficients.[4] An advantage of the random effect model is thus that the total residual variance can be partitioned into two components: the between-school variance σ_u^2 and the within-school (between-pupil) variance σ_e^2.

[3] This assumption should strictly be written as $E(e_{ij} \mid x_{kij}) = 0$, which is not quite equivalent to cov(x_{kij}, e_{ij}) = 0 because circumstances exist where the covariance can be zero even if $E(e_{ij} \mid x_{kij}) \neq 0$; ditto for cov($x_{kij}$, u_j) = 0 and $E(u_j \mid x_{kij}) = 0$. However, we take these two statements to be synonymous because a) the conditions under which zero covariance does not imply zero conditional mean are very much the exception rather than the rule, and b) the concept of two random variables being uncorrelated is more intuitive than that of a conditional mean equalling zero.

[4] Iterative or feasible generalized least squares estimators of random effects models are normally distributed in large samples, whether or not the residuals are normally distributed. However, note that most implementations of maximum likelihood and Markov chain Monte Carlo estimators do assume residual normality.

Random effects models have at least two major advantages over fixed effect models: 1) the possibility of estimating *shrunken* residuals; 2) the possibility of accounting for differential school effectiveness through the use of random coefficients models. We discuss each of these in turn.

In school effectiveness research, estimates of the school effects u_j are of major interest, but the accuracy of such estimates depends on the sample size for each school. Shrunken residuals have been proposed to allow for uncertainty in estimates for small schools (e.g. Aitkin and Longford, 1986; Goldstein, 1997). These may be estimated as:

$$\hat{u}_{j,RE} = c_j(\bar{y}_j - \bar{\mathbf{x}}'_j\hat{\boldsymbol{\beta}}_{RE}) \qquad (3)$$

where $\bar{y}_j$ and $\bar{\mathbf{x}}_j$ are school means of y_{ij} and $\mathbf{x}_{ij}$ respectively, and $\hat{\boldsymbol{\beta}}_{RE}$ is the random effects estimator of $\boldsymbol{\beta}$. The difference between the observed and predicted mean achievement for school j is multiplied by a constant c_j called the *shrinkage factor*. The shrinkage factor is defined as:

$$c_j = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + (\hat{\sigma}_e^2 / n_j)},$$

which will always be less than or equal to 1. When the number of pupils in a school (n_j) is small or the within-school variance $\hat{\sigma}_e^2$ is large relative to the between-school variance $\hat{\sigma}_u^2$, c_j will be noticeably less than 1 so that $\hat{u}_{j,RE}$ is pulled or 'shrunken' towards zero (the mean of u_j). The $\hat{u}_{j,RE}$ are therefore precision-weighted, taking into account the reliability of the residual estimate $\bar{y}_j - \bar{\mathbf{x}}_j\hat{\boldsymbol{\beta}}_{RE}$. In practice what this means is that "rogue" school effect estimates, based on very few pupils, are not given undue weight. The superiority of shrunken residuals as predictors of true school effects is long-established (e.g., Efron and Morris, 1973), and is crucially important if researchers are to interpret the school effect itself with a greater degree of confidence.

Various modifications to the standard 'random intercept' model (2) have also been proposed. One such extension, which we will make use of in this paper, is to include school-level covariates in $\mathbf{x}_{ij}$ to explore the extent to which between-school differences can be explained by observed school characteristics. A further extension is the use of 'random coefficient' models – in which the effects of pupil and family-level covariates are permitted to vary across schools – in order to study differential school effectiveness. While we do not pursue this approach in our paper (as our focus is on the effects of specific individual characteristics, rather than the school effects themselves), it is widely used in educational research (see, for example, Nuttall et al., 1989; Sammons et al., 1993).

However, whilst random effects models open up a range of interesting avenues for research and are more efficient (generate narrower confidence intervals) than fixed effects models, their use comes at the cost of an important additional assumption. This requirement mirrors the earlier regression assumption (for e_{ij}) and further states that u_j must be uncorrelated with the pupil, family and school characteristics represented by the covariates $\mathbf{x}_{ij}$ ($\text{cov}(x_{kij}, u_j) = 0$). This is often referred to as the *random effects assumption* and is of great concern to economists. In the case of modelling the determinants of pupil achievement, this assumption implies that unobserved characteristics of the school u_j that influence achievement, such as ethos or teacher quality, are not correlated with pupil, family or school characteristics that are included in the model, such as whether a pupil is eligible for free school meals or has special educational needs. We return to this issue again below.

2.2 Fixed school effects

Unlike in a random effects model, in a fixed effects model, no assumptions are made about u_j and so the school effects are treated as a nuisance. There are two alternative fixed effects estimators: one based on using dummy variables for the school effects and another in which the school effects are 'differenced' out.

The first approach involves including dummy variables for schools as additional predictors in an analysis of covariance (ANCOVA) model:

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \sum_{j=1}^{J} \gamma_j D_j + e_{ij} \qquad (4)$$

where D_j is the school *j* dummy with coefficient γ_j (j = 1,…, J). For identification, either the intercept or one of the school dummies, must be constrained to equal zero. Depending on this choice, γ_j is interpreted either as the effect of school *j* or the difference between the effects of school *j* and the reference school.

For completeness we also describe a second equivalent approach, namely, where we 'difference out' the school effects by subtracting the school means of the outcomes ($\bar{y}_j$) and covariates ($\bar{\mathbf{x}}_j$) from y_j and $\mathbf{x}_{ij}$, respectively. The model for the differences is thus

$$y_{ij} - \bar{y}_j = (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)'\boldsymbol{\beta} + e_{ij} - \bar{e}_j, \qquad (5)$$

which does not involve the school effects; the school effects are said to be 'differenced out' of the model. In applications to panel designs, where repeated measures on an individual over time (level 1) are nested within individuals (level 2), the above transformation is referred to as time demeaning (e.g. Wooldridge, 2002). We therefore refer to (5) as the *pupil demeaned* model.

The two fixed effects estimators are equivalent and obtained by estimating either (4) or (5) using ordinary least squares (OLS). As we have already noted, the main attraction of the fixed effects model is that no assumptions about u_j are required. However, the price for avoiding these assumptions is that the research questions that can be explored using a fixed effects approach are limited. The most important restriction is that, by design, the effects of school-level covariates are treated as nuisances and cannot be estimated: in (4) the school-level covariates are confounded with γ_j, while in (5) these covariates will be differenced out along with u_j.

Estimates of the school effects u_j may be obtained directly from estimation of γ_j in (4) or derived from the parameter estimates of (5) as:

$\hat{u}_{j,FE} = \bar{y}_j - \bar{\mathbf{x}}_j \hat{\boldsymbol{\beta}}_{FE}$.

where $\hat{\boldsymbol{\beta}}_{FE}$ is the fixed effects estimator of $\boldsymbol{\beta}$. In contrast to the random effects estimator (3), $\hat{u}_{j,FE}$ are not precision-weighted and may be very unreliable when n_j is small or the within-school variance is large relative to the between-school variance. Where analyses are based on population administrative data, these limitations may be inconsequential, but in analyses where some schools have only data on small numbers of pupils, these estimates can be poor because sampling variability will lead to some estimates being extremely small or extremely large relative to the true effect (Goldstein, 1997). In contrast, the equivalent shrunken residual (see above) for these schools will be close to zero to indicate that the data contain little information about its true size.

2.3 The relationship between random and fixed effects estimators

By expressing the fixed effects model in its pupil-demeaned form (5), the fixed effects estimator can be written

$$\hat{\boldsymbol{\beta}}_{\mathrm{FE}} = \left(\sum_i \sum_j (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)'\right)^{-1}\left(\sum_i \sum_j (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)(y_{ij} - \bar{y}_j)\right), \qquad (6)$$

where $\boldsymbol{\beta}$ is the vector of regression coefficients in (2). The fixed effects estimator is often called the *within* estimator because it is based on deviations of $\mathbf{x}_{ij}$ and $\mathbf{y}_{ij}$ from their group (school) means $\bar{\mathbf{x}}_j$ and $\bar{\mathbf{y}}_j$. To see clearly the relationship between the fixed and random effects estimators, we turn to the analysis by Wooldridge (2002) in which he shows that the random effects estimator can be written as

$$\hat{\boldsymbol{\beta}}_{RE} = \left(\sum_{i,j} (\mathbf{x}_{ij} - \hat{\lambda}\bar{\mathbf{x}}_j)(x_{ij} - \hat{\lambda}\bar{\mathbf{x}}_j)'\right)^{-1}\left(\sum_{i,j} (x_{ij} - \hat{\lambda}\bar{\mathbf{x}}_j)(y_{ij} - \hat{\lambda}\bar{y}_j)'\right), \qquad (7)$$

where $\hat{\lambda} = 1 - \sqrt{1/[1 + n(\hat{\sigma}_u^2 / \hat{\sigma}_e^2)]}$, and each school contains the same number of pupils ($n_j = n$). Clearly, this expression has the same form as (6). Comparing (7) with (6), it can be seen that the fixed and random effects estimators will be equal when $\hat{\lambda}$ = 1. In practice, the two estimators

will almost certainly not be equal, but it is instructive to consider scenarios under which they will be very close. First, it follows that $\hat{\boldsymbol{\beta}}_{FE} \approx \hat{\boldsymbol{\beta}}_{RE}$ if the between-school variation is small, i.e. $\hat{\sigma}_u^2 \approx 0$, in which case both estimators behave like the OLS estimator of model (2) without u_j. More generally, $\hat{\boldsymbol{\beta}}_{FE}$ will be close to $\hat{\boldsymbol{\beta}}_{RE}$ in the relatively unusual situations when either: a) n_j (the number of pupils within schools) is large, or b) when $\hat{\sigma}_u^2$ is very large relative to $\hat{\sigma}_e^2$ (when between-school variation is large relative to within-school variation, which may be likely with a high degree of sorting into schools, as in the UK). Clearly in the UK context, n_j will be large when analysing pupil achievement using the population administrative data commonly available to researchers, suggesting a situation in which these estimators will not produce wildly different results.

It should be noted that some random effects models can also be made robust to the random effects assumption. Both Skrondal and Rabe-Hesketh (2004) and Snijders and Berkhof (2004) have shown that, under certain conditions,[5] the following models are robust to the random effects assumption:

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \bar{\mathbf{x}}'_j\boldsymbol{\alpha} + u_j + e_{ij} \qquad (8)$$

and

$$y_{ij} = (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)'\boldsymbol{\beta}^* + \bar{\mathbf{x}}'_j\boldsymbol{\alpha}^* + u_j + e_{ij} \qquad (9)$$

where $\bar{\mathbf{x}}_j$ is again a vector containing the school-level covariate means and its associated regression coefficients in models (8) and (9) are $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$, respectively. The drawback in practice is that, while these random effects models can be used to obtain consistent estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}$, respectively, the interpretation of the intercept term and the level 2 random effects in

[5] If $\mathbf{z}_j$ represents the vector of all omitted variables correlated with $\mathbf{x}_{ij}$ then (8-9) can be used to obtain consistent estimates provided that: (i) the covariance between the pupil demeaned covariates and $\mathbf{z}_j$ is zero; and (ii) the model for the regression of the level 2 covariate means on $\mathbf{z}_j$ is linear.

both cases is **not** equivalent to that for model (2). In fact, the random effects in (8-9) cannot be interpreted straightforwardly as the school-level residual with the effects of the omitted variables correlated with $\mathbf{x}_{ij}$ removed. Fielding (2004) discusses why this approach is not widely used in more detail, and we consider a more transparent strategy for handling failure of the random effects assumption in the next section.

3. Policy-relevant inference and selection

As we have already mentioned, the fixed effects approach is used by some researchers, almost by default, on the basis that the random effects assumption is often a strong one. However, we will argue that this decision is often made too hastily and that the more flexible modelling strategies permitted by some random effects models can be used to offset failure of the random effects assumption. Certainly there are crucial efficiency advantages to using the random effects approach. Perhaps more importantly, we will also argue that failure of the random effects assumption is not the biggest barrier facing the researcher seeking to make pragmatic, policy-relevant inferences, and that we should be more concerned about failure of the regression assumption described earlier. In this section, we aim to clarify how hierarchical models can help to address these problems, and show how the choice of fixed or random effects should be dominated by background knowledge and the richness of the available data.

Before proceeding, it is worthwhile clarifying what we mean here by the term 'policy-relevant' inference. Ideally, policy-relevant inferences are causal inferences about average treatment effects (in our examples, the treatments are special educational needs status and eligibility for free school meals). Causal inferences tell us what would happen if we intervened and changed how things are currently done. It is well known that there are problems with making causal inferences using observational studies where the treatment is allocated non-randomly. Within the regression modelling framework, and in the absence of experimental or quasi-experimental data, these problems can only be overcome using modelling assumptions which are often unverifiable. Given this, estimating the treatment effect in an unbiased manner is problematic. However, a realistic goal is to produce policy-relevant estimates that may be biased, but are not sufficiently so as to lead to misleading policy recommendations.

To begin the comparison between the fixed and random effects approaches, recall that random effects models require that the random effects assumption holds, namely, the school-level residual is uncorrelated with any of the covariates. An important reason for the failure of the random effects assumption is the non-random selection of pupils into schools. If each school had drawn its pupils at random from the pupil population, then the random effects assumption would hold.[6] In reality, however, a non-random selection mechanism operates through which parents choose schools and some schools select which children to accept. Thus, the probability of selecting a particular school varies systematically according to a series of factors characterising the child, his/her family, the school itself, and higher levels like local education authority. Some (but not all) of these factors will also be associated with pupil attainment, either directly or indirectly through a mediating mechanism. It is these factors that are crucial and which we denote by the set F. This set can be partitioned as $F = \{F^{\text{school}}, F^{\text{other}}\}$: the first subset F^{school} contains those factors which vary only between schools, and F^{other} contains all other remaining factors.

Now consider a simple hierarchical regression model with treatment as the only covariate:

$$y_{ij} = \beta_0 + \text{Treat}_{ij}\tau + u_j + e_{ij},$$

where Treat_{ij} is a binary indicator for the 'treatment' received by pupil *i* in school *j*, and τ is the average effect of the treatment. In our subsequent examples, treatment corresponds to indicators for whether the pupil is eligible for free school meals or has special education needs. The school-level residual u_j is the sum of: a) all effects on attainment of the school-level factors (which are correlated with the treatment) in F^{school}, and b) the effects of all other school-level influences on attainment which are uncorrelated with the treatment. Hence, the random effects assumption fails if F^{school} is not empty, i.e. it contains school-level factors which are correlated both with the probability of a child 'selecting' treatment (e.g. receiving free school meals or having special educational needs) and attainment, because this induces correlation between attainment and the

[6] Provided that the regression model is correctly specified, the covariates are measured without error, and that the outcome and covariates are not simultaneously determined by the same process.

treatment variable itself. The random effects estimator of τ is therefore biased if F^{school} is not empty.

So should we conclude that the fixed effects approach is always preferable? The answer is "yes" and "no" depending on circumstances. The fixed effects estimator for $\boldsymbol{\beta}$ is robust to failure of the random effects assumption when F^{school} is not empty. However, if we have some knowledge about the school selection mechanism and can include measures of these factors in the model as 'controls', then we can also estimate the average treatment effect using the random effects model:

$$y_{ij} = \beta_0 + \text{Treat}_{ij}\tau + \mathbf{x}'_j(F^{\text{school}})\boldsymbol{\beta}_2 + u_j + e_{ij},$$

where $\mathbf{x}_j(F^{\text{school}})$ is a covariate vector containing school-level variables measuring the factor F^{school}, and $\boldsymbol{\beta}_2$ is its associated vector of regression coefficients.

We have written the covariate $\mathbf{x}_j = \mathbf{x}_j(F^{\text{school}})$ to emphasize that the choice of control variables should be based strictly on the researcher's belief about the school selection mechanism. If all the relevant factors are known, and provided that the 'control function' $\mathbf{x}'_j(F^{\text{school}})\boldsymbol{\beta}_2$ is correctly specified (e.g. using proxy variables, non-linear terms or dummy variables), then the school-level residual will contain only the omitted school-level factors described in b). As these are uncorrelated with the treatment, the random effects assumption holds and τ can be reasonably interpreted as the average treatment effect.

Of course, unless the selection mechanism is perfectly understood and the study contains reliable measures of every factor comprising F^{school} then the random effects assumption will not hold perfectly. However, this is not necessarily fatal for the purpose of producing policy-relevant inferences. If the data source is rich (in terms of school characteristics) and the selection mechanism is sufficiently understood by the researcher such that $\mathbf{x}_j(F^{\text{school}})$ includes *important*

factors from F^{school}, then the random effects assumption need not preclude the use of random effects models for policy relevant inference. This implies that the richness of data being used and theoretical understanding of school selection mechanisms will be crucial in determining which model the researcher is likely to be able to adopt. Given good data and understanding of selection effects, the efficiency advantages of random effects and its ability to consider differential school effects (through the use of random coefficient models) would make it the preferred approach.

In fact, we argue that the most important barrier to producing policy-relevant inferences is the other key assumption introduced in section 2: the regression assumption. If the treatment is correlated with any factor in F^{other} (the unobserved pupil or family characteristics in e_{ij}) then the advantage of the fixed effects approach is lost: both the random and the fixed effects models will give biased results because of the correlation between the treatment and e_{ij}. This is more than a theoretical possibility: Burgess et al. (2009) study the reasons that parents in England and Wales choose schools for their children. They identify the dominant reasons as: a) proximity/ease of travel to school; b) whether a sibling attends the same school; c) whether wider family members or friends attend the same school; d) school reputation; and e) pre-school and childcare facilities offered by the school. Given that selection depends on both the schools' selection (or non-selection) policies and the parents' choices, the effects of all of these factors work through pupils/families and so are in F^{other}, as well as in F^{school}. For example, from the family perspective, d) and e) depend on their perception and needs, respectively, and so vary between families, and hence become part of the between-pupil variation in the two-level hierarchical model. So ideally, a further set of covariate adjustments should be added to the model to give:

$$y_{ij} = \beta_0 + \text{Treat}_{ij}\tau + \mathbf{x}'_{ij}(F^{\text{other}})\boldsymbol{\beta}_1 + \mathbf{x}'_{j}(F^{\text{school}})\boldsymbol{\beta}_2 + u_j + e_{ij},$$

where $\mathbf{x}_{ij} = \mathbf{x}_{ij}(F^{\text{other}})$ is again written to emphasize that the choice of control function should be driven by understanding of the selection mechanism.

Another hypothesis is that parents select schools with large school effects. In reality, parents cannot select u_j directly because it will not have manifested itself until the children have actually spent time at the school. Hence, within this framework, selection should be modelled by including factors in F which parents use to *predict* u_j for their children; again, these will involve factors relating to the pupil, family and school.

In summary, if the regression and random effects assumptions both hold then the school selection process is 'ignorable'. In practice, however, it is unrealistic to expect either assumption to hold exactly because the selection mechanism is only partially understood and perfect measures of all the factors driving selection are rarely available. While this may seem less than satisfactory to some, we again emphasize that it is important to be realistic about what can be achieved using these analyses. Even if selection is ignorable, it cannot be guaranteed that τ equals the treatment effect: a strictly causal interpretation can really only be given if the additional assumptions embodied in the model can be shown to hold.[7] This is why the limit of our ambitions must remain producing policy-relevant inferences. Whether such policy-relevant estimates are good enough depends on the researcher's prior knowledge of the selection process and the richness of the available data in containing sufficient measures of F to ensure this.

4. Choice of treatments

Having considered the assumptions behind the two alternative models, we now introduce the two illustrative examples that we will be using to show the practical implications of using different modelling approaches in specific contexts: the relationship between pupil achievement and: 1) special educational needs status; 2) eligibility for free school meals. We discuss each of these examples in turn.

[7] In addition to the assumption that selection is ignorable conditional on the model covariates, estimation of the average treatment effect using the regression model above requires that the following assumptions hold: (a) the treatment is not constrained to have the same effect on each pupil, but the average effect of the treatment on pupils characterised by the same covariate values ($\mathbf{x}_j$ and $\mathbf{x}_{ij}$) must equal the average treatment effect itself and be the same among those who are treated as it is among those who are not; and (b) the mean treatment-free outcome (the counterfactual mean for those who are treated as well as the mean for those who are not) is linearly related to the covariates $\mathbf{x}_j$ and $\mathbf{x}_{ij}$. The average treatment effect can be estimated non-parametrically without these assumptions only if there are no population subgroups within which everyone receives treatment, or within which no-one receives treatment; see for example Wooldridge (2002, chapter 18) for a more complete account of average treatment effect estimation.

4.1 Special Educational Needs

Our first choice of illustrative example is an analysis of the gains in pupil achievement across the primary phase amongst students with special educational needs (SEN). Understanding the impact of having special educational needs on pupil achievement is not only a useful exemplar of many of the issues we want to discuss in this paper, but is also of crucial public policy importance.[8] By the end of primary school, one in four pupils in England is identified as having some kind of special educational need, with just under 4% of pupils having a statement of SEN, indicating more severe needs. Not only does SEN affect a large number of children, but it also represents a significant public investment. In 2003-04, the period relevant to our data, approximately £1.3billion was allocated to special educational needs in primary schools (excluding special schools), amounting to around £1,600, on average, per pupil with SEN.[9] Moreover, the recent Lamb Inquiry (Lamb Inquiry, 2009) recommended that the school accountability system, particularly OFSTED, should take greater account of the progress of SEN children and that judgement about the effectiveness of different schools should include greater consideration of SEN. Hence, accurately measuring the relative progress of pupils with SEN is clearly of pressing policy importance.

When modelling the relationship between SEN status and pupil achievement, there are a number of issues around causality that must be considered. Ideally we would like to measure the impact of having a specific SEN "treatment" on pupil performance. However, achieving this is problematic since SEN children receive a range of different interventions, and the level of additional support for SEN pupils will vary across schools and across type of SEN. For instance, some pupils (typically those with more severe needs) will attend special schools or have a full time teaching assistant, while others may simply be monitored more closely. In this paper, we focus on children with non-statemented (less severe) special educational needs. However, even amongst this more homogenous group, the treatment they receive will be extremely heterogenous. Furthermore, it is not possible to identify the nature of the intervention received by particular pupils using available data; we only know whether a pupil is identified as having SEN

[8] The specifics of SEN and its impact on achievement are considered in more detail in a sister paper (Crawford and Vignoles, 2010).
[9] Source: authors' calculations based on Section 52 financial returns data. See Crawford and Vignoles (2010) for more details.

or not.[10] Hence, we cannot model SEN as a conventional "treatment" in the usual evaluation sense. Our models are therefore limited to estimating the relationship between being identified as having non statemented SEN and academic progression, recognising the heterogeneous nature of the SEN programme as it is being implemented among the school population.

More fundamentally, a hierarchical analysis of the effects of SEN on pupil achievement is particularly appropriate because much of the variation in SEN labelling is at the school level due to variation in school and local authority policies (see Lamb Inquiry, 2009, for a full exposition of this issue). This variation in policy highlights a key methodological issue in relation to measuring the impact of a SEN treatment on pupil achievement: a priori, we might be concerned that pupils with SEN are more likely to attend schools that have particular unobserved characteristics, such as a more supportive ethos, and that such characteristics may also be correlated with pupil achievement. Such selection, if inadequately captured by the inclusion of observable school characteristics in a random effects model, would lead to a failure of the random effects assumption, rendering such models inappropriate for estimating the impact of SEN on pupil achievement.

4.2 Free School Meals

The second illustrative example we use is an analysis of the impact of eligibility for free school meals (FSM) status – an indicator of very low family income[11] – on educational achievement. From a policy perspective, this is a crucial area of research, as we know that gaps in educational achievement between richer and poorer pupils are substantial in the UK, at least by international standards.[12] Whether or not a child is eligible for FSM is the primary indicator of their socio-economic status recorded in schools administrative data and is regularly used by the government to assess the extent of the socio-economic gap in educational achievement (DfES, 2006). Moreover, much of the academic evidence base on the relationship between family background

[10] In more recent data, it is possible to identify which class of intervention non-statemented pupils receive – school action or school action plus – however this information is not available for our cohort of interest.
[11] Pupils are *entitled* to free school meals if their parents receive income support, income-based jobseeker's allowance, or child tax credit with a gross household income of less than £15,575 (in 2008–09 prices). They are *eligible* for free school meals if they are both entitled and registered as such with their local authority.
[12] http://www.oecd.org/dataoecd/60/1/39727764.ppt#2390,21, Slide 21.

and education achievement relies on this indicator (Chowdry et al. 2008; Sammons et al., 1997; Strand, 1999).

While previous work has suggested that FSM is an imperfect proxy for family income (Hobbs and Vignoles, 2008; Shuttleworth, 1995), we put this issue to one side in this paper and focus instead on the specific issue of how one might model the impact of FSM status on pupil achievement in the primary years. The "treatment" effect we focus on, therefore, is a comparison between pupils' attainments if they had low socio-economic status, as indicated by FSM eligibility, compared to their attainments under higher socio-economic status.

As in the case of our model of the impact of SEN status on educational achievement, a hierarchical model which takes account of the fact that children are clustered in schools is appropriate, because much of the variation in FSM status is across schools, as poorer children tend to be clustered in the same neighbourhoods and geographical area, and hence in the same schools. However, our major methodological concern in this example is the possibility that pupils' FSM status may be correlated with other unobserved pupil or family characteristics, such as parental aspirations, that also determine their achievement. If this is the case – and we are not able to adequately control for such factors in our model – then neither the fixed nor random effects models will provide unbiased estimates of the impact of FSM on achievement, and we would need to turn instead to alternative methods – such as instrumental variables – to obtain policy-relevant estimates.

5. Application: Analysis of Pupil Progress in Primary School

We now illustrate our applications of fixed and random effects models in an analysis of national achievement test results at Key Stage 2 (KS2) (sat at age 11), adjusting for performance at Key Stage 1 (KS1) (age 7). As described above, the parameters of interest are the effects of two pupil-level covariates: non-statemented special educational needs (SEN) status and eligibility for free school meals (FSM).

5.1 Data sources

We use the Avon Longitudinal Study of Parents and Children (ALSPAC) for our analysis.[13] ALSPAC is a longitudinal survey which has followed the children of around 14,000 pregnant women whose expected date of delivery fell between 1st April 1991 and 31st December 1992, and who were resident in the Avon area of England at this time. These sample restrictions mean that ALSPAC cohort members were typically born in one of three academic years: 1990-92 (sitting KS2 in 2001-02), 1991-92 (sitting KS2 in 2002-03) and 1992-93 (sitting KS2 in 2003-04).

ALSPAC cohort members and their families have been surveyed via high frequency postal questionnaires from the time of pregnancy onwards, with information collected on a wide range of family background characteristics, including mother's and father's education and occupational class, income, housing tenure, and so on (see Table 1 for details). Key Stage test results (at ages 7, 11, 14 and 16), plus limited personal characteristics from the annual school census – including our two covariates of interest –have also been linked in from administrative sources.

These characteristics are typically available in most English longitudinal surveys. In addition, however, ALSPAC cohort members have also been monitored through a number of hands-on clinics, during which staff have administered a range of detailed physical, psychometric and psychological tests. This provides us with a wealth of additional information which is not generally available in other longitudinal studies, including IQ and various measures of non-cognitive skills, such as the Strengths and Difficulties Questionnaire.[14] Furthermore, the teachers and head-teachers of ALSPAC cohort members have also been asked to complete questionnaires on more than one occasion, providing us with more detailed school characteristics than are typically available in other longitudinal studies, including head-teacher tenure.

Our sample comprises those individuals for whom we observe KS1 and KS2 scores, plus our covariates of interest (taken from administrative data at age 11) and information about the school that they attend in Year 3 (at the start of the Key Stage 2 curriculum period). This leaves us with

[13] See http://www.bristol.ac.uk/alspac/sci-com/ for more details on the ALSPAC data resource.
[14] See http://www.sdqinfo.com/b1.html for more details.

a sample of 5,417 pupils. Table 2 provides some selected descriptive statistics of this sample, and a comparison between our sample and the relevant school-age population in England.

5.2 Analysis strategy

The objective of our analysis is to illustrate the key methodological points raised in Section 3. There are two main issues that we focus on: the circumstances under which a) the random effects assumption fails and b) the regression assumption fails.

To address these issues, we consider a range of random and fixed effects models for KS2 standardised scores[15], starting with models including only KS1 standardised scores and our treatment (either SEN or FSM), and then successively adding pupil and family-level characteristics and finally (to random effects models) school-level variables. By doing this we effectively illustrate the bias arising as we move from a sparse data source through to a richer source. As shown in Table 1, we distinguish three types of data source for indicators of pupils' family background: administrative data, data available from typical longitudinal surveys, and rich cohort study data (such as that available in ALSPAC). For each model, we compare the estimated coefficients of SEN and FSM obtained from random and fixed effects models.

If we find differences between the random and fixed effects estimates of the effects of SEN and FSM on attainment then it would suggest that unmeasured but important school influences on progress are correlated with SEN/FSM, in which case the random effects assumption that u_j is uncorrelated with SEN/FSM is invalid. Of particular interest to us is the extent to which any difference between fixed and random effects estimates of SEN and FSM can be reduced by including school-level variables available from administrative and survey sources. It is also possible that adjusting for pupil-level characteristics may bring fixed and random effects estimates closer together, if these characteristics are correlated with unmeasured school-level predictors of progress.

As a secondary question we also examine whether, using either fixed or random effects models, the effects of SEN and FSM are explained by pupil and family level characteristics available

[15] We standardise scores within sample.

from different data sources. For example, is the effect of SEN status on progress explained by ability measures not captured by the KS1 score (such as parental education and children's IQ)? Is the effect of FSM explained by indicators of family circumstances (such as income and housing tenure) during childhood? This will enable us to better understand whether the regression assumption is likely to hold in practice when estimating pupil achievement models using sparser administrative data with limited covariates.

5.3 Effects of special educational needs status on progress

Table 3 shows the distribution of the proportion of non-statemented SEN pupils across primary schools in our ALSPAC sample. On average 20.4% of pupils are recorded as having non-statemented special educational needs, but there is substantial variation across schools, with a quarter of schools having fewer than 15% of pupils identified as SEN and a further quarter with more than 24% of pupils with SEN. A key question to be addressed in this paper is whether the factors driving differences between schools in the proportion of pupils labelled as SEN are correlated with unmeasured school-level influences on academic progress, after controlling for pupil characteristics and longitudinal indicators of family circumstances. If any correlation between these two sets of factors cannot be explained by available school-level measures, then the random effects estimator of the effect of SEN on progress in primary school will be biased.

Turning to our estimates (Table 4) we find a large negative effect of SEN status on progress, regardless of whether school effects are treated as fixed or random, such that pupils recorded as having non-statemented SEN score around 0.3 standard deviations lower at Key Stage 2 than pupils (with the same level of prior attainment) who do not have special educational needs. Furthermore, the effect of SEN remains substantial even after controlling for an array of child and family measures, with the addition of rich cohort measures (mainly IQ) leading to the largest reduction (M4 vs M3). This suggests, somewhat surprisingly perhaps, that the rich measures of family background, IQ and non-cognitive skills that are available in ALSPAC are not particularly highly correlated with the likelihood of having special educational needs (or attainment), and thus that the regression assumption is likely to hold in practice. However, controlling for additional pupil and family background characteristics accounts for a large amount of the between-school variation in progress, with the proportion of residual variance

found at the school level falling from 16.1% when administrative variables are included to 7.6% when richer longitudinal measures are added.[16]

Whatever set of explanatory variables is considered, the random and fixed effects estimates for SEN are very similar, with a relative difference of no more than 2.2%; this difference falls to 0.6% with the inclusion of school-level variables (M5 versus M4). The similarity of the random and fixed effects estimates suggests that the variation in SEN across schools (shown in Table 3) is driven by factors that are not associated with KS1-KS2 progress. It also implies that the between-school effect of SEN is negligible, which is confirmed by a small and non-significant effect of the school-level mean of SEN when added to M5 (results not shown). Although small, the between-school effect of SEN is negative which, when combined with a negative within-school effect, causes the random effects estimate (a weighted average of the two) to be greater in magnitude than the fixed effects (within) estimate.

In summary, our findings suggest that both the regression and random effects assumptions are likely to hold in this example. This suggests that we should prefer the random effects model here, as it is likely to produce unbiased (policy-relevant) estimates of the effect of SEN on pupil progress, even with only very limited information on pupil, family and school characteristics.

5.3 Effects of eligibility for free school meals on progress

As for our analysis of special educational needs, there is a large amount of between-school variation in the proportion of pupils eligible for free school meals, with figures ranging from 5.6% to 17.7% for schools in the middle 50% of the distribution.

Table 5 shows estimated effects of eligibility for FSM on progress for the same set of models considered in the analysis of SEN. In contrast to the results for SEN, we find that, for both fixed and random effects models, the effect of FSM on attainment becomes increasingly smaller as more pupil and family background controls are introduced. The reduction in the effect of FSM from M1 to M2 is largely explained by ethnicity and the indicator for English as an additional

[16] $\hat{\rho}_{RE}$ may also be interpreted as the correlation between the KS2 scores of two randomly pupils from the same school and who have the same values for KS1 and the other covariates in the model.

language, while subsequent decreases are observed after adjusting for various measures of family socioeconomic circumstances during childhood. This suggests that models which are only able to control for a limited range of background characteristics, as is the case when using administrative data, are likely to produce biased estimates of the effect of FSM on attainment, because of the failure of the regression assumption.

Even after we have added in the richest set of family background controls, however, there remains a negative association between eligibility for free school meals and academic progress, with FSM pupils scoring around between 0.09 (FE) and 0.1 (RE) standard deviations lower at Key Stage 2 than pupils who are not eligible for FSM (with similar prior attainment). This suggests that FSM has some value as an indicator of current economic hardship, over and above detailed information about family income, housing tenure and so on during early childhood.

There are also larger differences between the fixed and random effects estimates of the impact of FSM eligibility than there are for SEN. When only pupil and family characteristics are considered (M1 to M4), the relative difference between the two sets of estimates ranges from 11.5% (M1) to 15.7% (M3). In contrast to the SEN analysis, the random effects estimates are all greater in magnitude than the fixed effects estimates, which is due to a negative within-school effect combined with a negative between-school effect of FSM. The estimates become much closer when school-level variables are added (M5). On closer examination, however, this is found to be due entirely to the addition of the school-level proportion eligible for FSM. Although this measure is derived from administrative data on all pupils in a school (rather than from our survey data), it is highly correlated with the proportion of FSM among ALSPAC children in a school. Thus the random effects version of M5 has a similar form to (7) with separation of the within and between-school effects of FSM (with the within effect approximated by the coefficient on FSM and the between effect approximated by the coefficient on the school level mean of FSM). As noted in Section 5.3, inclusion of the school mean $\bar{x}_j$ of a pupil-level covariate x_{ij} is one way of removing the correlation between x_{ij} and u_j, in which case the fixed and random effects estimators (of the within school effect of FSM on achievement) will

coincide. If we do not include $\overline{\text{FSM}}_j$ in M5, the estimated coefficient of FSM_{ij} changes little from its estimate in M4, which suggests that there remains correlation between FSM_{ij} and u_j.

In summary, we find that without the inclusion of school characteristics (particularly the school-level proportion of students eligible for FSM) in our model, the random effects assumption is likely to fail in this case, perhaps leading us to favour a fixed effects approach here. However, we also find that without very rich data on individual characteristics, the regression assumption is likely to fail, such that neither fixed nor random effects models will produce unbiased, policy-relevant estimates of the relationship between FSM eligibility and attainment, particularly when using sparse administrative data.

6 Discussion

The primary aim of this paper has been to highlight the key issues that should be considered when deciding whether to use fixed or random school effects in models of pupil achievement. To illustrate these issues, we have examined the determinants of progress in primary schools amongst pupils with non-statemented special educational needs (SEN) and pupils who are eligible for free school meals (FSM). Two issues have been of paramount concern throughout:

1) the presence of correlation between unobserved individual characteristics (that are relevant for attainment) and our treatment indicators (SEN/FSM); both fixed and random effects models assume that there is zero correlation, which we have referred to throughout as the regression assumption;

2) the presence of correlation between unobserved school characteristics (that are relevant for attainment) and our treatment indicators; random effects models (but not fixed effects models) assume that there is zero correlation, which we have referred to throughout as the random effects assumption.

In essence, both of these concerns are related to the non-random selection of pupils (with certain characteristics) into schools, and the extent to which estimates from hierarchical models can be

interpreted as policy-relevant or simply measures of association stems from the ability of the researcher to account for these different types of selection in their model.

We take a pragmatic view of what can reasonably be achieved by analysing data from observational studies, be it by random or fixed effects estimation of hierarchical regression models. Causal inferences require randomised interventions or, failing that, quasi-randomised experimental designs involving instrumental variables or regression discontinuity. Ultimately, we argue that a realistic aim of analyses based on observational studies should be the production of inferences that are 'policy-relevant'; that is, estimates of average treatment effects affected by bias sufficiently small that misleading policy recommendations are difficult to make. Even this limited aim requires modelling the selection process (i.e. accounting for factors which are correlated both with our treatment effect and pupil achievement) as described in Section 3. In the context of education research questions such as those considered here, we highlight that this judgement should be based on knowledge of the mechanism through which parents select schools and schools select children (on the basis of individual characteristics), and the richness of the available data with which to adjust for the factors driving this selection. This fits in with recent recommendations by Rubin (2008) on how to justify that estimates have a causal interpretation. He goes further and recommends that all studies pertaining to causal inference are preceded by an analysis in which the postulated selection model is tested to ensure it satisfies some important but necessary requirements. However, even here, the infallibility of any conclusions cannot be guaranteed, but the assumptions under which any policy recommendations are made can be highlighted to policy-makers and subjected to critical scrutiny. Furthermore, articulating one's assumptions should form the basis of a common 'language' for discussion and criticism in areas of multidisciplinary research such as education.

While the production of policy relevant estimates requires both fixed and random effects models to account for selection on the basis of *individual* characteristics, the use of random effects models also requires adjustments for selection to be made on the basis of *school* characteristics. Thus, random effects models can be used with confidence in scenarios in which the selection mechanism (on the basis of both individual and school characteristics) is well understood and the available data has measures of the important factors influencing it. In such circumstances,

economists should consider using these models to take advantage of 'shrunken' estimates of school effects which downplay unreliable estimates, and the ability to model differential school effectiveness using random coefficients models. For instance, in the present study, we might expect variation between schools in the effect of being eligible for free school meals on achievement. Having established that the effect of FSM does indeed vary across schools, we might then consider whether this variation can be explained by interactions between FSM and a school covariate: for example, the proportion of children eligible for FSM in a particular school. This approach constitutes an elegant and practicable way of exploring the variation in covariate effects between schools; while differential school effects can be explored using the fixed effects approach by including interactions with school dummies, effects in schools with few pupils or few FSM or SEN pupils will be poorly estimated. We have focussed on 'random intercept' models to allow a direct comparison with fixed effects models here, but there are many examples illustrating the benefits of random coefficient models in an educational context (e.g., Aitkin and Longford, 1986; Nuttall et al., 1989; Sammons et al., 1993; Goldstein, 1997).

The very different findings of our two example analyses (on the relationship between special educational needs status and eligibility for free school meals, and pupil attainment) clearly indicate that different approaches are appropriate in different contexts, and that being overly reliant on one methodological approach is restrictive. The fixed effects approach will be preferable in scenarios where the primary interest is in policy relevant inference but selection is insufficiently understood or the data have a limited range of variables with which selection can be modelled: its robustness to the random effects assumption is attractive and educational researchers should consider using a fixed effects model in such scenarios, even if only to assess the robustness of estimates from an equivalent random effects model. On the other hand, when the selection mechanism is well understood and the researcher has access to rich data, the random effects model should be preferred both for its efficiency over the fixed effects approach, as well as its use of shrunken residuals and its ability to model differential school effectiveness.

In the context of educational research, however, it may be the regression assumption that is the greatest barrier to producing policy-relevant inferences, because many of the important factors influencing selection (of pupils into schools) are at the individual level. Indeed, this was borne

out by our analysis of the relationship between FSM eligibility and educational progress excluding many of the individual characteristics provided by our rich survey data, the results of which suggest that the failure of the regression assumption – particularly when using sparse administrative data – is likely to be the greater source of bias.

The potential failure of the regression assumption also raises some doubts about the use of the Hausman test as a statistical tool for determining whether a fixed or random effect model is most appropriate (e.g. Wooldridge, 2002). This is because the test itself is based on the difference between the regression coefficients for equivalent fixed and random effects models (i.e. with the same covariates included in the model) under the null hypothesis that both models are correctly specified. If the test is significant then the alternative hypothesis is often interpreted as evidence for failure of the random effects assumption. In panel data analysis, and other scenarios where the regression assumption can be taken to hold, it is reasonable to use this test to choose between the fixed or random effects approaches. However, if the regression assumption fails then a significant Hausman test result cannot be so easily interpreted; the alternative hypothesis confounds failure of the random effects assumption with all other aspects of model misspecification, and so cannot reliably be used to choose between approaches (Fielding, 2004).

To conclude, we hope to have clarified for educational researchers that economists tend to prefer fixed effects models because of their focus on causal – or, more realistically, policy-relevant – inference of the effects of individual characteristics on some outcome of interest, not least because the fixed effects approach is able to account for school-level factors influencing the selection of pupils into schools, even if these factors are unknown or unavailable in the available data. (It may also stem partly from their use in panel data analysis, with repeated observations for the same individual over time, where the random effects assumption is very unlikely to hold.) Conversely, we hope to have clarified to economists working in education that fixed effects models are not a panacea for causal inference: fixed effects are not robust to failure of the regression assumption; and policy-relevant inferences can also be obtained from random effects models if the researcher has a sufficiently good knowledge of the selection process, as well as rich enough data on both individual and school characteristics to account for this process. Moreover, random effects models offer distinct advantages over fixed effects models in terms of

their efficiency, their ability to calculate shrunken residuals and their ability to account for differential school effectiveness using random coefficient models. To help build inter-disciplinary understanding, therefore, researchers of all persuasions should clearly state the assumptions they have made about the selection of pupils into schools, and demonstrate that these assumptions have been modelled appropriately.

Finally, while the methodological points we raise are not new, we hope to have made clear that while the fixed effects approach has advantages, it limits the types of research question which can be addressed: if rich data are available, then random effects models have qualities very close to those of fixed effects models, and allow researchers to address a wider range of research questions. By making these trade-offs clear, we hope that this paper encourages researchers from all disciplines not to be constrained by dogma in their choice of method.

References

Aitkin, M. and Longford, N. (1986), Statistical modelling in school effectiveness studies, *Journal of the Royal Statistical Society*, Series A, vol. 149, pp. 1-43.

Burgess, S., Greaves, G., Vignoles, A. and Wilson, D. (2009), *What parents want: school preferences and school choice*, CMPO Working Paper No. 09/222.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A. and Vignoles, A. (2008), *Widening Participation in Higher Education: analysis using linked administrative data*, Institute for Fiscal Studies Report No. R69.

Crawford, C. and Vignoles, A. (2010), *An analysis of the educational progress of children classified as having special educational needs*, Institute of Education, mimeo.

DfES (2006), http://www.dcsf.gov.uk/rsgateway/DB/SFR/s000640/SFR09_2006.pdf

Efron, B. and Morris, C. (1973), Stein's estimation rule and its competitors – an empirical Bayes approach, *Journal of the American Statistical Association*, vol. 68, no. 1, pp. 117-130.

Fielding, A. (2004), *The role of the Hausman test and whether higher level effects should be treated as random or fixed*, Multilevel modelling newsletter, vol. 16, no. 2.

Goldstein, H. (1997), Methods in school effectiveness research, *School Effectiveness and School Improvement*, vol. 8, no. 4, pp. 369-395.

Hobbs, G. and Vignoles, A. (2009), Is children's free school meal 'eligibility' a good proxy for family income?, *British Educational Research Journal*, 1469-3518.

Lamb Inquiry (2009), http://www.dcsf.gov.uk/lambinquiry/downloads/Lamb%20Inquiry%20Review%20of%20SEN%20and%20Disability%20Information.pdf

Nuttall, D. L., Goldstein, H., Prosser, R. and Rasbash, J. (1989), Differential school effectiveness, *International Journal of Educational Research*, vol. 13, pp. 769-776.

Rubin, D. B. (2008), For objective causal analysis, design trumps analysis, *Annals of Applied Statistics*, vol. 2, no. 3, pp. 808-840.

Sammons, P., Nuttal, D. and Cuttance, P. (1993), Differential school effectiveness: results from a reanalysis of the Inner London Education Authority's Junior School Project data, *British Educational Research Journal*, vol. 19, no. 4, pp. 381-405

Sammons, P., West, A. and Hind, A. (1997), Accounting for variations in pupil attainment at the end of Key Stage 1, *British Educational Research Journal*, vol. 23, no. 4, pp. 489-511.

Shuttleworth, I. (1995), The Relationship between Social Deprivation, as Measured by Individual Free School Meal Eligibility, and Educational Attainment at GCSE in Northern Ireland: A Preliminary Investigation, *British Educational Research Journal*, vol. 21, no. 4, pp. 487-504.

Skrondal, A. and Rabe-Hesketh, S. (2004), *Generalized latent variable modelling: multi-level, longitudinal and structural equation modelling*, Chapman & Hall/CRC, Florida.

Snijders, T. and Berkhof, J. (2004), Diagnostic checks for multilevel models, in de Leeuw, J. and Kreft, I. (eds), *Handbook of Quantitative Multilevel Analysis*, Sage.

Strand, S. (1999), Ethnic group, sex and economic disadvantage: associations with pupils' educational progress from Baseline to the end of Key Stage 1, *British Educational Research Journal*, vol. 25, no. 2, pp. 179-202.

Todd, P. and Wolpin, K. I. (2003), On the Specification and Estimation of the Production Function for Cognitive Achievement, *Economic Journal*, vol. 113, no. 485, pp. F3-33.

Townsend, T. (2007), International Handbook of School Effectiveness and Improvement Review, Reflection and Reframing, *Springer International Handbooks of Education*, vol. 17, XIV, 998.

Wooldridge, J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Tables

Table 1 Explanatory variables by type of data source

Administrative[(i)]	
Eligible for free school meals (FSM)	Ethnicity (white vs. non-white)
Special educational needs (SEN) status	English as an additional language (EAL)
Month of birth	Gender
Standardised Key Stage 1 scores	
Typical longitudinal survey[(ii)]	
Child circumstances around birth	
Birth weight	Mother's age at birth
Multiple birth indicator	Number of older siblings
Ever breastfed	Mother's marital status at birth
Parental characteristics	
Mother's and father's occupational class	Mother's and father's level of education
Family circumstances during childhood	
No. younger siblings (at 81 months)	
Mean household income (at 33 and 47 months)	
Ever in financial difficulties (during pregnancy, or at 8, 21 or 33 months)	
Ever lived in council or housing association rented accommodation	
Lived in owner-occupied housing since birth	
Child cognitive and behavioural measures	
Child has internal locus of control at age 8	Self-perception of reading ability at age 9
Child has external locus of control at age 8	Self-perception of maths ability at age 9
Whether child likes school at age 8	Played truant by age 7 (teacher report)
Rich cohort study data[(iii)]	
Mother's and partner's parenting scores at 6, 18, 24 and 38 months	
Mother/partner reads to child	
IQ at age 8 (WISC scale)	
SDQ score at age 6 (reported by mother) and age 7 (reported by teacher)	
Depression score at age 10	
School characteristics[(iv)]	
School size	Percentage eligible for FSM
Mean Key Stage 2 class size	Percentage non-white
School type	Percentage EAL
Duration head teacher in post	Percentage non-statemented SEN

Notes: (i) We use administrative data from the Pupil Level Annual Schools Census (PLASC); (ii) These measures are available in typical longitudinal surveys, although our measures are taken from ALSPAC; (iii) We use variables from ALSPAC; (iv) School-level data are mainly available from administrative sources (PLASC) with information on the duration that a headteacher has been in post taken from ALSPAC.

Table 2 Our sample

	Our sample	Pupils in England taking KS2 in 2001-02, 2002-03 or 2003-04*
Administrative		
Achieved expected level at KS2	83.6%	80.9%
Eligible for FSM	10.3%	17.7%
Non-statemented SEN	16.9%	20.0%
Non-white ethnic origin	4.4%	19.0%
Typical longitudinal survey		
Mother has at least O-level qualifications	67.2%	
Mother has a degree	8.6%	
Partner has at least O-level qualifications	63.4%	
Partner has a degree	12.7%	
Child has ever lived in a single parent family	9.3%	
Child has ever lived in social housing	20.0%	
Child was breastfed	70.0%	
Rich cohort study data		
Mother frequently reads to child	62.6%	
Partner frequently reads to child	28.2%	
School characteristics		
Attends a community school	66.3%	64.9%
Average school size	294 pupils	325 pupils
Average KS2 class size	29 pupils	24 pupils
Observations	5,417	1,911,767

* Source: authors' calculations using the National Pupil Database.

Table 3 Distribution of school-level proportions of SEN and FSM pupils

					Quartiles			
	Mean	St Dev	Min	Max	25%	50%	75%	n
SEN	0.204	0.072	0	0.396	0.153	0.200	0.244	200
FSM	0.144	0.134	0	0.632	0.056	0.100	0.177	200

Table 4 Estimated effects of SEN status on progress between KS1 and KS2 for various fixed and random effects model specifications

Model	Fixed effects		Random effects			
	$\hat{\beta}_{FE}$	(se)	$\hat{\beta}_{RE}$	(se)	$\hat{\rho}_{RE}$ [(i)]	% difference[(ii)]
M1. KS1 average point score only	-0.335	(0.025)	-0.330	(0.025)	0.175	1.5
M2: M1 + administrative data	-0.347	(0.025)	-0.342	(0.025)	0.161	1.4
M3. M2 + typical survey data	-0.355	(0.025)	-0.349	(0.024)	0.086	1.7
M4: M3 + rich cohort data	-0.321	(0.024)	-0.314	(0.024)	0.076	2.2
M5: M4 + school-level data	-0.321	(0.024)	-0.319	(0.024)	0.064	0.6
Number of pupils	5,417					
Number of Schools	200					

Notes: (i) $\hat{\rho}_{RE}$ is the intra-school correlation estimated from the random effects model as $\hat{\sigma}_u^2/(\hat{\sigma}_u^2+\hat{\sigma}_e^2)$, (ii) the relative difference between the fixed and random effects estimates is calculated as $100\times(\hat{\beta}_{RE}-\hat{\beta}_{FE})/\hat{\beta}_{FE}$

Table 5 Estimated effects of FSM eligibility on progress between KS1 and KS2 for various fixed and random effects model specifications

Model	Fixed effects		Random effects			
	$\hat{\beta}_{FE}$	(se)	$\hat{\beta}_{RE}$	(se)	$\hat{\rho}_{RE}$	% difference
M1. KS1 average point score only	-0.157	(0.028)	-0.175	(0.028)	0.145	11.5
M2: M1 + administrative data	-0.122	(0.028)	-0.138	(0.027)	0.161	13.1
M3. M2 + typical survey data	-0.089	(0.029)	-0.103	(0.028)	0.086	15.7
M4: M3 + rich cohort data	-0.089	(0.028)	-0.102	(0.028)	0.076	14.6
M5: M4 + school-level data	-0.089	(0.028)	-0.095	(0.028)	0.064	6.7
Number of pupils	5,417					
Number of Schools	200					

*Calculated as $100\times(\hat{\beta}_{RE}-\hat{\beta}_{FE})/\hat{\beta}_{FE}$