Using Compulsory Mobility to Identify Heterogeneous School Quality and Peer Effects*

Francis Kramarz[†] Stephen Machin[‡] Amine Ouazad [§]

December 2010

Abstract

Nonexperimental estimates of educational production functions typically identify school quality using pupil mobility across schools. However, in many cases, pupils can stay in the same school for multiple grades, and only a selected subsample of movers identifies school quality and peer effects. We focus on a unique subset of English primary schools in which pupils have to move between grade 2 and grade 3. In this subset, all pupils contribute to the identification of school quality and peer effects, and the paper shows that this strongly reduces endogenous mobility bias. Peer effects are identified using cohort-to-cohort variation in grade composition. We use this strategy to estimate a full-fledged educational production function where school quality and peer effects depend on student characteristics, i.e. there is a "match" between schools, peers, and pupils. The model incorporates long-run effects of school quality, pupil background, and peer effects. Results show that there is substantial heterogeneity in peer effects and school quality across pupils, and question the idea that there are good or bad schools, or good or bad peers regardless of student characteristics.

*We would like to thank audiences of various seminars and conferences. In particular, John Abowd, Sandra Black, Eric Maurin, Thomas Piketty, Jean-Marc Robin suggested helpful improvements. Previous versions of this paper were presented at the May 2006 COST meeting, the Jamboree of the European Doctoral Program, the Lunch Seminar of Paris-Jourdan Sciences Economiques, the CREST internal seminar, the 2006 IZA prize in labor economics conference, the summer research program at Boston College, the networks seminar at Cornell University, the Frontiers of Science Symposium of the National Academy of Sciences, and the Matching and Sorting conference in Sciences Po Paris. The authors acknowledge financial and computing support from INSEAD, the Centre for Economic Performance, the Centre for the Economics of Education, CREST-INSEE, and the Economics of Education and Education Policy in Europe (EEEPE) program. The econometric packages `a2group`, `a2reg`, `xtlreg`, and `xtlreg2` are available by email or through the Statistical Software Components library at Boston College.

[†]CREST, CEPR and IZA

[‡]University College London, Centre for Economic Performance, London School of Economics and CEPR

[§]INSEAD, Centre for Economic Performance at the London School of Economics, and CREST; INSEAD, Boulevard de Constance, 77300 Fontainebleau Cedex, amine.ouazad@insead.edu

1 Introduction

The measurement of school quality and peer effects in the classroom has been the subject of intense scrutiny in recent years (Hoxby & Weingarth 2007, Graham 2008, Rivkin, Hanushek & Kain 2005, Sacerdote 2010, Epple & Romano 2010, Kane & Staiger 2002). Are peer effects significant? Are peer effects small? Is school quality larger or smaller than peer effects? Are schools equally effective for all students? Should we invest in schools' environment or, rather, in pupils' early environment? Indeed, getting estimates of school quality and peer effects is crucial for an understanding of the determinants of educational achievement.

Accurately estimating school quality and peer effects requires use of educational production functions that capture the effect of inputs in both observable and unobservable dimensions. Proxying school quality by various observable indicators like the teacher-pupil ratio, teacher education, teacher experience, teacher salary or expenditures per pupil is not sufficient to capture the entire contribution of schools to educational achievement. In the U.K. (Levacic & Vignoles 2002) and in the U.S. (Hanushek 1986, Betts 1995, Hanushek, Rivkin & Taylor 1996, Hanushek 2003), school resources and teacher characteristics tend to have a small effect on achievement. It is therefore important to capture the observed and the unobserved effects of schools on achievement. However, restrictive assumptions are needed both for the identification of school quality and for the identification of peer effects using nonexperimental data.

The identification of school quality typically relies on pupil mobility across schools. Gibbons (2007) and Dobson, Henthorne & Lynas (2000) have shown that family break-ups, unemployment, and labor market opportunities are the main driving forces of pupil mobility in England. Moreover, pupil mobility has a negative impact on educational achievement (Hanushek, Kain & Rivkin 2004). Hence, estimates of school quality may be severely biased if identification relies on this selected subsample of pupils who move because of events that affect their educational achievement, i.e. if there is dynamic sorting (Rothstein 2009, Rothstein 2010). If, for instance, students who experience negative family events move to worse schools, the difference between good and bad schools will typically be overestimated.

Peer effects can be identified if three difficult sets of issues are solved (Manski 1993, Manski 2000): first, the identification strategy should disentangle selection bias from actual peer effects.

Second, it should avoid simultaneity biases arising from students influencing each other. Third, the identification strategy should disentangle the effect of peers' characteristics from the effect of peers' achievement.

On the schools' side, this paper offers a sound identification strategy in the spirit of the displaced workers literature (Jacobson, LaLonde & Sullivan 1993, Dustmann & Meghir 2005). In some parts of England, there are two types of primary schools: Primary schools that cater only for grade 1 and grade 2 pupils, called Infant schools, and primary schools that cater for grade 3 to grade 6 pupils, called Junior schools. In these areas, students have to move between grade 2 and grade 3. These *compulsory movers* all change school, and hence provide us with a unique opportunity to estimate school quality on a larger set of students rather than on a self-selected minority of movers. In an overwhelming majority of cases, compulsory movers do not move to the school across the street, but they move to a variety of schools with different postcodes and names.

We show in a general econometric model that when mobility is compulsory, the endogenous mobility bias due to dynamic sorting is smaller than when mobility is noncompulsory, i.e. when pupils can stay in the same school for multiple grades. Also, the estimates obtained on compulsory movers are informative about the overall sample since the characteristics of pupils in schools where mobility is compulsory are close to the characteristics of other pupils.

On the peers' side, the paper applies Hoxby's insights that variation of peers composition within a school and a grade from one year to the next are a credibly random source of identification (Hoxby 2000). Our data shows that actual variations are very close to random demographic shocks. The paper estimates peer effects using pupils' average unobservable effects and suggests that most peer effects are due to peers' characteristics.

We use compulsory mobility and random demographic shocks to estimate a full-fledged educational production function with school quality and peer effects that depend on student characteristics, long run and short run effects of educational inputs. The paper provides a full set of proofs that compulsory mobility lowers the endogenous mobility bias – which also shows the benefits of focusing on displaced workers to estimate wage equations (Dustmann & Meghir 2005) –, as well as the consistency of all the effects in a full-fledged educational production function.

In the back of our mind lies the key question for education policy: which of the many educational inputs - including social background, schools, peers and teachers - should be targeted ? Because investment in educational inputs is efficient when it targets the most cost-efficient input, the relative standard deviations of pupil effects, pupil effects, and school effects provide us with upper-bounds on the relative prices of these inputs. For instance, since previous seminal research tends to show that desegregation/tracking has small effects on the actual variance of educational achievement, and since its relative price is high, investment in desegregation/tracking is not likely to be the most efficient policy investment.

Peer effects tend to be rather unstable across countries and contexts. While peers' race and gender seems to have a significant and large effect in Texas in a linear-in-means model and using random demographic variations in grade composition (Hoxby 2000), other estimates of peer effects suggest much smaller effects. Participants of the Boston METCO busing program (Angrist & Lang 2002) seem to benefit from better peers, but the effect is small and short-lived. In a very different context, Israeli pupils do not seem to be significantly affected by immigrant peers (Gould, Lavy & Paserman 2004).

Not all students are equally affected by their peers (Epple & Romano 2010). Low-achieving Israeli pupils seem to be significantly more affected by their peers than other, higher-achieving pupils (Gould et al. 2004). In North Carolina, results of the estimation of peer effects using conditionally random assignment of pupils to schools show very significant heterogeneity in peer effects, with low-achieving pupils benefiting much less from high-achieving peers than other pupils (Hoxby & Weingarth 2007).

Our econometric model will allow for the estimation of heterogeneous peer effects, i.e. peer effects that depend on students' gender, ethnicity, and free-meal status. Also, our educational production function accounts for school quality that depends on student characteristics, and hence questions the idea that there are good and bad schools regardless of student characteristics, or good and bad peers regardless of student characteristics. The ranking of schools depends on student characteristics. A school that is more effective than the median school for male pupils may be below the median for female pupils.

This is not the first paper to look at the effectiveness of educational inputs for different subgroups. Indeed, some teachers are more effective with some students than others. For instance, same-race and same-gender teachers are beneficial (Dee 2005). Also, there is considerable heterogeneity in the way parents value school inputs (Bayer, Ferreira & McMillan 2007). This paper estimates the heterogeneity of observable and unobservable school quality and peer effects, in a fashion comparable to the match effects of Dustmann & Meghir (2005).

We implement the empirical framework using an administrative database of English pupils. We use three cohorts of English pupils in state primary schools.[1] The data set follows all pupils from primary to secondary education. England also has a national curriculum with an associated national testing schedule. The outcomes we consider are national test scores (Key Stage 1, taken in grade 2 at age 6/7, and Key Stage 2, taken at the end of primary school in grade 6 at age 10/11). The grades achieved at the end of these Key Stages are particularly important instruments for both parents and the English education authorities. In particular, government uses them to set targets and parents can freely read about them in performance league tables, published on the web or in the popular press.

Estimation results show the standard deviation of pupil effects to be about four times the standard deviation of school quality. In turn, the standard deviation of school quality is more than three times larger than the standard deviation of peer effects. Moreover, a given school and a given set of peers do not raise student achievement equally across different groups of pupils. For instance, female pupils are more affected by underprivileged peers than male students. Also, minorities are more affected by minority peers than whites. Focusing on compulsory movers is essential to obtain these results. Finally, school quality and peer effects seem to have short-lived effects, but the background of the pupil has an effect both on the level of educational achievement and on the progress of the pupil. This suggests that gaps in educational achievement may grow without corresponding gaps in the quality of schools or peers. This also suggests that, assuming that the goal is to raise all grades by one standard deviation, the choice of the cost-effective policy must compare the cost per student of raising each pupil effect by one standard deviation with four times the cost per student of raising each school effect by one standard deviation.

The outline of the rest of the paper is as follows. Section 2 presents the identification strategy which tackles dynamic sorting using compulsory mobility. Section 3 presents the various specifications we consider, as well as the estimation strategy we implement for each specification. In each case, discussion of the specification is related to relevant papers in the associated literature. Section

[1] The data are census data on pupils in state schools - this comprises the vast majority of English school children. Only around 7 percent receive their education outside of the state sector in private schools (and the percentage is even lower in the primary schooling stage we study).

4 introduces the reader to the specific English policy context, describes the data set, and discusses extensively our identification strategies. Section 5 analyzes the regression results, discusses the robustness of the estimation, examines the mobility patterns, and gives some public policy implications of the results. Section 6 concludes.

2 Identification Strategy

2.1 Dynamic sorting and the Endogenous mobility bias

Estimating school quality using nonexperimental data typically relies on student mobility across schools. In this section, we introduce the most basic specification[2] – a full-fledged educational production function is presented in section 3. Students have the option of staying in the same school for multiple grades. A student's test score is determined by the pupil effect, the school effect, and a set of unobservable factors.

$$y_{i,t} = \theta_i + \psi_{J(i,t)} + x_{i,t}\beta + \varepsilon_{i,t} \tag{1}$$

where i indexes students, $J(i,t)$ is the school of student i in year t, $y_{i,t}$ is the test score of student i in year t, θ_i is the individual effect, $\psi_{J(i,t)}$ is the school effect, $x_{i,t}\beta$ is a set of controls including grade dummies which will be fully specified in section 3, and $\varepsilon_{i,t}$ is the residual. There are two years of observation $t = 1, 2$, and $j = 1, 2, \ldots, J$ schools.

We start by considering the typical identification strategy of the literature. Model 1 is typically identified using pupil mobility across schools, and identification relies on exogenous mobility $E(\varepsilon_{i,t}|i, j, t, x_{i.t}) = 0$ for all pupils i and all schools j. Mobility between the two years of observation will be driven by non time-varying factors, such as the student's gender, ethnicity, race, and also by time-varying unobservable factors which may be part of the residual $\varepsilon_{i,t}$, such as parental unemployment, divorce, noise in test scores, and other factors. Dobson et al. (2000) is one of the few studies that examine the nature and causes of pupil mobility in England on six LEAs. Their analysis identifies two key factors: migration (for several reasons including jobs, for reasons of career

[2]The more advanced models presented in section 11 account for peer effects, lagged effects, and control for cohort and grade effects.

progression, and among refugee families) and family break-ups. They also stress that mobile pupils are predominantly from low income families. As in many other studies that use administrative educational data, we do not observe these events which affect educational achievement and may or may not affect mobility across schools.

Hence, the residual is decomposed into a time-varying unobservable $u_{i,t}$ that affect pupil mobility across schools, and time-varying unobservables $\eta_{i,t}$ that do not affect mobility.

$$\varepsilon_{i,t} = -\delta u_{i,t} + \eta_{i,t}$$

The time-varying unobservable factor $u_{i,t} \in \{0,1\}$ captures favorable ($\delta < 0$) or adverse effects ($\delta > 0$) [3] such as parental unemployment, divorce, mobility-inducing shocks in test scores [4], and $E(\eta_{i,t}|u_{i,t}, i, j) = 0$.

There are two periods $t = 1, 2$. In this section, pupils who move in-between period 1 and period 2 pay a mobility cost $c > 0$ which represents a number of factors such as the disruptive effect of changing schools and the cost of search. Pupils' mobility decision is represented by a latent utility model.

$$U_{ij} = a_{ij} + d_j \cdot u_{i,2} + \nu_{i,j}$$

Utility depends on the match between the student and the school, and on the time-varying factor u_{i1} that affects test scores. Pupil i stays in her school j in period $t = 2$ if the cost of mobility is greater than the utility gain $U_{ik} - U_{ij}$ when moving to any other school k. And pupil i moves to school k in period $t = 2$ if U_{ik} is greater than $U_{ij} - c$ and U_{ik} is maximum. There is dynamic sorting whenever $d_j \neq 0$, i.e. whenever mobility depends on the time-varying unobservable factor $u_{i,2}$. This unobservable factor determines both the decision to move and the direction of mobility.

When pupil mobility is not driven by time-varying events that affect grades, $d_j = 0$ for all j, it is possible to identify and estimate the individual effect θ_i of each student and the school quality ψ_j of each school, up to a constant, consistently and without bias.

Proposition 1. *When there is no dynamic sorting, i.e. when $d_j = 0$ for all schools j, the OLS*

[3] Hanushek et al. (2004) mentions that mobility can be driven by positive or negative events.

[4] There is substantial evidence that shocks on test scores are significant(Kane & Staiger 2002, Chay, McEwan & Urquiola 2005).

estimates of θ and ψ are consistent regardless of the value of δ. When there is dynamic sorting, i.e. when $d_j \neq 0$ for at least one school and $\delta \neq 0$, the OLS estimates of θ and ψ are biased and inconsistent.[5]

With dynamic sorting, $E(\varepsilon_{it}|i, j, t, x_{i,t}) \neq 0$ and the OLS estimates are biased and inconsistent.

2.2 Identification using Compulsory Mobility

In the previous subsection, students could stay in the same school for multiple grades, as in many educational settings. In contrast, in this section, students have to change school between grades. We show that, when students have to move between grades, the endogenous mobility bias due to dynamic sorting is substantially alleviated.

English education offers us such a source of mobility, in the spirit of the displaced workers literature (Jacobson et al. 1993, Neal 1995). In some schools, mobility is compulsory after grade 2. Hence, all students must change school, whereas in other schools, where mobility is not compulsory, students only change school if they are affected by some important event that motivates the move.[6] Many more important details on the English educational context are provided in subsection 4.3, on page 20. We show here that considering only compulsory movers reduces the endogenous mobility bias and the bias on the variance of the effects.

The intuition is simple: when all students have to move to a different school, the sample of movers, which identifies school quality, is more representative of the overall population, i.e. its time-varying unobservables are more comparable to the time-varying unobservables of the overall sample.

Proposition 2. *Consider the case where $a_{ij} = a_j$ for all i. The endogenous mobility bias on each school effect estimate is smaller with compulsory mobility, i.e. when students have to move and always incur the mobility cost c, than without compulsory mobility, i.e. when students can stay in the same school and avoid incurring the mobility cost c.*

[5] We can also express the bias on the individual effects and the school effects by writing the educational production function in matrix form $Y = D\theta + F\psi + \delta U + \varepsilon$, with Y the vector of test scores, D the design matrix for pupil effects, θ the vector of pupil effects, F the design matrix for school effects, ψ the vector of school effects, U the vector of unemployment shocks and ε the residual. The estimates are clearly biased: $\hat{\theta} = \theta + \delta(D'M_F D)^{-1}D'M_F U$, and $\hat{\psi} = \psi + \delta(F'M_D F)^{-1}F'M_D U$. The bias is zero only if the effect of time-varying unobservables is zero $\delta = 0$, or if time-varying unobservables do not determine mobility, $D'M_F U = 0$ or $F'M_D U = 0$.

[6] Section 4 describes compulsory mobility in more detail.

Proof. We prove the proposition with any number of schools J. See Appendix. □

To illustrate that, we restrict attention to three schools $j = 1, 1', 2'$, and two periods $t = 1, 2$.[7] Students are indexed by i. Initially, in period $t = 1$, all students attend school $j = 1$. Students of $j = 1$ can either stay in school j to move in the second period. Either to school $j = 1'$ or to school $j = 2'$, which are both open only for $t = 2$ students.

The probability of a time-varying shock is $P(u_{i,t} = 1) = p$. The residual is orthogonal to the other sources of mobility hence $E(\varepsilon_{i,t}|i, j) = 0$ for all i, j.

Proposition 3. *In the model with three schools, the estimates $\hat{\psi}_{1'}$ and $\hat{\psi}_{2'}$ of $\psi_{1'}$ and $\psi_{2'}$ are biased. The difference between $\hat{\psi}_{1'}$ and $\hat{\psi}_{2'}$ is overestimated. Specifically,*

$$\begin{aligned}
\hat{\psi}_{2'} &= \psi_{2'} - \delta(P(u_{i,2}|j = 2') - p) \\
\hat{\psi}_{1'} &= \psi_{1'} - \delta(P(u_{i,2}|j = 1') - p)
\end{aligned}$$

and $\hat{\psi}_{1'} - \hat{\psi}_{2'} = \psi_{1'} - \psi_{2'} + \delta(P(u_{i,2}|j = 2') - P(u_{i,2}|j = 1'))$, where $P(u_{i,2}|j = 2')$ is the probability that pupils who have moved to school $j = 2'$ have experienced the shock $u_{i,2}$, and similarly for $P(u_{i,2}|j = 1')$.

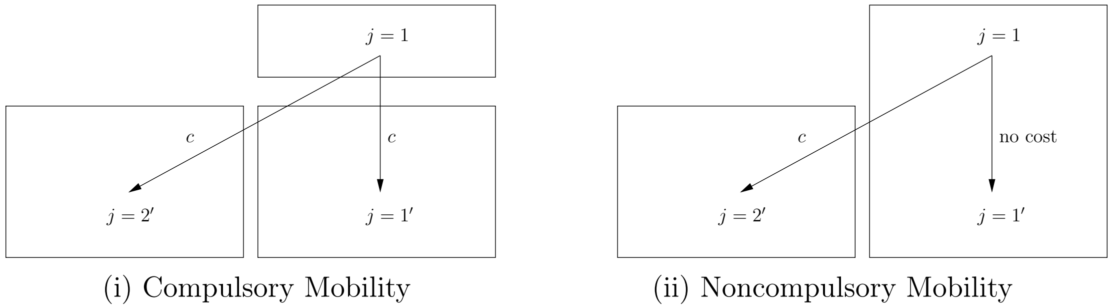
Proof. See appendix. □

The proposition formalizes the statement of the introduction that if, for instance, students who experience negative family events move to worse schools, the difference between good and bad schools will be overestimated.

Compulsory vs Noncompulsory mobility: In the noncompulsory mobility case, schools $j = 1$ and $j = 1'$ are in the same building, share the same resources, and families incur no cost in going from $j = 1$ to $j = 1'$, whereas they incur a cost c if the child goes from $j = 1$ to $j = 2'$.[8] In the compulsory mobility case, schools $j = 1$ and $j = 1'$ are separate schools, with different buildings, and families and children incur a cost c regardless of the school their child attends at $t = 2$.

[7] We focus on three schools without loss of generality. Results are general: compulsory mobility reduces the endogenous mobility bias even in the case of a large number of schools in period $t = 1$ and a large number of schools in period $t = 2$.

[8] The cost includes both the direct cost of mobility and all other costs, whether they are psychological or of any other nature.



With compulsory mobility, families have to move between Key Stage 1 and Key Stage 2, and incur the mobility cost c regardless of their choice. With noncompulsory mobility, families can avoid the mobility cost c by staying in school $j = 1$. Schools $j = 1$ and $j = 1'$ are in the same building, school $j = 1$ stands for the school effect in Key Stage 1, and school $j = 1'$ stands for the school effect in Key Stage 2.

Figure 1: A Model of Compulsory Mobility vs Noncompulsory Mobility

Families choose the school of period $t = 2$ by comparing the utility of attending school 2' and the utility of attending school 1'. School $j = 2'$ attracts students who have experienced the endogenous shock $u_{i,t}$, hence $U_{i2'} = a_{2'} + d_{2'}u_{i,1} + \eta_{i,2'}$ and $U_{i1'} = a_{1'} + d_{1'}u_{i,1} + \eta_{i,1'}$, where $\eta_{i,1'}$ and $\eta_{i,2'}$ have zero mean. Further details are provided in the appendix. The shock $u_{i,1}$ drives pupils to school $2'$, $d_{2'} > d_{1'}$.

In the noncompulsory mobility case, children move to school $j = 2'$ if $U_{i2'} - c > U_{i1'}$. In the compulsory mobility case, children move to school $j = 2'$ if $U_{i2'} - c > U_{i1'} - c$. In both cases, school $j = 1'$ is the school that attracts the larger share of students, i.e $P(j = 1') > \frac{1}{2}$. The consequences are summarized in the following proposition:

Proposition 4. *Compulsory mobility reduces the bias in $\hat{\psi}_{2'}$ and the bias on $\hat{\psi}_{1'}$. Formally,*

$$P(u_{i,1} = 1 | j = 2'; \textit{noncompulsory mobility}) > P(u_{i,1} = 1 | j = 2'; \textit{compulsory mobility}) > p$$
$$p > P(u_{i,1} = 1 | j = 1'; \textit{compulsory mobility}) > P(u_{i,1} = 1 | j = 1'; \textit{noncompulsory mobility})$$

Proof. See Appendix □

Intuitively, this result illustrates the fact that with compulsory mobility, the sample changing school is more representative of the overall population. It says that with compulsory mobility, the time-varying shocks have affected pupils in school $j = 2'$ and in school $j = 1'$ more equally than when mobility is noncompulsory. This result is applicable to the labor economics literature on the decomposition of wages using displaced workers (Dustmann & Meghir 2005).

3 The Full Educational Production Function

We use compulsory mobility to estimate a full-fledged educational production function which features heterogeneous school quality and peer effects, long-run and short-run effects of educational inputs, and the effect of the student's background. Our results show substantial heterogeneity in school quality and in peer effects.

3.1 Specifications

In this section we introduce the various econometric models from which we extract estimates of school quality, pupil effects, and peer effects. As such the specifications incorporate a number of features present in the literature on educational production functions. We show in the next subsection that educational production functions are identified under the assumptions of sufficient and exogenous mobility of pupils across schools. The plan is to implement the specifications using unique features of the English educational system, which in some settings provides exogenous sources of pupil mobility across schools.

The compulsory school careers of English children are organized into four Key Stages: Key Stages 1 and 2 which take place in primary school; and Key Stages 3 and 4 in secondary school. Our focus is on primary schools where Key Stage 1 examinations are taken at age 6/7 (grade 2) and Key Stage 2 examinations at the end of primary school at age 10/11 (grade 6).[9]

Our educational production functions combine a number of important features : (i) test scores are a function of the effect of pupil background, the effect of school quality, and the effect of peers; (ii) school quality and peer effects can vary over time (iii) the whole history of inputs shapes each pupil's experience as inter alia Rivkin et al. (2005) and Todd and Wolpin (2003) point out; and (iv) schools and peers do not affect all pupils equally. We present five different specifications that

[9]Ideally, we would like to estimate the model up to key stage 4, but we would need to follow at least two cohorts from key stage 1 to key stage 4 and currently the data set follows only one cohort all the way through school careers. Future releases of the National Pupil Database will include multiple cohorts from key stage 1 to key stage 3 and above.

incorporate these features one after another, as depicted below:

$$y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + \theta_i + \psi_{J(i,t)} + \varepsilon_{i,f,t} \tag{SE}$$

$$y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + \theta_i + \varphi_{J(i,t),g(i,t),t} + \varepsilon_{i,f,t} \tag{SGYE}$$

$$y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + \theta_i + \varphi_{J(i,t),g(i,t),t} + \lambda\varphi_{J(i,t-1),g(i,t-1),t-1} + \varepsilon_{i,f,t} \tag{PSGYE}$$

$$y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + (1 + \lambda(t-1)) \cdot \theta_i + \varphi_{J(i,t),g(i,t),t} + \lambda\varphi_{J(i,t-1),g(i,t-1),t-1} + \varepsilon_{i,f,t} \tag{PIE}$$

And, finally, the last specification where school quality and peer effects depend on student characteristics:[10]

$$y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + \theta_i + \varphi_{J(i,t),g(i,t),t}(\mathbf{z}_i) + \varepsilon_{i,f,t} \tag{HSGYE}$$

In each of these specifications, there are two Key Stage periods $t = 1, 2$[11], i denotes the N pupils with $i = 1, \ldots, N$; j denotes the J schools with $j = 1, \ldots, J$. $y_{i,f,t}$ is the test score of pupil i at time t in examination topic f. $J(i,t)$ denotes the school pupil i attended at time t and $g(i,t)$ denotes the grade in which pupil i attends in year t. $x_{i,f,t}$ is a vector of dummies controlling for the topic, the cohort and the grade. $\mathbf{z}_i$ is a vector of observable non-time varying pupil characteristics: gender, ethnicity, month of birth, free meal status, and special needs.

The first specification (SE) is the same as that analyzed in the worker-firm study of Abowd et al. (1999). In (SE) the test score is decomposed into a pupil effect θ_i, a school effect $\psi_{J(i,t)}$, $\mathbf{x_{i,f,t}}\beta$ is the effect of the K time-varying covariates, and $\varepsilon_{i,f,t}$ the residual[12]. The covariates are controls for cohorts, years and exam subjects. The main advantage of specification (SE) is its simplicity. However, it does not take into account the fact that school-specific inputs can vary over time. For instance, teachers may be different from one year to the other and the student body of the school changes. To take into account this feature, specification (SGYE) therefore generalizes this by positing that achievement is the sum of a student effect and a school-grade-year effect that incorporates all time-varying school inputs and peer effects.

Specification (PSGYE) extends specifications (SE) and (SGYE) in that it incorporates the effect

[10] In the heterogeneous specification, we did not include the effect of past inputs on current achievement. Indeed, our estimate of λ appears to be very close to zero, or at least very imprecisely estimated.

[11] Note that, as we are pooling test scores for Key Stage 1 and Key Stage 2, we checked that using the percentile instead of the standardized test score does not affect the result.

[12] Of course, in Abowd et al. (1999) the effects were firms (our schools) and workers (our pupils).

of past inputs on current achievement. For example, the quality of the teachers or the environment at the initial stage can affect future outcomes. This specification addresses the issues of the estimation of dynamic panel data models as raised in Nickell (1981) without the need for instruments. Thus, the test score is the sum of a student effect, the current school-grade-year effect and the past school-grade-year effect discounted by λ. Notice that, in the same fashion as in value-added models, this specification assumes that the current and past effect of schools to be proportional (Todd & Wolpin 2003).

The even more general specification (PIE) allows the child's progress $\Delta y_{i,f,t}$ to depend on changes in school quality *and* on his/her ability. The background of the pupil can affect both the level of achievement and the amount of progress in each year.

Overall, both specifications (PSGYE) and (PIE) give estimates of the long-run effects of educational inputs. If λ is estimated to be nonzero, schools have an effect not only on current achievement but also on achievement in the next period (i.e. grade 6 at the end of Key Stage 2 in our study context).

In the last model (HSGYE), we ask whether schools are equally effective for all pupils regardless of their characteristics. It also questions whether peer effects are independent of student characteristics. Moreover, a model with homogeneous effects of educational inputs such as (SE), misses an essential efficiency implication: in such a model the reallocation of students across schools has no effect on average test scores. In contrast, if, as we will show later on, some schools are more effective for female pupils than for male pupils, increasing the fraction of female pupils in these schools would increase average achievement. Also, the literature suggests that tracking affects average performance (Hanushek & Wössmann 2006).

The heterogeneous school-grade-year effect model (HSGYE) incorporates school-grade-year effects that depend on observable characteristics. In a second step, we will disentangle peer effects and school quality for each gender/ethnicity/free meal status subgroup of pupils.

3.2 Identification Hypotheses: Sufficient and Exogenous Pupil Mobility

The identification of specifications (SE), (SGYE), (PSGYE), (PIE), (HSGYE) requires both *sufficient mobility* and *exogenous mobility*, both of which are defined in this sub-section, in addition to the traditional exogeneity of the other covariates. In addition to this, we will assume that at least one of our five specifications is correct. For instance, we rule out interactions between unobservable student characteristics and either school quality or peer effects. [13]

Mobility is defined as sufficient when the mobility graph for pupils and schools is connected (Abowd, Kramarz & Margolis 1999, Abowd, Creecy & Kramarz 2002). Two schools are connected if and only if at least one pupil has attended both schools in different years. The set of all these connections is the mobility graph, and we say that it is connected when it has only one connex component. This is illustrated in figure 7 of the Appendix.

Moreover, exogeneity assumptions specific to our model with pupil and school effects are required (Abowd et al. 1999, Abowd et al. 2002). A threat to identification arises if, for instance, unmeasured unemployment shocks affect mobility and have an effect (e.g. through reduced income) on outcomes (Hanushek & Rivkin 2003). For this example, assume those families who experience an unemployment shock between the two periods are more likely (i) to make their child move to a bad school and (ii) to experience lower test scores due to their parents' joblessness, then the difference between bad and good schools might be underestimated.

To better understand why sufficient and exogenous mobility are jointly needed, consider the set of pupils who attend school j in period 1 and school j' in period 2. This mobility is clearly necessary since these movements allow the identification of the relative effectiveness of school j with respect to school j'. Ignoring the effect of covariates for the sake of clarity, in the (SE) model, this gives:

$$\begin{aligned} E[\Delta y_{i,f}|i, J(i,2)=j', J(i,1)=j] &= \psi_{j'} - \psi_j \\ &+ E[\Delta \varepsilon_{i,f}|i, J(i,2)=j', J(i,1)=j] \end{aligned} \tag{2}$$

with $\Delta y_{i,f} = y_{i,f,2} - y_{i,f,1}$ and $\Delta \varepsilon_{i,f} = \varepsilon_{i,f,2} - \varepsilon_{i,f,1}$. With exogenous mobility, $E[\Delta \varepsilon_{i,f}|i, J(i,2) = j', J(i,1) = j] = 0$, the last term of the sum drops. This is potentially where unemployment shocks, divorce, etc. could make trouble. Exogenous mobility rules this possibility out. Now, given this assumption, if ψ_j is identified then ψ_j' is identified.

It is then easy to see that when the mobility graph has only one connex component, choosing one arbitrary school $\jmath$ and one arbitrary pupil $\imath$ and setting their effects $\psi_\jmath$ and $\theta_\imath$ to zero identifies

[13] For interactions between unobservables, see Woodcock (2007) and Woodcock (2008). Match effects can be identified at the cost of additional assumptions on the correlation structure between match effects and other fixed effects.

all school and pupil effects. A formal definition of the mobility graph and the identification of the model are detailed in Abowd et al. (2002).

So far our intuition has been built on the simplest specification (SE). Introducing school-grade-year effects in specifications (SGYE) and (PSGYE) not only allows more flexibility in the measurement of school effectiveness, it also introduces more pupil mobility, since pupils necessarily change year-group between the two periods. In addition, because school-grade-year effects time-vary, the exogeneity assumptions are less constraining than before. Model (SGYE) is a particular case of (PSGYE) in which the discounting factor is set to 0. Again, identification conditions are identical.

Appendix C formalizes the proof of the existence and consistency of the estimators of θ and ψ either when $J \to \infty$ for I fixed or for $I \to \infty$ for J fixed.

3.3 Identification of Peer Effects

Once we have estimated school-grade-year effects, we would also like to disentangle the effect of the social composition of the school-grade-year from the other inputs under plausible identifying conditions. Identification of such peer effects is challenging. The main issues are described in Manski (1993). First, students may be sorted partly based on unobservable characteristics – for instance, teachers and students may not be randomly matched. Second, students influence each other which means it is hard to disentangle the effect of one on the other; in other words, there is a simultaneity bias. And third, it is hard to identify the effect of peers' characteristics from the effect of peers' behavior.

We assume, as in Hoxby (2000), Gould et al. (2004), and Lavy & Schlosser (2007), that the year-to-year variations in school-grade-year composition are exogenous. For instance, in the case of ethnic peer effects, we assume that plus or minus one black caribbean student in a given year is close to an idiosyncratic variation. Section 4.5 shows that these changes are indeed very close to random demographic shocks. We therefore regress the school-grade-year effect on a school identifier and the composition of the school-grade-year.

$$\varphi_{j,g,t} = \psi_j + \hat{E}(\mathbf{z}|j,g,t)\gamma + \nu_{j,g,t} \quad (3)$$

For each school-grade-year j, g, t, $\hat{E}(\mathbf{z}|j, g, t)$ denotes the vector of average observed student characteristics,[14] for instance the fraction of boys, the fraction of blacks, the fraction of free school meal students. This strategy is likely to capture most of the bias due to non-random sorting of students between schools, essentially assuming that there is no correlation between changes in school-grade-year composition and unobservable school inputs.

Formally, year-to-year variations in school-grade-year student composition should be exogenous conditional on the school-by-grade fixed effect. Variations in school-grade-year composition should not be correlated with unobserved time-varying school characteristics.

$$E[\nu_{j,g,t=2} - \nu_{j,g,t=1} | \hat{E}(\mathbf{z}|j, g, t = 2) - \hat{E}(\mathbf{z}|j, g, t = 1)] = 0 \tag{4}$$

$\hat{E}(\mathbf{z}|j, g, t = 2) - \hat{E}(\mathbf{z}|j, g, t = 1)$ is the year-to-year variation in school-grade-year composition. $\nu_{j,g,t=2} - \nu_{j,g,t=1}$ represents unobserved time-varying shocks in school quality.

This hypothesis adresses the issue of the selection bias. The simultaneity bias is adressed through the use of a common school-grade-year effect for all students. Students "share" the same local public good, which includes peer-effects.

Also, note that the model does not allow us to separately identify the effect of peers' characteristics and the effect of peers' behavior (Manski 1993), hence γ is the reduced-form peer effect, which captures both the effect of peers' characteristics and the effect of peers' behavior. Additional results, available from the authors, suggest that endogenous effects are small.[15]

3.4 Heterogeneous peer effects & school quality

Model (HSGYE) challenges the assumption that school quality and peer effects do not depend on student characteristics. Also, the linear-in-means model, where peer effects are constant, masks considerable heterogeneity (Sacerdote 2010).

In specification (HSGYE), the school-grade-year depends on student characteristics. From that

[14] Arcidiacono, Foster, Goodpaster & Kinsler (2007) innovates by estimating a model with peer effects through unobservables, i.e. $y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + \theta_i + \gamma\overline{\theta}_{-i,t} + \varepsilon_{i,f,t}$. The identification also relies on between-cohort changes in grade composition, since it implies $E(\varepsilon_{i,f,t} - \varepsilon_{i,f,t-1}|i, \mathbf{x_{i,f,t}} - \mathbf{x_{i,f,t-1}}, \overline{\theta}_{-i,t} - \overline{\theta}_{-i,t-1}) = 0$.

[15] To understand that point, consider the following model, with $y_{i,t} = dE(y|j, g, t) + \theta_i + \psi_{J(i,t)} + cE(z|j, g, t) + c'E(\theta|j, g, t) + \varepsilon_{i,t}$. d is the endogenous effect, and c, c' the exogenous effects. Then, following Manski (1993), we solve the social equilibrium and the coefficient γ is equal to $c/(1 - d)$. In additional regression results, we included $E(\theta|j, g, t)$ in the peer effects regression and found that $c' + d \simeq 0$, suggesting that both endogenous effects d and the effects of average unobserved characteristics c' are small, under the reasonable assumption that $c', d \geq 0$.

estimate, we extract school quality estimates and peer effects for each subgroup $\mathbf{z}$ of students. The initial step is,

$$y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + \theta_i + \varphi_{J(i,t),g(i,t),t}(\mathbf{z}) + \varepsilon_{i,f,t} \tag{HSGYE}$$

where $\mathbf{z}$ is a vector of observable characteristics. In a second step, peer effects are estimated on the previous heterogeneous effect, and yield our measure of peer effects for each subgroup of pupils:

$$\varphi_{j,g,t}(\mathbf{z}) = \psi_j(\mathbf{z}) + \hat{E}(\mathbf{z}|j,g,t)\gamma(\mathbf{z}) + \nu_{j,g,t}(\mathbf{z}) \tag{5}$$

So that $\gamma(\mathbf{z})$ is the reduced-form peer effect for students with characteristic $\mathbf{z}$. For instance, if a school is more effective with free meal students than with other students, then $\psi_j(\text{Free Meal}) > \psi_j(\text{Other students})$. As in the previous section, peer effects are estimated assuming random demographic shocks on school composition, conditional on pupils' observable characteristics.

4 Data Set and Identification Strategy: Compulsory Pupil Mobility in England

We use two features of English schools to identify our educational production functions. First, some schools cater for grade 1 and 2 pupils for but not for upper primary school grades; pupils have to change school. This compulsory mobility is a source of exogenous moves. Second, a significant number of schools close. As in the displaced workers literature in both cases – compulsory mobility and school closures – there is no stayer. This provides us with sufficient and reasonably exogenous mobility.

4.1 The English Educational Context

The English educational system currently combines market mechanisms (many of which were introduced in the Education Act of 1988) in different types of schools with a centralized assessment operating through a National Curriculum (Machin & Vignoles 2005). Therefore it has the advantage of providing us with fairly different management and funding structures and, at the same time, national exam results for all students.

The assessment system features a National Curriculum which sets out a sequence of Key Stages through the years of compulsory schooling: in primary school Key Stage 1 (from ages 5 to 7) and Key Stage 2 (from ages 7 to 11); and in secondary school Key Stage 3 (from ages 11 to 14) and Key Stage 4 (from ages 14 to 16). At the end of each Key Stage, pupils are assessed in the core disciplines: Mathematics, English and Science (not for Key Stage 1). These tests are nationally set and anonymously marked by external graders.

Two primary school systems coexist in England. In some areas, schools cater only for Key Stage 1 students (Infant schools) or for Key Stage 2 students (Junior schools). In other areas, schools cater for both Key Stage 1 and Key Stage 2 students (Infant and Junior schools). This means that in some areas, students have to change school after age 7. This will constitute our first source of credibly idiosyncratic pupil mobility.

Another relevant feature of English education is that a number of schools close (Gibbons 2007). Schools close mainly to restructure due to population change in the area. We identified 1,703 schools that experience closure at some point in the sample. This will constitute the second source of credibly idiosyncratic pupil mobility.

The English primary schooling system is also characterized by a variety of different management structures and funding sources (Appendix, Table 11). Community schools and voluntary controlled schools, which cater for more than half of the student body, are controlled by the Local Education Authority (LEA), of which there are 150 in England. In the case of community and voluntarily controlled schools, the LEA owns the buildings and employs the staff. On the other hand, in voluntary aided and foundation schools, teachers are employed by the school governing body and the LEA has no legal right to attend proceedings concerning the dismissal or appointment of staff.[16] Funding also varies across school types. While most state schools are funded by the government, voluntary aided schools contribute around 10% of the total capital expenditure.[17] These management and funding differences are likely to generate various kinds of incentives, and thus different educational outcomes for pupils.

[16]Code of Practice on LEA Schools Relationships, DfES 2001

[17]Source: Eurybase, 2006.

4.2 The National Pupil Database

The National Pupil Database (NPD) is a comprehensive administrative register of all English pupils in state schools. Data is collected by the Department for Children, Schools and Families and it is mandatory for all state schools to provide accurate data on pupils, who are followed from year to year through a Pupil Matching Reference. Thus panel data can be built by stacking consecutive years of the National Pupil Database.

The data set provides rich information on pupils' characteristics: gender, free school meal status, special educational needs, and the ethnicity group. Pupils who receive free meals are the 15 to 20% poorest pupils. The ethnicity variable of our sample encodes the main ethnic code: White, Black Caribbean, Other Black, Pakistani, African Black, Mixed background, Bangladeshi, Indian, Chinese, Other Background.[18] It also provides some information on school structures and types (i.e. whether they are community schools, foundation schools, voluntary aided or voluntary controlled schools, and so forth).

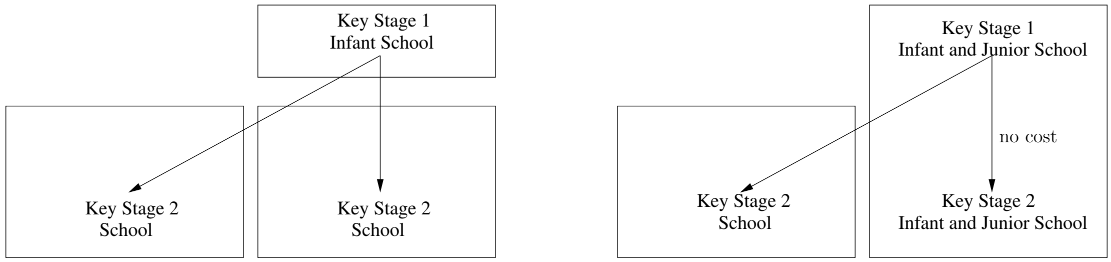
Test scores in English, Maths and Science are available - the latter Science test only for Key Stage 2. These tests are externally set and marked. We have standardized test scores to a mean of 50 and a standard deviation of 10 to make results comparable from one level to the other and from year to year.

Three cohorts of pupils are followed in grade 1 and 2 (Key Stage 1) and in grade 3, 4, 5 and 6 (Key Stage 2). Each cohort has approximately 600,000 pupils, with two observations in Key Stage 1 (Maths and English), and three observations in Key Stage 2 (Maths, English, and Science). The sample is made of 8,660,468 pooled test score observations. Free meal pupils, which is a proxy for social deprivation, comprise 17% of the observations.

4.3 Compulsory Movers

Compulsory movers are children who move between grade 2 and grade 3 because their Key Stage 1 school does not cater for Key Stage 2 children (asalready noted the key distinction is between Infant and Junior schools, some of which are in the same location, others of which are notthus facilitating a need to move school). There are at least three important conditions for compulsory mobility to be

[18]Coding of the ethnicity variable has changed in the period we consider. It was therefore necessary to recode it in a time consistent manner.



(i) Compulsory Mobility (ii) Noncompulsory Mobility

With compulsory mobility, pupils have to move between Key Stage 1 and Key Stage 2. With noncompulsory mobility, pupils can stay in the same school for multiple grades and only a selected subsample of movers identifies school quality. This is pupil mobility to schools with different locations, postcodes and different names.

Figure 2: Primary School Curriculum - Compulsory and noncompulsory moves

a credibly idiosyncratic source of identification: (i) compulsory movers should not be significantly different from noncompulsory movers; (ii) as compulsory mobility is expected by parents, we need to get evidence that Key Stage 1 only schools are not particular schools – either better or worse schools; (iii) compulsory mobility provides us with a exogenous reason to move, but it does not per se give an exogenous direction of mobility; children may still sort endogenously into schools.

Figure 3 shows that compulsory movers are present in most Local Education Authorities. There are 172 Local Education Authorities in England. The most frequent fraction of compulsory movers is a $1/3$ of all pupils in the Local Education Authority. More detailed analysis suggests that compulsory movers seem to be located in specific geographic areas within Local Education Authorities.

Table 2 shows that compulsory movers are very comparable to other students. Genders, months of birth, languages and ethnicities are very similar between compulsory movers, and other pupils. Differences between the two categories appear in the fraction of special needs students and free school meals. The fraction of free school meal students is higher among noncompulsory movers than among compulsory movers, but it very similar between compulsory movers and stayers. On the whole, there are only slight differences between the three categories of mobility.

Focusing on compulsory movers also provides us with sufficient mobility as defined in section 3.[19] The mobility graph has indeed one connex component.[20] To illustrate that, Table 3 presents some basic statistics on mobility. A sizeable fraction of compulsory movers change school and change LEA.

[19] Additional results show that pupils move to a variety of different schools, and not to the school across the street. Pupils move to schools with different postcodes, super output codes, and different names. Results available from the authors.

[20] Finding the connex components of the school mobility graph involves the algorithm `a2group`, available from the corresponding author.

Table 4 shows that compulsory movers move from and to all school types, i.e. from and to a variety of management and capital structures. 83% of compulsory movers move from community schools, 7.4% move from Voluntary Aided schools, and 7.0% move from Voluntary Controlled schools. That compares to 69.9% of the overall sample moving from community schools, 17.7% moving from Voluntary aided schools, and 9.7% moving from Voluntary Controlled schools.

The last issue we need to address is whether the direction of mobility is likely to be an identification issue. We define the most frequent school pupils go to. For each school j, the number of pupils who move from school j to school j' is computed. The most frequent school pupils from school j go to in the next period is noted $M(j)$. Among compulsory movers, 63% move to the most frequent school (Table 3). Therefore compulsory movers mainly tend to go to the 'usual' school, and the direction of their mobility is not likely to be mainly explained by individual unobserved time varying variables.

4.4 Estimation on School Closures

We also considered school closures as another source of exogenous mobility. There are about 1,703 schools closing in the time period which allows estimation on a limited subsample of children.

For each of the 1,703 schools, there are typically cohorts that went to the school without experiencing the closure and cohorts that experienced the closure. We restrict the sample to the latter cohorts of students. Anecdotal evidence suggests that most English school closures happen for administrative reasons other than demographic trends or failure; Indeed, a regression of the probability of closure on past school size and past test scores does not suggest that demographics or test scores are a major determinant of closure.[21]

We estimated the model on school closures and a couple of stylized facts emerged from the analysis. First, schools that close are much more heterogeneous than schools that do not close. Second, the basic stylized fact that pupils are much more heterogeneous than schools is still present in the data. However, given the small number of school closures compared to the number of compulsory mobility schools, our preferred method of estimation is compulsory mobility.

[21] Results available on request.

4.5 Identifying Peer Effects through Idiosyncratic Demographic Variations in Grade Composition

The effect of grade composition on school quality is estimated by looking at how year-to-year variations affect school-grade-year effects. This actually requires that year-to-year variations in grade composition are not correlated with other changes in school inputs, such as changes in teacher quality and school funding. One way of addressing this identification issue is to compare year-to-year variations to truly random variations around school average composition.

Formally, if changes in grade composition are truly exogenous, they must be some random fluctuation around the average school composition. Identification relies on the assumption that grade composition in a given year is a finite size approximation of the school's equilibrium composition.

$$\hat{E}(\mathbf{z}|j,g,t) = E(\mathbf{z}|j,g) + u_{j,g,t} \tag{6}$$

Notations as before. $\hat{E}(\mathbf{z}|j,g,t)$ is the empirical school-grade-year composition in school j, grade g and year t. This is a vector containing the percentage of each ethnicity, the percentage of boys, free meals and special needs. $E(\mathbf{z}|j,g)$ is average school composition across the three cohorts. The shocks $u_{j,g,t}$ are approximately normal with variance $\text{Var}\, E(\mathbf{z}|j,g)/n_{j,g,t}$.

The data set only contains the empirical composition of grades. Therefore school average composition is just an estimate of the true composition.

$$\hat{E}(\mathbf{z}|j,g) = E(\mathbf{z}|j,g) + v_{j,g,t} \tag{7}$$

and $\text{Var}\, E(\mathbf{z}|j,g)/n_{j,g}$ is a good first-order approximation of the variance of the error term. Therefore, finally, $\hat{E}(\mathbf{z}|j,g,t) = \hat{E}(\mathbf{z}|j,g) + u_{j,g,t} - v_{j,g,t}$.

Figure 4 compares the results of simulations to actual year-to-year variations in school-grade-year compositions. For boys, free meals and three important ethnic groups year-to-year variations are remarkably similar to random variations, as in Hoxby (2000) and Lavy & Schlosser (2007). This suggests that trends in school-grade-year composition are not likely to explain the results of peer effects regression. On the other hand, year-to-year varations in the fraction of special needs is bigger in the data set than what would be expected if it were purely random. There may be trends in the fraction of special needs students in schools, which is likely to be due to evolving support for special needs students in English elementary schools. Broadly speaking, apart from special needs students, variations in gender and ethnic compositions are similar to random variations around average school composition.

5 Results on Compulsory Movers

We estimate the full educational production function by focusing on compulsory movers only[22]. This substantially dampens the endogenous mobility bias, following section 3.

Results are informative about crucial policy questions: what is the relative efficiency of investment in different educational inputs? Policymakers in education face a trade-off between investing in school quality, investing in pupil effects, and investing in fostering tracking or desegregation plans.

5.1 Efficient Investment

Optimal investment in educational inputs is based both on the relative effects and on the relative costs of educational inputs. Models SE to PIE yield the relative effectiveness of educational inputs. Relative standard deviations give us an upper bound for the relative price of educational inputs: Above this relative price, investment in the educational input is less cost-efficient than investment in another educational input (Ouazad 2008).

So, how can an educational policymaker raise the achievement of a child by one standard deviation? The per-pupil cost of a one standard deviation increase in θ is p_θ. The standard deviation of θ is between 8.5 and 9.1 depending on the specifications. The per-pupil cost of a one standard deviation increase in ψ is p_ψ. The standard deviation of ψ is between 2.1 and 2.8 depending on the specifications. The cost of raising achievement y by one standard deviation if investing only in θ is between $p_\theta/0.85$ and $p_\theta/0.91$, and is between $p_\psi/0.21$ and $p_\psi/0.28$ if investing only in school quality ψ. Thus, if the relative cost of school quality is above $0.21/0.85 \simeq 25\%$ of the cost of rising pupils' backgrounds, investment in school quality is less efficient than investment targeted at improving pupils' backgrounds.

[22]Results for the whole population are available from the corresponding author.

5.2 Pupils, Schools, or Peers

Our first question is simple, even simplistic. Do schools matter more than pupils? Over the years, this question has been extensively discussed in the literature(Rivkin et al. 2005).

- **Explanation of the results - Stylized facts** The estimation of specifications SE to PIE yields pupil fixed effects, school-grade-year effects and school effects. School-grade-year effects are not available for the simple school effects model (specification SE). Some very clear stylized facts come out from considering the correlations reported in Table 5[23]: (i) first of all, pupils are much more heterogeneous than either school or school-grade-years effects (ii) individual effects explain a much larger share of the variance of test scores than school or school-grade-year effects.

- **The standard deviation of pupil fixed effects is 4 times the standard deviation of school effects** The first key finding is of a higher variation of pupil fixed effects as compared to school-grade-year effects and school effects can be seen from Table 5, where the standard deviation of pupil effects is between 3.3 and 4.3 times larger than the standard deviation of school effects. This, perhaps not surprisingly, suggests that pupils are more heterogeneous than schools. The same finding, of a higher variation, is is also true when comparing the standard deviation of pupil effects and the standard deviation of school-grade-year effects. Nevertheless, pupil fixed effects are less precisely estimated than school-grade-year effects or school effects. Indeed, at most five observations per child are available whereas on average 250 observations per school-grade-year are available. To address this potential issue, we look at the correlation between test scores and the school or school-grade-year effects. Individual effects are imprecisely estimated but this actually lowers the correlation between individual effects and test scores.[24]

- **Pupil fixed effects explain a larger share of the variance of achievement than school effects** The correlation between test scores and individual effects is seen to be between 0.78 and 0.83. This is between 5 and 6.7 times larger than the correlation between school effects and test scores. Also, the covariance of pupil effects and test scores is much larger than the

[23]Covariance Tables are available from the authors.
[24]Results available on demand.

covariance of school effects and test scores.

- **What explains pupil effects** Therefore baseline results strongly suggest that pupil effects are a more important determinant of test scores than school effects. Analyzing pupil effects differs from looking at raw test scores: pupil effects are free of peer effects and free of the correlation between school quality and observable characteristics. How can one interpret these pupil effects? From our perspective, it seems reasonable to think of them as picking up the whole range of educational experiences before age 7: this includes parental background, childcare and kindergarten. Considered in this way, the fact that pupil effects explain most of the variance of test scores puts recent seminal contributions in perspective (including, inter alia, Heckman & Masterov (2007), Currie (2001) and Garces, Thomas & Currie (2002)).

- **Analysis of pupil effect – observable and unobservable components** Pupil effects can be correlated with a range of individual characteristics (like ethnicity, gender, free school meal status, special needs and the child's month of birth). These correlations will capture the causal effects of these characteristics as well as determinants of educational achievement correlated with these characteristics but different from school quality and peer composition. From our analysis it is evident that pupil effects are reasonably well explained by observable characteristics. For example, the R-Squared from regressions of the fixed effects on observable characteristics is around 40% (Table 7). In these regressions, the estimated coefficients are remarkably robust to different specifications. Moreover, the results are in line with descriptive statistics on test scores. The pupil fixed effects of free school meal children are 40 to 41% of a standard deviation lower. Family disadvantage is important, with free school meal pupils being the 10 to 20% poorest children in England. Chinese pupils are the best performing pupils, with a fixed effect 16% of a standard deviation higher than white pupils. Interestingly, Indian pupils have a lower fixed effect than white pupils (6.4 to 7.8% of a standard deviation lower), whereas the test scores of indian pupils are higher than the test scores of whites. This suggests that basic descriptive statistics do not disentangle the effect of ethnicity from the effect of peers and the effect of school quality. Finally, male pupils have a higher fixed effect, by 2.5 to 2.6% of a standard deviation. This is certainly due to the fact that regressions were carried out by pooling all subjects together. Boys in primary school are better at mathematics and science, whereas girls are better at English. There are therefore around three observations for which boys are better – maths in grades 2 and 6, science in grade 6 – and 2 observations for which girls are better – English in grades 2 and 6[25].

- Table 7 confirms that pupil effects are well explained by observable characteristics, even if one is unable to offer a causal interpretation for the reported findings. This analysis allows us to disentangle pure individual effects from the social context working through peers and school quality.

- **School Quality – Management Structures** Are some schools better than others ? Are schools that are organized under particular structures better than others? This is an important question, even if school quality explains a small share of the variance of test scores, since there may be scope for improvement in the way schools are structured. As has already been noted, English schools can have a variety of different organizational structures. Some schools are able to hire and dismiss their staff, while in other schools the staff is recruited and dismissed by the Local Education Authority (See Appendix, Table 11).

- **School Quality – Overview of the Results** The regressions shown in Table 10 do indeed show significant differences by school type. There is evidence of beneficial effects of local recruitment of staff coupled with external control of the school board. Indeed, in both grade 2 and grade 6, voluntary controlled schools perform worse than community schools; like community schools, they cannot locally manage their human resources and do not own their assets. The main difference with community schools is that they are mostly religious Church of England schools.

- Table 10 shows that voluntary aided schools, who can hire and dismiss staff locally, have a higher school fixed effect in grade 6, by 1.5% of a standard deviation. The effect is not significant in grade 2 (Table 10). Other types of schools can hire locally, e.g. foundation schools. Various other types of schools have a significantly lower school effect: voluntary controlled in grade 2, community special, non-maintained special, and foundation special in grade 6.

[25] Another version of the tables, available upon request, discarded science test scores. Male fixed effects are then not significantly higher.

- School management structures are not the whole story, though. The R-Square of the regressions of school effects on school type dummies is small, being not more than 1%, a finding in line with the school effectiveness literature we cited earlier. There are therefore many other determinants of school quality that, unfortunately, are not observed in the data set we utilize.

5.3 School Quality by Subgroup

Results of the estimations of the HSGYE model are presented in figure 5 and in table 9.

Figure 5 shows the distribution of the school-grade-year effects (a, b, and c), and the distribution of the school effects (d, e, and f). Overall, there are few differences in the overall distribution of school quality and school-grade-year quality between whites and nonwhites, boys and girls, free meal students and non free meal students. The flatter distributions correspond to school quality measures estimated on a smaller number of observations, thus having more measurement error. Hence, when accounting for these differences in measurement error, the distributions are similar. As later results will show, this does not mean that schools are equally effective for all students.

The distribution of school quality for nonwhite pupils is skewed to the right. There is a number of school-grade-years that are higher for nonwhite pupils than for white pupils (Figure 5, a). This skewness is primarily due to differential peer effects. School effects do not display the same skewness (Figure 5, d). Overall, the distribution of school quality is similar for different demographic subgroups.

A look at school quality measures school by school shows that school quality depends on student characteristics. When ranking schools by their school effect measured in the last specification, the correlation between the school quality rank for boys and the school quality rank for girls is 0.51 in Key Stage 1 and 0.41 in Key Stage 2. The correlation between the school quality rank for whites and the school quality rank for nonwhites is 0.31 in Key Stage 1 and 0.25 in Key Stage 2. The correlation between the school quality rank for free meals and the school quality rank for non free meals is 0.34 in Key Stage 1 and 0.29 in Key Stage 2. This weak correlation is only partly due to measurement error to heterogeneity.[26] These results strongly suggest that school quality is not independent of student characteristics.

[26]Indeed, asymptotically, estimates of the school effects follow the central limit theorem. Simulating multiple estimates of the school effects and correlating the ranks of simulated effects gives a rank correlation above 0.9.

5.4 Peer Effects

In all specifications, the standard deviation of school-grade-year effects and test scores is 30% larger than the standard deviation of school effects (Table 5[27]). These 30% capture both time-varying inputs – changes in teacher quality, school resources, principals – and peer effects. Peer effects explain at least 20% of the time-varying variation in school-grade-year effects. Also, most of these peer effects operate through observable characteristics, little effect of peers' average pupil effect is observed.[28]

Table 8 estimates peer effects assuming peer effects are equal for all subgroups.[29] It shows the effect of the fraction of different social groups on school-grade-year effects, separately for grade 2 and grade 6. Of course, the fraction of different social groups was calculated including both compulsory and noncompulsory movers. In grade 2, in the full-fledged specification, increasing the fraction of male students by 10% makes school-grade-year effects fall by 1.14% of a standard deviation[30] (table 8, column 3). This is a reasonable but large effect compared to the average effects of the literature [31]. This effect is robust to different specifications. In grade 6, the effect of the fraction of boys is positive, i.e. increasing the fraction of boys by 10 percentage points increases test scores by 0.68% of a standard deviation (table 8, column 6). The difference between grade 2 and grade 6 gender composition effects is likely to be due to the fact that grade 2 exams are in English and Maths, whereas grade 6 exams are in English, Maths and Sciences. Boys are better than girls in both science and maths, but not better in English. This effect is robust to the inclusion of past school-grade-year effects and past individual effects in the baseline specification. Most papers in the literature find a negative effect of boys on achievement both in English and in Maths, e.g. Hoxby (2000).

The fraction of free meal children has a detrimental effect on school-grade-year effects in grade 6; increasing the fraction of free meal children by 10 percentage points lowers school-grade-year effects by roughly 0.6% of a standard deviation (table 8, column 6).

[27]Note that school composition has been computed using both compulsory and noncompulsory movers in each school-grade-year.
[28]Results available on demand.
[29]Note that the peer group is correctly calculated here, i.e all pupils where put back in the dataset, compulsory and noncompulsory movers in Key Stage 2.
[30]The standard deviation of test scores is 10.
[31]For a review of the order or magnitudes of peer effects, see Ammermueller & Pischke (2009).

In grade 6, ethnic composition has an effect on school quality. Chinese and Indian children exert a positive contextual effect. The effects are large: increasing the fraction of chinese students by 10 percentage points increases fixed effects by 4% of a standard deviation (table 8, column 6). These results are in line with the intuition that being surrounded by high performing peers is good for your test scores; chinese children are at the top of the test score distribution. Interestingly, black children and black caribbean children also exert a positive contextual effect in grade 2. The effect is large. However, and in contrast to Chinese and Indian children, black and black caribbean children are at the lower tail of the tests distribution. All other effects are not significant.

5.5 Peer Effects by Subgroup

The sign and magnitude of peer effects depends on pupils' characteristics (Table 9). Estimating peer effects by demographic subgroup strongly changes the estimate of the effect of minority peers on individual achievement. In grade 6, increasing the fraction of black carribean peers lowers the achievement of minority pupils by 8.5% of a standard deviation and increasing the fraction of black peers (non carribean) lowers the achievement of minority pupils by 11% of a standard deviation. White pupils are not significantly affected by black peers. This suggests that lower levels of segregation by ethnicity are beneficial. Although the difference is not significant, free meal peers seem to be more affected by free meal peers than other students, which suggests that lower levels of segregation by free meal status are beneficial.

Other results point toward more complex optimal patterns of peer composition. For instance, female students are more adversely affected by free meal peers than male students.

5.6 Matching of Pupils to Schools

In all specifications except HSGYE, schools are equally effective for all students, that is, there is no complementarity between pupils and schools. If the educational production function is truly specified as in specifications SE to PIE, the model does not predict any particular matching of pupil effects and school effects at equilibrium. Matching patterns are indeed determined by the complementarity between pupil effects and school effects, following Becker (1973). In the world of specifications SE to PIE, the model predicts zero correlation between pupil effects and school-grade-year fixed effects[32].

However, some of the correlations between child effects and school effects in Table 5 are negative. Does it mean that pupils with a high pupil effect are structurally matched with low school-grade-year effects? The correlation between estimated pupil effects and estimated school effects is actually downward biased. In general, the measurement errors of child and school effects are negatively correlated (Abowd & Kramarz 2004). The intuition behind this result is that (i) pupils who change school get a better estimated effect but school effects are less precisely estimated (ii) pupils who do not change school have a less well estimated effect but their associated school effect is more precisely estimated.[33]

Simulations using artificial data with zero correlation show that there is indeed a substantial bias in the estimated correlations. This has been pointed out in the context of worker-firm matched panel data sets (Abowd & Kramarz 2004).

Even in the absence of a true correlation between pupil effects and school effects, the correlation between estimated effects is negative.[34] The true correlation between school-grade-year and individual effects is therefore likely to be close to -0.1, with school-grade-year effects explaining little of the variance of test scores. Most papers find a zero or negative correlation (Abowd & Kramarz 2004, Abowd et al. 1999). But these papers do not include a match effect that could account for the complementarity between pupil and school-grade-year effects or worker and firm effect(Woodcock 2007).

[32]This of course, assumes a particular form of preferences and special market conditions. The housing market should be perfect, parents should know the educational production function as specified in equation PSGYE and the only reason for location decisions should be the level of test scores.

[33]The correlation between estimates of pupil and school effects are the sum of the correlation between measurement errors and the true covariance between the effects.

$$Cov(\hat{\theta}, \hat{\varphi}) = Cov(\hat{\theta} - \theta, \hat{\varphi} - \varphi) + Cov(\hat{\theta} - \theta, \varphi) + Cov(\hat{\varphi} - \varphi, \theta) + Cov(\theta, \varphi) \tag{8}$$

θ is the individual effect, φ is the school-grade-year effect, $\hat{\theta}$ is the estimated individual effect, $\hat{\varphi}$ is the estimated school-grade-year effect. Typically, $Cov(\hat{\theta} - \theta, \hat{\varphi} - \varphi) \leq 0$.

[34]We generate pupil effects who have a normal distribution with the same variance as the estimated pupil effects. We also generate school effects the same way. In the simulated data set, pupil effects and school effects are uncorrelated. We then generate simulated test scores using the specification with past and current school-grade-year and individual effects.

5.7 Matching of Pupils to Peers

A pupil's individual effect θ_i is strongly correlated with his peers' mean individual effect $\overline{\theta}_i$[35]. The correlation is between 0.4 and 0.6 depending on the specification. Furthermore simulation results indicate that a large share of this correlation is actual sorting and not measurement error.[36] θs are not randomly sorted across schools, but there is strong assortative matching between pupils and schools. Good students are matched together.

Also there is both sorting in observable and unobservable dimensions (Table 5). The correlation between the observable part of pupil effects and peers' average effect is 0.2 and between the unobservable part of pupil effects and peers' average effect is 0.4 in the full model. Interestingly, there is sorting on unobservable dimensions even though unobservable dimensions of peers' background do not affect individual achievement.

Results on peer effects (Table 9) suggest that the matching pupils to peers on unobservable dimensions is suboptimal. Indeed, in most tables, underprivileged pupils are more affected by their peers than other pupils, and assortative matching lowers actual levels of educational achievement (Epple & Romano 2010).

5.8 Long-Run Effects of Background, Peers, and Schools

In specifications PSGYE and PIE, grade 1 and grade 2 school quality and peer effects affect educational achievement in grade 6. The discounting factor λ measures the long term effect of school quality and of pupil background. Two broad conclusions emerge: (i) the background of the pupil strongly determines both the level and the progress (ii) schools and peers have a substantially lower effect in the long run than in the short run.

The decay rate λ is imprecisely estimated. For the 1998-2002 cohort and the 2000-2004 cohort, the optimal discounting factor is zero. The school-grade-year effect and individual effect specification (equation PSGYE) is not rejected[37]. For the cohort in-between, the optimal discounting factor is 0.1. The confidence interval around λ includes 0.9[38]. This is evidence that school quality and peer effects

[35] In labor economics, this correlation has been computed by de Melo (2009).

[36] See simulation results, Table 13 of the Appendix.

[37] Due to the large number of computations, we decided to estimate λ at a 0.1 precision.

[38] Under the null hypothesis that λ is equal to the optimal lambda, e.g. $\lambda^* = 0$, the statistic $2 \cdot ln(L(\lambda)/L(\lambda^*))$ converges to a χ^2 statistic (Hoel 1962).

may have little long run effects. The literature has found that the benefits of increased spending per student and reductions of class size are short-lived(Prais 1996, Krueger 1999, Hanushek 2003). Similarly, Angrist & Lang (2002) suggest that the beneficial effect of a better peer group are not permanent.[39] Our finding generalizes these estimates by showing that the average long-run effect of school-related inputs and of peer effects is small.

Confidence intervals for specification PIE are much smaller. Since specification PIE estimates the long-run effect of pupil effects, as well as school quality and peer effects, this shows that pupil effects have a much stronger role in the dynamics of achievement. This result is fundamental, because it shows that achievement gaps may grow without an increasing gap in the quality of educational inputs.

6 Conclusion

In this paper we consider a detailed econometric model which evaluates the importance of pupil and school factors in determining children's educational achievement. We match pupils to schools, using very rich administrative data on English primary school children. We decompose children' test scores into the effect of the pupil's background, the effect of the peers and the effect of the schools.

Educational production functions are typically hard to identify because estimates are confounded by unobservables and by endogenous mobility bias. We use a credibly idiosyncratic source of mobility: In some areas, because of primary school availability at Infant and Junior level, students have to move after grade 2. All affected pupils are movers rather than a self-selected minority whose moves may be driven by, for example, family events. This provides an exogenous source of mobility, but does not offer an exogenous direction of mobility. An important contribution of this paper is to prove that compulsory mobility alone substantially reduces the endogenous mobility bias.

The main finding from the estimation of the detailed econometric model is that pupils' backgrounds is four times more variable than school quality, and that the standard deviation of peer effects are around 20% the standard deviation of school quality. Either inherited ability, early educational experiences (acquired before the age of seven) and family background play a very important

[39] In this paper we focus on test scores, whereas there may be long-run effects on other outcomes, see for instance Black, Devereux & Salvanes (2010).

part in the educational process. School time-invariant inputs are the second most important input, but prove to be far less important than pupil effects. Peer effects may be the least important input, most effects being small.

The conclusions of this paper should be of substantial interest to policy makers seeking to spend public funding into the right inputs either to increase efficiency or to narrow educational inequalities. First of all, they do have clear implications for education policy and design. The relative standard deviations of the effects give upper bounds on the relative prices of educational investment to be efficient. Because tracking/desegregation is costly relative to investment in other inputs, investment in desegregation or tracking is unlikely to be efficient compared to investments in school quality or in remedial education, or in policies that target families. Also, our model shows that school quality is a quantity that depends on the characteristics of the students. Schools are not ranked the same way when they are assessed on the subset of boys, girls, whites, minorities, or free meal students.

Finally, and importantly, this paper applies and refines methods from the literature on matched worker-firm data to pupils in schools. Our econometric results on bias-reduction with compulsory mobility are straightforwardly applicable to other literatures using dynamic matched data (e.g. the decomposition of wages when workers are mobile, or occupational mobility).

References

Abowd, J., Creecy, R. & Kramarz, F. (2002), Computing person and firm effects using linked longitudinal employer-employee dataset, Technical report, US Census Bureau.

Abowd, J. M. & Kramarz, F. (2004), Are good workers employed by good firms? A simple test of positive assortative matching models, Econometric Society 2004 North American Winter Meetings 385, Econometric Society.

Abowd, J. M., Kramarz, F. & Margolis, D. N. (1999), 'High wage workers and high wage firms', *Econometrica* **67**(2), 251–334.

Ammermueller, A. & Pischke, J.-S. (2009), 'Peer effects in european primary schools: Evidence from the progress in international reading literacy study', *Journal of Labor Economics* **27**(3), 315–348.

Angrist, J. D. & Lang, K. (2002), How important are classroom peer effects? Evidence from Boston's METCO program, NBER Working Papers 9263, National Bureau of Economic Research, Inc.

Arcidiacono, P., Foster, G., Goodpaster, N. & Kinsler, J. (2007), Estimating spillovers using panel data, with an application to the classroom. unpublished manuscript.

Bayer, P., Ferreira, F. & McMillan, R. (2007), 'A unified framework for measuring preferences for schools and neighborhoods', *Journal of Political Economy* **115**(4), 588–638.

Becker, G. S. (1973), 'A theory of marriage: Part I', *Journal of Political Economy* **81**(4), 813–46.

Betts, J. R. (1995), 'Does school quality matter? Evidence from the national longitudinal survey of youth', *The Review of Economics and Statistics* **77**(2), 231–50.

Black, S. E., Devereux, P. J. & Salvanes, K. G. (2010), 'Under pressure? the effect of peers on outcomes of young adults', *NBER Working Paper Series* .

Chay, K. Y., McEwan, P. J. & Urquiola, M. (2005), 'The central role of noise in evaluating interventions that use test scores to rank schools', *American Economic Review* **95**(4), 1237–1258.

Currie, J. (2001), 'Early childhood education programs', *Journal of Economic Perspectives* **15**(2), 213–238.

de Melo, R. L. (2009), Sorting in the labor market: Theory and measurement. unpublished manuscript.

Dee, T. S. (2005), 'A teacher like me: Does race, ethnicity, or gender matter?', *American Economic Review* **95**(2), 158–165.

Dobson, J. M., Henthorne, K. & Lynas, Z. (2000), 'Pupil mobility in schools: Final report', *Migration Research Unit at University College London* .

Dongarra, J., Duff, I., Sorensen, D. & van der Vorst, H. (1991), *Solving Linear Systems on Vector and Shared Memory Computers*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, United States.

Dustmann, C. & Meghir, C. (2005), 'Wages, experience and seniority', *Review of Economic Studies* **72**(1), 77–108.

Epple, D. & Romano, R. E. (2010), *Handbook of Social Economics*, Elsevier, chapter Peer Effects in Education: A Survey of the Theory and Evidence.

Garces, E., Thomas, D. & Currie, J. (2002), 'Longer-term effects of Head Start', *American Economic Review* **92**(4), 999–1012.

Gibbons, S. (2007), Mobility and school disruption. Centre for the Economics of Education Discussion Papers.

Gibbons, S., Machin, S. & Silva, O. (2008), 'Choice, competition and pupil achievement', *Journal of the European Economic Association* **6**, 912–947.

Gould, E. D., Lavy, V. & Paserman, M. D. (2004), Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence, NBER Working Papers 10844, National Bureau of Economic Research, Inc.

Graham, B. S. (2008), 'Identifying social interactions through conditional variance restrictions', *Econometrica* **76**(3), 643–660.

Hanushek, E. A. (1986), 'The economics of schooling: Production and efficiency in public schools', *Journal of Economic Literature* **XXIV**, 1141–1177.

Hanushek, E. A. (2003), 'The failure of input-based schooling policies', *Economic Journal* **113**(485), F64–F98.

Hanushek, E. A., Kain, J. F. & Rivkin, S. G. (2004), 'Disruption versus tiebout improvement: The costs and benefits of switching schools', *Journal of Public Economics* **88**.

Hanushek, E. A. & Rivkin, S. G. (2003), Does public school competition affect teacher quality?, *in* C. M. Hoxby, ed., 'The Economics of School Choice', National Bureau of Economic Research, chapter 1.

Hanushek, E. A., Rivkin, S. G. & Taylor, L. L. (1996), 'Aggregation and the estimated effects of school resources', *The Review of Economics and Statistics* **78**(4), 611–27.

Hanushek, E. A. & Wössmann, L. (2006), 'Does educational tracking affect performance and inequality? differences-in-differences across countries', *The Economic Journal* **116**(510), C63–C76.

Heckman, J. & Masterov, D. V. (2007), 'The productivity argument for investing in young children', *Review of Agricultural Economics* **29**(3), 446–493.

Hoel, P. (1962), *Introduction to Mathematical Statistics*, 3rd edn, Wiley.

Hoxby, C. (2000), Peer effects in the classroom: Learning from gender and race variation, NBER Working Papers 7867, National Bureau of Economic Research, Inc.

Hoxby, C. M. & Weingarth, G. (2007), Taking race out of the equation.

Jacobson, L. S., LaLonde, R. J. & Sullivan, D. G. (1993), 'Earnings losses of displaced workers', *American Economic Review* **83**(4), 685–709.

Kane, T. & Staiger, D. (2002), 'The promise and pitfalls of using imprecise school accountability measures', *The Journal of Economic Perspectives* .

Krueger, A. B. (1999), 'Experimental estimates of education production functions', *The Quarterly Journal of Economics* **114**(2), 497–532.

Lavy, V. & Schlosser, A. (2007), Mechanisms and impacts of gender peer effects at school, NBER Working Papers 13292, National Bureau of Economic Research, Inc.

Levacic, R. & Vignoles, A. (2002), 'Researching the links between school resources and student outcomes in the U.K.: A review of issues and evidence', *Education Economics* **10**(3), 313–331.

Machin, S. & Vignoles, A. (2005), *What's the Good of Education? The Economics of Education in the United Kingdom*, Princeton University Press, Princeton, New Jersey, United States.

Manski, C. F. (1993), 'Identification of endogenous social effects: The reflection problem', *Review of Economic Studies* **60**(3), 531–42.

Manski, C. F. (2000), 'Economic analysis of social interactions', *Journal of Economic Perspectives* **14**(3), 115–136.

Matthews, K. (1991), *Linear Algebra Notes*, University of Queensland.

Neal, D. (1995), 'Industry-specific human capital: Evidence from displaced workers', *Journal of Labor Economics* **13**(4), 653–77.

Newey, W. (1991), 'Uniform convergence in probability and stochastic equicontinuity', *Econometrica* **59**(4), 1161–1167.

Nickell, S. J. (1981), 'Biases in dynamic models with fixed effects', *Econometrica* **49**(6), 1417–26.

Ouazad, A. (2008), Inequalities and discriminations: Essays in the economics of education and in urban economics. Ph.D. thesis.

Prais, S. (1996), 'Class-size and learning: the Tennessee experiment – what follows?', *Oxford Review of Education* **22**(4), 399–414.

Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005), 'Teachers, schools, and academic achievement', *Econometrica* **73**(2), 417–458.

Rothstein, J. (2009), 'Student sorting and bias in value-added estimation: Selection on observables and unobservables', *Education Finance and Policy* **4**(4), 537–571.

Rothstein, J. (2010), 'Teacher quality in educational production: Tracking, decay, and student achievement', *Quarterly Journal of Economics* **125**(1), 175–214.

Sacerdote, B. (2010), 'Peer effects in education: How might they work, how big are they and how much do we know thus far?', *Handbook of the Economics of Education* **3**.

Strang, G. (1980), *Linear Algebra and Its Applications*, Academic Press.

Todd, P. E. & Wolpin, K. I. (2003), 'On the specification and estimation of the production function for cognitive achievement', *Economic Journal* **113**(485), F3–F33.

Woodcock, S. (2007), Match effects. unpublished manuscript.

Woodcock, S. (2008), 'Wage differentials in the presence of unobserved worker, firm, and match heterogeneity', *Labour Economics* **15**(3), 492–514.

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

Table 1: Descriptive Statistics – Overall Sample (Compulsory and Noncompulsory Movers)

	Number	Percentage
Number of pupils	1,783,255	
Number of schools	20,705	
Number of School-Grade-Years	96,154	
Sample size	8,660,468	(100.00 %)
Key Stage 1 Observations	3,403,195	(39.30%)
... 1998	1,122,018	(12.96%)
... 1999	1,158,413	(13.38%)
... 2000	1,122,749	(12.96%)
Key Stage 2 Observations	5,257,273	(60.70%)
... 2002	1,703,250	(19.67%)
... 2003	1,821,849	(21.04%)
... 2004	1,732,189	(20.00%)

	Number	Percentage
Male	4,413,006	(0.51)
Free School Meal	1,486,477	(0.17)
Special Needs	1,966,470	(0.23)
English spoken at home	7,892,989	(0.91)

	Mean	(Std. Dev.)	Min.	Max.
All Test Scores	49.89	(9.96)	12.68	80.57
Key Stage 1 Test Scores	49.95	(9.95)	12.93	80.57
English	49.94	(9.95)	21.02	80.57
Maths	49.96	(9.95)	12.93	78.00
Key Stage 2 Test Scores	49.86	(9.96)	12.68	74.26
English	49.82	(9.96)	20.48	74.13
Maths	49.86	(9.98)	23.99	66.45
Science	49.88	(9.95)	12.68	74.26

Figure 3: Fraction of compulsory movers by Local Education Authority

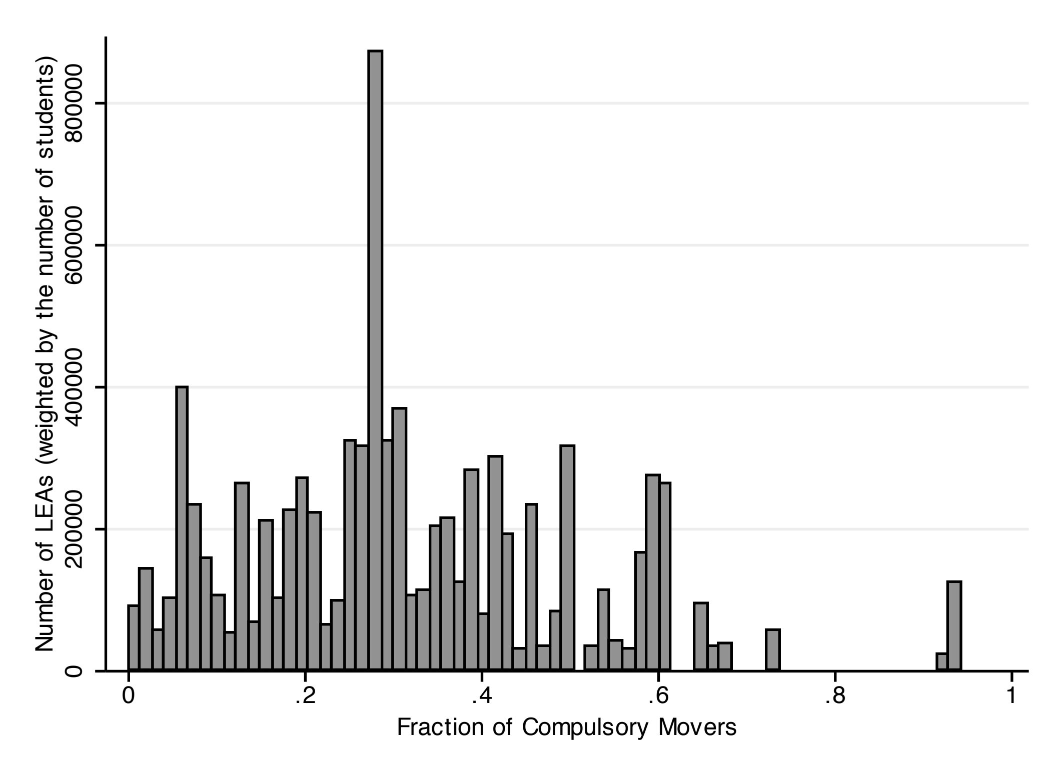


Table 2: The characteristics of compulsory movers vs other students

	Compulsory Movers	Can Stay in the Same School
Moving to the most frequent school (among movers)	0.775	0.151
	(0.418)	(0.102)
Male	0.508	0.509
	(0.500)	(0.500)
Month Of Birth	6.532	6.514
	(3.557)	(3.607)
Special Needs	0.213	0.229
	(0.409)	(0.420)
Free School Meal	0.156	0.174
	(0.363)	(0.379)
English spoken at home	0.928	0.910
	(0.259)	(0.286)
White	0.868	0.847
	(0.338)	(0.360)
Black Carribean	0.010	0.015
	(0.101)	(0.120)
Black, Other	0.004	0.005
	(0.064)	(0.071)
Pakistani	0.022	0.025
	(0.146)	(0.158)
Black African	0.008	0.014
	(0.087)	(0.119)
Mixed	0.017	0.019
	(0.128)	(0.136)
Bangladeshi	0.007	0.010
	(0.084)	(0.097)
Indian	0.023	0.021
	(0.151)	(0.143)
Chinese	0.003	0.003
	(0.050)	(0.054)

Compulsory movers: the pupil had to move because his Key Stage 1 does not cater for Key Stage 2 pupils.
Source: National Pupils Database, Department for Education and Skills.

Table 3: Descriptive Statistics on Mobility – Overall sample of Compulsory and Noncompulsory movers

Number of Pupils	1,783,255	pupils	(100.00 %)
. . . with 2 years of observation	1,663,431	pupils	(93.28 %)
. . . changing school	744,303	pupils	(41.74 %)
			Percentage of moving pupils
Compulsory move	542,646	pupils	(72.91 %)
Noncompulsory move	201,657	pupils	(27.09 %)
Changing School Type	198,586	pupils	(26.68 %)
. . . among compulsory movers	116,055	pupils	(15.59 %)
. . . among non-compulsory movers	82,531	pupils	(11.09 %)
Changing LEA	111,861	pupils	(15.03 %)
. . . among compulsory movers	37,230	pupils	(5.00 %)
. . . among non-compulsory movers	74,631	pupils	(10.03 %)
Moving to the most frequent school	470,041	pupils	(63.15 %)
. . . among compulsory movers	421,442	pupils	(56.62 %)
. . . among non-compulsory movers	48,601	pupils	(6.53 %)

Compulsory movers: the pupil had to move because his Key Stage 1 does not cater for Key Stage 2 pupils.Source: National Pupils Database, Department for Education and Skills.

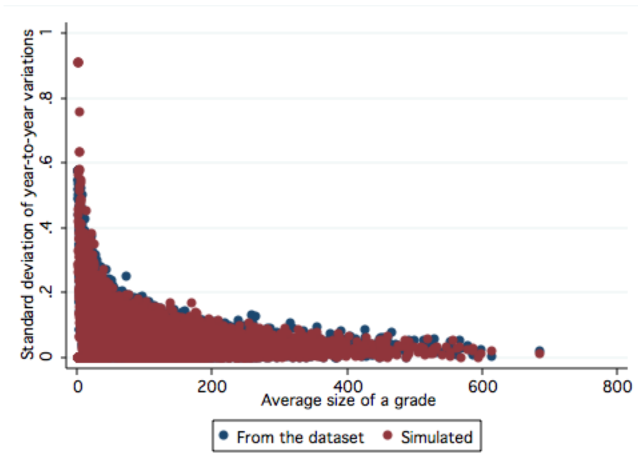
Table 4: The direction of mobility of compulsory movers and other pupils

	Compulsory Movers		Overall	
	Come from	Go to	Come from	Go to
Community School	0.830	0.764	0.699	0.677
Voluntary Aided	0.074	0.101	0.177	0.184
Voluntary Controlled	0.070	0.084	0.097	0.101
Foundation	0.021	0.046	0.021	0.031
Community Special	0.001	0.003	0.004	0.006
Non Maintained Special	0.000	0.000	0.000	0.000
Foundation Special	0.000	0.000	0.000	0.000
Other	0.003	0.001	0.001	0.001

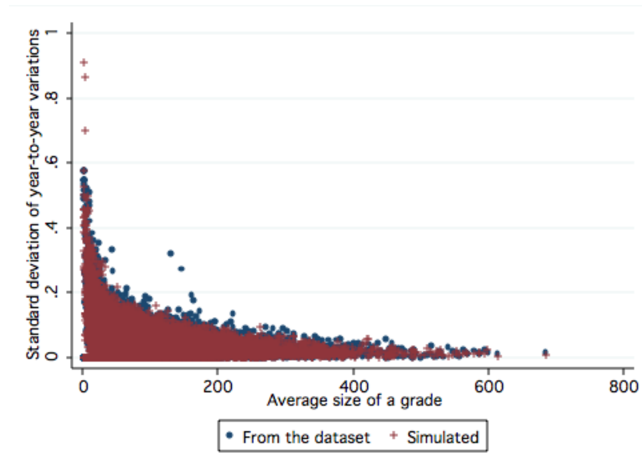
Compulsory movers: the pupil had to move because his Key Stage 1 does not cater for Key Stage 2 pupils.Source: National Pupils Database, Department for Education and Skills.

Figure 4: Year to year variations in grade composition - Realized vs simulated deviations – Estimated on Compulsory Movers Only

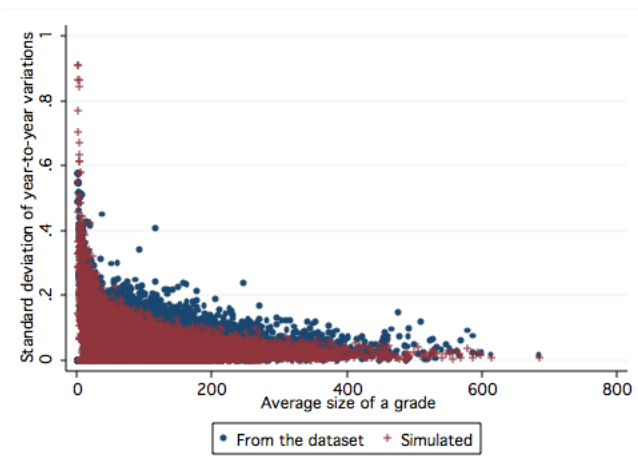
(a) Fraction Male



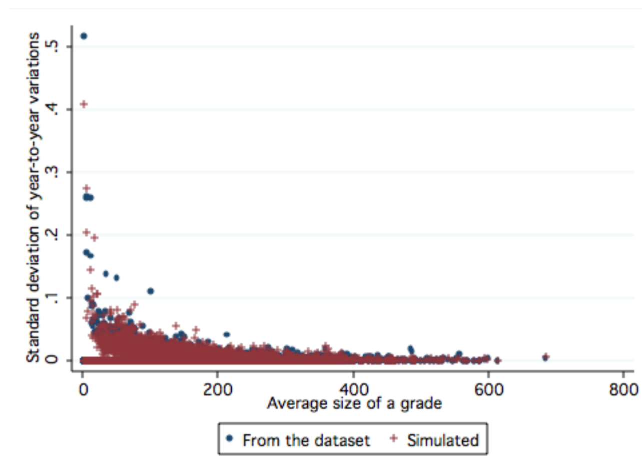
(b) Fraction on Free School Meals



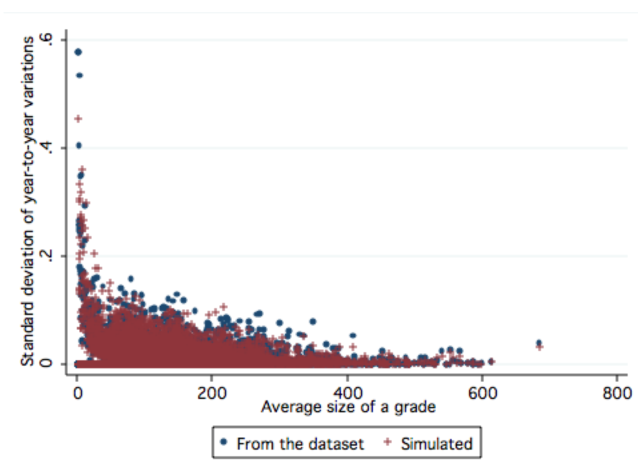
(c) Fraction of Special Needs



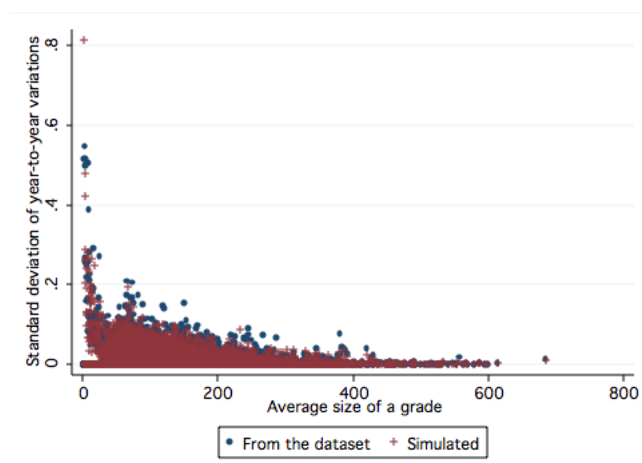
(d) Fraction Chinese



(e) Fraction Indian



(f) Fraction Black African



See section 4.5 for the description of the simulation procedure. Inspired by Lavy & Schlosser (2007).

Table 5: Correlation Tables – Estimated on Compulsory Movers Only

	School Effect and Individual Effect (SE)								Past and Current School-Grade-Year Effect, Individual Effect (PSGYE)							
	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	$\overline{\theta}_{-i}$	φ	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	$\overline{\theta}_{-i}$	φ
y, Standardized grade	50.067	9.739	1.000						50.067	9.739	1.000					
θ, Pupil Effect	0.000	9.123	0.801	1.000					0.000	9.172	0.781	1.000				
θ^*, Observed Pupil Effect	0.000	6.208	0.462	0.681	1.000				0.000	5.266	0.536	0.574	1.000			
$\theta^\perp$, Unobserved Pupil Effect	0.000	6.685	0.663	0.733	-0.000	1.000			0.000	7.510	0.578	0.819	-0.000	1.000		
$\overline{\theta}_{-i}$, Peers' Average Effects,	0.000	4.889	0.235	0.520	0.513	0.233	1.000		0.000	4.881	0.200	0.515	0.184	0.500	1.000	
φ, School-Grade-Year Effect	-	-	-	-	-	-	-	-	0.000	3.753	0.088	-0.098	-0.005	-0.116	-0.184	1.000
ψ, School Effect	0.000	2.114	0.057	-0.024	-0.014	-0.020	-0.045	-	0.000	2.814	0.129	-0.181	-0.002	-0.220	-0.340	0.230
ε, Residual	-0.000	4.532	0.465	-0.000	-0.000	0.000	0.000	-	0.000	4.419	0.454	0.000	-0.000	0.000	0.000	0.000
	School-Grade-Year Effect and Individual Effect (SGYE)								Past School-Grade-Year Effect and Past Individual Effect (PIE)							
y, Standardized grade	50.067	9.739	1.000						50.067	9.739	1.000					
θ, Pupil Effect	0.000	9.323	0.778	1.000					0.000	8.537	0.826	1.000				
θ^*, Observed Pupil Effect	0.000	6.314	0.454	0.677	1.000				-0.000	5.141	0.537	0.602	1.000			
$\theta^\perp$, Unobserved Pupil Effect	0.000	6.859	0.640	0.736	-0.000	1.000			0.000	6.816	0.630	0.798	-0.000	1.000		
$\overline{\theta}_{-i}$, Peers' Average Effects,	0.000	5.242	0.205	0.548	0.511	0.274	1.000		0.000	3.978	0.253	0.446	0.213	0.398	1.000	
φ, School-Grade-Year Effect	0.000	2.913	0.075	0.132	0.285	-0.083	0.235	1.000	0.000	3.106	0.106	-0.150	-0.006	-0.183	-0.321	1.000
ψ, School Effect	0.000	2.207	0.137	-0.089	-0.002	-0.119	-0.157	0.284	0.000	2.326	0.129	-0.120	-0.004	-0.147	-0.257	0.519
ε, Residual	0.000	4.452	0.457	0.000	-0.000	0.000	0.000	0.000	0.000	4.412	0.454	0.000	0.001	-0.000	0.000	0.000

Estimated with a2group, a2reg, xtlreg and xtlreg2. Available through the corresponding author Amine Ouazad.Reading: Test Scores have a standard deviation of 10 and a mean of 50.Source: National Pupils Database, Department for Education and Skills.

Table 6: Decomposition of Variance – Estimated on Compulsory Movers Only

	Overall Variance	Between Schools	Between LEAs	Between School Types
Key Stage 1				
Test Scores	94.729	12.530	1.226	0.796
	(100.0 %)	(13.2 %)	(1.3 %)	(0.8 %)
Individual Effects	74.756	13.632	1.950	0.606
	(100.0 %)	(18.2 %)	(2.6 %)	(0.8 %)
School-Grade-Year Effects	9.680	9.680	0.642	0.048
	(100.0 %)	(100.0 %)	(6.6 %)	(0.5 %)
School Effects	5.543	5.543	0.642	0.050
	(100.0 %)	(100.0 %)	(11.6 %)	(0.9 %)
Key Stage 2				
Test Scores	94.792	17.432	1.215	2.357
	(100.0 %)	(18.4 %)	(1.3 %)	(2.5 %)
Individual Effects	71.606	17.258	1.918	1.832
	(100.0 %)	(24.1 %)	(2.7 %)	(2.6 %)
School-Grade-Year Effects	9.626	9.626	0.641	0.056
	(100.0 %)	(100.0 %)	(6.7 %)	(0.6 %)
School Effects	5.323	5.323	0.643	0.060
	(100.0 %)	(100.0 %)	(12.1 %)	(1.1 %)

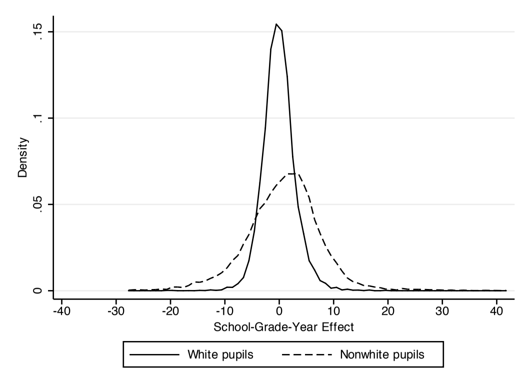
Source: National Pupils Database, Department for Education and Skills.

**: Significant at 1%. *: Significant at 5%.

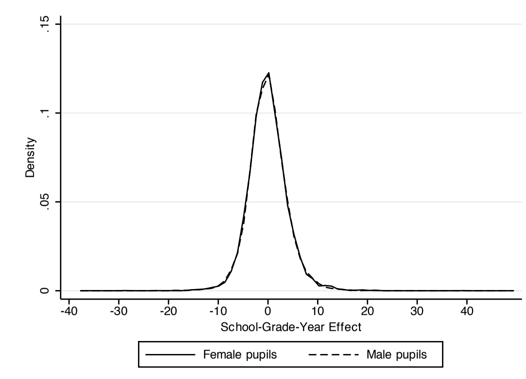
Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Figure 5: School Effects: Distribution for subgroups of Pupils – Estimated on Compulsory Movers Only

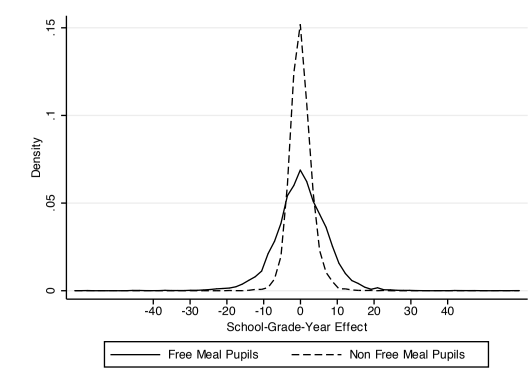
(a) By Ethnicity



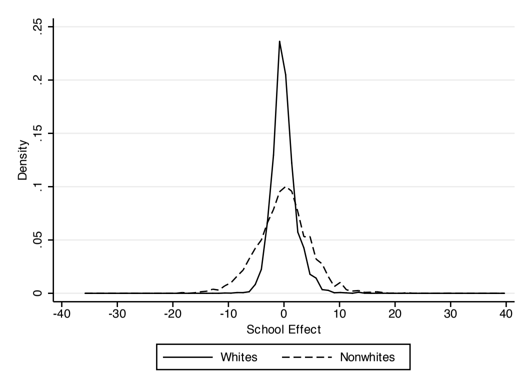
(b) By Gender



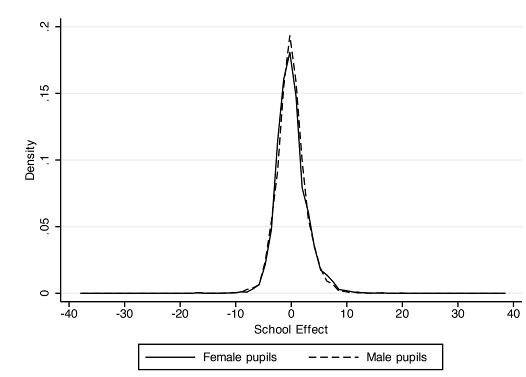
(c) By Free Meal Status



(d) By Ethnicity



(e) By Gender



(f) By Free Meal Status

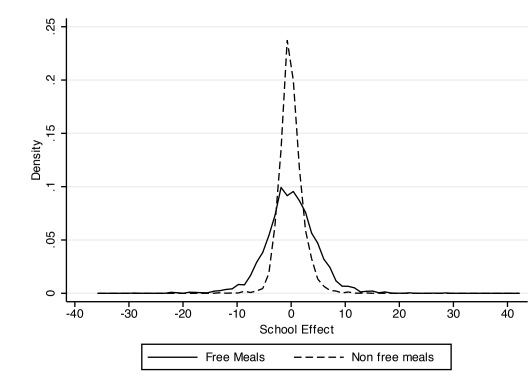


Table 7: Analysis of Pupil Fixed Effects – Estimated on Compulsory Movers Only

	Dependent variable: Pupil Fixed Effect Sample: Key Stage 1 and Key Stage 2 Specification			
	School Effects	School-Grade-Year f.e.	Past S.-G.-Y. f.e.	Past Individual f.e.
Male	0.301**	0.294**	0.307**	0.296**
	(0.018)	(0.018)	(0.020)	(0.018)
Free School Meal	-3.981**	-3.949**	-4.177**	-3.887**
	(0.026)	(0.027)	(0.029)	(0.027)
Special Needs	-11.108**	-11.099**	-11.085**	-10.899**
	(0.025)	(0.025)	(0.027)	(0.025)
Month Of Birth	-0.293**	-0.295**	-0.291**	-0.288**
	(0.003)	(0.003)	(0.003)	(0.003)
Ethnicity				
Chinese	2.266**	2.202**	2.324**	2.193**
	(0.168)	(0.175)	(0.193)	(0.172)
Mixed	0.631**	0.647**	0.740**	0.635**
	(0.072)	(0.073)	(0.082)	(0.072)
Indian	-0.532**	-0.574**	-0.508**	-0.525**
	(0.058)	(0.060)	(0.065)	(0.059)
White	Ref.	Ref.	Ref.	Ref.
Bangladeshi	-3.259**	-2.951**	-3.060**	-2.993**
	(0.113)	(0.117)	(0.128)	(0.117)
Black African	-0.567**	-0.659**	-0.847**	-0.654**
	(0.106)	(0.109)	(0.119)	(0.108)
Pakistani	-3.517**	-3.554**	-3.666**	-3.478**
	(0.066)	(0.068)	(0.075)	(0.067)
Black, Other	-0.758**	-0.878**	-0.926**	-0.851**
	(0.138)	(0.142)	(0.152)	(0.141)
Other ethnicity	-0.330**	-0.314**	-0.153**	-0.274**
	(0.048)	(0.049)	(0.054)	(0.049)
Black Carribean	-1.122**	-1.182**	-1.334**	-1.169**
	(0.090)	(0.092)	(0.100)	(0.091)
R Squared	0.46	0.46	0.33	0.36
F Statistic	23,531.90	23,512.43	13,706.62	15,366.73
Number of Pupils	556,762	556,762	556,762	556,762

Source: National Pupils Database, Department for Education and Skills.**: Significant at 1%. *: Significant at 5%.Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 8: Peer Effects in Schools - Analysis of School-Grade-Year Effects – Estimated on Compulsory Movers Only

	Dependent variable: $\varphi_{j,g,t}$ School-Grade-Year Effect					
	Grade 2			Grade 6		
Specification	(SGYE)	(PSGYE)	(PIE)	(SGYE)	(PSGYE)	(PIE)
Fraction in Grade						
Male	-1.058**	-1.732**	-1.142**	0.686**	0.686**	0.669**
	(0.364)	(0.615)	(0.406)	(0.157)	(0.164)	(0.173)
Free School Meal	-0.217	-0.439	-0.225	-0.640**	-0.592**	-0.516*
	(0.584)	(0.976)	(0.635)	(0.217)	(0.228)	(0.236)
Special Needs	0.422	0.906	0.247	-0.147	-0.102	0.215
	(0.411)	(0.644)	(0.426)	(0.194)	(0.195)	(0.200)
White	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Chinese	-4.500	-5.513	-5.802	4.133**	4.632**	4.544**
	(4.034)	(6.678)	(4.191)	(1.406)	(1.449)	(1.497)
Mixed	-1.148	-1.479	-0.876	0.149	0.221	0.255
	(1.213)	(1.980)	(1.278)	(0.567)	(0.570)	(0.586)
Indian	1.413	2.593	1.488	1.632*	1.632*	1.709*
	(1.897)	(2.968)	(1.980)	(0.741)	(0.753)	(0.767)
Bangladeshi	-5.595	-5.364	-4.574	-1.149	-0.917	-0.616
	(3.689)	(5.712)	(3.810)	(1.254)	(1.286)	(1.376)
Black African	-1.142	-0.030	-0.503	1.063	0.903	0.840
	(2.303)	(4.161)	(2.433)	(0.809)	(0.822)	(0.849)
Pakistani	3.902	6.741*	4.000	1.461	1.364	1.322
	(2.150)	(3.398)	(2.274)	(0.756)	(0.759)	(0.781)
Black, Other	6.602**	8.481*	6.507*	-0.122	-0.357	-0.711
	(2.487)	(4.073)	(2.539)	(1.031)	(1.037)	(1.061)
Black Carribean	4.001*	7.415*	4.449*	-0.880	-0.906	-1.074
	(1.929)	(3.162)	(2.013)	(0.800)	(0.796)	(0.818)
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R Squared	0.88	0.94	0.83	0.91	0.60	0.62
F Statistic	2,550.92	8,235.92	1,643.25	4,252.44	98.12	222.76
Number of school-grade-years	12,111	12,111	12,111	31,596	31,596	31,596
Number of schools	4,474	4,474	4,474	13,994	13,994	13,994

Source: National Pupils Database, Department for Education and Skills.**: Significant at 1%. *: Significant at 5%.Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 9: Peer Effects for Subgroups of Pupils – Estimated on Compulsory Movers Only

	Dependent variable: $\varphi_{j,g,t}(\mathbf{z})$ School-Grade-Year Effect					
Subsample $\mathbf{z}$	Male	Female	Whites	Nonwhites	F.S.M.	Non free meal
Fraction in Grade 6						
Male	-0.315	0.695	0.373	1.097	-0.128	0.367
	(0.540)	(0.569)	(0.546)	(1.533)	(1.159)	(0.436)
Free School Meal	-0.867	-1.765**	-2.050*	-1.039	-1.354	-0.884
	(0.561)	(0.604)	(0.823)	(1.792)	(1.487)	(0.642)
Special Needs	-0.032	0.081	0.040	-1.706	-0.846	0.019
	(0.443)	(0.461)	(0.523)	(1.516)	(1.029)	(0.423)
Chinese	2.958	5.712	4.296	16.540	13.391	7.798
	(3.849)	(3.928)	(3.544)	(9.144)	(12.421)	(4.533)
Indian	3.983*	2.821	2.356	-0.060	3.395	1.243
	(1.720)	(2.015)	(2.112)	(2.982)	(3.766)	(1.551)
Bangladeshi	-2.045	-1.282	3.573	-3.373	-2.946	-1.373
	(3.157)	(2.633)	(3.852)	(5.051)	(4.813)	(2.877)
Black African	1.937	1.002	0.394	3.259	8.350	0.928
	(2.211)	(2.456)	(2.480)	(3.795)	(5.083)	(1.988)
Pakistani	3.825*	2.187	1.591	-0.431	0.296	4.258*
	(1.920)	(1.992)	(2.547)	(2.976)	(3.459)	(1.672)
Black, Other	1.076	0.829	0.348	-10.941**	-6.143	1.878
	(2.513)	(2.489)	(2.528)	(4.227)	(6.354)	(2.178)
Black Carribean	-2.660	0.196	-1.709	-8.478*	-2.553	0.893
	(1.936)	(2.164)	(3.273)	(3.442)	(4.091)	(1.705)
Other controls	School Fixed Effects and Year Dummies					
R Squared	0.74	0.87	0.91	0.67	0.70	0.90
F Statistic	741.08	2,386.73	1,983.50	209.10	383.48	2,655.92
Number of school-grade-years	28,753	28,753	7,391	7,391	22,582	22,582

Table 10: The Analysis of School Effects in grades 2 and 6 – Estimated on Compulsory Movers Only

	Dependent variable: ψ_j School Effect							
	Grade 2				Grade 6			
Specification	(SE)	(SGYE)	(PSGYE)	(PIE)	(SE)	(SGYE)	(PSGYE)	(PIE)
School Status								
Community	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Voluntary Aided	0.039	0.115	0.058	-0.092	0.314**	0.355**	0.337**	0.332**
	(0.161)	(0.173)	(0.305)	(0.226)	(0.088)	(0.092)	(0.094)	(0.099)
Voluntary Controlled	-0.517**	-0.492**	-0.723*	-0.735**	-0.092	-0.172	-0.201	-0.220
	(0.152)	(0.163)	(0.288)	(0.214)	(0.103)	(0.106)	(0.109)	(0.114)
Foundation	-0.501	-0.494	-0.969	-0.838	0.056	0.102	0.206	0.151
	(0.390)	(0.418)	(0.738)	(0.546)	(0.240)	(0.248)	(0.254)	(0.265)
R Squared	0.01	0.03	0.00	0.01	0.03	0.03	0.02	0.01
F Statistic	6.48	19.18	2.85	4.49	46.88	45.23	37.81	22.79
Number of observations	4,474	4,474	4,474	4,474	13,994	13,994	13,994	13,994

Source: National Pupils Database, Department for Education and Skills.**: Significant at 1%. *: Significant at 5%.Reading: Test Scores have a standard deviation of 10 and a mean of 50.

A Appendix: Tables

Table 11: Primary School Categories in England – **Supplementary material, not for publication**

Type	Faith	Governors	Admissions Authority	Assets Owned By	Employer
Non-Majority Controlled Schools					
Community Schools	Secular	Parents $>$ 30 % LEA 20% Staff $<$ 30% Community 20%	LEA	LEA	LEA
Voluntary Controlled	Mostly Church of England, Some other faiths Some secular	Foundation $<$ 25% Parents $>$30% LEA $<$ 20% Staff $<$ 30% Community 10%	LEA	LEA	LEA
Foundation	Mostly secular some Church of England	Foundation $<$ 25% Parents $>$ 30% LEA $<$ 20% Staff $<$30% Community 10%	Governors	Curch or charity	Governors
Majority Controlled Schools					
Voluntary Aided	Mostly Church of England Catholic, some secular	Foundation $>$ 50 % Parents $>$ 30% LEA $<$10% Staff $<$30%	Governors	Church or Charity	Governors

This table is taken from Gibbons, Machin & Silva (2008)

Table 12: The Analysis of School Effects in grades 2 and 6 – Estimated on Compulsory Movers Only – **Supplementary Material, Not For Publication**

	Dependent variable: ψ_j School Effect					
Specification	School-Grade-Year f.e.					
Subsample	Male	Female	Whites	Nonwhites	F.S.M.	Non free meal
Grade 2						
Community	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Voluntary Aided	0.209	0.336*	0.128	-0.244	-0.316	0.324*
	(0.177)	(0.166)	(0.189)	(0.371)	(0.331)	(0.159)
Voluntary Controlled	-0.160	-0.603**	-0.134	-0.206	-0.530	-0.416**
	(0.167)	(0.157)	(0.178)	(0.348)	(0.312)	(0.149)
Foundation	-0.736	-0.525	-0.770	-1.527*	0.333	-0.497
	(0.423)	(0.398)	(0.397)	(0.776)	(0.760)	(0.364)
Grade 6						
Community	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Voluntary Aided	0.728**	0.428**	0.478**	-0.030	0.407	0.322
	(0.155)	(0.147)	(0.185)	(0.289)	(0.254)	(0.166)
Voluntary Controlled	-0.290	-0.397*	-0.678**	-0.136	-0.228	-0.239
	(0.173)	(0.164)	(0.227)	(0.354)	(0.299)	(0.196)
Foundation	0.350	-0.212	0.510	-0.686	0.872	0.152
	(0.339)	(0.322)	(0.367)	(0.572)	(0.514)	(0.337)
R Squared	0.01	0.00	0.02	0.00	0.00	0.02
F Statistic	14.06	6.70	14.24	1.99	3.36	19.89
Number of observations	8,723	8,723	4,523	4,523	6,568	6,568

Source: National Pupils Database, Department for Education and Skills.**: Significant at 1%. *: Significant at 5%.Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 13: Correlation Tables – Simulations **Supplementary material, not for publication**

– First Simulation –

	Mean	Std. Dev.	y	θ	$\overline{\theta}$	φ
y, Standardized grade	49.893	9.958	1.000			
θ, Pupil Effect	0.000	9.271	-0.001	1.000		
$\overline{\theta}_{-i}$, Peers' Average Pupil Effect	0.000	2.356	-0.010	0.219	1.000	
φ, School-Grade-Year Effect	2.187	6.008	0.012	-0.033	-0.143	1.000
ε, Residual	17.198	26.010	0.113	0.000	0.000	-0.000

– Second Simulation –

	Mean	Std. Dev.	y	θ	$\overline{\theta}$	φ
y, Standardized grade	49.893	9.958	1.000			
θ, Pupil Effect	0.000	9.277	0.000	1.000		
$\overline{\theta}_{-i}$, Peers' Average Pupil Effect	0.000	2.379	0.010	0.222	1.000	
φ, School-Grade-Year Effect	4.002	6.354	-0.011	-0.033	-0.178	1.000
ε, Residual	17.201	26.010	0.113	0.000	0.000	0.000

– Third Simulation –

	Mean	Std. Dev.	y	θ	$\overline{\theta}$	φ
y, Standardized grade	49.893	9.958	1.000			
θ, Pupil Effect	-0.000	9.271	0.001	1.000		
$\overline{\theta}_{-i}$, Peers' Average Pupil Effect	0.000	2.333	0.001	0.216	1.000	
φ, School-Grade-Year Effect	-2.960	6.664	0.020	-0.032	-0.176	1.000
ε, Residual	17.216	26.027	0.112	0.000	0.000	0.000

Source: National Pupils Database, Department for Education and Skills.
Estimated with a2group, a2reg, xtlreg and xtlreg2. Available through the corresponding author Amine Ouazad.
Reading: Test Scores have a standard deviation of 10 and a mean of 50.

B Appendix : Bias Reduction with Compulsory Mobility

B.1 The Model With Endogenous Mobility

There are three schools $j = 1, 1', 2'$, and two periods $t = 1, 2$. Students are indexed by i. Initially, in period $t = 1$, all students attend school $j = 1$. Students of $j = 1$ have to move in the second period. Either to school $j = 1'$ or to school $j = 2'$, which are both open only for $t = 2$ students.

Family events affect individuals in the first and the second period.

$$y_{i,t} = \theta_i + \psi_{J(i,t)} - \delta u_{i,t} + \eta_{i,t}$$

with $P(u_{i,t} = 1) = p$ for $t = 1, 2$. The residual $\varepsilon_{i,t}$ is made of time-varying unobservables that do not affect mobility, hence $E(\varepsilon_{i,t}|i, j) = 0$ for all i, j.

Solving for the OLS estimates of $\hat{\psi}_{2'}$

$$\begin{aligned}
\hat{\psi}_{2'} &= \psi_2 - 2\delta[P(u|j = 1')P(j = 1') + P(u|j = 2)(1 - P(j = 1')) - p] \\
\hat{\psi}_{1'} &= \psi_{1'} - 2\delta[P(u|j = 1')(1 - P(j = 2')) + P(u|j = 2')P(j = 2') - p]
\end{aligned}$$

Using $P(u|j = 1')P(j = 1') + P(u|j = 2')P(j = 2') + P(u|j = 1)P(j = 1) = p$, and $P(u|j = 1) = p$, $P(j = 1) = 1/2$, this simplifies into:

$$\begin{aligned}
\hat{\psi}_{2'} &= \psi_{2'} - \delta(P(u|j = 2') - p) \\
\hat{\psi}_{1'} &= \psi_{1'} - \delta(P(u|j = 1') - p)
\end{aligned}$$

B.2 Noncompulsory Mobility: Schools 1 and 1' are in the same building, with the same staff, no move required

In period $t = 1$, all students attend school $j = 1$. In period $t = 2$, parents consider which school their child will attend. The latent utility of attending school j is U_{ij}.

$$\begin{cases}
U_{i2'} = a_{2'} + d_{2'} \cdot u_{i2} + \nu_{i,2'} \\
U_{i1'} = a_{1'} + d_{1'} \cdot u_{i2} + \nu_{i,1'}
\end{cases}$$

We assume $a_{ij} = a_j$ for all i. The cost of moving to school 2' is c. We define $U = U_{i2'} - U_{i1'} - c = a_{2'} - a_{1'} - c + (d_{2'} - d_{1'}) \cdot u_{i,2'} + \nu_{i,2'} - \nu_{i,1'}$. We define $a \equiv a_{2'} - a_{1'}$, $d \equiv d_{2'} - d_{1'}$, $\nu_i \equiv \nu_{i,2'} - \nu_{i,1'}$. We assume that the adverse event drives students to school $j = 2'$, i.e. $d > 0$.

Hence,

$$\begin{aligned}
P(J(i, 2) = 2'|J(i, 1) = 1) &= P(U_{i2'} - c > U_{i1'}) \\
P(J(i, 2) = 1'|J(i, 1) = 1) &= 1 - P(U_{i2'} - c > U_{i1'})
\end{aligned}$$

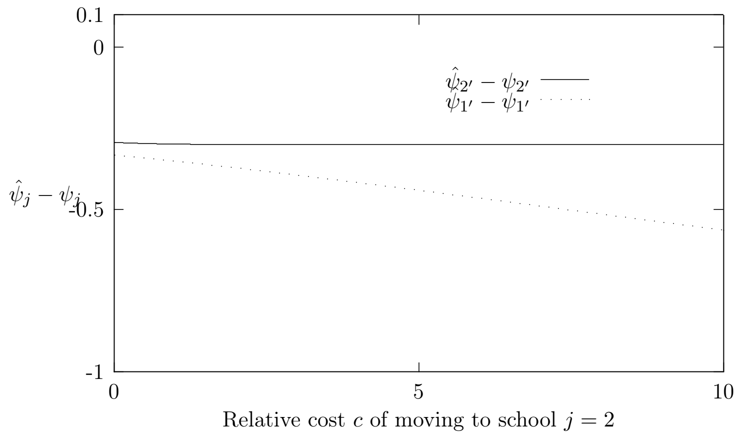


Figure 6: Bias Reduction with Compulsory Mobility, $a = 0$, $d = 1$, $p = 0.05$, $\delta = 1.0$

And, $P(u|j = 2') > p > P(u|j = 1')$ since the adverse shock drives pupils to the other school 2'.

B.3 Compulsory Mobility: Schools 1 and 1' are separate schools

When mobility is compulsory, students pay the mobility cost whether they go to school 1' or to school 2'.

$$\begin{aligned} P(J(i,2) = 2'|J(i,1) = 1) &= P(U_{i2'} - c > U_{i1'} - c) \\ P(J(i,2) = 1'|J(i,1) = 1) &= 1 - P(U_{i2'} - c > U_{i1'} - c) \end{aligned}$$

B.4 Comparing the biases

Proposition 5. *With compulsory mobility, the bias on the mean of $\hat{\psi}$ is smaller, the bias on the variance of $\hat{\psi}$ is smaller. Indeed, the bias of the estimate $\hat{\psi}_{2'}$ declines strongly, whereas the bias of $\hat{\psi}_{1'}$ increases only moderately.*

Computations are straightforward, and the result is illustrated in figure 6.

Note that, with noncompulsory mobility:

$$P(u|j = 2') = \frac{1}{1 + \frac{F(a-c)}{F(a-c+d)} \frac{1-p}{p}}, \quad P(u|j = 1', i) = \frac{1}{1 + \frac{1-F(a-c)}{1-F(a-c+d)} \frac{1-p}{p}}$$

And with compulsory mobility:

$$P(u|j = 2', i) = \frac{1}{1 + \frac{F(a)}{F(a+d)} \frac{1-p}{p}}, \quad P(u|j = 1', i) = \frac{1}{1 + \frac{F(a)}{F(a+d)} \frac{1-p}{p}}$$

Given the properties of the c.d.f F of the normal distribution, the bias on $E(\hat{\psi})$ and on $Var(\hat{\psi})$ are smaller with compulsory mobility.

B.5 Extension: many schools

Compulsory mobility also reduces the bias of the estimates of school effects in a model with a large number of schools.

To make that point, we consider $j = 1, 2, \dots, J$ schools in period $t = 1$ and $j' = 1', 2', \dots, J'$ schools in period $t = 2$. In period $t = 1$, pupils are equally split into the J schools, and in period $t = 2$, pupils' mobility depends on a rational choice based on latent utilities. In period $t = 2$, pupils can be struck by a shock u with probability p.

The educational production function is similar to the previous framework. The only difference with the previous model is that there are more than 2 schools. The latent utility $U_{ij'}$ of attending school j' in period $t = 2$ is:

$$U_{ij'} = a_{ij'} + d_{ij'} u_{i,2} + \varepsilon_{ij'}$$

where $\varepsilon_{j'}$ is extreme-value distributed, so that the probability of attending school j' conditional on u is:

$$P(j'|u_{i,2}, t = 2, i) = \frac{e^{a_{ij'} + d_{ij'} u_{i,2}}}{\sum_{k'=1'}^{J'} e^{a_{ik'} + d_{ik'} u_{i,2}}}$$

under compulsory mobility, and:

$$P(j'|u_{i,2}, t = 2, i) = \frac{e^{a_{ij'} + d_{ij'} u_{i,2}}}{\sum_{k'=1', k' \neq j}^{J'} e^{a_{ik'} - c + d_{ik'} u_{i2}} + e^{a_{ij'} + d_{ij'} u_{i,2}}}$$

for $j \neq j'$ and a similar expression for $j = j'$. The ordinary least squares estimates of the $J + J' - 1$ estimates are determined by:

$$\hat{\Psi} = (Id + 2A)(\delta_2 \cdots \delta_J \delta_{1'} \cdots \delta_{J'})'$$

where δ_j is defined as before, and $A = (P(j = 2) \cdots P(j = J)\, P(j = 1') \cdots P(j = J')) \otimes (1\,1\,1 \cdots 1)'$. From this,

Proposition 6. *Compulsory mobility lowers the endogenous mobility bias on all school effects estimates.*

C Appendix: Identification of The Individual and School Effects Model

C.1 Framework

In the simplest (SE) model, we put forward the identification and estimation framework. In matrix form,

$$Y = D\theta + F\psi + \varepsilon \tag{C-9}$$

where Y is an N vector, D is an $N \times I$ matrix, θ is an I vector, F is an $N \times J$ matrix, ψ is a J vector and ε is an N vector.

C.2 Estimation under Sufficient and Exogenous Mobility

We introduce a few notations. $J(i,t)$ is the school of pupil i in period t. n_i is the number of observations of pupil i, n_j is the number of observations of school j, and $n_{i,j}$ is the number of observations of pupil i in school j.

Definition 7. The *mobility graph for schools G_J* is an undirected graph $(\mathcal{J}, \mathcal{T})$ where $\mathcal{J}$ is the set of schools and $\mathcal{T} \subset \mathcal{J} \times \mathcal{J}$ is the set of transitions. Two schools (j, j') are connected if and only if there is a pupil i that has attended both school j and school j', i.e. $\exists i$ s.t. $n_{i,j} > 0$ and $n_{i,j'} > 0$.

The estimation of θ and ψ proceeds under the assumptions of sufficient mobility and exogenous mobility.

Definition 8. The adjacency matrix of the mobility graph for schools is $A_J = (F'F)^{-1}F'D(D'D)^{-1}D'F$. The element of the j-th row and the j'-th column is the probability of moving to school j' when attending school j.

Assumption 1. [Sufficient mobility] The mobility graph for schools has one connex component.

Assumption 2. [Exogenous mobility] $E(\varepsilon|D, F) = 0$.

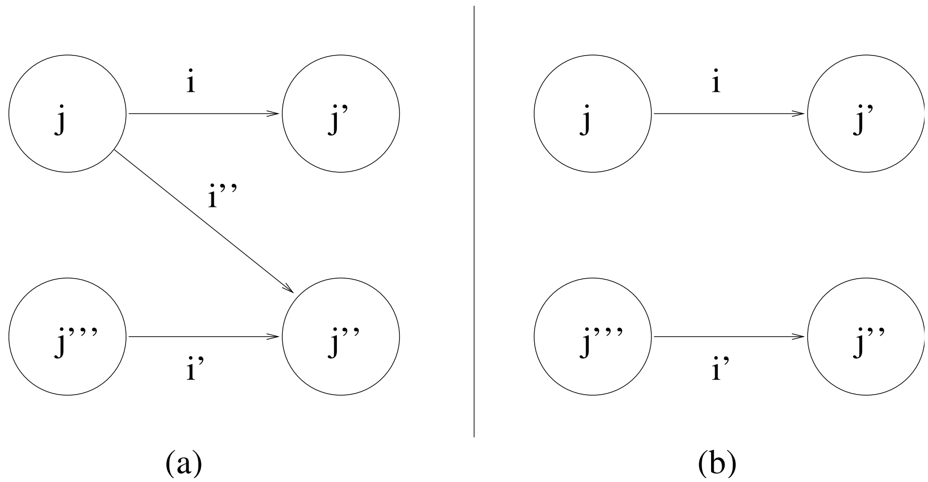
Proposition 9. *$\hat{\theta}$ and $\hat{\psi}$ are unbiased estimators of θ and ψ, up to a constant.*

Proof. Notice that $\hat{\theta}$ and $\hat{\psi}$ satisfy:

$$\begin{cases} (Id - (F'F)^{-1}F'D(D'D)^{-1}D'F)\hat{\psi} &= C \\ (Id - (D'D)^{-1}D'F(F'F)^{-1}F'D)\hat{\theta} &= D \end{cases} \tag{C-10}$$

where C and D are two matrices. Now A_J is a Markov matrix since each row j lists the probabilities of moving to school j' conditional on being in school j. Since the graph is connected, there exists a power k such that $A_J^k > 0$, i.e A_J^k has all elements positive and is called a regular or primitive Markov matrix. By the Frobenius-Perron theorem for Markov matrices (Matthews 1991), 1 is an eigenvalue and the corresponding vector space has dimension 1. Thus the solution $\hat{\psi}$ of C-11 belongs to a one-dimensional vector space. Fixing $\psi_1 = 0$, $\hat{\psi}$ is uniquely defined.

The same reasoning applies to $\hat{\theta}$. □



Notations: i, i', i'' are pupils. j, j', j'', j''' are schools or school-grade-years. Reading: i connects school j and school j' because i attends school j in the first period, and i attends school j' in the second period.

Figure 7: (a) Sufficient mobility – The mobility graph has only one connex component .
(b) Mobility is not sufficient – The mobility graph has two connex components, $\{j, j'\}$ and $\{j'', j'''\}$.

C.3 Consistency under Asymptotic Sufficient and Exogenous Mobility

The consistency of the estimators $\hat{\theta}$ and $\hat{\psi}$ can be established when the number of observations becomes large and the number of schools is fixed.

Assumption 3. θ, ψ, I, J are fixed, and $N \to \infty$.

Consistency of $\hat{\psi}$ can also be proven when θ is stochastic and $I \to \infty$, under similar conditions.

Assumption 4. plim $\frac{1}{N}D'D$, plim $\frac{1}{N}F'F$ exist and are nonsingular. $\text{plim}(D'D)^{-1}D'F$ and $\text{plim}(F'F)^{-1}F'D$ exist.

And we need, similarly, the asymptotic mobility graph to be connected.

Definition 10. The *asymptotic mobility graph* for schools G_J is an undirected graph $(\mathcal{J}, \mathcal{T})$ where $\mathcal{J}$ is the set of schools and $\mathcal{T} \subset \mathcal{J} \times \mathcal{J}$ is the set of transitions. Two schools (j, j') are connected if and only if the probability of a pupil i attending both school j and school j' is strictly positive $\exists i$ s.t. $P(J(i,t) = j) > 0$ and $P(J(i,t) = j') > 0$.

And, similarly, we assume,

Assumption 5. The asymptotic mobility graph for schools has one connex component.

Then

Proposition 11. *$\hat{\theta}$ and $\hat{\psi}$ are consistent estimators of θ and ψ.*

Proof. The proof follows from:

$$\begin{cases} (Id - \text{plim}(F'F)^{-1}F'D \, \text{plim}(D'D)^{-1}D'F) \, \text{plim} \, \hat{\psi} & = \text{plim} \, C \\ (Id - \text{plim}(D'D)^{-1}D'F \, \text{plim}(F'F)^{-1}F'D) \, \text{plim} \, \hat{\theta} & = \text{plim} \, D \end{cases} \tag{C-11}$$

The asymptotic mobility graph for schools is $\text{plim}(F'F)^{-1}F'D\,\text{plim}(D'D)^{-1}D'F$ and the asymptotic mobility graph for pupils is $\text{plim}(D'D)^{-1}D'F\,\text{plim}(F'F)^{-1}F'D$. The same arguments as for the finite sample case, and the fact that $\text{plim}\,\frac{1}{N}D'\varepsilon = \text{plim}\,\frac{1}{N}F'\varepsilon = 0$ prove consistency. □

From the proof, we see that under similar conditions, $\hat{\psi}$ is a consistent estimator of ψ when $I \to \infty$ and $N \to \infty$.

D Appendix: Identification of The Current and Past School Effects Model

The full specification of the paper is a model where test scores depend on current and past individual effects and school quality (Specifications PSGYE and PIE). This model is a contribution to the literature on dynamic panel data: it allows for a short- and long-run effect of school quality on achievement without relying either (i) on the value-added model (which constraints the long-run effect to be equal to the short-run effect) or (ii) on a model with the lagged dependent variable as a regressor, which can be severely biased (Nickell 1981).

In this section, I develop the theory (estimator, iterated estimator, consistency of the estimator) for the estimation of a model with a single current and past fixed effect. The theory can be easily extended to multiple current and lagged fixed effects. The estimation technique for the full model (Specification PIE) is presented in the next section of this appendix.

D.1 The model

In the lagged fixed effects specification, test scores are a function of current school effects, and past school effects. There are two periods, $t = 1, 2$.

$$\begin{cases} y_{i,1} &= \psi_{J(i,1)} + \varepsilon_{i,1} \\ y_{i,2} &= \psi_{J(i,2)} + \lambda \psi_{J(i,1)} + \varepsilon_{i,2} \end{cases} \tag{D-12}$$

where $y_{i,t}$ is the test score of individual i in period t, ψ_j is the school effect of school j, $J(i,t)$ is the school of individual i in period t. λ is called the discount factor. The model can be written in an equivalent matrix form.

$$\begin{cases} Y_1 &= F_1\psi + \varepsilon_1 \\ Y_2 &= F_2\psi + \lambda F_1 \psi + \varepsilon_2 \end{cases} \tag{D-13}$$

where Y_t is an N_t vector, $t = 1, 2$, F_t is an $N_t \times J$ matrix, ψ is a J-vector, ε_t is an N_t-vector. We write the model in a compact matrix form, $Y = (F + \lambda F_{-1})\psi + \varepsilon$, where F and F_{-1} are $N \times J$ matrices, $N = \sum_t N_t$. By convention, $\psi_0 = 0$.

The following part of the section develops an iterated estimator of ψ as a contraction, a least-squares estimator for λ, and shows that the estimator is consistent when $N_t \to \infty$. The estimator is developed in the case of a balanced panel dataset.

Assumption 6. The dataset is balanced, $N_1 = N_2$.

We introduce additional notations. $n_{j,t}$ is the number of observations of school j in period t, $n_j = \sum_t n_{j,t}$. P_λ is an N_1 square matrix, the projection matrix on the image of $F_2 + \lambda F_1$. $\|x\|$ is the euclidean norm of vector x, and $\|\|A\|\|$ is the norm of matrix A (Strang 1980, p280).

D.2 The estimator

The estimation of the school effects and the discount factor relies on two assumptions.

Assumption 7. $F'F$ is nonsingular.

Assumption 8. The discount factor belongs to a compact subset of $(-1, 1)$, i.e. $\exists \underline{\lambda}, \overline{\lambda}$ s.t. $\lambda \in [\underline{\lambda}, \overline{\lambda}] \subset (-1, 1)$.

Under these assumptions, the estimators minimize the sum of squared differences:

$$\hat{\psi}(\lambda) = \text{argmin } \|Y - F\psi - \lambda F_{-1}\psi\|^2 \quad \text{(D-14)}$$

$$\hat{\lambda} = \text{argmin } \|Y - F\hat{\psi}(\lambda) - \lambda F_{-1}\hat{\psi}(\lambda)\|^2 \quad \text{(D-15)}$$

The following sections present a feasible estimator for $\psi(\lambda)$, and prove the consistency of $\hat{\psi}(\lambda)$ and $\hat{\lambda}$.

D.3 An iterated estimator for $\hat{\psi}(\lambda)$

The estimation of $\psi(\lambda)$ is a high-dimensional problem. The complexity of the variance-covariance matrix leads to a computationally difficult problem. Using an iterated estimator for $\hat{\psi}(\lambda)$ offers new theoretical insights and simplifies the calculation. The corresponding software, `alreg` is available from the corresponding author. Section E presents an estimation of the model that has been programmed in `xtlreg`, a Stata procedure available from the corresponding author.

Define the following iterated estimator, for $k = 0, 1, 2, \ldots, \infty$:

$$\hat{\psi}^0 = 0 \quad \text{(D-16)}$$

$$\hat{\psi}^{k+1} = (F'F)^{-1}F'(Y - \lambda F_{-1}\hat{\psi}^k) \quad \text{(D-17)}$$

This iterated estimator converges to the OLS solution of D-14.

Proposition 12. *1.* $\hat{\psi}^k(\lambda) \to_{k\to\infty} \hat{\psi}(\lambda)$.

2. $\hat{\psi}(\lambda) = (Id + \lambda(F'F)^{-1}F'F_{-1})^{-1}(F'F)^{-1}F'Y$.

Proof. Write

$$\hat{\psi}^{k+1} = f(\hat{\psi}^k) \quad \text{(D-18)}$$

Notice that $\hat{\psi}(\lambda)$ is a fixed point of f. Moreover, f is a contraction for any $\lambda \in (-1, 1)$. Indeed, $\|f(\hat{\psi}^{k+1}) - f(\hat{\psi}^k)\| \leq |\lambda|\|(F'F)^{-1}F'F_{-1}\|\|\hat{\psi}^{k+1} - \hat{\psi}^k\|$. $\|(F'F)^{-1}F'F_{-1}\|$ is a Markovian matrix, hence its dominant eigenvalue is one, and $\|f(\hat{\psi}^{k+1}) - f(\hat{\psi}^k)\| \leq |\lambda|\|\hat{\psi}^{k+1} - \hat{\psi}^k\|$. This proves that f is a contraction and, since, $\hat{\psi}(\lambda)$ is a fixed point of f, by standard arguments, $\hat{\psi}^k(\lambda)$ converges to $\hat{\psi}(\lambda)$ for a given realization of the observations. □

D.4 Consistency under the Assumption of Exogenous Mobility

We now prove the consistency of $\hat{\lambda}$ and $\hat{\psi}(\hat{\lambda})$, when the number of schools and the number of periods is fixed and the number of observations per school becomes large. The key assumption is that the probability of staying in the same school in the two periods of observation does not converge to 1.

Assumption 9. The number of schools J is fixed. ψ is fixed. F, F_{-1} and ε are stochastic.

The probability of belonging to any given school should not converge to zero.

Assumption 10. $\text{plim}\, \frac{1}{N}(F'F)$ exists and is invertible.

Finally, the probability of staying in a school should not converge to 1.

Assumption 11. The probability of staying in the same school in period 1 and in period 2 does not converge to one. Formally,

$$\exists \eta > 0, \quad \forall \lambda \in [\underline{\lambda}, \overline{\lambda}], \qquad \text{plim}\, \frac{1}{N} \|F_{-1}\psi - P_\lambda F_{-1}\psi\|^2 > \eta$$

These assumptions imply the consistency of the estimators of the discount factor and the school effects. The proof follows the standard arguments of the consistency of M-estimators (Wooldridge 2002).

Theorem 13. *$\hat{\lambda}$ is a consistent estimator of λ.*

Proof. Define:

$$g_N(\lambda) \equiv \frac{1}{N} \|Y - (F + \lambda F_{-1})\hat{\psi}(\lambda)\|^2 \tag{D-19}$$

and $g(\lambda) = \text{plim}_{N\to\infty} \quad g_N(\lambda)$.

We will prove that (i) λ is the only minimizer of $g(\lambda)$, and that (ii) $g_N(\lambda)$ converges uniformly to $g(\lambda)$ on $[\underline{\lambda}, \overline{\lambda}]$.

Notice that $(F + \lambda F_{-1})\hat{\psi}(\lambda) = P_\lambda Y$, the linear projection on the image of $F + \lambda F_{-1} = M_\lambda$. Then $g_N(\lambda) = \frac{1}{N}\|Q_\lambda Y\|^2$, where Q_λ is the projector on the orthogonal space of $F + \lambda F_{-1}$. Let λ_0 be the true value λ.

$$\begin{aligned} g(\lambda) &= \text{plim}\, \frac{1}{N} \|Q_\lambda M_{\lambda_0}\psi_0 + Q_\lambda \varepsilon\|^2 \\ &= \text{plim}\, \frac{1}{N} \left(\|Q_\lambda M_{\lambda_0}\psi_0\|^2 + 2 < Q_\lambda M_{\lambda_0}\psi_0, Q_\lambda \varepsilon > + \|Q_\lambda \varepsilon\|^2 \right) \end{aligned}$$

ψ_0 and ε are independent, and $E(\|Q_\lambda \varepsilon\|^2) = \text{tr}(Q_\lambda)\sigma_\varepsilon^2$ by the usual tricks. The trace of the projection matrix Q_λ is the dimension of the orthogonal space of $F + \lambda F_{-1}$, i.e. $\frac{1}{N} tr(Q_\lambda) \to 1$.

$$g(\lambda) \quad = \quad \text{plim}\, \frac{1}{N} \|Q_\lambda M_{\lambda_0}\psi_0\|^2 + o(1) + \sigma_\varepsilon^2$$

Now, notice that $M_{\lambda_0} = M_\lambda + (\lambda_0 - \lambda)F_{-1}$, and that, moreover, $Q_\lambda M_\lambda = 0$ by definition of Q_λ. Hence,

$$g(\lambda) \quad = \quad (\lambda_0 - \lambda)^2 \operatorname{plim} \frac{1}{N} \|Q_\lambda F_{-1} \psi_0\|^2 + o(1) + \sigma_\varepsilon^2$$

λ_0 is a minimum of this expression, and, by assumption 11:

$$\exists \eta > 0, \quad \forall \lambda \in [\underline{\lambda}, \overline{\lambda}], \qquad \operatorname{plim} \frac{1}{N} \|Q_\lambda F_{-1} \psi_0\|^2 > \eta$$

λ_0 is therefore the only minimizer of $g(\lambda)$.

Now, $g_N(\lambda)$ converges uniformly to $g(\lambda)$. Indeed, $g_N(\lambda)$ converges pointwise to $g(\lambda)$, $g(\lambda)$ is equicontinuous, and $g_N(\lambda)$ is stochastically equicontinuous. Therefore (Newey 1991), $g_N(\lambda)$ converges uniformly to $g(\lambda)$. □

Theorem 14. *$\hat{\psi}(\hat{\lambda})$ is a consistent estimator of ψ.*

$$\operatorname{plim} \hat{\psi}(\hat{\lambda}) = \psi \tag{D-20}$$

Proof. From the properties of OLS estimation, $\operatorname{plim} \hat{\psi}(\lambda) = \psi$. Now, $\operatorname{plim} \hat{\lambda} = \lambda$ and $\hat{\psi}(\lambda)$ is a continuous function of λ. By the Slutsky theorem, the result follows. □

E Appendix: Estimation of the Current and Past School and Pupil Effects

In this section of the appendix, we present the estimation of the full model with covariates, current and past individual and school effects. The estimation is carried out by an iterative technique. The corresponding program is `xtlreg`, available from the corresponding author.

E.1 Matrix formulation of the Model

We write the specification in matrix form to get the normal form equations and proceed to the estimation by conjugate gradient. The number of students is N. The number of schools is J. The number of covariates is K. The number of observations in the ith period is n_i, and n is the total number of observations, $n = \sum_{t=1}^{T} n_t$.

$$\begin{aligned} Y &= X\beta + D_\lambda\theta + \Phi_\lambda\psi + U \\ \text{with} &\quad \Phi_\lambda = F + \lambda F_{-1} + \cdots + \lambda^T F_{-T} \end{aligned} \tag{E-21}$$

Observations are ordered such that the vector of observations Y contains observations of the first period, then the observations of the second period.

$$Y = (Y_1'|Y_2'|\cdots|Y_T')' \qquad Y \in \mathbb{R}^n$$

Y_i is a column vector with n_i elements. The design matrix of individuals D is a matrix of $n \times N$ elements. Again, I decompose the design matrix into a first period matrix and a second period matrix.

$$D_\lambda = (D_1'|(1+\lambda)D_2'|\cdots|\sum_{k=0}^{T}\lambda^k D_k')' \qquad D \in \mathcal{M}_{n,N}(\mathbb{R})$$

The design matrices for units are linked. Indeed, if

$$F = (G_1'|G_2'|\cdots|G_T')' \qquad F \in \mathcal{M}_{n,J}(\mathbb{R})$$

Then

$$F_{-1} = (0|G_1^{-1'}|G_2^{-1'}|\cdots|G_{T-1}^{-1'})'$$

$$F_{-2} = (0|0|G_1^{-2'}|G_2^{-2'}|\cdots|G_{T-2}^{-2'})'$$

With these notations the identification hypothesis – namely that no unobserved time-varying shock should be correlated with the covariates, student and/or school effects – can be translated in matrix form, ie:

$$E(U|D_\lambda, \Phi_\lambda, X) = 0 \tag{E-22}$$

The normal equations follow:

$$\begin{aligned}
X'(Y - X\beta - D_\lambda\theta - \Phi_\lambda\psi) &= 0 && \text{(E-23)}\\
D'_\lambda(Y - X\beta - D_\lambda\theta - \Phi_\lambda\psi) &= 0 && \text{(E-24)}\\
\Phi'_\lambda(Y - X\beta - D_\lambda\theta - \Phi_\lambda\psi) &= 0 && \text{(E-25)}
\end{aligned}$$

Which can be written:

$$A_\lambda b = \begin{pmatrix} X'X & X'D_\lambda & X'\Phi_\lambda \\ D'_\lambda X & D'_\lambda D_\lambda & D'_\lambda \Phi_\lambda \\ \Phi'_\lambda X & \Phi'_\lambda D_\lambda & \Phi'_\lambda \Phi_\lambda \end{pmatrix} \begin{pmatrix} \beta \\ \theta \\ \psi \end{pmatrix} = \begin{pmatrix} X'Y \\ D'_\lambda Y \\ \Phi'_\lambda Y \end{pmatrix} \tag{E-26}$$

b is the parameter vector. A_λ is non-singular under conditions exposed in section 2. Moreover A_λ is a symmetric, positive definite matrix, and therefore the problem of finding b can be solved by the conjugate gradient algorithm.

Let us denote by $n_{i,j,t}$ the number of observations of student i in school j in year t.

$$\begin{aligned}
D'_\lambda D_\lambda &= Diag(n_{1,\cdot,\cdot} + \lambda n_{1,\cdot,1}, n_{2,\cdot,\cdot} + \lambda n_{2,\cdot,1}, \ldots, n_{N,\cdot,\cdot} + \lambda n_{N,\cdot,1})\\
F'F &= Diag(n_{\cdot,1,\cdot}, n_{\cdot,2,\cdot}, \ldots, n_{\cdot,J,\cdot})\\
F'_{-1}F_{-1} &= Diag(n_{\cdot,1,1}, n_{\cdot,2,1}, \ldots, n_{\cdot,J,1})
\end{aligned}$$

And,

$$\Phi'_\lambda \Phi_\lambda = F'F + \lambda F'F_{-1} + \lambda F'_{-1}F + \lambda^2 F'_{-1}F_{-1}$$

$$\begin{aligned}
D'F &= [n_{i,j,\cdot}]_{i,j} & D'F_{-1} &= [n_{i,j,1}]_{i=1,\ldots,N,j=1,\ldots,J}\\
F'F_{-1} &= (0|G'_2G_1) & G'_2G_1 &= [m_{j,j'}]_{j,j'=1,\ldots,J}
\end{aligned}$$

Where $m_{j,j'}$ is the number of observations that move from unit j to unit j' between period 1 and 2.

The estimation of the normal form equations uses a conjugate gradient estimator. Recursive sequence $b'_n = (\beta'_n \quad \theta'_n \quad \psi'_n)$ is defined such that $b_0 = 0$, and b_n is built from b_{n-1} by the conjugate gradient algorithm described in Dongarra, Duff, Sorensen & van der Vorst (1991).[40]

[40] The conjugate gradient is mathematically an exact method. Due to rounding errors in the computational process, it is practically an approximation of the true solution. The speed of convergence depends on the condition number of matrix A_λ, ie the ratio of its highest eigenvalue and its lowest eigenvalue. Reducing the condition number increases the convergence speed. This is the purpose of preconditioning, for details see source code of `xtlreg`.

E.2 Estimating the discounting factor λ

The estimation of λ is done using grid search. Select H values $\underline{\lambda} = \lambda_1 < \lambda_2 < \cdots < \lambda_H = \overline{\lambda}$ and set

$$\hat{\lambda}_H = \text{argmin}_h \quad \|Y - F\hat{\psi}(\lambda_h) - \lambda_h F_{-1}\hat{\psi}(\lambda_h)\|^2 \tag{E-27}$$

Then $\hat{\lambda}_H \to \hat{\lambda}$ as $H \to \infty$.

F Appendix: Measurement Error in the Correlation of Pupil and School Effects

In all specifications except HSYGE, schools are equally effective for all students, that is, there is no complementarity between pupils and schools. If the educational production function is truly specified as in specifications PSGYE, the model does not predict any particular matching of pupil effects and school effects at equilibrium. Matching patterns are indeed determined by the complementarity between pupil effects and school effects, following Becker (1973)[41]. In such a world, the model predicts zero correlation between pupil effects and school-grade-year fixed effects.

However, some of the correlations between child effects and school effects in Table 5 are negative. Does it mean that pupils with a high pupil effect are structurally matched with low school-grade-year effects? The correlation between estimated pupil effects and estimated school effects is actually downward biased and we perform simulations to estimate the magnitude of the bias, suggesting that the correlation is likely to be close to zero.

The correlation between estimated effects is downward biased. This has been pointed out in the context of worker-firm matched panel datasets (Abowd & Kramarz 2004). To make this clear, let us decompose the correlation between child and school effects. This correlation can be written as the sum of the correlation between measurement errors and the true covariance between the effects.

$$Cov(\hat{\theta}, \hat{\varphi}) = Cov(\hat{\theta} - \theta, \hat{\varphi} - \varphi) + Cov(\hat{\theta} - \theta, \varphi) + Cov(\hat{\varphi} - \varphi, \theta) + Cov(\theta, \varphi) \tag{F-28}$$

θ is the individual effect, φ is the school-grade-year effect, $\hat{\theta}$ is the estimated individual effect, $\hat{\varphi}$ is the estimated school-grade-year effect.

The estimation of $Cov(\theta, \varphi)$ therefore requires the estimation of $Cov(\hat{\theta} - \theta, \varphi)$, $Cov(\varphi, \hat{\theta} - \theta)$, $Cov(\hat{\theta} - \theta, \hat{\varphi} - \varphi)$. In general, the measurement errors of child and school effects are negatively correlated (Abowd & Kramarz 2004). The intuition behind this result is that (i) pupils who change school get a better estimated effect but school effects are less precisely estimated (ii) pupils who do not change school have a less well estimated effect but their associated school effect is more precisely estimated.

Simulations can assess the order of magnitude of the downward bias of the correlation. We

[41] This of course, assumes a particular form of preferences and special market conditions. The housing market should be perfect, parents should know the educational production function as specified in equation PSGYE and the only reason for location decisions should be the level of test scores.

generate pupil effects who have a normal distribution with the same variance as the estimated pupil effects. We also generate school effects the same way. The point here is that pupil effects and school effects are uncorrelated. We then generate simulated test scores using the specification with past and current school-grade-year and individual effects.

The results of simulations[42] suggest that even in the absence of a true correlation between pupil effects and school effects, the correlation between estimated effects is negative. The true correlation between school-grade-year and individual effects is therefore likely to be close to -0.1, with school-grade-year effects explaining little of the variance of test scores.

Generally speaking, estimating the correlation between pupil effects and school effects remains a difficult challenge in pupil-school or worker-firm fixed effects specifications. Most papers find a zero or negative correlation (Abowd & Kramarz 2004, Abowd et al. 1999). But these papers do not include a match effect that could account for the complementarity between pupil and school-grade-year effects or worker and firm effect.

[42] Available on request. See supplementary tables.