Online Appendix for "Using Compulsory Mobility to Identify School Quality and Peer Effects"

Francis Kramarz* Stephen Machin† Amine Ouazad ‡

April 2013

1 Data Appendix: National Pupil Database, Edubase

Each student's test score, demographic characteristics, and geographic location are collected in the National Pupil Database, a comprehensive administrative register of all English pupils in state schools. Data are made available by the UK Department for Education. Reporting is mandatory for all state schools, and students are followed from year to year via *pupil matching reference* numbers. Longitudinal data can thus be built by stacking consecutive years of the National Pupil Database.

The data set provides rich information on pupils' characteristics and location: gender, free school meal status, special educational needs, ethnicity group, school identifier, and school postcode. The postcode is obtained by merging the UK EduBase with the school identifier of the National Pupil Database. The geographic coordinates of the postcode's centroid are obtained using the Ordnance Survey's database.

Pupils who receive free meals are from families in the lowest quintile of income (17%). The ethnicity variable of our sample encodes the main ethnic groups: White, Asian (Pakistani, Bangladeshi, Indian, Chinese), Black (Black Caribbean, Other Black, African Black), mixed ethnicity, other. Coding of the ethnicity variable changed during the period we consider, so it was necessary to recode this variable in a time consistent manner, following procedures used by Gibbons (2007). We thank the authors for providing this code. The data also provide

*CREST, CEPR, and IZA.
†University College London, Centre for Economic Performance, London School of Economics, and CEPR.
‡INSEAD, Centre for Economic Performance, and CREST.

some information on school structures and types (i.e. whether they are community schools, foundation schools, voluntary aided or voluntary controlled schools, etc.).

We use test scores in English and mathematics (thereafter "Math"), which are available in both Key Stage 1 and Key Stage 2. These tests are constructed and scored externally. Test scores are standardized — to a mean of 50 and a standard deviation of 10, before the deletion of missing observations — in order to make results comparable across grades and fields.

Three cohorts of pupils are followed in grades 1 and 2 (Key Stage 1) and in grades 3, 4, 5 and 6 (Key Stage 2): the 1998–2002 cohort, the 1999–2003 cohort, and the 2000–2004 cohort. Each cohort comprises about 600,000 pupils, and for each pupil there are two observations (Math and English test scores) in Key Stage 1 and Key Stage 2. The sample consists of 1,705,300 students and 21,360 schools.

2 Further Discussion: Compulsory Mobility

2.1 Characteristics of Infant-Only Schools and "All-Through" Schools

If the infant and the linked (most frequent destination) junior school are essentially the same school, there will be little mobility towards different schools. We therefore merged our data set with Edubase (from 1998 to 2000 for Key Stage 1 schools, and from 2002 to 2004 for Key Stage 2), to identify: the name of the headteacher, the postcode, county, town of the school. Table 2 presents the statistics. An infant and a junior school are said to share a head when the first and last name of the headteacher of the infant are similar to the first and last name of the headteacher of the junior school.[1] The first column shows the fraction of compulsory movers who moved from an infant school to a junior school that had the same headteacher name, postcode, county, town, or comes from the most likely infant school. 0.34% of such compulsory movers transferred to a junior school whose headteacher had the same name as the infant school's headteacher. 29.28% of the compulsory movers moved to a school with the same postcode (full postcode), 94% to a school in the same county – there are 83 counties in England, so this suggests short- to medium-distance moves –, 87.21% to a school in the same town – when applicable – and 73.49% of the students came from the most likely infant school. The second column presents similar statistics for compulsory movers outside the Greater London area. Finally, the third column presents the statistics considering only

[1] In this table, the headteachers can have different middle names, which may overstate the fraction of infant schools sharing a head with the corresponding junior school. Otherwise we checked that the fraction was robust to the use of an inexact match using the Levenshtein distance, which allows for small spelling differences and omission of the middle name (Levenshtein 1966).

students who come from the most likely infant school (hence the 100% statistic in the 5th row). As expected the fraction of moves to a school sharing the same headteacher, postcode, county, or town increases. About 39.8% of such compulsory movers transfer to a school with the same postcode.

Figure 1 shows the density of the distance of moves in miles for compulsory an non-compulsory movers. The solid line is the distribution of the distance of moves for compulsory movers. The dashed line is the same distribution for non-compulsory movers. The distribution clearly suggests that both the mean distance and the median distance are greater for *non-compulsory movers* than for compulsory movers. As pointed out in Table 5 of the paper, the median distance for compulsory movers is 1.33 miles and 20.73 miles for non-compulsory movers. Such differences are not likely due to a handful of outliers, but rather to the large distances of non-compulsory movers' mobility. The mode of their distribution is between 4.5 and 5 miles, and the distribution has a "plateau" from 5 miles onwards. Interestingly, the distribution of compulsory movers' distances follows a distribution that would be expected if students move randomly within urban clusters; while the distribution of non-compulsory movers' distances follows a distribution that would be expected if they move across urban clusters, from city to city. Such observation would be consistent with the hypothesis that non-compulsory movers tend to move to schools in other urban areas, for instance because of parental unemployment and/or family events such as divorce, as described in Gibbons (2007).

Overall the statistics suggests that the distance between the infant and the junior school is small, but the two schools seem to be distinct administrative entities. Although such statistics do not provide definitive evidence as to whether the two schools have a common admission procedure, it suggests that a substantial fraction of students did not attend the linked infant school.

2.2 External Validity of the Estimations

As the variance of school effects is estimated on a subset of schools, the representativeness of the sample of students is key to the generalizability of our estimation results. Table 1 of the paper suggested that non-compulsory movers' characteristics are comparable to the characteristics of the overall sample of students. However, the paper did not provide such table for the characteristics of schools. Table 3 shows the distribution of school types across infant-only and infant-and-junior schools. Infant-only schools are more likely community schools, less likely to be voluntary controlled or voluntary aided schools, and as likely to be foundation schools. Figure 2 presents the distribution of urban/rural status for infant-only schools (compulsory movers) and for infant-and-junior schools (non-compulsory movers). There are a little bit more compulsory movers in "urban >10k, less sparse areas" (the difference is about 3 percentage points), and a little bit more non-compulsory movers in "village, less sparse" areas.

The distributions of characteristics are close but are not identical. Therefore we reestimated the variance and standard deviation of the school effects by weighting the sample so that the distribution of observable characteristics is similar to the distribution of observable characteristics in the overall sample.

To account for differences in the distribution of student characteristics x_i (free meal, nonwhite, male) in the sample of compulsory movers and in the overall sample, we first divide the sample into C cells of students with identical observable characteristics x_i.[2] In each cell $c = 1, 2, ..., C$, we compute the variance of school effects $Var(\text{School}_j|c)$. Because the distributions of observables are different for compulsory movers and all students, we note $n_c^{\text{compulsory}}$ the number of compulsory movers in cell c and $n_c^{\text{all students}}$ in the entire sample. The variance of school effects in the sample of compulsory movers is:

$$\text{Var}(School_j) = \sum \frac{n_c^{\text{compulsory movers}}}{n^{\text{compulsory movers}}} \cdot \text{Var}(School_j|c)$$

We then estimate the *corrected* variance of school effects, which accounts for the different distribution of observables in the entire sample; it uses weights of the sample of all students, but keeps the estimates based on compulsory movers:

$$\text{Corrected Var}(School_j) = \sum \frac{n_c^{\text{all students}}}{n^{\text{all students}}} \cdot \text{Var}(School_j|c)$$

The standard deviation in such a case is 2.05, which is 0.4 percentage points of a s.d. of test scores, higher than the variance of 2.01 when not adjusting for the different distribution of the covariates. This is very much consistent with the paper's statements on the non-compulsory mobility bias when estimating the variance of school effects.

2.3 Geographic Distribution of Infant-Only Schools

Although compulsory movers are not exactly evenly distributed, all Local Education Authorities (LEAs) have students in infant-only schools. Across Local Education Authorities, the median fraction of infant-only school students is 31.4%, 80% of Local Education have more than 14% of such students, and 20% have more than 62% of such students.

[2] There are thus $C = 8$ cells for 3 binary characteristics.

Figure 3 takes the example of the Sheffield Local Education Authority to display the location of infant-only schools (blue dots) and of all-through schools (red dots). It suggests that infant-only schools are relatively evenly distributed across the Local Education Authority.

2.4 Random Shocks to Peer Group Composition

Here we provide evidence that both the randomness of school demographics and the finite number of students in a given grade provide exogenous variations in grade composition for the identification of peer effects. Because this paper focuses on estimating an educational production for compulsory movers only, the grade-6 and the grade-2 school are always different. Therefore, in paper's main specification the school fixed effect is a *school*×*grade* fixed effect. Hence, peer effects are estimated using year-to-year variations in grade composition, as in other frameworks that follow Hoxby (2000).

To illustrate that point, we take the within-grade transformation of the simple education production function, augmented with peer effects:

$$\begin{aligned} Test\ score_{i,f,t} - E(Test\ score_{i,f,t} \mid School\ j) \quad = \quad & \gamma \left\{Peers_{-i,g(i,t)} - E(Peers_{-i,g(i,t)} \mid School\ j)\right\} \\ & + Student_i - E(Student_i \mid School\ j) \\ & + \varepsilon_{i,f,t} - E(\varepsilon_{i,f,t} \mid School\ j) \end{aligned} \tag{1}$$

Now observe that $E(Peers_{-i,g(i,t)} \mid School\ j) = E(Students_{j,t} \mid School\ j) = Students_j$, where $Students_{j,t}$ is the demographic composition of grade-6 school j in year t and $Students_j$ is the average demographic composition of school j. So conditional on a school effect, the identification assumption for peer effects implies that year-to-year variations in grade composition, $Peers_{-i,g(i,t)} - Students_j$, are not correlated with students' time-varying changes in unobservables, $\varepsilon_{i,f,t} - E(\varepsilon_{i,f,t} \mid School\ j)$.

One way of checking for whether changes in grade composition are truly random is to compare year-to-year variations with truly random variations in school average composition. For instance, if the fraction of boys and girls in coeducational schools is nearly 51% then small changes in the fraction of boys and girls will be due to the small number of students in each grade rather than to the endogenous sorting of boys and girls into schools. Formally, if changes in grade composition are truly exogenous, they must exhibit some random fluctuation around the average school composition. Identification relies on the assumption that grade composition in a given year is a finite-sized approximation of the school's equilibrium composition.

$$Students_{j,t} = Students_j + Shock_{j,t} \tag{2}$$

As before $Students_j$ is the average composition of school-grade j and $Students_{j,t}$ is the composition of school-grade j in year t. The latter term is a vector that contains the percentage of of boys, of each ethnicity, and of free-meal students. $Students_j$ is average school composition across the three cohorts. Since peer composition is a finite mean that fluctuates around $Students_j$, by virtue of the central limit theorem (Wooldridge 2002), it follows that the shocks $Shock_{j,t}$ should be approximately normal with variance $\mathrm{var}(Students_j))/Grade\ Size_{j,t}$. Therefore, the variance of deviations of peer composition around the school average $Students_j$ should decline in inverse proportion to grade size if the fluctuations are random.

This decline is illustrated in Figure 5, which compares actual year-to-year variations in grade composition to the variations that would be expected if grade composition truly varied randomly around a school-specific average. For boys, free-meal students, and three important ethnic groups, year-to-year variations are remarkably similar to random variations. This finding indicates that trends in grade composition neither occur on a large scale nor introduce significant bias in the estimations. The only simulation-based evidence of deviation from actual year-to-year varations is the variance in the fraction of special-needs students. The actual variance is greater in the data set than what would be expected if it were purely random, indicating the possible existence of trends in the fraction of special-needs students. Such a trend may well be due to evolving support for special-needs students in elementary schools in England. After reestimating our education production function on the subset of schools with *no* special-needs students, we find no significant difference in the estimated peer effects for populations other than such students.

3 Estimation Method

3.1 Estimator of the Education Production Function

Our specification combine a number of important features: (i) test scores are a function of the effects of pupil background, school quality, and peers; (ii) school quality and peer effects can vary over time; and (iii) the whole history of inputs shapes each pupil's experience (c.f. Rivkin, Hanushek & Kain 2005, Todd & Wolpin 2003). Formally, in grade 2, we have:

$$Test\ score_{i,f,t} = Controls_{i,f,t}\beta + Peers_{-i,g(i,t)}\gamma + Student_i + School_{j(i,t)} + \varepsilon_{i,f,t}$$

In the subsequent grade, past inputs have an impact on current achievement:

$$\begin{aligned} Test\ score_{i,f,t} &= Controls_{i,f,t}\beta + Peers_{-i,g(i,t)}\gamma + Student_i + School_{j(i,t)} \\ &\quad + Peers_{-i,g(i,t-1)}\gamma_{-1} + \lambda^{school} School_{j(i,t-1)} + \lambda^{student} Student_i + \varepsilon_{i,f,t} \end{aligned}$$

Our estimators of the student and school effects, the discount factors and the coefficients on the covariates minimizes the sum of squared residuals.

$$
\begin{aligned}
RSS &= \sum_{i=1}^{N}(Test\ score_{i,f,2} - Controls_{i,f,t}\beta + Peers_{-i,g(i,t)}\gamma + Student_i + School_{j(i,t)})^2 \\
&\quad + \sum_{i=1}^{N}(Test\ score_{i,f,6} - Controls_{i,f,t}\beta - Peers_{-i,g(i,t)}\gamma - Student_i - School_{j(i,t)} \\
&\qquad - Peers_{-i,g(i,t-1)}\gamma_{-1} - \lambda^{school} School_{j(i,t-1)} - \lambda^{student} Student_i)^2
\end{aligned}
$$

Hence, if the discount factors are zero, the estimates are simple OLS estimates with dummy variables for student and school effects. If discount factors are nonzero, the estimates are non-linear least squares estimates. The identification assumption is that the residuals are mean-independent of the covariates and student and school effects.

We do not make assumptions regarding the finite-sample distribution of the residuals.

The specifications for grade 2 and for grade 6 can be stacked and written in vector form.

$$
\begin{aligned}
TS &= X \cdot \beta + \Gamma \cdot \Pi + D \cdot S + F \cdot \Sigma \\
&\quad + \Gamma_{-1} \cdot \Pi_{-1} + \lambda^{school} \cdot F_{-1} \cdot \Sigma + \lambda^{student} \cdot D_{-1} \cdot S + U
\end{aligned} \tag{3}
$$

TS is a vector of size $n = 2NT$. N is the number of students, T the number of time periods, 2 the number of subjects. We assume here the panel is balanced but calculations and properties are robust to a unbalanced panel as long as attrition and entry into the panel are exogenous. We sort observations by time period first, then by student, and then by subject, so that the first $2N$ observations are for the first period of observation, and the first two observations for the first student. X is the matrix of observable controls $Controls_{i,f,t}$. Π is the n vector of peers' characteristics. Γ is a vector with peer effects' coefficients. S is the N vector of student effects, Σ is the J vector of school effects. D is a $\{0,1\}$ matrix of size $2NT \times N$ which maps student effects into observations. D is called the student effects' design matrix following Abowd, Kramarz, Margolis (1999). F is similarly called the school effects' design matrix. Π_{-1} is the vector of peers' characteristics in the previous school. Γ_{-1} is the vector of lagged peer effects' coefficients. D_{-1} is the design matrix for previous student effects, so that the first NT rows of D_{-1} are empty. $\lambda^{student}$ is the discount factor for student effects, and λ^{school} is the discount factor for school effects. U is the stacked vector of residuals.

We assume, as in the Ordinary Least Squares framework (Wooldridge 2002), that residuals are mean-independent of the observable covariates and the effects:

$$E(U|X, \Pi, S, \Sigma, \Pi_{-1}, F_{-1}, D_{-1}) = 0$$

Whether this assumption is satisfied is the focus of the paper; the paper argues that focusing on the subset of compulsory movers makes this assumption more plausible than when focusing on the entire set of students.

3.2 Estimation Technique

The estimation of our main specification (3) is more difficult than a usual ordinary least squares regression for essentially two reasons. First, lagged effects are multiplied by the discount factors, and *both* the discount factors and the effects should be estimated in the same specification. The model is nonlinear. Second, even assuming that the discount factors are known, the specification includes four fixed effects: student effects, school effects, lagged student effects, and lagged school effects. Moreover, it is assumed (to ensure identification of the model) that the effect of a school on test scores in grade 6 is equal to the effect of a school on test scores in grade 2, multiplied by the discount factor. This means that, conditional on the discount factors, we face the problem of estimating a specification with four fixed effects with the constraint that lagged school effects are equal to current school effects.

We proceed in the following way. Assuming the values of the school effects and student effects are known, estimates of the discount factors $\widehat{\lambda^{student}}$ and $\widehat{\lambda^{school}}$ are the coefficients of the lagged student and school effects when regressing the test score on the covariates, the lagged student $D_{-1}S$ and school $F_{-1}\Sigma$ effects, controlling for student and school fixed effects.

We first regress the test score on the covariates, student, and school effects – but with no lagged student or school effects. The predicted student and school effects provide us with a first-step estimate of the lagged student and school effects. We then proceed iteratively by repeating the following two steps:

1. We regress the test score on the covariates, the lagged student and school effects, controlling for student and school effects. The coefficient of the lagged student effect is an estimate of $\lambda^{student}$ and the the coefficient of the lagged school effect is an estimate of λ^{school}.

2. We predict new estimates of the student and school effects.

Empirically, the estimates of the discount factors are stable after around 30 iterations. The first 3 digits of the discount factors are unchanged thereafter. This is illustrated in Figure 6.

Finally, standard errors for $\lambda^{student}$ and λ^{school} are obtained by bootstrap. We draw M samples with replacement from the original sample. The unit of clustering is the student level. We estimate all parameters with each sample separately. This yields estimates $\widehat{\lambda^{student}}_m$ and $\widehat{\lambda^{school}}_m$ for each sample $m = 1, 2, \ldots, M$. The 5th (resp., the 95th) percentile of the distribution of $\widehat{\lambda^{student}}_m$ gives the lower bound (resp., upper bound) of the 95% confidence interval for the estimate of $\widehat{\lambda^{student}}$. Similar reasoning applies for $\widehat{\lambda^{school}}$.

3.3 Discount Factors

Estimates of λ^{school} and $\lambda^{student}$ together with their standard errors are presented in table 5.

These estimates suggest that students differ more in the level of test scores than in their progress in test scores. A 1 s.d. increase in the student effect raises grade-2 scores by 1 s.d., and grade-6 scores by 1.024 s.d. Thus, students with a higher student effects have higher average test scores and make more progress between grade 2 and grade 6. But the impact on progress is smaller than the impact of student effects on levels. By the end of grade 2 therefore, inequalities between students' test scores thus also broadly reflect inequalities between students' test scores in grade 6.

School effects' discount factor is interpreted in the following way: a 1 s.d. increase in the grade-2 school effect raises grade-6 test scores by 8% of a standard deviation. Standard errors suggest that the previous (grade-2) school has a significant and positive impact on later (grade-6) achievement.

Although the discount factor is estimated, it may be interesting to see the impact of different values of the discount factor on the estimated variance of the student and school effects. Table 6 constrains the discount factor on students and schools to be identical and suggests that, as λ (the discount factor) goes from 0.024 to 0.5, the standard deviation of the estimated student effects goes down, from 9.2 to 8.9 and the standard deviation of school effects goes up from 2.6 to 3.4. This is not surprising as a larger discount factor makes the variance of student effects explain a larger share of the standard deviation of test scores. The R-squared is highest in the table where the discount factor is equal to its estimated value, as expected given the estimation procedure.

4 Monte Carlo Analysis of Baseline Results

The estimator presented in Section 3.2 of this Appendix is consistent, i.e. converges in probability to the true values, when the number of observations N and the number of time periods T goes to infinity (Blundell & Robin 1999). There are essentially two concerns regarding the empirical validity of this theorem. First, although the estimator of the parameters is consistent for $N \to \infty$ and $T \to \infty$, it may be biased in a finite sample of fixed N and T — as are non-linear least squares estimators (Wooldridge 2002). The second concern is that our data set of students and schools has a limited number of time periods.

We analyze the finite-sample properties of our estimator using a Monte Carlo procedure (Davidson & McKinnon 1993). We generate $S = 20$ data sets where:

- The number N of students, the number J of schools, and the number of time periods T is equal to the corresponding numbers in the original data set.

- The assignment of students i to schools $j(i,t)$ is kept as in the original data set, thus preserving the mobility patterns of students across schools.

- Covariates $X_{i,t}$ are kept as in the original sample.

And the outcome variable – Test Score of student i in year t – is generated using the following parameters:

- Student effects $Student_{i,s}$ in each Monte Carlo sample $s = 1, 2, \ldots, 20$ are drawn from a central normal distribution with standard deviation equal to the standard deviation (8.720) of the student effects in the main table 9 of results of the paper.

- School effects $School_{j,s}$ are similarly drawn from a central normal distribution with standard deviation equal to the standard deviation (2.018) of the school effects in the main table 9 of results of the paper.

- The discount factor λ is equal to the discount factor 0.024 of table 10 of the paper.

- Coefficients β on covariates $X_{i,t}$ are kept as in table 9 of the paper.

- Residuals $\varepsilon_{i,t}$ are drawn from a central normal distribution with standard deviation such that the R-squared of the regression equals the R-squared of the main regression of table 9 of the paper.

We then estimate the student effects $\widehat{Student}_{i,s}$ and school effects $\widehat{School}_{j,s}$ using the estimation technique presented in section 3.2. If such estimation technique has desirable properties, estimated student effects $\widehat{Student}_{i,s}$ in each Monte Carlo sample s should be highly correlated with generated student effects $Student_{i,s}$ (resp., school effects $School_{j,s}$). Also, the estimated discount factor $\widehat{\lambda}_s$ should not be statistically different from the actual discount factor λ.

This is what we observe in the results presented in Figure 7. The average of the estimates $\hat{\lambda}_s$ of the discount factor λ is equal to 0.021, which is only 0.003 below the actual discount factor. None of the estimated discount factors is statistically different from the true value λ. On average across Monte Carlo samples $s = 1, 2, \ldots, 20$, estimated student effects $\widehat{Student}_{i,s}$ and student effects $Student_{i,s}$ have a correlation of 0.93. Similarly, estimated school effects $\widehat{School}_{j,s}$ and school effects $School_{j,s}$ have an average correlation across Monte Carlo samples of 0.86.

Finally, the correlation between the measurement error $\widehat{Student}_{i,s} - Student_{i,s}$ and the student effect $Student_{i,s}$ is not statistically different from zero (the correlation has pvalue 0.402), indicating classical measurement error, and similarly for school effects.

References

Blundell, R. & Robin, J. M. (1999), 'Estimation in large and disaggregated demand systems: an estimator for conditionally linear systems', *Journal of Applied Econometrics* **14**(3), 209–32.

Davidson, R. & McKinnon, J. G. (1993), *Estimation and inference in econometrics*, Oxford University Press, Oxford, United Kingdom.

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

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

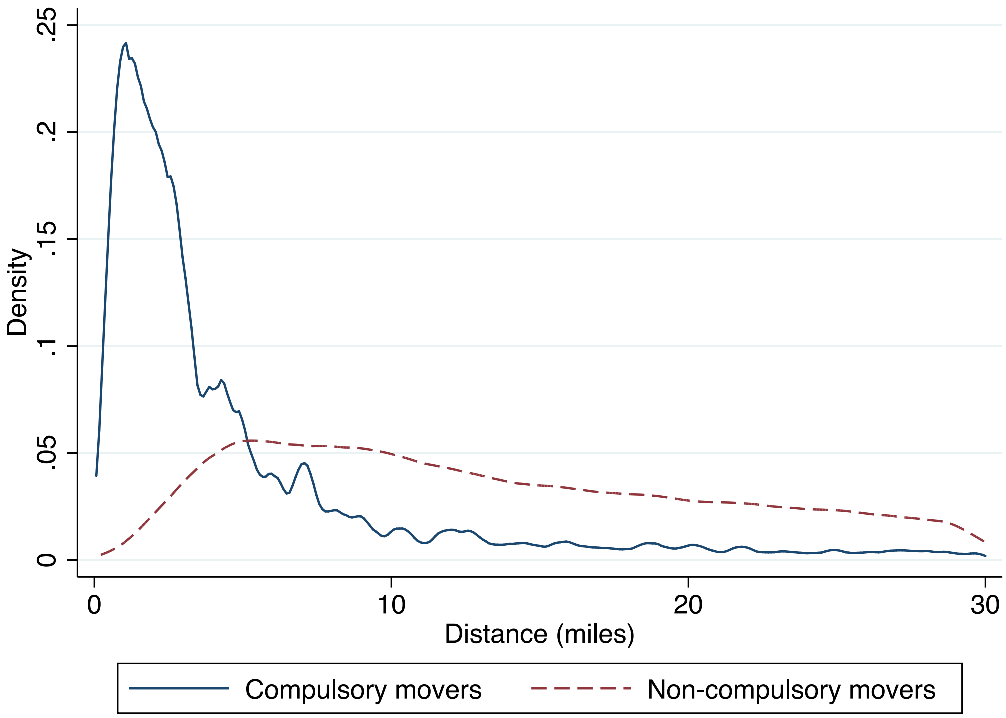
Levenshtein, V. I. (1966), 'Binary codes capable of correcting deletions, insertions, and reversals', *Soviet Physics Doklady* .

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

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.

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

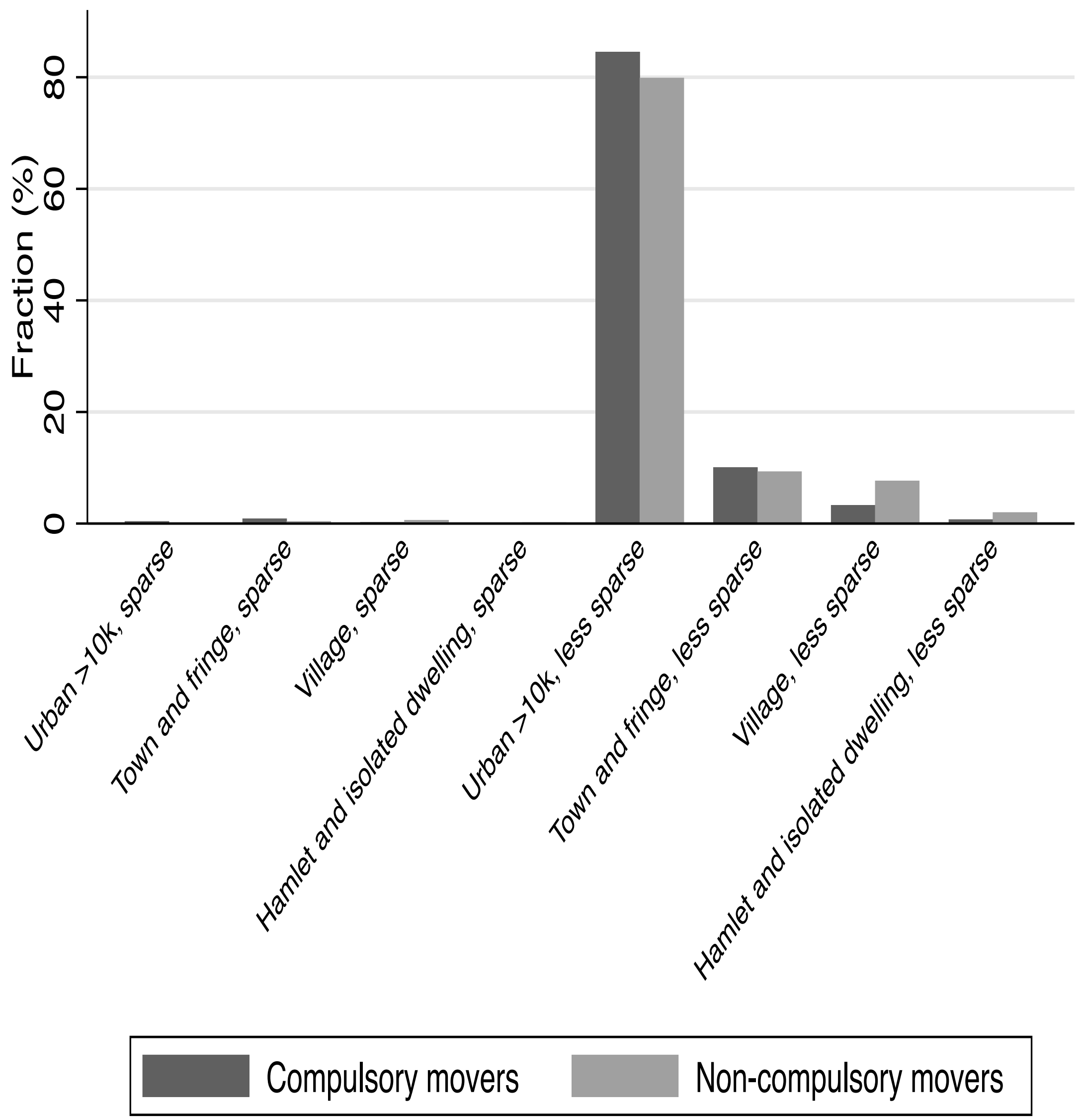
Figure 1: Distribution of Distances of Mobility: Compulsory and Non-Compulsory Movers



Notes: The distance between the origin and the destination school is taken by calculating the distance in miles between the centroid of the origin postcode and the centroid of the destination postcode.

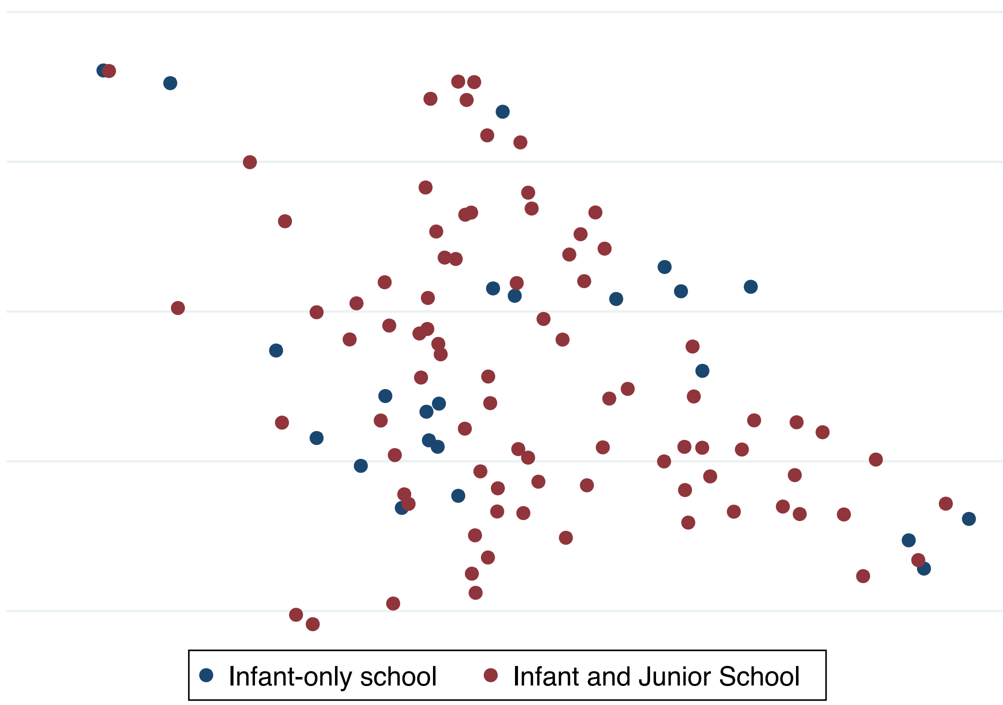
Source: Pupil Level Annual School Census, Register of educational establishments, Edubase, and National Pupil Database, 1999 to 2004.

Figure 2: Urban/Rural Status of Compulsory Mobility Infant Schools



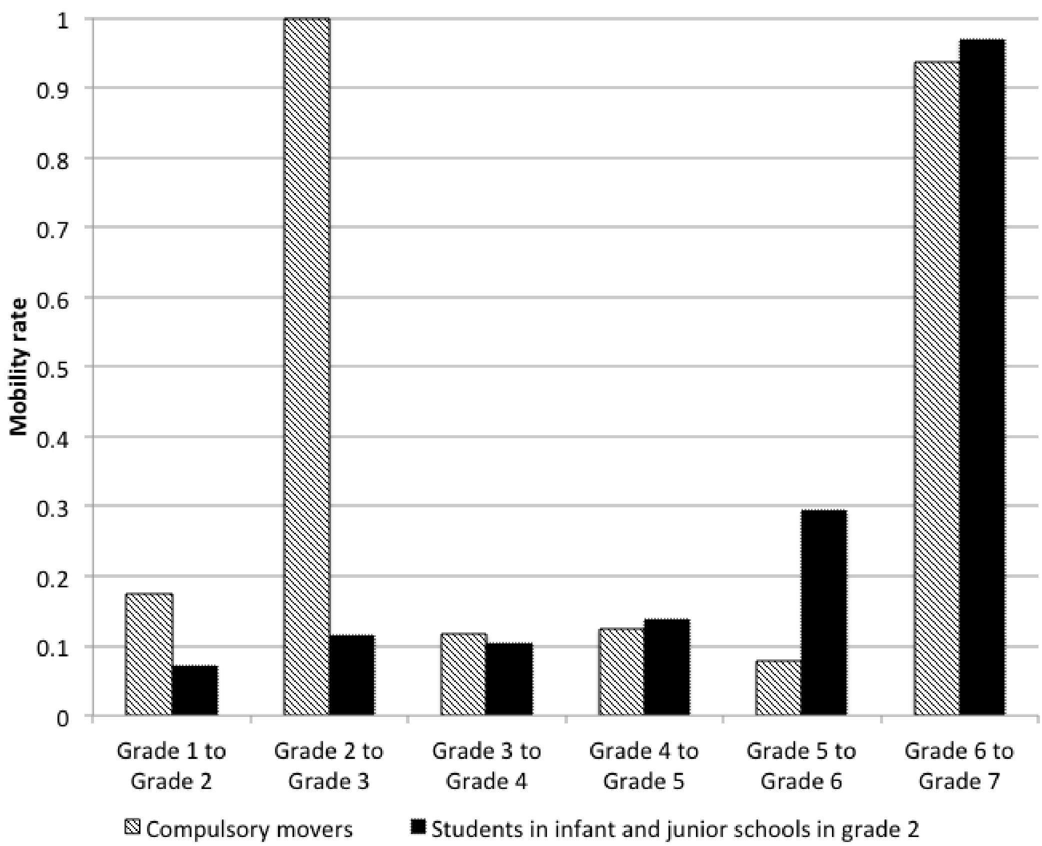
Note: Urban/rural status as defined by the Office for National Statistics, ONS Postcode Directory User Guide.

Figure 3: Spatial Distribution of Infant-Only and Infant-and-Junior Schools in Sheffield LEA



Location of Infant-only Schools (Red Dots), and of Infant and Junior Schools (Blue Dots) in Sheffield LEA (Code 373)
Source: Edubase, and Ordnance Survey Postcode Centroids.

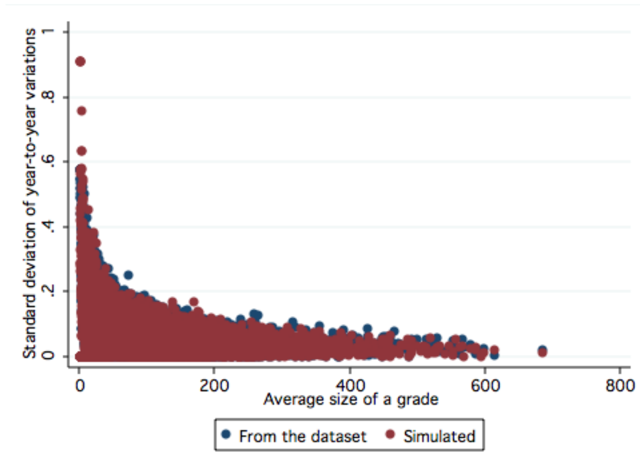
Figure 4: Grade-to-grade Mobility Rates



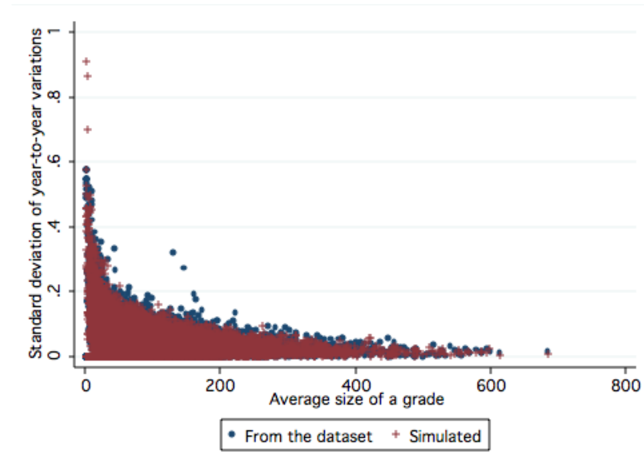
Source: Pupil Level Annual School Census, 1999 to 2004. Compulsory movers are students who started school in an infant-only school. The mobility rate between, say, grade 2 and grade 3 is the number of students of grade 2 who are in a different school in the grade 3 divided by the number of students in grade 2.

Figure 5: Year-to-year variations in grade composition - Realized vs Simulated Deviations - For Compulsory Movers

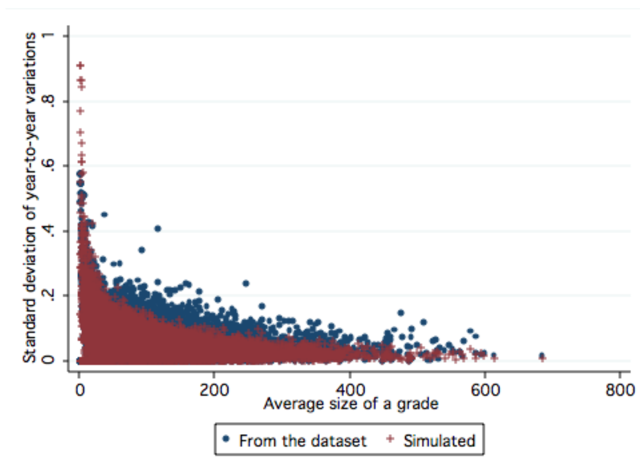
(a) Fraction Male



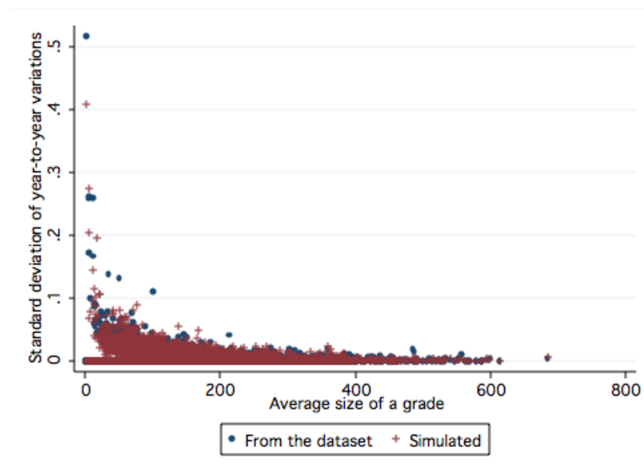
(b) Fraction of Free School Meals



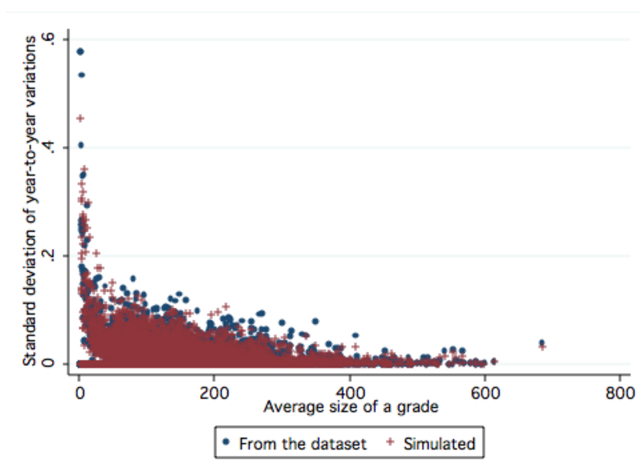
(c) Fraction of special-needs



(d) Fraction Chinese



(e) Fraction Indian



(f) Fraction Black African

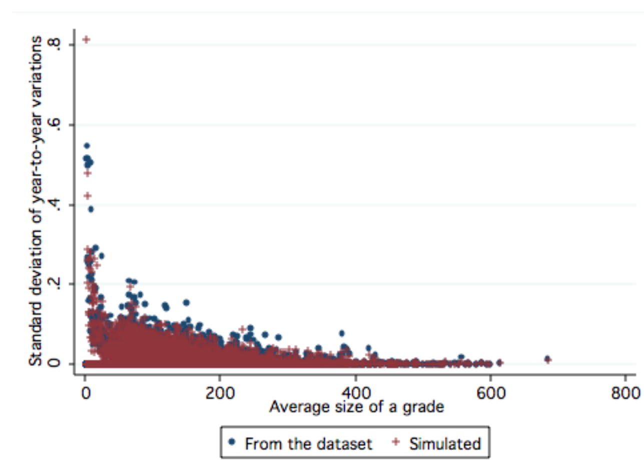
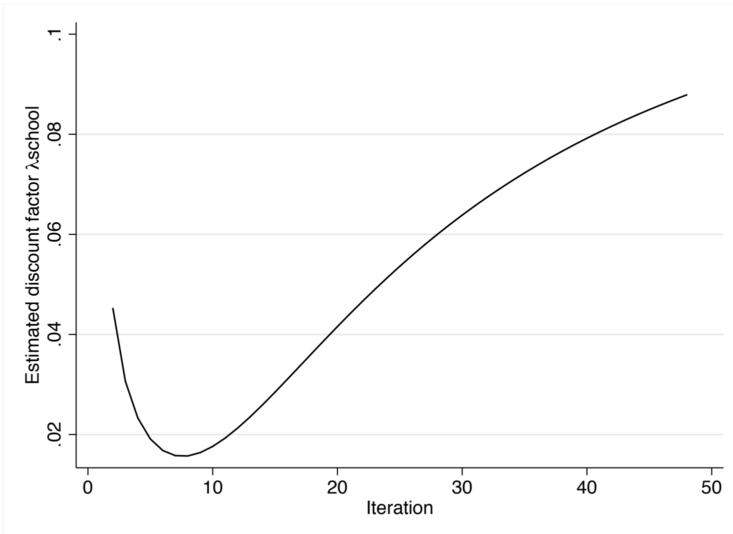


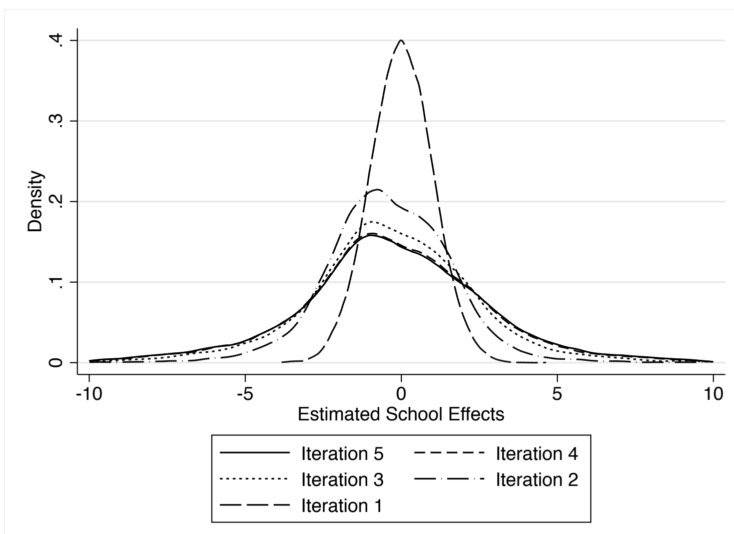
Figure 6: Convergence of the Estimation Process

(a) Convergence of School Discount Factor Estimates



This graph shows the estimate of the school discount factor at each iteration as described in Section 3.2.

(b) Convergence of School Effect Estimates

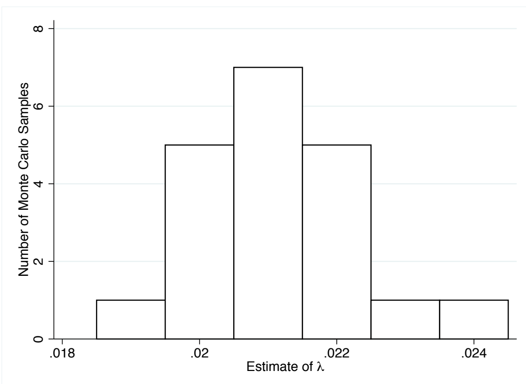


This graph shows the distribution of the school effect estimates for iterations 1–5. The estimation starts with initial guesses for school effects, taken from a standard normal distribution (dashed line). As the standard deviation of the final estimates is larger than 1, the distribution of school effects expands. The distribution of the estimates is also asymmetric. The distribution converges fast, with little change between iteration 4 and iteration 5.

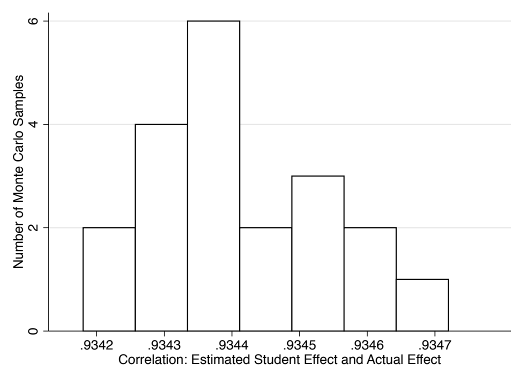
Figure 7: Properties of the Estimators of Student and School Effects and of the Discount Factor — Monte Carlo Analysis

This table presents the estimates of the discount factor, the school and student effects for estimations performed on each of the $S = 20$ Monte Carlo samples, as described in Section 4 of this Appendix. The Monte Carlo samples were generated using the results of the main estimation (Table 9 of the main paper). School effects were generated to have a normal distribution with standard deviation (s.d.) 2.018 (Column 1, Row 1 of Table 9). Student effects were generated to have a normal distribution with s.d. 8.720 (Column 1, Row 2 of Table 9). The discount factor λ used for generating the data was 0.024 (Column 2, Row 4 of Table 10).

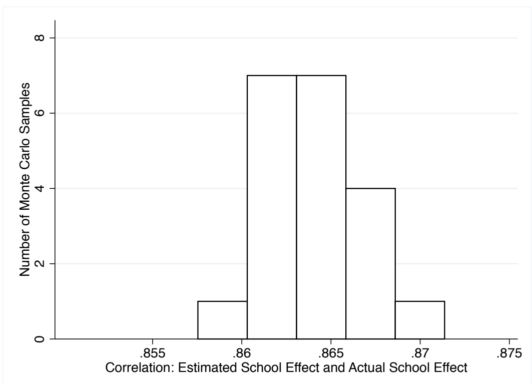
(a) Distribution of Estimated Discount Factors $\widehat{\lambda}_s$



(b) Distribution of corr($\widehat{Student}_{i,s}, Student_{i,s}$)



(c) Distribution of corr($\widehat{School}_{j,s}, School_{j,s}$)



	Mean	S.D.
$\hat{\lambda}_s$	0.021	0.001
Corr($\widehat{Student}_{i,s}, Student_{i,s}$)	0.934	0.001
Corr($\widehat{School}_{j,s}, School_{j,s}$)	0.864	0.003

Table 1: Distribution of Infant Schools, and Infant and Junior Schools

	Observations	%
Schools	21,360	100.00
Infant-only schools	4,479	21.70
Infant and junior schools	13,689	66.34
Junior-only schools	3,192	19.77
Grade-2 students	1,705,300	100.00
In infant-only schools	556,947	32.66
In infant and junior schools	1,148,353	67.34
Movers from grade 2 to grade 6	750,888	44.03
Noncompulsory movers	204,119	11.97
Grade-6 students	1,752,541	100.00
In junior-only schools	560,118	31.96
In infant and junior schools	1,192,423	68.04
Noncompulsory movers to junior-only schools	536,551	30.62
Noncompulsory movers to infant and junior schools	214,337	12.23

Table 2: How Similar are Grade 2 and Grade 6 Schools?

	All Compulsory Moves	Excluding London	Linked Junior School
Sharing a Head	0.34%	0.35%	0.41%
Same Postcode	29.28%	28.82%	39.84%
Same County	94.00%	94.00%	97.93%
Same Town	87.21%	87.03%	95.93%
Coming from most likely Infant school	73.49%	73.06%	100.00%
Observations	542,655	524,710	371,910

Source: Edubase, National Pupil Database, and Office of National Statistics' Postcode Directory.

Table 3: School Types

	Infant-only schools	Infant and Junior Schools
Community Schools	69.6%	53.9%
Voluntary Controlled Schools	11.2%	15.7%
Voluntary Aided Schools	9.7%	24.9%
Foundation Schools	1.6%	1.7%
Other School Types	7.9%	3.8%

Table 4: Non-Compulsory Movers' Position in the Test Score Distribution

	Dependent variable: Grade 2 Score – School-Grade Average
Noncompulsory mover	–2.166 **
	(0.026)
Observations	2,291,286
F statistic	1,705.4

Table 5: Estimation Results with a Discount Factor for School Effects and for Student Effects

	Dependent variable: Test Score	
Specification	(1)	(2)
S.d.(Student)	9.203	9.196
	(0.102)**	(0.102)**
S.d.(School)	2.592	2.679
	(0.996)**	(0.947)**
Discount factor for Students ($\widehat{\lambda^{student}}$)	0.024	
	(0.001)**	
Discount factor for Schools ($\widehat{\lambda^{school}}$)	0.087	
	(0.005)**	
Single discount factor ($\widehat{\lambda}$)		0.025
		(0.001)**
Observations	2,201,308	2,201,308
R Squared	0.80	0.81

Table 6: Impact of the Discount Factor λ on Estimated Variances

In the following regressions we fix the discount factor to a given value (either the value coming from our main regression 0.024, or higher values 0.03, 0.04, 0.05) and estimates the main regression with such a fixed discount factor.

	Dependent variable: Test Score			
$\lambda =$	0.024 (Estimated)	0.03	0.04	0.05
S.d.(Student)	9.203	9.196	8.554	8.987
	(0.102)**	(0.102)**	(0.082)**	(0.039)**
S.d.(School)	2.592	2.679	3.538	3.437
	(0.996)**	(0.947)**	(0.932)**	(0.019)**
Corr(Student,School)	0.073	0.083	0.322	0.303
	(0.001)**	(0.002)**	(0.002)**	(0.002)**
Observations	2,201,308	2,201,308	2,201,308	2,201,308
R Squared	0.81	0.80	0.77	0.76