

Valuing English Primary Schools

Steve Gibbons* and Stephen Machin**

July 2002

* Department of Geography and Environment and Centre for Economic Performance,
London School of Economics, Houghton Street, London, WC2A 2AE, Tel: +44 (0) 20 7955 6245,
email: s.gibbons@lse.ac.uk

** Department of Economics, University College London, Centre for the Economics
of Education and Centre for Economic Performance, London School of Economics

Abstract

This paper provides the first empirical evidence for the UK on the effect of primary school performance on property prices. We find that, on average, a one percentage point increase in the neighbourhood-proportion of children reaching the target grade pushes up neighbourhood property prices by 0.67%. At 2000 property prices, we calculate the social valuation of a sustained 1% improvement in primary school performance to be up to £90 per child per year. Our IV and semi-parametric approach avoids the problems of school quality endogeneity and omitted neighbourhood variables in typical property value models.

JEL Classification: I2.

Keywords: House prices; School performance.

Acknowledgements

We would like to thank the Editor and two anonymous referees, Evan Davis, Raquel Fernandez, Jonathan Haskel, Paul Johnson, Dennis Leech, Hamish Low, Andrew Oswald and participants in the CEPR Economics of Education conference in Bergen, Anna Vignoles and others at the CEP Annual Conference, the Royal Economic Society Conference, the Public Economics seminar and a UCL Applied Microeconomics seminar for a number of helpful comments and suggestions.

1. INTRODUCTION

Severe inequalities in the measured performance of English primary schools have parents clamouring to get their children into the best schools. The issue of school attendance and its link with house prices has rapidly become a key public policy issue, particularly since the introduction of class size constraints and tougher rationing in the 1990s. Indeed, many parents are prepared to move house to try to secure admission to a good school, and it seems that they are often prepared to pay a high premium on property price to do so. Stories of soaring house prices close to good schools are commonplace. For sure, moves to particular areas may buy more than just better education as good schools are typically located in neighbourhoods that are better in other ways (e.g. they tend to have lower crime rates, quieter neighbours, cleaner streets, and better local amenities). But it seems likely that some component of any premium paid for a re-location from a bad-school neighbourhood to a good-school neighbourhood may well be attributable to the price of an improvement in school quality.

This phenomenon is by now widely recognised in the US, and many attempts have been made to quantify it (see [1], [3] and [7] for an up-to-date list of references). For Britain, the issue has received much less attention in the academic arena, despite being discussed a great deal in the media, and amongst politicians and parents. A first aim of this paper is to fill this gap. We do so by extending upon the recent US literature that tries to estimate the connection between local school quality and house prices. One interesting difference in our analysis compared to the US is that, in many areas, school attendance zones are not well-defined. This means that the link between property location and school attendance is not as tight as in the US, and we must be careful to build this into the development of our empirical specifications.

Our analysis estimates the premium attracted by differentials in primary school quality in England from 1996 to 1999. We build upon hedonic property price models, but offer a different approach to try to circumvent the problems of collinearity, endogeneity and model selection problems that plague the traditional approach.¹ We use instrumental variables for primary school performance and exploit the co-variation in house prices and school performance within narrowly defined spatial groups, thereby removing spatial fixed effects, thereby reducing the need for inclusion of a large set of covariates. This methodological innovation allows hedonic methods to be applied when some property and neighbourhood characteristics are unknown.

The US literature is rich in efforts to evaluate school performance, or school characteristics, through house price models. But school characteristics are, in part, determined by neighbourhood socio-economic composition and hence by house prices. Also, in the US, school funding is determined by the local property tax base. Recent work uses a variety of strategies to try to correct for biases in property value models induced by this endogeneity. Researchers in the US compare properties in similar neighbourhoods on either side of school district boundaries [2], [1], use repeat sales when school district boundaries are re-drawn [3], or include an extensive set of neighbourhood characteristics as controls [7]. By contrast, the literature for the United Kingdom is very thin. It is hard to construct data-intensive estimators of the type favoured in the US, because geocoded data on individual property transactions is scarce, and because school admissions districts are not rigidly enforced. Instead, in this paper we use strategies that compare average school performance and house prices in proximate neighbourhoods and employ exogenous, permanent school characteristics as instruments.

¹ Such problems are often induced by the inclusion of an excess of highly correlated explanatory variables in the property characteristics matrix.

To our knowledge, our analysis is the first to value *primary* school performance in England. Other researchers have looked at the value house buyers attach to secondary schools. Rosenthal [18], for example, finds rather low elasticities of house prices with respect to school performance, in line with the US work of Judd and Watts [13].² Cheshire and Sheppard [5] estimate the value of location within specific school catchment areas in Reading and Darlington. Leech and Campos [14] do the same for Coventry. Neither study relates value to school performance measures and nor do any of these studies look at primary school performance³. We think this is important, as in the British context hedonic methods relying on spatial associations to link properties to schools may be poor for evaluating secondary school performance, except in few and far between special cases where catchment area boundaries are well defined and exclusive. In Britain teenagers are a fairly mobile group and can travel long distances to secondary schools. In contrast, primary age children typically attend schools which are within walking distance (at least in urban areas), and catchment areas can shrink down to just a few blocks for those in the highest demand.

Our interest in primary schools also has a sound theoretical basis. We would expect primary school performance to be the principal object of choice by parents seeking to improve the life chances of their offspring. For a start, there is evidence that attainments in the early years are positively correlated with later academic and economic success [8-10]. If gains made in the primary years reap rewards in terms of achievements at secondary school, then the payoff for the investment is higher if the investment is made early on in a child's life. What is more, in Britain investment in good primary education may be a pre-requisite of admission to selective secondary schools. Given

² Haurin and Brasington [12] however, find a 0.52% increase in price for a 1% increase in the proportion of 9th graders passing all sections of the 1990 proficiency test in Ohio.

³ Their later report [4] prices age 16 examination rates in Reading, Darlington and Nottingham, typically finding small valuations, and with only the coefficient on student performance for Reading being statistically significant. Even more recently [6], they find evidence of secondary and primary school effects for the Reading area.

the high fixed costs of moving house, a rational parent will make a once and for all locational choice when their first child enters the education system.

Our focus in this paper is therefore upon the associations between local house prices and primary school performance. The rest of the paper proceeds as follows. In Section 2 we discuss the usefulness of the hedonic framework in the current context. This section also discusses a critical issue underpinning our work, namely the extent to which location matters for admission to primary schools. Development of the empirical model is undertaken in Section 3 where we also discuss a number of relevant econometric issues. Section 4 moves on to discuss the data we use and, as the data comes from several sources, the matching procedures we adopt. Section 5 presents econometric estimates of our house price models. Section 6 concludes.

2. ENGLISH PRIMARY SCHOOLS AND THE HOUSING MARKET

2.1. *The hedonic approach and the basic empirical model*

We begin with the standard hedonic property value framework. This framework has been employed frequently in the environmental, land and urban economics literatures to price local environmental amenities (see Rosen [17], for the classic exposition, or Sheppard [19] for a modern survey). An estimate of the implicit price of school productivity is available from a simple regression of property prices on local school performance measures, assuming school admissions are restricted to local residents, and that school quality can be taken as exogenous.

Most of our analysis is based upon a Government Land Registry property price data set in which individual transactions are aggregated to provide an average of prices in four property-type categories (flat/maisonette, detached, semi-detached, or terraced) at *postcode sector* level. The postcode sector is a geographical postal area containing an average of 2500 households. For the hedonic price function, we specify a semi-log functional form, with an unknown function $g(\mathbf{c}_i, t)$

mapping geographical location given by the co-ordinate pair \mathbf{c}_i to house prices in each time period t . Our specification of the log-price of a house of type r in neighbourhood i at time t is then:

$$\ln P_{irt} = \alpha + \beta_x x_{it} + \gamma' \mathbf{z}_{irt} + g(\mathbf{c}_i, t) + u_{irt} \quad (1)$$

where P_{irt} is the mean property price of property type r in postcode sector i at time t . The vector \mathbf{z}_{irt} includes exogenous property and neighbourhood characteristics. The variable x_{it} is the effectiveness of the primary school in producing educated children. Specifically, we assume this is the probability that a child sent to the school reaches the age-11 target level in national assessment tests. These *Key Stage 2* tests are standard across all schools in the state sector. This success probability is unobserved, but we will proxy it by the proportion of children reaching this grade, as published in the DfEE school league tables. This is exactly what parents do when they inspect these tables as the basis for school choice.

So the basic structure results in an empirical model relating average house prices in a postcode sector to school performance in that sector, the mix of housing types, the number of social tenants and mean owner-occupied house size (to control for neighbourhood housing structures and exogenous socio-economic stratification), plus the spatial fixed effect function $g(\mathbf{c}_i, t)$ to capture unobserved neighbourhood factors. The econometric issues that relate to the latter are deferred to Section 3. But the assumption of exogenous school quality predicated in the standard hedonic pricing approach needs to be discussed here. In particular, whilst some determinants of school quality may well be exogenous to local community characteristics (like physical and institutional differences, and possibly teacher skills and head leadership styles) it seems reasonable that a substantial component of school performance is determined by the education, earnings and other characteristics of parents in the local community. This results in a potential endogeneity that in turn presents a problem for estimating the above model. A causal link from school inputs to pupil achievement and a causal link from local family incomes to pupil achievement are observationally

equivalent in terms of data on local incomes or house prices and pupil test success rates. Careful discussion of identification of the demand for school characteristics from both an economic and statistical viewpoint is therefore an important part of our analysis. The next sub-section considers the former, whilst the latter is addressed with other modeling issues in Section 3.

2.2. *Does location matter for primary school admission?*

Any attempt at valuation of schooling using the hedonic technique requires some method of linking property prices in a given area to the performance of schools available to residents in those properties. In specifying the above house price equation there is an implicit assumption that geographical proximity to a school is an important criterion for admission. However, whilst geographical proximity is one criterion, it is certainly not the only one. In England, Local Education Authorities (LEAs) are the administrative bodies responsible for the provision of state schooling. There are around 174 LEAs in England. They are largely funded by central government grants, so there is no link between local taxation and school funding, but these LEAs *do* operate their own systems of prioritising applications to a primary school. Legal precedent (the Rotherham Judgement, 1997) has determined that parental preference *must* be the LEA's first consideration. However, good primary schools are usually oversubscribed, so the admissions authority must employ some system for ranking applications in order of priority. Typically, for LEA administered schools, priority is assigned according to the following over-subscription criteria:

- i) those with siblings at the school;
- ii) those with special educational or medical needs;
- iii) those resident in a local "catchment" or "neighbourhood area";
- iv) children of those employed in the school;
- v) those ranked first by other geographical criteria (*e.g.* walking distance to the school).

The exact details and order vary across Local Education Authorities. For religious schools, some statement or evidence of religious affiliation is usually the first criterion that has to be met. Even then, parents must attend the local church regularly, or the school must be the nearest of the same denomination for children to be eligible to attend. Although catchment area boundary data might be helpful where catchment areas are well defined, we view close proximity to a primary school as a reasonable proxy for meeting geographical criteria for admission.⁴

What is clear is that choosing a location within the LEA and close to a school will maximise the chances of school admission for a family moving house for this purpose, whatever other criteria have been met. It will also minimise the costs of delivering children to school. But a more specific statement of our underlying assumption about the link between residential location and school access might be helpful at this stage. We maintain that location of residence defines the probabilities of admission to a number of schools, depending on the residence-to-school distance. Expected school quality at a specific residential location is thus an admission-probability weighted average of the performance of local schools.

Since we are interested in the price premium generated by those actively seeking school quality, and since catchment areas are non-exclusive, we argue that the relationship between mean neighbourhood property prices and mean neighbourhood school performance will provide just as much information as data based on individual schools and catchment areas. We use the association between property prices and primary school performance averaged at the postcode sector level. We also test this assumption by comparison with fixed effect models that rely on property price and school performance differences between adjacent postcode sectors separated by Local Education Authority boundaries.

⁴ The London Borough of Hackney publishes information on the maximum distance of residence for successful applicants. The median distance in 1999/2000 amongst 27 schools was 580m. Weighting by the difference between applications and intake gives a demand-adjusted median of 450m.

3. ECONOMETRIC METHODOLOGY

3.1. Estimation

Estimation of a full structural specification of the mapping of location to house prices requires data on local amenities, local housing characteristics, the proximity of neighbourhoods to transport services, local labour demand, environmental quality and other unknown local goods. The general function $g(\mathbf{c}_i, t)$ in equation (1) could then be replaced by a specific function of known variables. The difficulty of this approach is knowing exactly what to include in the hedonic price function. Researchers often include a large set of property characteristics in the regression, alongside an ad-hoc selection of socio-economic characteristics to proxy unobserved local characteristics. This is a poor strategy, since neighbourhood socio-economic composition is determined by sorting processes that are driven to a large extent by the housing market. *Neighbourhood composition is endogenous in a property value model.* As a result, parameter estimates obtained by OLS regressions will be biased.

What we do here instead is allow for general effects of location on property prices using non-parametric estimates of the impact of location on our model variables – in effect abstracting from the unobserved area-specific effects on prices, $g(\mathbf{c}_i, t)$. A simple and approximate solution would be include geographical area dummy variables to proxy unobserved area effects. However, a drawback of this area fixed-effect approach is its reliance on an arbitrary specification of the comparison neighbourhood group. In our case we could assume postcode *district* fixed effects since postcode sectors are nested within postcode districts. But there is no theoretical basis for believing that postcode districts are the appropriate controls, and using noisy measures of area fixed effects will lead to inconsistent estimates of the model parameters.

Instead we compute spatially weighted means of the variables in our model at each observation in the data, in which the nearest observations receive the highest weights. These

averages capture general, unobserved, area and amenity impacts on the housing market, *centred at the location of the unit of observation*. We then transform the data into deviations from these spatially weighted, means, and use the transformed variables in our regressions.⁵ But we still need to specify how rapidly the weights in our spatial averages decay as we move away in space from a given observation at location. We use a weighting function $m(\cdot | \mathbf{c}_i, b)$, with a bandwidth b that specifies how rapidly the weights decay with distance.⁶ Expressing the model in deviations from these estimated spatial averages, the regression model becomes:

$$[\ln P_{irt} - m(\ln P_{irt} | \mathbf{c}, b)] = \beta[x_{it} - m(x_{it} | \mathbf{c}, b)] + \gamma'[\mathbf{z}_{irt} - m(\mathbf{z}_{irt} | \mathbf{c}, b)] + \bar{\omega}_{irt} \quad (2)$$

Parameters β, γ and their variance covariance matrix can then be estimated by applying ordinary least squares to the transformed variables. This *smooth spatial effects* (SSE) estimator is an application of the semi-parametric partial linear model (see, for example [11],[16],[21])

The choice of bandwidth in (2) is important because we have no way of knowing, other than by casual empiricism, what geographical area comprises the correct reference group. We therefore experimented with a number of choices of b . A bandwidth of near zero is equivalent to a fixed effects estimator with postcode sector fixed effects. In this case, the relationship between school performance and house prices would be identified by changes over time alone. This is

⁵ In practice we also allow time effects via a separate non-parametric surface for each period, so we have: $g(\mathbf{c}_i, t) = \sum_t d_t \cdot g_t(\mathbf{c}_i)$ where d_t is a time dummy. This allows for differential growth in house prices across geographical space.

⁶ Formally, the spatially weighted mean is defined by the Nadaraya-Watson estimator

$$m(y | \mathbf{c}_i, b) = \frac{\sum_i^N y \times k\left\{(\mathbf{c} - \mathbf{c}_i)' \mathbf{B}^{-1}(\mathbf{c} - \mathbf{c}_i)\right\}}{\sum_i^N k\left\{(\mathbf{c} - \mathbf{c}_i)' \mathbf{B}^{-1}(\mathbf{c} - \mathbf{c}_i)\right\}}$$

where \mathbf{B} is a 2 x 2 bandwidth matrix (e.g. $b^2 \times \mathbf{I}_2$) and $k\{\cdot\}$ is a multivariate kernel. For the Gaussian kernel this is $k\{\mathbf{v}\} = 2\pi^{-1} \exp\{-0.5\mathbf{v}\}$. For details of multivariate kernels, see Silverman [20].

unsatisfactory since the most sought after schools tend to be near the top of the 0-100% performance range and show low performance growth relative to others. Moreover, in a short panel, this removes nearly all of the useful cross-sectional variation in prices and performance. We retain some cross-sectional variation by including neighbouring postcode sectors with non-zero weights. However, the land area and household density of postcode sectors is far from constant in our sample. Postcode sectors in rural locations are much larger than in urban locations, reflecting lower population densities in rural locations. To compensate for this, we vary the bandwidths in inverse proportion to the square root of the local household density as recorded in the 1991 census. Our main results use a bandwidth corresponding to approximately 3400 households. This bandwidth choice process is discussed in more detail in Appendix A.

3.2. The potential endogeneity of school performance

As we have already noted, school performance is likely to be related to local house prices through factors other than sorting by parents on good schools. For example, a relationship between neighbourhood property prices and school performance could arise through differences in property-tax based LEA funding. At a more localised level, differences may be generated by heterogeneity in family and community inputs. Wealthy parents purchase homes in neighbourhoods with bigger houses and with better amenities. Schools in these neighbourhoods perform better because the parents have more resources to devote to their children.

We can deal with the simultaneity between property prices and performance if we can obtain plausible instrumental variable(s) for school performance. We need characteristics of schools that influence performance but are unaffected by local property prices or neighbourhood socio-

economic status. For this, we draw on school characteristics available in the school performance tables.⁷ We use *historically determined* school-type characteristics as instruments.

Primary schools in England fall into three main categories: Community, Voluntary Aided or Controlled. Voluntary schools are almost always church schools (mainly Catholic and Church of England). Community schools are distinct as they are non-religious and that the Local Education Authority employs the staff and administers admissions procedures. Primary school age ranges do vary. Some take pupils before the compulsory school age of 5 years. Others are “Middle” or “Junior” schools that take pupils at age 8 or 9 who have attended separate first-stage primary schools. School performance measured in age-11 tests varies across all these groups, even though there is no selection on the basis of academic aptitude. We assume these differences reflect organisational, teaching quality or ‘ethos’ differences that impact on success rates in the national age-11 tests. Indeed, the better performance of church schools is widely recognised by parents. Differences between school age range categories may arise for a number of reasons: there may be benefits from continuity in education between nursery, infants and junior stages; children in neighbourhoods with nursery units may benefit from earlier introduction to school life; and schools accommodating a wider age-range are larger, offering potential economies of scale.

For all these reasons we believe that age-range and Community status provide good predictors of school performance. This is borne out in the first stage regressions we present (in Appendix C). These are fairly permanent characteristics so are good indicators of expected *long-run* performance. We also argue that age-range and Community status are unaffected by *local* spatial variation in current house prices, or by related variation in incomes or other socio-economic status. Indeed, only around 2% of primary schools in the sample opened or changed status during the

⁷ We tried qualified teacher-pupil ratios as an instrument at school level, but the underlying relationship between school performance does not work in the direction that we, and we assume parents, would expect. This suggests that more teachers are assigned to bad schools or disadvantaged areas, or that classes are smaller in schools that are less in demand. Either case invalidates its use as an instrument.

previous 10 years, so it is unlikely that current house prices have much of an influence on these characteristics. But factors at a broad geographical level may have an impact – Local Educational Authority policy, for example. So, it is important that we incorporate the instruments within our *smooth spatial effects* framework.

3.3. *Measurement error in school performance*

Another empirical issue concerns the assumption (implicit in equation (1)) that parents move house on the basis of a single-year measure of school performance. This is unlikely. The fixed costs associated with housing transactions and family relocations make moves each year in response to league table results highly inefficient. Instead, parents look to longer run indicators of school productivity. They may seek further information from school visits, school inspection reports, teaching staff, and by talking to other parents. Indeed, results published in the national tables are noisy measures of long-run school quality, and parents are more likely to seek out schools with proven track records of high performance, or those which exhibit characteristics that are, on average, associated with good long run performance. On this basis, least squares estimation of β in equation (2) will lead to a *downward* biased estimate of the value of persistent differences in school performance. This problem can be partly addressed by using time averages of prior school performance. Using permanent school characteristics as instruments for school quality potentially eliminates the problem.⁸

⁸ This is simply an application of the classical measurement error model. If permanent school performance is functionally dependent on observable neighbourhood characteristics, such as the proportion of social tenants in the catchment area, then inclusion of this neighbourhood characteristic in the OLS regression will further downward bias the estimate of β .

3.4. *An alternative methodology that is closer to US work: Identification from differencing across local authority boundaries*

If, as in US work, there existed well-defined catchment area boundaries in England, we might do better in assigning property prices to schools.⁹ Without this information, our estimates on the price-performance response based on matching mean postcode sector prices to mean postcode sector school performance may well be lower bounds, due to the classical errors-in-variables problem induced by mean school performance in a postcode sector being a noisy measure of the mean school performance of the schools available to residents of that postcode sector.

However, to check whether absence of catchment area information presents a serious challenge to the credibility of our estimates, and as an alternative methodology to compare with our IV results, we focus in some detail on a subset of postcode sectors in the London metropolitan area for which we infer catchment area boundaries. We do this on the assumption that any Local Education Authority (LEA) boundary is also a primary school catchment area boundary. We make this assumption on the grounds that none of the LEAs we contacted drew their catchment or neighbourhood area boundaries to cross LEA boundaries (even though applicants from outside the LEA are not legally excludable). This approach is similar to that taken in Black's Boston study [1] and Leech and Campos's study [14] of secondary schools in Coventry, though they use detailed information on catchment area boundaries and property level data for a single, small geographic area.

We use a London sample of postcode sectors that share a Local Education Authority boundary with at least one other. The empirical model is as in equation (1), but the function $g(\mathbf{c}_i, t)$ is replaced by dummy variables. These dummy variables indicate pairs of postcode sectors that are

⁹ One solution to improving the match between schools and property prices is to average individual school performance within a given radius of the centroid of each postcode sector. Our initial estimates based on this approach were similar to those obtained by simple postcode sector matching. However, this procedure introduces an additional bandwidth selection problem, so was abandoned.

adjacent, but on either side of an LEA boundary, plus LEA dummies and time dummies. We exploit differences in performance and prices *within* pairs of postcode sectors that straddle an LEA border, allowing for general differences between LEAs and general differences over time. This assumes similarity in the unobserved attributes of immediately adjacent postcode sectors. Adjacent postcode sectors that adjoin LEA boundaries, but are separated by some major physical obstacle are excluded, because the assumption that they form homogenous neighbourhoods is likely to be violated..¹⁰

4. DATA AND EMPIRICAL ESTIMATES

4.1. Data

In Britain, confidentiality considerations dictate that data on individual properties and transactions is almost impossible to obtain. The data we use for our main national estimates comes from four sources, which we splice together at postcode sector level. We have house price data from the Government Land Registry. Our primary school performance data comes from the public primary school performance tables, available from the Department of Education and Employment. Additional data (on the proportion in social housing, household density and postcode sector grid references) is derived from the 1991 Census for England and Wales. Fuller details are available in Appendix B. We end up with an unbalanced panel with up to four property types and property

¹⁰ This includes, for example, all postcode sectors separated by the river Thames downstream of Richmond. LEA dummies remove differences in local council tax, housing and education policy. Unitary Authorities responsible for other aspects of local government are geographically coincident with LEAs in London.

prices in each postcode sector in each year. Household density, grid-references and the proportion in social housing vary across postcode sectors but are constant across years in our data.

In all, 7444 postcode sectors and 2060 postcode districts are represented in our matched house-price and primary-school sample for the years 1996 to 1999. The mean number of households is 2900 per sector, and 12900 per district. There are primary schools in 5681 sectors and 1888 districts. Postcode sectors for which we have house prices, but no school information (because there is no school present here, or because there of no successful match between house-prices and schools) are assigned zero Key Stage 2 results. We include a dummy variable to indicate these in our regressions.

Figure 1 illustrates the geographical relationship between postcode sectors, districts and primary schools. It shows one postcode district – E3 in the East End of London. This district, being an inner city area, has a higher density of housing and primary schools than average, but it illustrates the main features used in the analysis. The housing density in postcode sector E3 4 is 6000/km², so a bandwidth choice of 3400 households in our smooth spatial fixed effect estimator corresponds to a radius of 0.42km. Very little weight is attached to sectors beyond 2.5 bandwidths, so the spatial group for a given postcode sector, assuming a bandwidth of 3400 households is, roughly speaking, those postcode sectors whose centre is captured within a 1 km radius from the centre of the observation postcode sector. Each grid represents 0.5 km on this map.

The symbols in Figure 1 represent the school types. Black circles are Community schools covering all primary years. The grey circle is a Community school that takes children from compulsory school age (5 years) only. White circles are schools that take older children only, or have a separate organisational units for older children. Black triangles are Voluntary Aided (Church of England and Catholic) primary schools. In this example of an inner city postcode district we can see a considerable variety of school types and age range within quite localised areas.

In addition to this postcode sector data, we have obtained a property price data set for the London area from a property valuation firm¹¹. This gives property level information on sales prices and property characteristics for an area covering around 800km² of the London region. We use this data set so that we can compare our main results with those obtained using more traditional methods.

4.2. *Descriptive statistics*

We present our results separately for three broad geographical areas. These areas correspond to grouped Standard Statistical Regions. The grouping scheme was chosen to illustrate any broad regional differences in property markets, whilst retaining a mix of rural, urban and metropolitan areas within each area. The groupings we use are:

South East and East: London, South East and East Anglia

The North: East Midlands, Yorkshire and Humberside, North, North West

West and South West: West Midlands, South West

The upper panel of Table 1 reports summary statistics on our postcode sector property price data set. House price growth from 1998 to 1999 appears lower than might be expected, considering the media attention on soaring house prices in the South East. The figures show a growth of just over 11% in postcode sector mean house prices in the East and South East between 1998 and 1999. Land registry published figures suggest a growth of over 15% in the South East. The anomaly is in part due to our use of annual averages, rather than the last quarter prices. Also, our sample includes only those properties with recorded postcodes. Comparison with other data sources reveals that this

¹¹ Ekins, the valuation arm of Woolwich plc.

sub-sample slightly under represents newer, higher price properties, though the problem is not serious.¹²

The lower panel of Table 1 shows some summary statistics for postcode sector school performance data. The performance measures are fairly similar in each regional group in each year, though The North is always marginally below the other areas. Attainment at Key Stage 2 has improved since the introduction of the performance tables in 1996, though there was little change between 1997 and 1998.

School characteristics are also given in Table 1 (those recorded in 1999). It shows the East and South East has slightly larger schools, the North has more schools with pre-school and reception years, and the West and South West has more voluntary aided or controlled schools and fewer junior schools. Variation in the age range across areas is attributable to LEA policy – in some LEAs, primary schools take children from compulsory school age only. In others, primary schools take children from age 4, or even earlier if a nursery is attached to the school.

The raw correlation between house prices and school performance is illustrated in Figure 2 which tests our log-linear specification semi-parametrically using a kernel regression of the deviations of 1999 log house prices from postcode *district* means on the deviation of average 1996-1998 school performance from postcode district means. The relationship shows an upward sloping relationship between house prices and Key Stage 2 performance, and looks comfortably linear for all regions.

¹² The Land Registry confirmed that many new properties are registered without postcodes, so are missing from the postcode sector level data. The under-representation of these groups in the dependent variable has the potential to downward bias our regression estimates. Given that the difference between the means in the postcode sample and the full sample is only around 5% we do not expect this to be a serious problem.

4.3. Results

Econometric results are reported in Table 2 for the three area configurations. The Table reports three specifications of log-linear regressions of property prices on school performance, with smooth spatial fixed effects, for each area. Other controls, included in all specifications, are the observation property type indicator, the postcode sector in social housing, and the mean number of rooms occupied by home-owning households. Local high school performance was never significant and had no impact on the primary school coefficient, so we have not included it in the regressions.

To reiterate, these estimates are regressions using the deviations from the local spatial group means, where these are estimated non-parametrically for each period. The minimum, mean and maximum bandwidths are shown in the Table notes. The distribution of household density on which the bandwidth is based is right skewed, so the median bandwidth is around 1km. To illustrate the way in which the estimator works, Figure 3 shows the estimated function $g(\mathbf{c}_i, t)$ that defines the smooth spatial fixed effects surface for the London region.

Columns (1), (4) and (7) are the basic OLS models including the schools' lagged Key Stage 2 performance. Columns (2), (5) and (8) contain the same specifications, but regress 1999 property prices on Key Stage 2 performance averaged over 1996-98. Columns (3), (6) and (9) show results that instrument school performance by school type and age range dummies. The Table reveals a common pattern of results across all three regional groupings. In all cases there is a positive statistically significant association between house prices and school performance and across regions the estimates of the implicit price of Key Stage 2 performance are very close. Indeed, calculation of the minimum distance estimate of the OLS parameter, based on the separate regional regressions shows it to be .240 and we do not reject equality of the parameters across regions (p -value = .117).

Columns (2), (5) and (8) use the time-averaged school performance measure. As we expected, the 3-year means provide better measures of long-run performance than the year-to-year

results, which give downward biased estimates of the implicit price of good schooling. The minimum distance estimate based on 3-year means rises to .331 and we do not reject equality across regions (p -value = .695). Based on the 3-year performance averages, a 10% increase in the mean Key Stage 2 performance in a postcode sector is associated with a 3.4% premium on property prices – equivalent to £3300 on an average-price property in England in the last quarter of 2000.

We now turn to our instrumental variables estimates. Let us first consider the suitability of our instruments. As already discussed in some detail above, we use admissions age-range and school-type indicators as instruments for Key Stage 2 performance. To show that our instruments are good predictors of school performance, we ran the underlying first stage prediction regressions. These are given in the Table in Appendix C, which reports within-postcode-district regressions of postcode sector mean Key Stage 2 results on the proportion of community schools, the proportion of schools in each of two age ranges, plus all the exogenous variables in the property price equation. The Table shows the estimated coefficients on the variables excluded from the property price equation, and the F-tests for their exclusion from the school performance equation. We see a broadly similar pattern across regions, and the F-statistics and t-statistics are always high. The proportion of children at Community schools achieving Level 4 is between 4.6% and 6.6% lower than those at church schools, and overall the instruments generate a ± 10 percentage point range in predicted performance.^{13 14}

¹³ We should note that Neal [15] finds that the advantage of a catholic secondary education in the US varies by geographical location. He finds that it is only in urban areas that catholic schooling offers clear benefits. If this were true in our sample, we would have to question the usefulness of community/voluntary status as an instrument, without area interaction effects. We investigated whether religious school advantage we detect here varies by area, but found no clear pattern. In London, the religious school advantage rises to around 12.3% (s.e. 1.04%), but it is similar or even higher in some, predominantly rural postcode areas, for example Peterborough (13.5%, s.e. 4.1%) or Carlisle (11.3% s.e. 5.8%), and lower in other urban areas such as Manchester (7.0% s.e. 3.5%). Unfortunately, Neal's analysis of the impact of catholic schooling sheds little light on the endogeneity of religious status with respect to neighbourhood status, as he disregards sorting by parents on school quality.

For all three regions, the IV estimates of the implicit price of Key Stage 2 performance are *higher* than the estimates using three-year means. Note though, that the Sargan test statistics imply that our instruments are uncorrelated with the residuals from the regression.¹⁵ That the IV-SSE estimates are higher is, in part, surprising. We might expect the IV estimator to remove residual catchment-area-to-school-performance effects that upward bias the least squares estimates of willingness to pay for school quality. The results suggest that these effects are *not* the principal source of bias. Instead, the use of year-to-year performance measures, and even 3-year averages, seriously attenuates the estimated impact of long-run school quality on property prices due to the noise components of the raw, transitory performance measures. We have further evidence that this what is happening. As it stands, the models in Columns (1), (4) and (7) have period-specific smooth spatial surfaces allowing area specific time trends. Instead we could use a single cross sectional surface and allow for general time effects with time dummies. This allows area-specific time series variation a greater role in determining the coefficient estimates. What we find though is that the estimated impact on prices is halved; the conditional variance in school performance increases with the transient noise components without corresponding increases in the covariance with prices.

Based on the IV estimates in columns (3), (7) and (11) the minimum distance estimate across regions is now .663, with equality across regions (p -value = 0.859).¹⁶ This 6.9% premium amounts to about £6600 for a 10 percentage point performance improvement, in England in 2000 prices.

¹⁴ The performance advantage of church schools does not appear to be related to selective admissions procedures by Voluntary Aided schools (who may conduct interviews to determine religious convictions). Voluntary Controlled schools, where the LEA administers admissions, also have better pass rates.

¹⁵ The exclusion restrictions are rejected if we do not control for social housing. This result is consistent with our observation that Community schools are more likely to be located near local authority housing estates. Apparently, the poorer performance of Community schools relative to religious schools is in part due to their catchment areas containing a higher proportion of local authority tenants.

¹⁶ The results are robust to a number of sensitivity checks. First, if we actually directly enter postcode district geographical effects, the estimates are quite close to those from our smoothed spatial effects approach. For the IV specification coefficients (standard errors) were as follows: East and South East .566 (.114); North .672 (.116); West and South West .509 (.146). Second, any argument that our results reflect possible selection on

4.4. *Comparison with private sector fees*

It is natural to ask how these estimated implicit prices compare with private sector fees. Private sector fees should provide an upper bound to what households are prepared to pay via the property market. In the private sector, the equivalent of primary schools are “preparatory” and “pre-prep” schools, covering the age-range from nursery to age 13. The total number of accredited nursery, pre-prep and prep schools in England on the Independent Schools Information Service (ISIS) database is 717, with nearly 40% of these in London and the South East. The mean national average reported by ISIS for 515 prep and pre-prep schools is £6324 in 2001. Assuming this is paid for eight years, and discounting at a rate of 5%, the present value of the costs of this investment amount to about £38,000. Unfortunately we have no information on age-11 performance for private schools. But, we can guess that parents paying for private primary education would expect nearly everyone at the school to reach the equivalent Level 4 in Key Stage 2, implying a 25 percentage point advantage over the mean state sector primary school in 1999. In terms of property prices in the last quarter of 2000, this performance advantage would be worth roughly £17400 nationally, or around £35300 in London, where property prices are highest. This suggests that for families with only one child in London, the capitalised costs of state-sector primary education (over and above the unavoidable direct costs of taxation) are close to the costs of a private-sector primary education. For families with, or intending to have, more than one child of primary school age, and for those in other areas of the country, moving house is probably a cheaper option. Even in London, the state-

academic ability by voluntary aided schools is refuted by results that include controls for special educational needs (used to proxy for academic abilities). The parameters are not statistically different from those in the models without special needs. Third, the results are not sensitive to controlling for more demographic characteristics (school ethnic mix, proportion getting free school meals), nor to varying bandwidth (e.g. to 1700 or 5100 households). Fourth, we have also tested the effectiveness of our smooth spatial effects in removing spatial correlation in the data. Calculating Moran’s I coefficients for the South and South East regions on the residuals from a model without spatial give a value of $I = .3299$ (z -statistic = 15.576). This implies quite strong and significant spatial correlation, but using the residuals from the smooth spatial effects models, this falls to $I = -.0012$ (z -statistic = -0.854).

sector is cheaper in annual terms: the mortgage costs associated with a 25 percentage point improvement amount to around £2500 each year.

4.5. *London results and comparison estimates*

The results from the cross-LEA boundary model outlined in section 3.4 are presented in the top panel of Table 3. Given the detailed map-work and analysis required here, we focus on Greater London. We also compute the SSE estimates for all postcode sectors within the same geographical boundary. Of course, the samples are not the same. The cross-LEA method relies on a selected subsample of all the postcodes in the London area.

Remember that our comparable baseline parameter estimates were 0.240, 0.331 and 0.663 for the annual, 3-year average and IV-SSE models respectively. The comparable cross-LEA models for London are 0.107, 0.353 and 0.747. There is some deviation in the point estimates, but the standard errors are large for the cross London boundary model. The attenuated coefficient when we use transient single-year performance measures is, we suggest, due to the poor signal to noise ratio. But nothing here suggests that our imprecise definition of catchment areas in the SSE framework leads us to *underestimate* the response of property prices in our baseline results. In fact, *we do not reject equality with the baseline SSE estimates*: the minimum distance 3-year average estimate is 0.365, *p*-value 0.998; the minimum distance IV estimate is 0.641, *p*-value 0.880. Neither do we reject equality with the SSE estimates computed for London alone – which are close to the baseline estimates across all regions. Both methods are consistent, but the SSE approach is efficient relative to the cross-LEA boundary estimator.

We infer from this that precise mapping of the areas served by schools is relatively unimportant. The relationship between performance and property prices can be analysed in terms of spatial surfaces measuring expected price and performance. What we've done in our models is

estimate points on these surfaces defined by the postcode sector centroids, using postcode sector mean property prices and postcode sector mean school performance.

For further comparison, we estimate models using our property-level data for the London region. We match school performance to properties using spatially-weighted average performance of the nearest 8 primary schools. The lower panel of Table shows these results. Other controls in the regression are as described in the table notes. The first row in the lower panel provides parameter estimates using models with no spatial controls other than a quadratic in distance to the central business district, distances to various amenities and Local Education Authority dummy variables. Our primary school coefficients are somewhat *higher* than those in our main results. In the next two rows we apply our SSE techniques to the micro data to remove unobserved spatial impacts on prices. The resulting coefficients on primary schools, are very close to our results using postcode sector mean data with few property controls. From this, it seems that we do not need micro data to measure the extent of capitalisation of a neighbourhood-based good like school performance. But allowing for general neighbourhood effects on prices *is* important, even when the regression includes a fairly broad range of neighbourhood controls.

5. CONCLUSIONS

In this paper we ask how much parents in England are prepared to pay to get their children into better schools by moving house. We use postcode sector level data on house prices and primary school performance in England to estimate the magnitude of the association between primary school quality and local house prices. We eliminate the effects of catchment area wealth on pupils' achievements by concentrating on the effects within narrow geographical areas, and by instrumenting measured pupil achievements by characteristics of the school itself. Our best estimates imply a premium on postcode sector house prices of 6.9% for each 10% improvement in the proportion of children reaching the target level in Key Stage 2 tests at age 11. This translates

into capitalised values, at regional mean property prices, between £4500 in the North and £13500 in Greater London (all at 2000 property prices). Interestingly, our estimates of the primary school effect are of the same order as those obtained for suburbs of Boston, Massachusetts [1]. She finds that a 10% increase in primary school mean test scores attracts a 5% property price premium. Using a similar estimation technique and London data, we find a 3.6% property premium. Our lowest estimates based on OLS, within-area estimators put the figure at around 2.4% for a 10% school improvement.

The clear message that emerges is that households value improvements in primary school performance. Importantly, this valuation relates to differences attributable to exogenous schooling inputs, not simply to exogenous neighbourhood status. *From this we infer that school inputs must matter.* Lack of suitable data means we cannot empirically address the question of which inputs matter most. This is the appropriate question for policymakers who want a policy lever to apply, and more research on this question using detailed data on children and schools is vital. Nonetheless our findings are important as they show parents to strongly value better school performance.

Extrapolating from our results, we can say that any technology which raises primary school standards by one percentage point has a social valuation per household equivalent to between 0.33% and 0.67% of the local mean property price. For a national population of 21 million households, and mean property price of £96700 at the end of 2000, this implies a maximum aggregate social valuation of £13,600 million, or about £1.09 million per school. This means that a sustained one percentage point improvement in primary school performance scores has an annual value of somewhere around £90 for each child of primary school age or younger.¹⁷

¹⁷ These calculations assumes 12500 Key Stage 2 primary schools, 7.6 million children age 11 and under and a 5% social discount rate.

REFERENCES

1. Black, S.E., Do Better Schools Matter? Parental Valuation of Elementary Education, *Quarterly Journal of Economics*, **114**, p. 578-599 (1999).
2. Bogart, W.T. and B.A. Cromwell, How Much is a Good School District Worth?, *National Tax Journal*, **50**, p. 215-232 (1997).
3. Bogart, W.T. and B.A. Cromwell, How much is a Neighbourhood School Worth?, *Journal of Urban Economics*, **47**, p. 280-305 (2000).
4. Cheshire, P., I. Marlee, and S. Sheppard, Development of a Microsimulation Model for Analysing the Effect of the Planning System on Housing Choices, Department of Geography, London School of Economics, unpublished manuscript, (1999).
5. Cheshire, P. and S. Sheppard, On the Price of Land and the Value of Amenities, *Economica*, **62**, p. 247-67 (1995).
6. Cheshire, P. and S. Sheppard, Capitalising the Value of Free Schools: The Impact of Land Supply Constraints, Paper prepared for Conference on the Analysis of Urban Land Markets and the Impact of Land Market Regulation: Lincoln Institute of Land Policy, Cambridge MA, (2002).
7. Downes, T.A. and J.E. Zabel, The Impact of School Characteristics on House Prices: Chicago 1987-1991, *Journal of Urban Economics*, **52**, p.1-25 (2002).
8. Feinstein, L., The Relative Economic Importance of Academic, Psychological and Behavioural Attributes Developed in Childhood, Centre for Economic Performance Discussion Paper, DP0443, London School of Economics, (2000).
9. Feinstein, L. and J. Symons, Attainment in Secondary School, *Oxford Economic Papers*, **51**, p. 300-321 (1999).
10. Gregg, P. and S. Machin, Child Development and Success or Failure in the Youth Labour Market, in *Youth Employment and Joblessness in Advanced Countries*, (R.B. Freeman and D. Blanchflower, Ed.), University of Chicago Press, (2000).
11. Hardle, W., "Applied Nonparametric Regression", Cambridge, Cambridge University Press (1990).
12. Haurin, D.R. and D. Brasington, School Quality and Real House Prices, *Journal of Housing Economics*, **5**, p. 351-368 (1996).
13. Judd, G. and J. Watts, Schools and Housing Values, *Land Economics*, **57**, p. 460-470 (1981).
14. Leech, D. and E. Campos, Is Comprehensive Education Really Free? A Case Study of the Effects of Secondary School Admissions on House Prices, University of Warwick Economic Research Paper, 581, (2001).
15. Neal, D., The Effect of Catholic Secondary Schooling on Educational Achievement, *Journal of Labour Economics*, **15**, p. 109-123 (1997).
16. Robinson, P.M., Root-N-Consistent Semiparametric Regression, *Econometrica*, **56**, p. 931-954 (1988).

17. Rosen, S., Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *The Journal of Political Economy*, **82**, p. 34-55 (1974).
18. Rosenthal, L., The Value of Secondary School Quality in England, University of Keele, Department of Economics Working Papers, 2000/06, (2000).
19. Sheppard, S., Hedonic Analysis of Housing Markets, in Handbook of Urban and Regional Economics, (P. Cheshire, Ed.), Elsevier Science, (1999).
20. Silverman, B.W., "Density Estimation for Statistics and Data Analysis", London, Chapman Hall (1986).
21. Stock, J.H., Non-Parametric Policy Analysis: An Application to Estimating Hazardous Waste Cleanup Benefits, in Non-parametric and Semi-parametric Methods in Econometrics and Statistics, (W. Barnett, J. Powell, and G. Tauchen, Ed.), Cambridge University Press, (1991).

TABLE 1
Descriptive Statistics

	East and South East	North	West and South West
A. Property prices (£)			
1996 sector mean	86591	52286	63221
1997 sector mean	98701	55879	69852
1998 sector mean	112303	61045	78600
1999 sector mean	125757	63921	85040
Detached house mean	165532	94006	110870
Semi-detached mean	106521	53254	65343
Terraced mean	96183	40375	54178
Flat/maisonette mean	69881	40607	43510
Mean sales volume	140	95	114
Number of postcode sectors	2900	2998	1554
B: School characteristics			
1996 key stage 2, level 4 proportion	0.598	0.584	0.592
1997 key stage 2, level 4 proportion	0.667	0.656	0.657
1998 key stage 2, level 4 proportion	0.665	0.648	0.657
1999 key stage 2, level 4 proportion	0.745	0.733	0.739
Proportion community school	0.643	0.635	0.609
Proportion of schools with pre-school/reception	0.229	0.346	0.200
Proportion of schools with infants	0.467	0.444	0.572
School roll	310.1	282.3	285.9
Number of age 11 pupils present	53.6	45.5	48.2
Number of schools in postcode sector	1.9	2.2	2.0
Number of postcode sectors	2289	2242	1164

Property prices are matched to lagged Key Stage 2 results in estimation sample
Price means are means of postcode sector means (unweighted by sales volume)
1999 Key Stage 2 results reported for completeness (not used in estimation sample)
Key Stage 2 assessment tests are sat in Spring and results are released in Autumn.

TABLE 2:
Property price response to primary school performance: postcode-sector data, by region

	South East and East			The North			The West and South West					
	1-year (1)	3-year (2)	IV-SSE (3)	Mean (4)	3-year (5)	IV-SSE (6)	Mean (7)	3-year (8)	IV-SSE (9)	Mean		
Proportion reaching target grade, age-11	0.203 (0.023)	0.289 (0.048)	0.711 (0.134)	0.648 ¹	0.289 (0.027)	0.372 (0.055)	0.688 (0.144)	0.639 ¹	0.243 (0.036)	0.357 (0.075)	0.510 (0.197)	0.639 ¹
No primary school	0.183 (0.017)	0.244 (0.035)	0.524 (0.090)	0.165	0.238 (0.021)	0.305 (0.044)	0.502 (0.100)	0.178	0.230 (0.027)	0.304 (0.056)	0.404 (0.134)	0.215
Detached	0.479 (0.003)	0.481 (0.004)	0.481 (0.004)	0.229	0.524 (0.003)	0.539 (0.005)	0.539 (0.005)	0.261	0.500 (0.003)	0.501 (0.005)	0.501 (0.005)	0.264
Terraced	-0.185 (0.002)	-0.187 (0.003)	-0.187 (0.003)	0.266	-0.277 (0.003)	-0.277 (0.004)	-0.277 (0.004)	0.311	-0.188 (0.003)	-0.188 (0.005)	-0.188 (0.005)	0.275
Flat/Maisonette	-0.588 (0.003)	-0.582 (0.005)	-0.582 (0.005)	0.246	-0.399 (0.006)	-0.387 (0.010)	-0.389 (0.010)	0.116	-0.474 (0.005)	-0.480 (0.009)	-0.480 (0.009)	0.183
Mean rooms in owner-occupied housing	0.196 (0.011)	0.195 (0.013)	0.187 (0.011)	5.534	0.184 (0.014)	0.169 (0.026)	0.170 (0.026)	5.555	0.193 (0.016)	0.196 (0.029)	0.189 (0.028)	5.702
Proportion in social housing	-0.442 (0.054)	-0.439 (0.080)	-0.257 (0.092)	0.191	-0.436 (0.045)	-0.488 (0.086)	-0.397 (0.048)	0.223	-0.319 (0.056)	-0.273 (0.090)	-0.222 (0.112)	0.183
Within R ²	0.849	0.850	0.855	0.806	0.806	0.791	0.800	0.838	0.838	0.835	0.835	0.835
Overall R ²	0.934	0.934	0.934	0.887	0.887	0.881	0.880	0.904	0.904	0.904	0.904	0.904
Over-identification test p-value	-	-	0.196	-	-	-	0.120	-	-	-	0.141	-
Sample size	29606	9972	9972	25431	8544	8544	8544	15605	5274	5274	5274	5274
Years	1997-99	1999	1999	1997-99	1999	1999	1999	1997-99	1999	1999	1999	1999

Note: ¹ Mean is 1996-1998 average school performance

Dependent variable is log of property price

Instruments in columns 3, 6, 9 are Community school status and school admissions age-range dummies, weighted optimally according to clustered error structure

Min, mean, max bandwidth: .23 km, 1.43 km, 9.25 km in South East and East; 0.35km, 1.72km, 26.07 km in The North; 35 km, 1.75 km, 26.46 km in the West and South West

The sample centroids are, by region: (52835,18350), (41279, 41852), (36027, 20062)

Sample comprises unbalanced panel with up to 4 house-price observation types (Detached, Semi-Detached, Terraced, Flat) for each postcode sector, observed for 1997, 1998 and 1999 (1999 only in columns 2, 5, 8)

School performance measures are the proportion in the school obtaining Key Stage 2 at Level 4 and above. Key Stage 2 results are average of maths, reading and science scores. Standard errors corrected for clustering on postcode sectors (in parentheses).

TABLE 3:
Comparison Of Coefficients From Cross-Local Authority Boundary Effects, SSE
Estimator and Property-level Micro Data For Greater London Area

	1-Year	3-Yr Mean	IV
	(1)	(2)	(3)
<i>X-LEA model, postcode sector data</i>	0.107 (0.050)	0.353 (0.134)	0.747 (0.304)
Sample size ¹	4012	1351	1351
<i>SSE model on London area, postcode sector data</i>	0.195 (0.038)	0.231 (0.068)	0.596 (0.171)
Sample size ²	11051	5274	5274
<i>P-value of X-LEA & SSE parameter equality:</i>			
Hausman test (assuming same sample)	0.007	0.291	0.548
Assuming independent samples	0.161	0.417	0.665
<i>Traditional property-level micro data model</i>			
Performance in nearest primary schools	0.454 (0.055)	0.664 (0.068)	1.041 (0.232)
<i>Property-level micro data with spatial effects</i>			
Performance in nearest primary schools	0.174 (0.079)	0.263 (0.092)	0.756 (0.321)
Sample size ³	8067	8067	8067

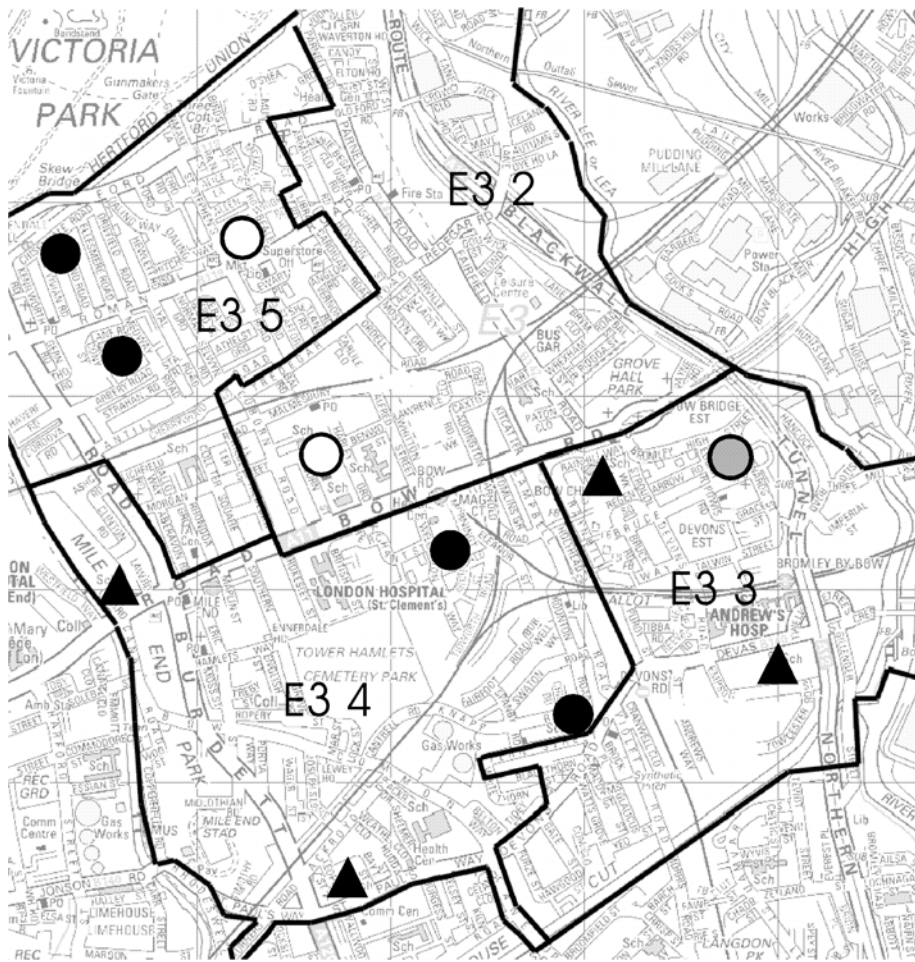
Samples:

¹ Postcode sectors adjoining local authority boundaries, on the Geoplan Greater London postcode sector map

² All postcode sectors within the boundary in ¹

³ Micro data covering 800km² of the London region in 2001, provided by Ekins surveyors

Controls in micro-data models are property style dummies (12), number of rooms, number of floors, garage, year built, floor area, distance to: CBD, CBD², local town centres, green spaces, underground stations, police stations; plus mean neighbourhood house size, proportion in social housing, housing density, population density, density of purpose built flats, incidents of criminal damage per km²,



All
 No pre-school
 Separate unit for primary school
 Church school

FIG 1: Example Postcode District, Postcode Sectors and Geographical Distribution of School Types 7+

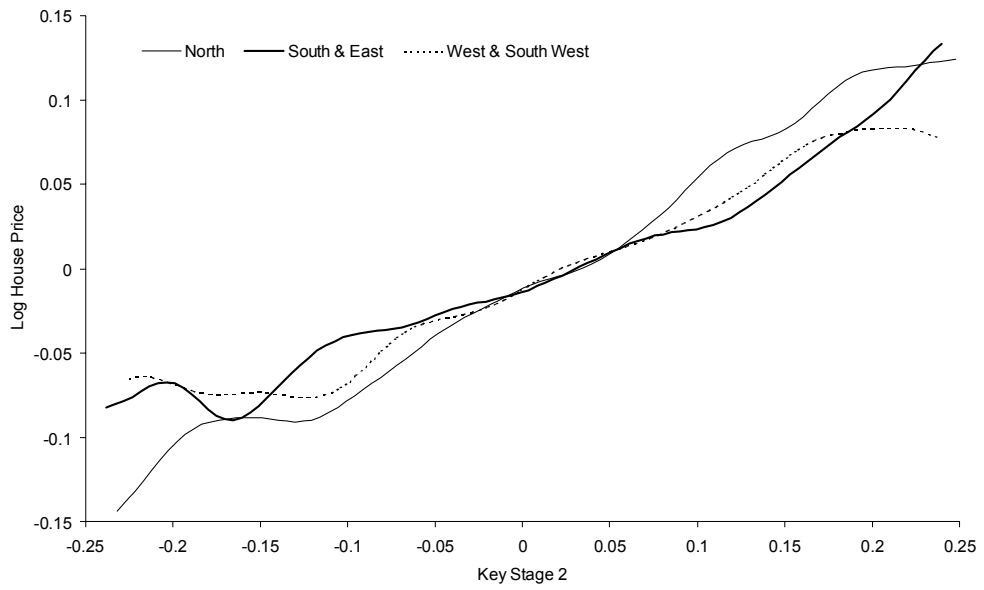


FIG 2: Relationship Between Log House Prices in 1999 and Mean 1996-1998 Primary School Performance – Deviations From Postcode Sector Means

Figure illustrates kernel regression of within-postcode-district variation in log house prices on within-postcode-district Key Stage 2 performance. Bandwidths in accordance with Silvermann’s rule of thumb (0.02 for North and South and East, 0.025 for West and South West)

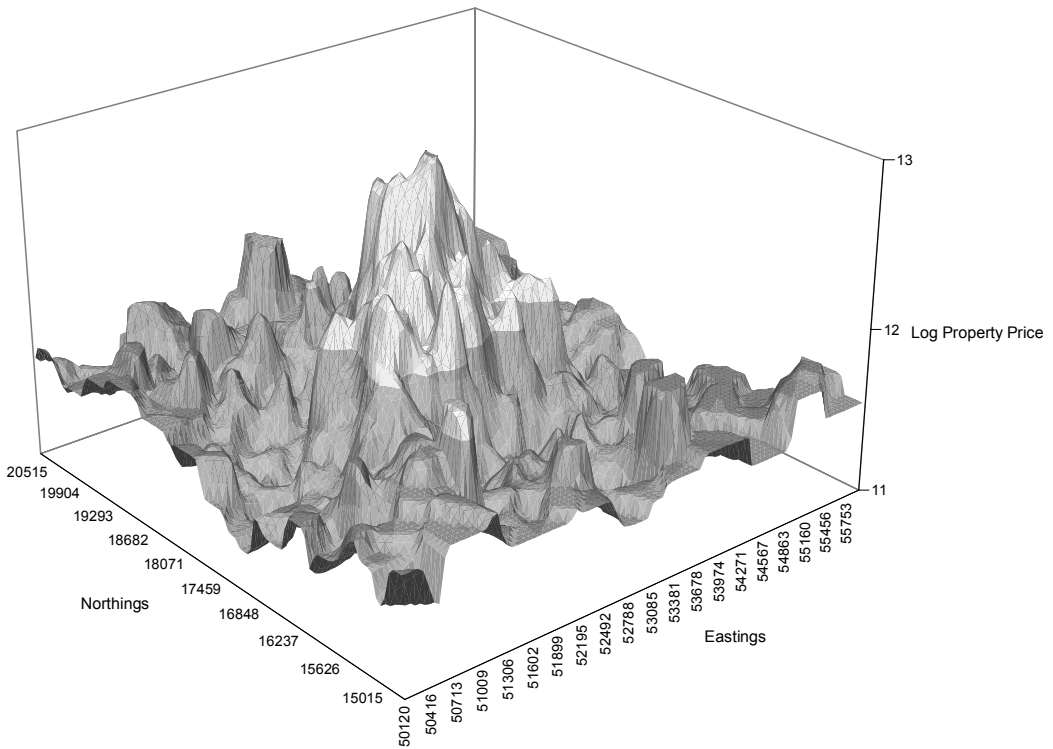


FIG 3: Example House Price-Location Surface For London, From Smooth Spatial Effect Model

APPENDIX A: BANDWIDTH CHOICE

The choice of bandwidth for the kernel in our SSE estimator is important, since the appropriate comparison needs to encompass more than one postcode sector, without averaging over too broad an area. Since the appropriate area (in terms of geographical distance) depends on local household density, we need to take this into account. Postcode district HG4, just north of Harrogate, has an area of roughly 270 km² and postcode sector household densities that range from 20 to 1300 per km². By contrast, E3 around Bow and Tower Hamlets in east London has an area of roughly 4.5 km² and household densities between 6000 and 6800 per km². No fixed bandwidth can accommodate this variation: a suitable bandwidth choice at HG4 will average over much of the London area if applied to a sector in E3. A bandwidth suitable for E3 if used in HG4 will apply virtually no weight to any observations beyond the postcode sector. Consequently, we weight the neighbourhood bandwidth using data on household density matched in from the 1991 Census.

Fixing the number of households n in a circular spatial group of radius b , gives us a bandwidth weighting rule dependent on housing density h :

$$b = \sqrt{\frac{n}{\pi h}} \quad (0-1)$$

In order to choose a bandwidth regulator h , it is useful to know how our postcode sectors relate to primary school catchment areas. This is made more difficult by the fact that we could obtain almost no information on this from our enquiries to LEAs, as catchment areas are rarely precisely defined, and vary with demand. Data on addresses of pupils actually attending is considered confidential, and is usually held only by the schools themselves. We were unable to obtain this. From our primary school performance data, the total number on the school role of primary schools recorded in the 1999 performance tables is 3.77 million and the total number of imputed households in our CACI data is 20.1 million. The ratio of households to primary school children is 5.33, implying an average catchment area of around 1400 households, which is about half a postcode sector. This is consistent with the fact that there are, on average, two primary schools per postcode sector in the school performance tables. Choosing bandwidths corresponding to groups of roughly one, two and three postcode sectors and adjusting downwards by 40% to compensate for the use of a Gaussian kernel (which applies non-zero weights to observations outside the bandwidth window), suggests corresponding household groups of roughly 1700, 3400 and 5100 respectively. The main results we present use bandwidths corresponding to 3400 households, but comparisons are made with other bandwidth choices.

APPENDIX B: DETAILS OF THE DATA SOURCES

Data on individual housing transactions is unavailable in Britain, so we have used the best available alternative: house prices aggregated to postcode sector level. This data set covers the whole of England and Wales, and is available from 1995 to 2000. It contains mean house prices and total sales volumes at postcode sector for each postcode sector, where annual sales numbered 3 or more. Properties sold for under £10,000 and over £1,000,000 are excluded. This amounted to only 0.5% of all property sales in 1999.

In the UK, postcodes contain up to seven alphanumeric characters, and contain four hierarchical components. The first two alphabetic characters define the Postcode Area, the broadest postal zone. Examples are N, EX and YO representing North London, Exeter and York. Within Postcode Areas, the next level down is the Postcode District. This is defined by a single or two-digit number following the Postcode Area. Examples are N6, EX24, and YO10. A single letter further subdivides some postcode districts in central London. Below this, we have Postcode Sectors. This is the unit of observation in our house price data set.

The school performance tables for England compiled by the Department for Education and Employment (DfEE) provide the basis for our school performance measures. We have the 1999 primary and secondary school tables, which include background information on the schools in 1999, plus the performance measures for years 1996 to 1999 inclusive. We also have the original data for the years 1996-1998 which includes the school background characteristics for these years. The primary performance measures are proportion of pupils reaching Level 4 (the target level of attainment) in the Key Stage 2 standard assessment tests administered at age 11. We average the measures for Maths, Reading and English tests. We average these school performance measures and characteristics across schools within each postcode sector to provide a postcode sector level primary school performance indicator and characteristics. Here, we experimented with simple means and school-size weighted means, but opted for the former on the basis that weighting by school size conflates school size and school performance issues. In practice, the choice of scheme made little difference to our results.

We match postcode sector house prices to the postcode sector school performance and characteristics from the school data set, giving us up to four house prices (detached, semi-detached, terraced, flat/maisonette) for each postcode sector in each year.

Additional variables at postcode sector level are derived from the 1991 Census, and from the 1998 postcode to Census enumeration district directory, which relates 1998 postcodes to corresponding 1991 census area codes. These sources give us geographical data including the national grid reference, the proportion of social housing, and the density of households per kilometre-squared. Although postcode-sector aggregated census data is available, the postcodes relate to the 1991 postcode geography, so the census variables we use are means of the values in the enumeration districts which are wholly or partly included in a given postcode sector. Grid references are taken as the mid point between the maximum and minimum in each direction.

No population bases are available at the postcode sector level later than 1991, though we have household figures in our CACI data set on household incomes. The mean number of household addresses per postcode sector in the CACI data in 1999 is 2800. In the UK there are 26 million postal addresses, 2901 districts and 9624 sectors, so a crude average is 9000 per postcode district, 2700 per sector. These numbers change over time with changes in the postcode geography. In 1996, the number of households in England was 20.2 million, implying an average of around 9600 households per postcode district, and around 2560 in each postcode sector.

APPENDIX C: 1ST STAGE EQUATIONS FOR IV ESTIMATES

TABLE C1

Coefficients on instruments in first stage IV equation

	South East and East	North	West & South West
Community school – LEA appointed governors and admissions	-0.066 (0.006)	-0.046 (0.007)	-0.064 (0.008)
School has pre-school/ reception years	-0.034 (0.009)	-0.014 (0.009)	-0.032 (0.016)
School has infants and junior years	0.022 (0.006)	0.069 (0.008)	0.042 (0.009)
F-test of instruments	F(3, 843) = 72.0 P = 0.0000	F(3,713) = 67.6 P = 0.0000	F(3,502) = 37.5 P = 0.0000

Predicted Key Stage 2 performance, from identifying instruments (all areas):

s.d. = 0.02
max = 0.102
min = -0.097

Models include property type dummies, proportion of local social housing, and are estimated within postcode-district-year groups

Standard errors (and F-tests) corrected for clustering on postcode districts.

Results shown for illustration only; estimation of main models does not use 2-stage least squares method.