Sorting and school quality

Lars Nesheim

Centre for Microdata Methods and Practice

(CEMMAP)

University College London

and

the Institute for Fiscal Studies

October 2005

Abstract

This paper develops a theoretical and empirical model of household sorting and school quality. It characterizes an equilibrium in which heterogeneous households sort into different locations based on school quality and housing prices. It also develops the empirical implications of the model and estimates the impact of school quality on educational outcomes controlling for sorting.

The housing price, the distribution of school quality, and the distributions of agents within and across schools are determined in equilibrium.

The paper shows that the structural parameters are identified.

The empirical results indicate that sorting is important and that much of the correlation between school quality and test scores is not causal.

1 Introduction

The impacts of school quality on student outcomes like test scores are difficult to measure. Due to sorting, the set of students at a school is typically not a random sample of the population of students. For example, suppose students at schools in location A perform better on a test than those in location B. How does one know whether the schools in location A are better schools or whether the students in location A are better students? If parents of "good" students believe that their children will benefit more from attending schools in location A, they will try to gain access to schools in A. Generally, they will try harder to gain access to A than parents of "bad" students. If access to "high quality" schools is rationed, at least in part, through residential location, as it certainly is in the United States and in the United Kingdom (and as it is likely to be in other countries as well), then those households who are willing to pay the most for high quality schools, will buy houses near high quality schools. The result will be an equilibrium in which house prices in A are higher than B and students in location A are systematically different from students in location B. The students in A will be the children of parents who are willing to pay more for school quality. In some circumstances they will also be "better" students on average. In this case, it is difficult to determine how much of the high performance at A is due to the school and how much is due to the "quality" of the students.

In this paper, I adopt a direct approach to overcome this problem. I develop a model of parental location decisions that yields predictions for who chooses high and low quality schools in equilibrium. The model is sufficiently flexible that the households who choose high quality schools are heterogeneous. Some of them are rich and have children with relatively low ability. Some of them are poor and have relatively high ability children. Some have moderate income and moderate ability children but have a strong preference for or belief in the benefits of high quality schools. Observed background characteristics like education and income and unobserved characteristics like tastes and abilty all play roles. I use the model is used to measure the importance of each factor in the sorting of students into schools and to estimate how much of student performance at high quality schools is due to the quality of the schools and how much is due to the quality of the students. Finally, the model yields predictions for the equilibrium differences in housing prices across locations with different quality schools.

An important feature of the model is that the quality of schools in a location is endogenous. Quality is determined by the average test scores of the students at a school. This average in turn depends on the set of residents in the location. I discuss this feature in more detail in section 2.

While some of the details of solving the model are quite involved, the basic mechanism is simple and intuitive. To gain access to high quality schools, parents must purchase a house in a high quality school district. Since everyone prefers high quality school districts, housing prices in those locations are higher. Some people are more willing and more able to pay the higher housing prices. The people who end up in the high quality school districts are those who are willing to pay the most. In equilibrium, each location has a set of residents, a price of housing, and a level of school quality. People in each location are heterogeneous, they may have different amounts of education or income or they may have children with different levels of ability, but, all have the same willingness to pay for school quality.

The model yields many interesting insights. Rich patterns of sorting are possible. For example, the relationship between school quality and average income or average student ability is determined in equilibrium and need not be linear or monotonic. Similarly, statistics like the within school variance of parental income vary with school quality and are determined in equilibrium. Impacts of educational reforms that affect education production function parameter or parameters describing the distribution of income can be simulated.

The model also provides a system of empirical equations that can be estimated. These equations describe parental location decisions and student test score outcomes. The parameter estimates describe the importance of school quality, observable parental characteristics (like education and income), and unobservable characteristics (like student ability and parental preferences) in both location decisions and in producing test score outcomes.

I estimate the parameters of these equations using data from the US National Educational Longitudinal Survey (NELS) of 1988. The results show that accounting for sorting is important. If sorting is ignored and elasticities are estimated using national data, the elasticities of student reading test scores and student math test scores with respect to school quality are 0.671 and 0.664 respectively. (See Tables 2 and 5.) After controlling for sorting, these elasticities fall to 0.0498 and 0.266 (See Tables 3 and 6.). The evidence also indicates that there are important differences in parameters across groups defined by race or ethnicity and sex.

I structure the rest of the paper as follows. Sections 2 and 2.1 describe the details of the economic model and define equilibrium. Section 2.2 discusses a special case with a closed form solution. Section 2.3 then discusses computation of equilibrium in the general case. Section 3 analyses the empirical properties of the model and Section 4 describes the data and the empirical results. Section 5 concludes.

2 Model

An economy is populated with households each represented by a pair (x, d) where $x \in R^5$ is a vector of continuous variables and $d \in \{1, 2, ..., n_d\}$ is a discrete valued variable. The variables (x, d) measure characteristics or traits of households that affect educational outcomes and the utility obtained from schools of various qualities. The elements of x include $x_1 = \ln s$ where s is parental schooling attainment or education, $x_2 = \ln\left[\frac{y_H(y-y_L)}{y_H-y_L}\right]$ where $y \in (y_L, y_H)$ is parental income[1], $x_3 = \ln a$ where a is the ability of the household's child in school, and $x_4 = \ln\beta$ where β measures how much the household values a child's test score relative to consumption. Households with high values of β value their child's education relatively highly. The variable x_5 is a shock to the child's educational process that is realized after the parent makes their choice of location.

[1]This normalization is used for numerical stability. When $y_L = 0$ and $y_H = \infty$, $x_2 = \ln y$.

Each value of d indicates membership in a group defined by race or ethnicity and sex. White males are indicated by $d = 1$. White females by $d = 2$, etc. There are n_d groups. In the empirical section of the paper, I will use $n_d = 8$ since males and females in four racial/ethnic groups are represented in the NELS data.

Households with characteristics (x, d) maximize utility by choosing a location based on school quality $q \in R_+$ and house prices $p(q) \in R_+$. If they choose a location of quality q, then they must pay a housing price premium $p(q)$. Locations with higher quality schools have higher prices. Households value their own consumption, $y - p(q)$, and their child's educational outcome as measured by a test score $t \in R_+$. They value school quality because it affects their child's test score. Treating the price function $p(q)$ and the distribution of school qualities as given, they solve

$$\max_q \left\{ \frac{1 - e^{-\gamma(y-p(q))}}{\gamma} + \beta E(t|q, x, d) \right\} \tag{2.1}$$

where $E(t|q, x, d)$ is the expected test score of a child from a household with background characteristics (x, d) that attends a school of quality q. Parents don't know the precise outcome their child will obtain but choose schools based on their expectations as to the impact of the schools on their child's outcome. They trade off consumption and school quality. All else equal, households with higher levels of income will choose higher quality schools. However, heterogeneity in the expected returns from school quality and in preferences will lead households with equal income levels to make heterogeneous choices about school quality. The preference parameter β measures parents' marginal valuation of their child's expected test score. This parameter is a composite measure of several aspects of parental heterogeneity in preferences and beliefs. Its value reflects parental altruism toward children, an abstract preference for education, and beliefs about and/or discounting of future monetary returns accruing from test scores.

The quality of schools in a location is determined by the residents of the location. If

$X \times D$ is the set of all households in the economy and $F(q) \sqsubseteq X \times D$ is the set residing in locations with quality q, the quality of schools in q is determined by the **school quality production** function

$$q = E\left(t\,|(x,d) \in F(q)\right). \tag{2.2}$$

The quality of the schools in neighborhood q equals the average test score of the students attending schools in q.

This is a pure peer effect economy in which school quality is determined solely by peer effects. One could specify other school quality production functions. For example, this model could be generalised to explicitly study how peer effects and spending on teachers and other inputs jointly impact school quality. In this paper, I focus on a given measure of school quality and answer the questions how do students sort across schools and how much of the observed correlation between this measure of school quality and student outcomes is due to school quality and how much is due to sorting. As I discuss in section 4 below, there is a strong correlation between student test score outcomes and this measure of school quality. Ordinary least squares results from regressions of student test scores on school level average test scores and individual level background characteristics are reported in Tables 2 and 5. In this paper I answer the question, how much of this correlation is causal.

Regardless of the specification of the school quality production function, questions about sorting must be answered. This paper develops a methodology to answer these questions and presents empirical evidence as to the importance of the sorting process. The question of exactly what mechanism produces school quality is left for future research.

I assume that the population distribution of the household characteristics conditional on d is log-normal, i.e. $x\,|d \sim N(\mu(d), \Sigma(d))$.[2] The distribution of d is given by the probability mass function π. $\pi(d)$ is the probability that a household is in group d. The joint distribution of (x,d) describes the relative frequencies of different types of households in the population.

[2]The normality assumption is not required except to derive the specific results in Section 2.2 and to calculate the specific empirical results in Section 4. The assumption is made for ease of exposition. Nesheim (2005) studies models with more general distributions.

The distribution is one of the main determinants of the shape of the equilibrium because it determines both the relative demand for different locations in the economy as well as the relative supply of different types of households. In the analysis below, I discuss how the equilibrium outcomes depend on the parameters of the joint distribution.

School quality is important to households because it interacts with family background characteristics to produce children's test scores. I assume that a child's test score, which I denote by t, is a function of four inputs. It is a function of school quality q, the log of parental schooling attainment $x_1 = \ln s$, the log of the child's ability $x_3 = \ln a$, and the shock x_5. The **test score production** function is

$$t = q^{\eta_1} e^{\eta_0(d)+\eta_2 x_1 + x_3 + x_5} \tag{2.3}$$

When a parent chooses their residential location, they know the quality of the neighborhood q, their own education e^{x_1}, their child's racial and sexual group d, and their child's ability e^{x_3}. These variables affect their residential decision. They do not know the value of the variable x_5 but do know the distribution from which it is drawn. Income, y and the preference parameter $\beta = e^{x_4}$ do not directly affect the educational production function but indirectly affect educational outcomes through their effect on location choice. Subgroup membership affects test score outcomes explicitly through the term $\eta_0(d)$ and implicitly through the correlations of x on d. This will be made more clear in the empirical section.

To complete the specification of the model, I must describe an equilibrium. An equilibrium is defined by two objects, the price function $p(q)$ and the function $F(q)$ describing the set of households living in location q. This pair is an equilibrium if both are consistent with all households maximizing utility and if condition (2.2) is satisfied. This is formalized in the following definition.

Definition 1 ***Sorting Equilibrium****. Let $p(q)$ be the price of housing in location q and let $F(q)$ be the set of households living in location q. The pair, (p, F) is a sorting equilibrium if* *the following two conditions are satisfied:*

$$\begin{aligned}&\textit{1) } q \textit{ is optimal for household } (x,d) \textit{ if and only if } (x,d)\in F(q)\\&\textit{2) } q = E\left(t\,|(x,d)\in F(q)\right).\end{aligned}$$

The first condition ensures that every household chooses a location that maximizes utility. The second condition is a rational expectations equilibrium condition. It ensures that the average of the test scores in location q actually equals q. Parents expect the quality of schools in a location to equal q, and, in equilibrium, the quality in that location is q.

In the next section, I analyse the solution of the household's problem and the equilibrium of this model.

2.1 Equilibrium

To compute an equilibrium, I first solve the household's problem given an arbitrary price function $p(q)$. Then, I use this solution to characterize the correspondence $F(q)$ describing the set of people living in each location. For an arbitrary price function $p(q)$, this correspondence $F(q)$ need not satisfy the equilibrium conditions. However, the unique equilibrium pair (p,F) will satisfy these conditions. Finally, I will use the results to analyse the properties of the equilibrium F and p.

Each household chooses a location to maximise utility. After substitituting the conditional expectation of (2.3) into (2.1), the household problem is

$$\max_q\left\{\frac{1-e^{-\gamma(y-p(q))}}{\gamma}+A(d)\,q^{\eta_1}e^{\eta_2x_1+x_3+x_4}\right\}$$

where $A(d)=e^{\eta_0(d)+\mu_5(d)+0.5\sigma_{55}(d)}$. Differentiating the objective function, the first-order condition for this problem is

$$-e^{-\gamma(y-p(q))}\cdot p_q(q)+A(d)\,\eta_1q^{\eta_1-1}e^{\eta_2x_1+x_3+x_4}=0. \tag{2.4}$$

Assume (2.4) has a unique solution satisfying the second order conditions for maximization for all households. I will check that this is true in equilibrium.

For a single household (x, d), (2.4) describes the optimal choice of q. The optimal choice equates the marginal cost of increasing q to the marginal willingness to pay. More importantly from the equilibrium perspective, for each location q, this equation describes the set of people who choose location q. This set is the set of all households who satisfy (2.4) at the point q.

After rearranging and taking logarithms, (2.4) is equivalent to

$$v(q) = \ln \eta_1 + \ln A(d) + \eta_2 x_1 + x_3 + x_4 + \gamma y$$

where

$$v(q) = \gamma p(q) + \ln [p_q(q)] + (1 - \eta_1) \ln (q). \tag{2.5}$$

The expression $v(q)$ is an index that summarizes the net **value** of a location of quality q. This index depends on q and on the price function $p(q)$.

Define

$$w(x, d) = \ln \eta_1 + \ln A(d) + \eta_2 x_1 + x_3 + x_4 + \gamma y. \tag{2.6}$$

The expression $w(x, d)$ is an index that summarizes the **willingness to pay** for school quality a household. The set of households that choose location q is

$$F(q) = \{(x, d) \in X \times D : \ w(x, d) = v(q)\}. \tag{2.7}$$

Households that have high willingness to pay $w(x, d)$ will choose locations with high values $v(q)$. Households that have low willingness to pay $w(x, d)$ will choose locations with low values of $v(q)$. Note that $w(x, d)$ is not the **marginal willingness to pay** for quality, but rather, is a willingness to pay index useful for ranking households. All households with the same willingness to pay rank will choose the same location. Individual households may have high values of $w(x,d)$ because the parents are well educated (x_1 is high), because they have able children (x_3 is high), because they prefer high quality schools (x_4 is high), or because they are rich (x_2 is high).

The set $F(q)$ defined in (2.7) depends implicitly on the price function $p(q)$ through the index $v(q)$. For an arbitrary p, this function F will not be an equilibrium if the equilibrium condition (2.2) is not satisfied. Using (2.7) and imposing condition (2.2), an equilibrium price function must satisfy

$$q = E\left(t \,|w(x,d) = v(q)\right). \tag{2.8}$$

where $w(x,d)$ is defined in equation (2.6) and $v(q)$ is defined in equation (2.5). This equation is an ordinary differential equation in p. The unique solution is an equilibrium price function for this economy. In general, an approximate solution must be computed numerically.

2.2 Special case

A simple solution is obtained in the special case in which $\gamma = 0$. In this case,

$$\begin{aligned} w(x,d) &= \ln \eta_1 + \ln A(d) + \eta_2 x_1 + x_3 + x_4 \\ &= w_0(d) + {w_1}' x \end{aligned}$$

where $w_0 = \ln \eta_1 + \ln A(d)$ and $w_1 = (\eta_2, 0, 1, 1, 0)'$ and

$$v(q) = \ln\left[p_q(q)\right] + (1-\eta_1)\ln(q). \tag{2.9}$$

As a result, equation (2.8) simplifies and I have the following theorem.

Theorem 2.1 *If $\eta_1 \in (0,1)$, $k_1(d) > 0$ for all d, and $\gamma = 0$, the slope of the equilibrium* *price is uniquely determined and satisfies the differential equation*

$$q^{1-\eta_1} = \sum_d \left[\pi_d k_0 (d)\, q^{k_1(d)(1-\eta_1)} p_q^{k_1(d)}\right]$$
$$p(0) = p_0$$

where p_0 *is the price of a quality zero location,*

$$\ln k_0 (d) = (1 - k_1)(w_0 + (\eta_2\mu_1 + \mu_3)) - k_1 (\ln(\eta_1) + \mu_4) + 0.5 \left(\delta' \Sigma \delta - \frac{\left(\delta' \Sigma w_1\right)^2}{w_1' \Sigma w_1} \right),$$

$$k_1 (d) = \frac{\delta' \Sigma w_1}{w_1' \Sigma w_1},$$

and

$$\delta = (\eta_2, 0, 1, 0, 0)'.$$

Furthermore, if $n_d = 1$,

$$p(q) = p_0 + \frac{k_0^{\frac{-1}{k_1}} q^{\eta_1 + \frac{1-\eta_1}{k_1}}}{\eta_1 + \frac{1-\eta_1}{k_1}}.$$

Proof. See Appendix A. ■

If reservation utility is zero, then $p_0 = 0$. When $n_d = 1$, the price premium is a constant elasticity function of q with elasticity equal to $\eta_1 + \frac{1-\eta_1}{k_1}$. The variable η_1 is the elasticity of test score with respect to school quality. The variable k_1 is the regression coefficient from the regression $E\left(\delta' x \,|w_1' x, d\right)$. It measures the correlation of parental expectations of log test scores with the willingness to pay index $w(x, d)$. When k_1 grows large, parental expectations of log test scores will be the primary determinant of willingness to pay and the elasticity of housing prices with respect to quality will approach η_1. When k_1 grows small, variation in the parental preference parameter x_4 will be the primary determinant of willingness to pay and the elasticity of housing prices with respect to quality will grow large. In the former case, small price differentials across locations are required to separate households into groups with different levels of average expected test scores. In the latter, large price differentials are required. The exact price differential is given in the theorem. The key parameters that affect the price differential are the parameters of the test score production function and the parameters describing the distribution of consumer characteristics.

Further results in the simple case follow immediately using the same logic.

Theorem 2.2 *If* $\gamma = 0$, $n_d = 1$, *and* $k_1 > 0$, *the distribution of* x *conditional on* q *satisfies*

$$(x\,|q) \sim N\left[\mu + \left(\frac{(1-\eta_1)\ln q - k_2}{\delta'\Sigma w_1}\right)\Sigma w_1, \Sigma - \frac{\Sigma w_1 w_1'\Sigma}{w_1'\Sigma w_1}\right]$$

where

$$k_2 = \ln A + \delta'\mu + 0.5\left(\delta'\Sigma\delta - \frac{(\delta'\Sigma w_1)^2}{w_1'\Sigma w_1}\right).$$

Proof. See Appendix A. ■

In equilibrium, the distribution of consumer characteristics within schools is log normal. This means, for example, that the within-school mean log ability is

$$E(\ln a\,|q) = \mu_3 + \frac{(\eta_2\sigma_{13} + \sigma_{33} + \sigma_{34})}{\delta'\Sigma w_1}\left[(1-\eta_1)\ln q - k_2\right]. \tag{2.10}$$

Since $\eta_1 < 1$, average log ability will increase linearly with log quality if the correlation between log ability and willingness to pay is positive (if $\eta_2\sigma_{13}+\sigma_{33}+\sigma_{34} > 0$). This correlation will be positive if, as one might expect, log ability is positively correlated with both log education and preferences for education. If one or both of these correlations are negative and large enough in magnitude, average log ability will decrease linearly with log quality. The conditional dependence of log ability on log quality is the source of the bias in simple linear regressions of log test scores on log quality. The parameters in (2.10) determine how large that bias is.

The model also predicts the conditional variance of log ability. It is a constant and given by the formula in Theorem 2.2:

$$V(\ln a\,|q) = \sigma_{33} - \frac{(\eta_2\sigma_{13} + \sigma_{33} + \sigma_{34})^2}{w_1'\Sigma w_1}.$$

The ratio of the within school variance to the population variance is determined by the correlation of log ability with willingness to pay. The larger the magnitude of the correlation, the lower the within school variance of log ability.

Such simple to characterise results are not available when $\gamma \neq 0$ or $n_d > 1$. However, the qualitative features of the results remain. Consumers sort so that sets of people with higher values of a willingness to pay index $w(x,d)$ are matched with locations with value index $v(q)$. The equilibrium determines patterns of sorting through the function $F(q)$. Both the patterns of sorting and equilibrium prices depend on the distribution of households in the population and on the technological parameters of the model. In the simple case studied in this section, average log ability is likely to be higher in higher quality schools. In the more general model in which $\gamma \neq 0$ or $n_d > 1$, this may not be the case. Average log ability may be higher in moderate quality schools than in low quality schools, but may be lower in high quality schools than in moderate quality schools. Many other patterns are possible as well. Rather than enumerate them further, the next section briefly discusses computation of the equilibrium price when $\gamma \neq 0$ and $n_d > 1$. Then Section 3 discusses how to estimate the parameters of the model, introduces the US NELS dataset, and estimates the model using this data.

2.3 General case

The general form of equilibrium pricing equation (2.8) is

$$q = G(v, \theta) \tag{2.11}$$

where

$$G(v,\theta) = q^{\eta_1} \sum_{d} \pi(d) A(d) E\left(e^{\delta' x} \left| w(x,d) = v - w_0 \right.\right). \tag{2.12}$$

and θ represents the vector of all parameters in the model. No closed form solution is available for the right side of (2.12). Appendix B shows that it can be simplified and expressed as the ratio of two one dimensional integrals. In the empirical calculations, I compute numerical approximations to (2.12) using Gaussian quadrature.

An equilibrium price can then be computed in two steps. First compute the function $v(q)$ defined implicitly by (2.11). Then given the function $v(q)$, compute the price function using the equation

$$v(q) = \gamma p(q) + \ln[p_q(q)] + (1 - \eta_1)\ln(q) \tag{2.13}$$

with the initial condition $p(0) = p_0$. Nesheim (2002) shows that the first step has a unique solution $v(q)$ that is continuous and differentiable almost everywhere. The second step also has a unique solution since (2.13) is an ordinary differential equation that, assuming $p(0) = 0$, has a unique solution.

In the next section, I use this method to calculate the equilibrium functions $v(q,\theta)$ as part of an estimation procedure. The second step is not required because data on $p(q)$ are not currently available. Therefore, the empirical section focuses attention on the equilibrium equation (2.11) and the function $v(q,\theta)$. When data on prices are not available, this equation and this function contain all the information in the model that is relevant for analysing empirical location choices of households. I discuss this more in the next section. More details are contained in Nesheim (2002).

3 Econometric analysis

Assume that a large microdataset with observations on (t, q, x_1, x_2, d) is available but that no data on housing prices are available. This data is the data that is available in the NELS dataset described in Section 4.1. This data is sufficient to estimate the parameters of the model up to some normalizations. In particular, η_1 the impact of school quality on educational outcomes is identified as long as γ does not equal zero.

The empirical implications of the sorting model in this paper are completely described by the joint distribution of (x, d) and three equations: the test score production function equation (2.3), the location choice equation (2.12), and the equilibrium equation (2.11). These equations are

$$\ln t = \eta_0(d) + \eta_1 \ln q + \eta_2(d)\, x_1 + x_3 + x_5$$

$$v(q, \theta) = w_0(d) + \eta_2 x_1 + x_3 + x_4 + \gamma y$$

$$q = G(v, \theta).$$

In these equations, (t, q) are observable endogenous variables, (x_1, x_2) are observable exogenous variables, and (x_3, x_4, x_5) are unobservable variables. The function $v(q, \theta)$ is defined implicitly by the equilibrium equation.

Because it is possible that (x_3, x_4) are correlated with (x_1, x_2), I first need to reparameterise the system to remove this correlation.[3] Define

$$x_3 = \alpha_1(d) + \alpha_2(d)\, x_1 + \alpha_3(d)\, x_2 + \varepsilon_3$$
$$x_4 = \beta_1(d) + \beta_2(d)\, x_1 + \beta_3(d)\, x_2 + \varepsilon_4$$

where $(\varepsilon_3, \varepsilon_4)$ both have mean zero and $(\varepsilon_3, \varepsilon_4)\,|d \sim N(0, \Omega(d))$. This decomposition is done separately for each subgroup d. The variables $(\varepsilon_3, \varepsilon_4)$ are the components of (x_3, x_4) that are independent of (x_1, x_2). The parameters $(\alpha_2(d), \alpha_3(d))$ and $(\beta_2(d), \beta_3(d))$ reflect the correlation of (x_3, x_4) with (x_1, x_2) conditional on d. After making these changes, the

[3] One might expect that child's log ability x_3 is correlated with both parental education and income. Similarly, one might expect that a household's "taste" for school quality, measured by x_4, is correlated with both education and income.

system of empirical equations becomes

$$\ln t = \widehat{\eta}_0 (d) + \eta_1 \ln q + \widehat{\eta}_2 (d) x_1 + \alpha_3 (d) x_2 + \varepsilon_3 + x_5 \tag{3.1}$$

$$v (q, \theta) = \widehat{w}_0 (d) + \widetilde{\eta}_2 (d) x_1 + \widetilde{\eta}_3 (d) x_2 + \gamma y + \varepsilon_3 + \varepsilon_4 \tag{3.2}$$

$$q = G (v, \theta)$$

where $\widehat{\eta}_0 (d) = \eta_0 (d) + \alpha_1 + \mu_5$, $\widehat{\alpha}_2 = \eta_2 + \alpha_2$, $\widehat{w}_0 (d) = w_0 (d) + \alpha_1 + \beta_1$, $\widetilde{\eta}_2 (d) = \widehat{\eta}_2 (d) + \beta_2 (d)$, and $\widehat{\eta}_3 (d) = \alpha_3 (d) + \beta_3 (d)$. The first equation is the test score production function. The second is the location choice equation, and the third is the equilibrium sorting equation.

Let the elements of $\Omega (d)$ be given by $\omega_{11} (d) = var (\varepsilon_3 + x_5)$, $\omega_{12} (d) = cov (\varepsilon_3 + x_5, \varepsilon_3 + \varepsilon_4)$, and $\omega_{22} (d) = var (\varepsilon_3 + \varepsilon_4)$.

The system (3.1) and (3.2) is the system of empirical equations. The unobserved variables in this system are independent of the exogenous observable variables x_1 and x_2. It is immediate that some of the structural parameters cannot be identified. For example, the parameter η_2 cannot be identified because it affects the system solely through the composites $\widehat{\eta}_2 (d) = \eta_2 + \alpha_2 (d)$. If however, $\alpha_2 (1)$ is normalized to zero, then η_2 is identified and $\alpha_2 (d)$ is identified for $d > 1$. Less obviously, the parameters $(\widehat{w}_0 (d), \omega_{22} (d))$ are not identified for all d. These parameters must be normalized for one subgroup.

Theorem 3.1 *Assume $\widehat{w}_0 (1) = 0$, $\omega_{22} (1) = 1$, and either $\gamma > 0$ or $n_d > 1$. Then, the parameters*

$$\{\mu_1 (d), \mu_2 (d), \sigma_{11} (d), \sigma_{12} (d), \sigma_{22} (d), \pi (d)\}_{d=1}^{n_d},$$

(η_1, γ),

$$\{\widehat{\eta}_0 (d), \widehat{\eta}_2 (d), \alpha_3 (d), \omega_{11} (d), \widetilde{\eta}_2 (d), \widetilde{\eta}_3 (d), \omega_{12} (d)\}_{d=1}^{n_d},$$

and $\{\widehat{w}_0 (d), \omega_{22} (d)\}_{d=2}^{n_d}$ are identified.

Proof. See Appendix A. ■

The identification of $(\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})$ is trivial. These are the first and second moments of the observable variables x_1 and x_2. Identification of the other parameters follows from showing that there is a one-to-one mapping from the joint distribution of $(\ln q, x, d)$ to the parameter space. The fact that this mapping is one-to-one relies on the fact that generically $E(\ln q\,|x)$ is not a linear function of x. This is an implication of equilibrium sorting in the model. This result is a special case of the result in Ekeland, Heckman, and Nesheim (2004) that nonlinearity in hedonic models is non-generic. In this example, the result follows when either $\gamma > 0$ or $n_d > 1$. To see this, consider the unidentified case. When $\gamma = 0$ and $n_d = 1$, Theorem 2.1, implies that

$$v(q, \theta) = \frac{-\ln k_0 + (1 - \eta_1)\ln q}{k_1}.$$

In this case, the system (3.1) and (3.2) is a system of two linear equations with no exclusion restrictions. However, when either $\gamma > 0$ or $n_d > 1$, $v(q, \theta)$ is no longer a linear function of $\ln q$. Effectively, the nonlinearity allows one to use equation (3.2) to create a valid instrument for $\ln q$ in equation (3.1). There is effectively an exclusion restriction in that higher powers of x are excluded from the first equation but not from the second.

The normalizations $\widehat{w}_0(1) = 0$ and $\omega_{22}(1) = 1$ are necessary because the function $v(q, \theta)$ is never directly observed. The first is a "location" normalization. It is like the assumption that the mean of the unobservable in a probit model is zero. The second is a scale normalization. It is like the assumption that the variance of the unobserved term in a probit is one. If prices $p(q)$ were observed, then the second normalization would not be necessary and $\omega_{22}(1)$ could be estimated. Details are discussed in Nesheim (2002).

Given these normalizations, the model can be estimated by maximum likelihood. The likelihood function is detailed in appendix **??**. The parameters to be estimated are listed in the statement of Theorem 3.1. The parameter η_1 is the elasticity of test scores with respect to school quality. This is the variable of primary interest. The parameters $(\widehat{\eta}_0(d), \widehat{\eta}_2(d), \alpha_3(d))$

are the coefficients in the log test score equation. They measure the importance of parental background, race or ethnicity, and sex in producing test score outcomes conditional on school quality. The parameters $(\widehat{w}_0(d), \widetilde{\eta}_2(d), \widetilde{\eta}_3(d), \gamma)$ measure the importance of parental background and preferences in the sorting equation. Finally, the parameters $(\omega_{11}(d), \omega_{12}(d), \omega_{22}(d))$ measure variability of unobserved heterogeneity within each of the subgroups in the sample.

4 Empirical results

This section describes the data and presents the empirical results.

4.1 Data

The model is estimated using data from the 1988 cross-section of the US National Educational Longitudinal Survey (NELS). This sample contains information on 24,000 students in the 8th grade in the US in 1988 with an average of around 30 students per school. The sample includes information on parental education and income, school attended (whether public or private), race, and sex. Additionally, spring term test scores are available for cognitive tests in reading and mathematics. After dropping those with missing data and those in private schools, the sample size is 16,577. Summary statistics are detailed in Table 1. Parental education ranges from 10 to 21 years with a mean of 13.9. Parental income ranges from \$1 to \$300,000 with a mean of \$36,100. Reading and math test scores range from 0.317 to 0.706 and 0.339 to 0.772 respectively.

4.2 Empirical results

I estimated the model using all public schools in the US. This assumes that the "market" for schools is a national market and that the parameters of the model are the same in different regions as well as in urban, suburban, and rural areas.

I estimated equation (3.1) by generalized least squares (GLS) and the system of equations (3.1) and (3.2) by maximum likelihood. Table 2 contains the GLS results when the reading test score is the dependent variable. Tables 3 and 4 contain the maximum likelihood estimates of the system of equations. Tables 5-7 contain results for the case where the math test score is the dependent variable.

In Table 2, the estimated elasticity of the reading test score with respect to school quality is 0.671. The corresponding estimate controlling for sorting is reported in Table 3. After controlling for sorting, the estimated elasticity falls to 0.498. The estimated elasticities with respect tolog parental education and trnasformed income rise slightly after controlling for sorting. The tables also show that there are some differences in parameters across subgroups. The constant term $\widehat{\eta}_0$ is largest in magnitude for white males while the coefficients on log parental education and parental income are larger for white males. Table 4 indicates that white males have the smallest variance of unobserved heterogeeneity in preferences.

Tables 5-7 display analagous results for the math test score. Before controlling for sorting, the elasticity of math test score with respect to school quality is estimated to be 0.664. After controlling for sorting, this elasticity drops to 0.226. The coefficients on log parentla education and transformed income are larger after controlling for sorting. Other qualitative features of the math test score results are similar to the reading test score results.

5 Conclusions

This paper develops a theoretical and empirical model of household sorting and school quality. In a special case, I derived a closed form solution for the equilibrium and characterise sorting outcomes. In a more general case, I develop methods to compute equilibrium.

The empirical implications of the model are developed and it is shown that most of the structural parameters of the sorting model are identified. I then estimate the model using maximum likelihood using data from the NELS dataset. The results show that controlling for sorting is important. If the model is correctly specified, the results indicate that a large fraction of the correlation between school quality and test score outcomes is driven by selection. This is not the last word on evaluating the impact of school quality on test score outcomes. Further work is required to test the robustness of the results to assumptions about the school quality production function, the test score production function, and the locational sorting mechanism. Other work (Nesheim, 2005) under way is extending the model by allowing for a non-parametric specification and by adding detailed housing price data to the analysis. Further work is required to investigate alternative quality production functions, to investigate models in which parents choose locations based on school quality and on characteristics in addition to school quality, and to study environments in which the tie between locational choices and school quality is not so tight perhaps because of policiy's such as busing. Additional work is also required to evaluate how the structural parameters in these types of models respond to changes in government policy.

In order to be credible, such future work must take into account in some fashion the lessons learned from the analysis in this paper, that sorting matters and that proper modelling of the interaction between heterogeneity and household location choices can help disentangle the causal effects of quality from selection effects.

A Proofs

Proof of theorem 2.1. When $\gamma = 0$, equation (2.8) becomes

$$\begin{aligned} q &= E\left(t \,| v\left(q\right) = w_0 + w_1' x\right) \\ &= q^{\eta_1} \sum_d \left[\pi_d A\left(d\right) E\left(\exp\left(\delta' x\right) | w_1' x = v - w_0\right)\right]. \end{aligned}$$

where $\delta = (\eta_2, 0, 1, 0, 0)'$. The pair $(\delta' x, w_1' x)' \sim N(\nu, \Psi)$ where

$$\nu = (\delta' \mu, w_1' \mu)'$$

and

$$\Psi = \begin{bmatrix} \delta' \Sigma \delta & \delta' \Sigma w_1 \\ w_1' \Sigma \delta & w_1' \Sigma w_1 \end{bmatrix}.$$

So $(\delta' x \,| w_1' x) \sim N\left[\delta' \mu + \frac{\delta' \Sigma w_1}{w_1' \Sigma w_1}\left(w_1' x - w_1' \mu\right), \delta' \Sigma \delta - \frac{\left(\delta' \Sigma w_1\right)^2}{w_1' \Sigma w_1}\right]$. Hence

$$q = q^{\eta_1} \sum_d \left[\pi_d A\left(d\right) \exp\left(\delta' \mu + \frac{\delta' \Sigma w_1}{w_1' \Sigma w_1}\left(v - w_0 - w_1' \mu\right) + 0.5\left(\delta' \Sigma \delta - \frac{\left(\delta' \Sigma w_1\right)^2}{w_1' \Sigma w_1}\right)\right)\right].$$

Recall that

$$v\left(q\right) = \ln p_q + \left(1 - \eta_1\right) \ln q$$

and define

$$\ln k_0\left(d\right) = \left(1 - k_1\left(d\right)\right)\left(\ln A + \delta' \mu\right) - k_1\left(d\right)\left(\ln \eta_1 + \mu_4\right) + 0.5\left(\delta' \Sigma \delta - \frac{\left(\delta' \Sigma w_1\right)^2}{w_1' \Sigma w_1}\right).$$

$$k_1\left(d\right) = \frac{\delta' \Sigma w_1}{w_1' \Sigma w_1}.$$

Then the equilibrium can be rewritten as

$$q = q^{\eta_1} \sum_d \pi_d k_0 (d)\, q^{k_1(d)(1-\eta_1)} p_q^{k_1(d)} \tag{A.1}$$

as claimed. Define

$$G(v) = q^{\eta_1} \sum_d \pi_d k_0 (d)\, e^{k_1(d)v}.$$

When $k_0(d) > 0$ for all d, this function is monotonic in v and implicitly defines a function $v(q)$. Then the unique equilibrium price satisfies

$$\begin{aligned} p_q(q) &= q^{\eta_1 - 1} e^{v(q)} \\ p(0) &= 0. \end{aligned}$$

When $n_d = 1$, this simplifies to

$$p(q) = \frac{q^{\frac{1-\eta_1}{k_1}+\eta_1}}{k_0^{\frac{1}{k_1}} \left(\frac{1-\eta_1}{k_1} + \eta_1\right)}.$$

■

Proof of theorem 2.2. The variables $(x, w_1' x)$ are jointly normal. Hence,

$$(x \,|w_1' x = v - w_0) \sim N\left(\mu + \frac{(v - w_0 - w_1'\mu)}{w_1'\Sigma w_1}\Sigma w_1, \Sigma - \frac{\Sigma w_1 w_1' \Sigma}{w_1'\Sigma w_1}\right).$$

$$\theta = \mu + \frac{(v - w_0 - w_1'\mu)}{w_1'\Sigma w_1}\Sigma w_1$$

$$v = -\frac{\ln k_0}{k_1} + \frac{(1-\eta_1)}{k_1}\ln q.$$

$$\theta = \mu + \left(\frac{(1-\eta_1)\ln q - k_2}{k_1 w_1'\Sigma w_1}\right)\Sigma w_1$$

$$k_2 = \ln A + \delta'\mu + 0.5\left(\delta'\Sigma\delta - \frac{(\delta'\Sigma w_1)^2}{w_1'\Sigma w_1}\right)$$

■

Proof of theorem 3.1. The parameters

$$\{\mu_1(d), \mu_2(d), \sigma_{11}(d), \sigma_{12}(d), \sigma_{22}(d), \pi(d)\}_{d=1}^{n_d}$$

are trivially identified. They are the moments of the observable variables (x_1, x_2, d). Next consider identification of the remaining parameters. Let θ represent the vector of all remaining parameters listed in the statement of the theorem. Define $\widetilde{t} = \ln t$ and $\widetilde{q} = \ln q$. Let $F_{(\widetilde{t},\widetilde{q})|x,d}\left(\widetilde{t}, \widetilde{q}, x, d\right)$ be the joint distribution function of $\left(\widetilde{t}, \widetilde{q}\right)$ conditional on (x, d). Equations (3.1) and (3.2) imply that this distribution function satisfies

$$\begin{aligned} & F_{(\widetilde{t},\widetilde{q})|x,d}\left(\widetilde{t}, \widetilde{q}, x, d\right) \\ &= \Phi\left(\chi_1, \chi_2, 0, \Omega(d)\right) \end{aligned} \tag{A.2}$$

where Φ is the bivariate normal CDF with mean 0 and covariance matrix $\Omega(d)$ evaluated at the point (χ_1, χ_2) and (χ_1, χ_2) are given in equations (3.1) and (3.2). The left side is a non-parametrix function of $\left(\widetilde{t}, \widetilde{q}, x, d\right)$. The right side is a function of the same variables parameterised by θ. Consider a Taylor expansion of the right side of $(A.2)$ around the point θ_0.

$$\Phi\left(\widetilde{t}, \widetilde{q}, x, d\right) = \Phi_0\left(\widetilde{t}, \widetilde{q}, x, d\right) + \partial\Phi^*\left(\widetilde{t}, \widetilde{q}, x, d\right) \cdot \partial\theta \tag{A.3}$$

where $\partial\Phi^*\left(\widetilde{t}, \widetilde{q}, x, d\right)$ is the gradient of Φ with respect to θ. Inspection of the elements of this gradient proves that the elements are linearly independent when $\gamma > 0$ or $n_d > 1$ and different groups have different parameters. As a result, a local implicit function theoremt applies and there is a unique mapping from $F_{q|x}$ to θ in some neighborhood of θ_0. ■

B General case

The equilibrium differential equation is

$$G(v,\theta) = q^{\eta_1} \sum_d \pi(d) A(d) E\left(e^{\delta' x} \,|w(x,d) = v - w_0\right) \tag{B.1}$$

where

$$v(q) = (1-\eta_1)\ln q + \gamma p + \ln p_q - \ln(\eta_1).$$

Define the inverse transformation $x_i = z_i$ for $i = 1, 2, 3$ and

$$x_4 = z_4 + v - w_0(d) - \ln A(d) - \eta_2 z_1 - \gamma y(z_2) - z_3$$

with Jacobian

$$\left|\frac{\partial(x_1,x_2,x_3,x_4)}{\partial(z_1,z_2,z_3,z_4)}\right| = 1$$

Then, since $x\,|d \sim N(\mu,\Sigma)$, $z\,|d$ has the probability density function

$$\psi(z_1,z_2,z_3,z_4) = \phi_4\left(z_1,z_2,z_3,z_4 + v - w_0(d) - \ln A(d) - \eta_2 z_1 - \gamma y(z_2) - z_3, \widetilde{\mu}, \widetilde{\Sigma}\right)$$

where ϕ_4 is the four-dimensional normal probability density function and $\widetilde{\mu}$ and $\widetilde{\Sigma}$ are the mean vector and covariance matrix of the first four components of x.

Define $\delta_4 = (\eta_2, 0, 1, 0)'$. To compute the right side of (2.1), we must compute

$$\begin{aligned} E(t\,|v) &= \sum_d \pi(d) A(d) E\left(e^{\delta_4' z} \,|z_4 = 0, d\right) \\ &= \sum_d \pi(d) A(d) G(v,d) \end{aligned} \tag{B.2}$$

By definition

$$G(v,d) = \frac{\iiint e^{\delta_4' z}\phi_4\left(z_1,z_2,z_3,g,\widetilde{\mu},\widetilde{\Sigma}\right) dz_1 dz_2 dz_3}{\iiint \phi_4\left(z_1,z_2,z_3,g,\widetilde{\mu},\widetilde{\Sigma}\right) dz_1 dz_2 dz_3} \tag{B.3}$$

where $g = v - w_0(d) - \ln A(d) - \eta_2 z_1 - \gamma y(z_2) - z_3$. Factoring the normal density, equation $(B.3)$ becomes

$$G(v,d) = \frac{\int \phi(z_2,\mu_2,\sigma_{22}) \left(\iint e^{\delta_4' z} \phi_3(z_1,z_3,g,\xi,\Lambda)\, dz_1 dz_3 \right) dz_2}{\int \phi(z_2,\mu_2,\sigma_{22}) \left(\iint \phi_3(z_1,z_3,g,\xi,\Lambda)\, dz_1 dz_3 \right) dz_2} \tag{B.4}$$

where

$$\xi = \left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(z_2 - \mu_2), \mu_3 + \frac{\sigma_{23}}{\sigma_{22}}(z_2 - \mu_2), \mu_4 + \frac{\sigma_{24}}{\sigma_{22}}(z_2 - \mu_2) \right)'$$

and

$$\Lambda = \begin{bmatrix} \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} & \sigma_{13} - \frac{\sigma_{12}\sigma_{23}}{\sigma_{22}} & \sigma_{14} - \frac{\sigma_{14}\sigma_{24}}{\sigma_{22}} \\ \sigma_{13} - \frac{\sigma_{12}\sigma_{23}}{\sigma_{22}} & \sigma_{33} - \frac{\sigma_{23}^2}{\sigma_{22}} & \sigma_{34} - \frac{\sigma_{23}\sigma_{24}}{\sigma_{22}} \\ \sigma_{14} - \frac{\sigma_{14}\sigma_{24}}{\sigma_{22}} & \sigma_{34} - \frac{\sigma_{23}\sigma_{24}}{\sigma_{22}} & \sigma_{44} - \frac{\sigma_{24}^2}{\sigma_{22}} \end{bmatrix}$$

Let $I_1(z_2)$ represent the inner pair of integrals in the numerator of $(B.4)$. $I_1(z_2)$ can be rewritten as

$$I_1(z_2) = \iint_{z_1,z_3} \frac{e^{\delta_4' z - 0.5 h' D' \Lambda^{-1} D h}}{(2\pi)^{1.5} |\Lambda|^{0.5}} dz_1 dz_3$$

where $h = (z_1 - \xi_1, z_3 - \xi_2, v - w_0(d) - \ln A(d) - \eta_2 \xi_1 - \gamma y(z_2) - \xi_2 - \xi_3)'$ and

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\eta_2 & -1 & 1 \end{bmatrix}$$

Letting $T = D^{-1}\Lambda(D')^{-1}$ and factoring the integrand, this integral can be further simplified to

$$I_1(z_2) = \frac{e^{-0.5\frac{h_3^2}{\tau_{33}}}}{\sqrt{2\pi\tau_{33}}} \iint_{z_1,z_3} e^{\delta_4' z} \phi_2(z_1,z_3,\pi,\Xi)\, dz_1 dz_3$$

where $\pi = \left(\xi_1 + \frac{\tau_{13}}{\tau_{33}} h_3, \xi_2 + \frac{\tau_{23}}{\tau_{33}} h_3 \right)^T$ and $\Xi = \begin{bmatrix} \tau_{11} - \frac{\tau_{13}^2}{\tau_{33}} & \tau_{12} - \frac{\tau_{13}\tau_{23}}{\tau_{33}} \\ \tau_{12} - \frac{\tau_{13}\tau_{23}}{\tau_{33}} & \tau_{22} - \frac{\tau_{23}^2}{\tau_{33}} \end{bmatrix}$. Define $\delta_2 =$

$(\eta_2, 1)^T$. This integral reduces to the analytic expression

$$I_1(z_2) = \frac{e^{-0.5\frac{h_3^2}{\tau_{33}} + \delta_2^T \pi + 0.5\delta_2^T \Xi \delta_2}}{\sqrt{2\pi\tau_{33}}}$$

Thus, the conditional expectation in equation $(B.4)$ reduces to the ratio of two one-dimensional integrals

$$G(v, d) = \frac{\int \frac{e^{-\frac{1}{2}\frac{(z_2-\mu_2)^2}{\sigma_{22}}}}{\sqrt{2\pi\sigma_{22}}} \left(\frac{e^{-0.5\frac{h_3^2}{\tau_{33}} + \delta_2^T \pi + 0.5\delta_2^T \Xi \delta_2}}{\sqrt{2\pi\tau_{33}}} \right) dz_2}{\int \frac{e^{-\frac{1}{2}\frac{(z_2-\mu_2)^2}{\sigma_{22}}}}{\sqrt{2\pi\sigma_{22}}} \left(\frac{e^{-0.5\frac{h_3^2}{\tau_{33}}}}{\sqrt{2\pi\tau_{33}}} \right) dz_2}$$

Define $\delta_5 = (\eta_2, 0, 1, 0, 0)^T$. Then

$$G(v, d) = \frac{e^{\delta_5^T \mu + 0.5\delta_5^T \Sigma \delta_5} \int e^{-\frac{(z_2-\mu_2-\sigma_{12}\eta_2-\sigma_{23})^2}{2\sigma_{22}}} e^{-\frac{0.5(h_3-\eta_2\tau_{13}-\tau_{23})^2}{\tau_{33}}} dz_2}{\int e^{-\frac{1}{2}\frac{(z_2-\mu_2)^2}{\sigma_{22}}} e^{-0.5\frac{h_3^2}{\tau_{33}}} dz_2}$$

where

$h_3 = v - w_0(d) - \ln A(d) - \eta_2\xi_1 - \gamma y(z_2) - \xi_2 - \xi_3$

$v = \gamma p + \ln p_q + (1 - \eta_1) \ln q - \ln(\eta_1)$

$\xi_1 = \mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(z_2 - \mu_2)$

$\xi_2 = \mu_3 + \frac{\sigma_{23}}{\sigma_{22}}(z_2 - \mu_2)$

$\xi_3 = \mu_4 + \frac{\sigma_{24}}{\sigma_{22}}(z_2 - \mu_2)$

$\tau_{13} = \eta_2\Lambda_{11} + \Lambda_{12} + \Lambda_{13}$

$\tau_{23} = \eta_2\Lambda_{12} + \Lambda_{22} + \Lambda_{23}$

$\tau_{33} = \eta_2^2\Lambda_{11} + 2\eta_2(\Lambda_{12} + \Lambda_{13}) + \Lambda_{22} + \Lambda_{33} + 2\Lambda_{23}$.

$$\tau_{13} = (\eta_2 + \alpha_2 + \beta_2)\left(\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right)$$

$$\tau_{23} = \alpha_2(\eta_2 + \alpha_2 + \beta_2)\left(\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right) + \sigma_{\varepsilon_{33}} + \sigma_{\varepsilon_{34}}$$

$$\tau_{33}=(\eta_2+\alpha_2+\beta_2)^2\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right)+\sigma_{\varepsilon 33}+\sigma_{\varepsilon 44}+2\sigma_{\varepsilon 34}$$

$$\Lambda=\begin{bmatrix}\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}} & \alpha_2\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right) & \beta_2\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right)\\ \alpha_2\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right) & \alpha_2^2\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right)+\sigma_{\varepsilon 33} & \alpha_2\beta_2\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right)+\sigma_{\varepsilon 34}\\ \beta_2\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right) & \alpha_2\beta_2\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right)+\sigma_{\varepsilon 34} & \beta_2^2\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right)+\sigma_{\varepsilon 44}\end{bmatrix}$$

Then $(B.2)$ becomes

$$E\left(t\,|v\right)=$$
$$q^{\eta_1}\sum_d \pi\left(d\right)A\left(d\right)e^{w_0(d)}\frac{e^{\delta_5^T\mu+0.5\delta_5^T\Sigma\delta_5}\int e^{-\frac{(z_2-\mu_2-\sigma_{12}\eta_2-\sigma_{23})^2}{2\sigma_{22}}}e^{-\frac{0.5(h_3-\eta_2\tau_{13}-\tau_{23})^2}{\tau_{33}}}dz_2}{\int e^{-\frac{1}{2}\frac{(z_2-\mu_2)^2}{\sigma_{22}}}e^{-0.5\frac{h_3^2}{\tau_{33}}}dz_2}$$

This can be rewritten as

$$q^{1-\eta_1}=\sum_d \pi\left(d\right)\left(\frac{\begin{array}{l}e^{w_0(d)+\mu_5+0.5\sigma_{55}}\times\\ e^{(\eta_2+\alpha_2)\mu_1+\alpha_1+\alpha_3\mu_2}\times\\ e^{0.5\left((\eta_2+\alpha_2)^2\sigma_{11}+2\alpha_3(\eta_2+\alpha_2)\sigma_{12}+\alpha_3^2\sigma_{22}+\sigma_{\varepsilon 33}\right)}\times\\ \left[\begin{array}{c}\int e^{-\frac{(z_2-\mu_2-\sigma_{12}(\eta_2+\alpha_2)-\alpha_3\sigma_{22})^2}{2\sigma_{22}}}\times\\ e^{-\frac{\left[0.5\left(h_3-\left[(\eta_2+\alpha_2)(\eta_2+\alpha_2+\beta_2)\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right)+\sigma_{\varepsilon 33}+\sigma_{\varepsilon 34}\right]\right)^2\right]}{(\eta_2+\alpha_2+\beta_2)^2\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right)+\sigma_{\varepsilon 33}+\sigma_{\varepsilon 44}+2\sigma_{\varepsilon 34}}}\,dz_2\end{array}\right]\end{array}}{\int e^{-\frac{1}{2}\frac{(z_2-\mu_2)^2}{\sigma_{22}}}e^{-0.5\frac{h_3^2}{\tau_{33}}}dz_2}\right) \tag{B.5}$$

$$\begin{aligned}h_3=\,&v-w_0\left(d\right)-\mu_5-0.5\sigma_{55}-\alpha_1-\beta_1\\ &-\left(\eta_2+\alpha_2+\beta_2\right)\mu_1\\ &-\left(\alpha_3+\beta_3\right)\mu_2\\ &-\left(\left(\eta_2+\alpha_2+\beta_2\right)\frac{\sigma_{12}}{\sigma_{22}}+\alpha_3+\beta_3\right)\left(z_2-\mu_2\right)\\ &-\gamma y\left(z_2\right)\end{aligned}$$

Define

$$
\begin{aligned}
\widehat{\eta}_2 &= \eta_2 + \alpha_2 \\
\widehat{w}_0 &= w_0(d) + \mu_5 + \alpha_1 \\
\omega_{22} &= \sigma_{\varepsilon_{33}} + \sigma_{\varepsilon_{44}} + 2\sigma_{\varepsilon_{34}}
\end{aligned}
$$

Then rewrite $(B.5)$ as

$$
\begin{aligned}
q &= G(v, \theta) \\
&= q^{\eta_1} \sum_d \pi(d) \left(\begin{array}{l} e^{\widehat{w}_0 + 0.5\sigma_{55}} \times \\ e^{\widehat{\eta}_2 \mu_1 + \alpha_3 \mu_2} \times \\ e^{0.5\left(\widehat{\eta}_2^2 \sigma_{11} + 2\alpha_3 \widehat{\eta}_2 \sigma_{12} + \alpha_3^2 \sigma_{22} + \sigma_{\varepsilon_{33}}\right)} \times \\ \dfrac{\left[\begin{array}{c} \int e^{-\frac{(z_2 - \mu_2 - \sigma_{12}\widehat{\eta}_2 - \alpha_3 \sigma_{22})^2}{2\sigma_{22}}} \times \\ e^{-\frac{\left[0.5\left(h_3 - \left[\widehat{\eta}_2(\widehat{\eta}_2 + \beta_2)\left(\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right) + \sigma_{\varepsilon_{33}} + \sigma_{\varepsilon_{34}}\right]\right)^2\right]}{(\widehat{\eta}_2 + \beta_2)^2 \left(\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right) + \omega_{22}}} dz_2 \end{array} \right]}{\int e^{-\frac{1}{2}\frac{(z_2 - \mu_2)^2}{\sigma_{22}}} e^{-0.5\frac{h_3^2}{(\widehat{\eta}_2 + \beta_2)^2\left(\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right) + \omega_{22}}} dz_2} \end{array} \right)
\end{aligned}
$$

$$
\begin{aligned}
h_3 = v &- \widehat{w}_0 - 0.5\sigma_{55} - \beta_1 \\
&- (\widehat{\eta}_2 + \beta_2)\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(z_2 - \mu_2)\right) \\
&- (\alpha_3 + \beta_3) z_2 - \gamma y(z_2)
\end{aligned}
$$

C Tables

Table 1:

Summary statistics

Variable	Average	Standard deviation	Minimum	Maximum	Observations
Parental schooling	13.9	2.46	10	21	16,577
Parental income	3.61e+4	3.34e+4	1	3.0e+5	16,577
Male	0.503	0.500	0	1	16,577
White	0.706	0.456	0	1	16,577
Black	0.140	0.347	0	1	16,577
Hispanic	0.107	0.309	0	1	16,577
Other racial group	0.0466	0.211	0	1	16,577
Reading test score	0.4921	0.0989	0.3171	0.7055	16,577
Math test score	0.4944	0.0999	0.339	0.772	16,577
Quality (reading)	0.492	0.0413	0.363	0.637	16,577
Quality (math)	0.494	0.0480	0.379	0.692	16,577

Table 2:

GLS estimates of the test score equation.

Dependent variable: reading test score

	$\widehat{\eta}_0$	η_1	$\widehat{\eta}_2$	α_3	ω_{11}
White males	-0.923	0.671	0.264	0.0104	0.0329
	(5.09e-2)	(1.84e-2)	(1.70e-2)	(2.8e-3)	
White females	–0.889		0.272	0.0149	0.0299
	(6.56e-3)		(2.32e-2)	(3.7e-3)	
Black males	–0.748		0.174	0.00770	0.0244
	(9.87e-2)		(3.62e-2)	(4.3e-3)	
Black females	-0.739		0.189	0.0101	0.0254
	(1.02e-1)		(3.75e-2)	(4.5e-3)	
Hispanic males	-0.645		0.152	0.0124	0.0270
	(9.03e-2)		(3.33e-2)	(4.7e-3)	
Hispanic females	-0.595		0.137	0.0066	0.0266
	(8.82e-2)		(3.27e-2)	(4.1e-3)	
Other males	-0.8548		0.241	0.0161	0.0299
	(1.23e-1)		(4.35e-2)	(6.3e-3)	
Other females	-0.7099		0.200	0.0166	0.0276
	(1.22e-1)		(4.34e-2)	(5.2e-3)	

Table 3

Maximum likelihood estimates of the system.

Test score equation parameters.

Dependent variable: reading test score.

	$\widehat{\eta}_0$	η_1	$\widehat{\eta}_2$	α_3	ω_{11}
White males	-1.52	0.0498	0.331	0.0171	0.0347
	(2.03e-3)	(7.37e-4)	(4.83e-4)	(2.88e-4)	(5.02e-3)
White females	-1.48		0.336	0.0199	0.0318
	(1.93e-3)		(4.79e-4)	(3.05e-4)	(5.59e-3)
Black males	-1.38		0.233	0.0106	0.0263
	(4.75e-3)		(1.14e-3)	(2.98e-4)	(7.22e-3)
Black females	-1.32		0.231	0.0139	0.0275
	(4.67e-3)		(1.23e-3)	(3.12e-4)	(7.40e-3)
Hispanic males	-1.26		0.215	0.0179	0.0292
	(4.91e-3)		(1.03e-3)	(2.75e-4)	(8.01e-3)
Hispanic females	-1.23		0.206	0.0109	0.0285
	(4.68e-3)		(9.30e-4)	(1.97e-4)	(7.65e-3)
Other males	-1.52		0.333	0.0257	0.0334
	(6.78e-3)		(1.29e-3)	(3.58e-4)	(1.04e-2)
Other females	-1.36		0.282	0.0232	0.0326
	(6.64e-3)		(1.12e-3)	(2.33e-4)	(1.03e-2)

Table 4:

Maximum likelihood estimates of system.

Location choice equation parameters.

Dependent variable: school quality (school level average reading test score)

	$\widehat{w}_0$	$\widetilde{\eta}_2$	$\widetilde{\eta}_3$	γ	ω_{22}	ω_{12}
White males	0.00	1.49	0.0768	1.14	1.00	0.0444
	NA	(5.97e-3)	(3.28e-4)	(8.44e-2)	NA	(3.37e-4)
White females	-1.48	1.40	0.0512		1.00	0.0448
	(1.93e-3)	(4.94e-3)	(3.39e-4)		(9.62e-3)	(3.48e-4)
Black males	-1.38	1.38	0.0667		1.55	0.0550
	(4.75e-3)	(4.42e-3)	(2.78e-4)		(4.61e-2)	(2.73e-4)
Black females	-1.32	0.982	0.0781		1.67	0.0597
	(4.67e-3)	(3.85e-3)	(2.45e-4)		(4.78e-2)	(2.42e-4)
Hispanic males	-1.26	1.47	0.129		1.33	0.0548
	(4.91e-3)	(4.98e-3)	(3.45e-4)		(3.91e-2)	(3.43e-4)
Hispanic females	-1.23	1.57	0.107		1.35	0.0513
	(4.68e-3)	(4.63e-3)	(3.18e-4)		(3.83e-2)	(3.16e-4)
Other males	-1.52	2.13	0.186		1.78	0.0796
	(6.78e-3)	(4.68e-3)	(4.94e-4)		(6.70e-2)	(5.00e-4)
Other females	-1.36	1.94	0.186		1.58	0.0923
	(6.64e-3)	(2.87e-3)	(4.08e-4)		(5.89e-2)	(4.14e-4)

Table 5:

GLS estimates of the test score equation.

Dependent variable: math test score

	$\widehat{\eta}_0$	η_1	$\widehat{\eta}_2$	α_3	ω_{11}
White males	-0.895	0.664	0.267	0.0169	0.0291
	(4.75e-2)	(1.51e-2)	(1.60e-2)	(2.67e-3)	
White females	−0.978		0.295	0.0143	0.0267
	(6.18e-2)		(2.19e-2)	(3.48e-3)	
Black males	−0.705		0.161	0.00887	0.0200
	(9.00e-2)		(3.30e-2)	(3.92e-3)	
Black females	−0.655		0.145	0.0112	0.0190
	(9.02e-2)		(3.30e-2)	(4.04e-3)	
Hispanic males	−0.672		0.165	0.0114	0.0239
	(8.52e-2)		(3.14e-2)	(4.42e-3)	
Hispanic females	−0.496		0.0841	0.00588	0.0225
	(8.17e-2)		(3.03e-2)	(3.85e-3)	
Other males	−0.658		0.197	0.0235	0.0352
	(1.31e-1)		(4.63e-2)	(6.68e-3)	
Other females	−0.720		0.207	0.0125	0.0293
	(1.25e-1)		(4.43e-2)	(5.18e-3)	

Table 6

Maximum likelihood estimates of the system.

Test score equation parameters.

Dependent variable: math test score.

	$\widehat{\eta}_0$	η_1	$\widehat{\eta}_2$	α_3	ω_{11}
White males	-1.32	0.266	0.333	0.0245	0.0306
	(2.27e-3)	(1.46e-4)	(2.35e-4)	(1.54e-4)	(6.49e-4)
White females	-1.40		0.357	0.0202	0.0282
	(5.68e-4)		(4.32e-4)	(1.36e-3)	(3.37e-5)
Black males	-1.13		0.208	0.0123	0.0205
	(1.17e-3)		(1.95e-4)	(8.28e-4)	(3.63e-5)
Black females	-1.05		0.184	0.0159	0.0201
	(8.50e-4)		(1.14e-4)	(1.12e-5)	(1.67e-5)
Hispanic males	-1.05		0.205	0.0184	0.0257
	(1.70e-3)		(1.96e-4)	(9.57e-4)	(4.78e-5)
Hispanic females	-0.903		0.132	0.0109	0.0232
	(1.52e-3)		(1.01e-4)	(4.33e-7)	(4.47e-5)
Other males	-1.11		0.271	0.0330	0.0401
	(1.05e-3)		(2.06e-4)	(5.35e-4)	(7.44e-5)
Other females	-1.15		0.271	0.0185	0.0327
	(8.22e-4)		(3.83e-4)	(3.26e-4)	(1.04e-4)

Table 7:

Maximum likelihood estimates of system.

Location choice equation parameters.

Dependent variable: school quality (school level average math test score)

	$\widehat{w}_0$	$\widetilde{\eta}_2$	$\widetilde{\eta}_3$	γ	ω_{22}	ω_{12}
White males	0.00	1.71	0.0623	1.55	1.00	0.0389
	NA	(1.57e-2)	(2.16e-2)	(1.23e-3)	NA	(1.59e-4)
White females	0.259	1.61	0.0391		1.04	0.0394
	(7.23e-4)	(1.63e-2)	(3.14e-2)		(8.29e-4)	(2.54e-4)
Black males	-1.22	1.32	0.0358		1.34	0.0306
	(8.13e-4)	(2.94e-2)	(4.26e-2)		(8.65e-4)	(2.08e-4)
Black females	0.840	1.03	0.0809		1.43	0.0402
	(8.94e-4)	(2.44e-2)	(2.73e-2)		(6.54e-4)	(1.04e-4)
Hispanic males	1.34	1.03	0.130		1.20	0.0452
	(9.46e-4)	(1.88e-2)	(4.71e-2)		(1.18e-3)	(3.09e-4)
Hispanic females	0.403	1.33	0.0894		1.20	0.0304
	(5.23e-4)	(1.55e-2)	(1.38e-2)		(2.03e-3)	(5.38e-5)
Other males	-0.0675	1.77	0.144		1.50	0.0895
	(4.92e-4)	(2.50e-02)	(1.01e-2)		(1.14e-3)	(8.36e-5)
Other females	0.222	1.61	0.0781		1.38	0.0686
	(5.49e-4)	(1.94e-2)	(2.69e-2)		(2.86e-3)	(2.15e-4)

D References

Nesheim, Lars (2002), "Equilibrium distributions of heterogeneous consumers across locations: theory and empirical implications," CEMMAP working paper CWP08/02.

Nesheim, Lars (2005) "Nonparametric estimation of school quality effects in sorting models," unpublished manuscript.