Neighborhood Sorting, Endogeneity, and the Valuation of School Quality

Gregorio Caetano*

September 9, 2012

Abstract

In this paper I estimate the parental valuation of public school quality in a unit of measurement comparable to the tuition charged by private schools: per child, year and grade. I also develop a new method to identify the valuation of neighborhood amenities in the presence of neighborhood sorting. The method controls for unobservable neighborhood amenities, including those affected by sorting, which are likely to bias the valuation estimates. For the state of Minnesota, I found that parents are willing to pay yearly about $2400 per child for a one standard deviation (16%) increase in elementary school quality. The corresponding valuations are $1400 for middle school and $2700 for high school.

1 Introduction

In this paper, I use detailed microdata to estimate the parental value of an improvement in school quality. I argue the need to impose more structure in the analysis in order to (1) provide a more useful estimate to the policy maker; (2) solve the endogeneity problem, which I argue to be more complicated to be solved than previously thought.

*Assistant Professor, Department of Economics, University of Rochester. E-mail: gregorio.caetano@rochester.edu. I thank David Card for many helpful discussions and invaluable advice throughout the duration of this project. I also thank Patrick Bayer, Carolina Caetano, Fernando Ferreira, Josh Kinsler, Patrick Kline, Ashley Langer, Vikram Maheshri, Robert McMillan, Enrico Moretti, Romans Pancs, Ronni Pavan, Claudia Sitgraves, Melissa Tartari and seminar participants for their comments. Yusuke Jinnai and Daniel Ringo provided excellent research assistance. I am especially thankful to Angela Andrus, Arnie Reznek, Liliana Souza and Jon Stiles of the Census Bureau. The research results and conclusions expressed are mine and do not necessarily reflect the views of the Census Bureau. This paper has been screened to ensure that no confidential data are revealed. All remaining errors are my own.
I explore the notion that there are two different valuations of school quality that one could estimate. One valuation of school quality, the price capitalization, describes the effect in house prices caused by a change in school quality. Another valuation, the one I focus on in this paper, is the marginal willingness to pay (MTWP) for school quality. It takes the perspective of a policy maker who wants to perform cost-benefit analysis of a particular education policy that improves school quality.

It is important to distinguish between these two valuations, as the price capitalization parameter is intrinsically related to the housing market while the MWTP parameter is not. Indeed, ideally if there was a market to directly purchase public school quality, researchers would then estimate the MWTP parameter without relying on housing data, but would still use such data to estimate the capitalization parameter.

Any identification strategy for the MWTP parameter, therefore, must control for all factors intrinsic to the housing market and unrelated to the valuation of school quality per se. One such factor is the commonly studied endogeneity problem present in neighborhoods: since amenities are bundled in a neighborhood, some of them may be correlated with school quality and unobserved to the researcher. In addition to this factor, I focus on two others: the fact that families, not children, are the housing decision makers, and the fact that housing is a dynamic decision. Since a family may have more than one child, and its children may attend different grades, it is necessary to disentangle the valuation per child-grade from the total valuation of each family. Moreover, because of moving costs, it is necessary to disentangle the valuation of school quality per year-grade from the cumulative valuation, which is likely to span more than one year and one grade. These features of the housing decision are not related to school quality preference, nevertheless they make it hard for the policy maker to interpret the MWTP estimates using housing data. For instance, the horizon of people when they make their housing decisions is unrelated to the horizon of a particular education policy, so a policy maker who contemplates implementing a policy that improves school quality for one year needs an estimate that does not change depending on the expected time households intend to live in the neighborhood, which is a function of moving costs.

I provide two main contributions: I develop a novel approach to address the endogeneity problem that arises in the context of neighborhoods, and I estimate the valuation of

---

school quality at the same unit of measurement as the tuition charged by private schools: per year-child-grade (i.e., for each child enrolled in a particular grade during one year). This unit is particularly useful to the policy maker, as the education policy may be targeted to specific grades, to a specific duration of time and to a specific number of students. Moreover, the estimates may be useful as a benchmark for the implementation of alternative policies (e.g., private school voucher programs) and for a better understanding of how much pressure competing private schools exert on public schools.

The literature so far has not focused on providing MWTP estimates that are comparable to private school tuition, however it has focused substantially on the endogeneity problem. Several identification strategies to deal with the endogeneity problem have been suggested in the literature. Black (1999), Kane, Riegg & Staiger (2006), Bayer, Ferreira & McMillan (2007), Machin & Salvanes (2010) and Fack & Grenet (2010) use school attendance boundary fixed effects (BFE) to control for contemporaneous unobservables that may be correlated with school quality. Figlio & Lucas (2004), Reback (2005) and Fiva & Kirkeboen (2008) suggest a differences approach to estimate the change in prices before and after the introduction of a new policy, while Cellini, Ferreira & Rothstein (2010) suggest a differences-in-differences approach, where some neighborhoods were affected by a policy, and prices before and after the policy are compared among the treated neighborhoods and the untreated ones, with the additional desirable feature that the assignment to treatment is quasi-experimental. Finally, Bradbury, Mayer & Case (2001), Downes & Zabel (2002) and Gibbons & Machin (2003) suggest an instrumental variables approach, and Clapp, Nanda & Ross (2008) suggest a panel regression strategy with fixed effects at the neighborhood level.

In this paper, I offer a new perspective on the endogeneity issue present in neighborhoods. I argue that any procedure to treat this endogeneity problem that does not explicitly control for contemporaneous unobservables will in fact identify the price capitalization parameter, rather than the MWTP parameter. Indeed, changes in school quality, no matter how random the source, generally imply changes in the unobservables due to sorting along characteristics of neighbors. Thus, because the data is generally observed only after some neighborhood sorting has occurred, any attempt to measure the willingness to pay for school quality not explicitly controlling for contemporaneous unobservables identifies the total observed change in the valuation of the neighborhood due to the change in school quality, which is the sum of the original change in the valuation of the neighbor-
hood due to school quality per se (the direct valuation) and the change up to that point in
the valuation of the neighborhood due to the change in neighbors’ characteristics induced
by the original change in school quality (the indirect valuation).

To identify the MWTP for school quality, it is important to control fully for the indi-
rect valuation, not only because it is a function of valuations of other amenities instead,
but also because not doing so will make the estimates prone to the contamination bias
commonly studied in the program evaluation literature, where untreated units are affected
by treatment (e.g., [Heckman, LaLonde & Smith (1999) and Miguel & Kremer (2004)].

The contamination bias arises because any positive change in the indirect valuation of
some neighborhood is associated with negative changes in the indirect valuation of some
other neighborhood due to sorting. As an illustration, when a rich household decides to
sort to one neighborhood because of school quality, its decision not only increases the
average income of neighbors in that neighborhood; it also reduces the average income of
neighbors in its former neighborhood. The gap between the valuations of the treated and
untreated neighborhoods due to school quality would then be doubly overestimated.

Of the identification strategies proposed in the literature, the boundary fixed effects
(BFE) approach is the only one that attempts to control for contemporaneous unobserv-
able amenities, as it explores only variation within time to identify the parameter. In
that approach, prices of houses close to each other but on different sides of the school
attendance boundary are compared, with the underlying assumption that unobservables
do not change across the boundary. However, [Kane, Riegg & Staiger (2006) and Bayer,
Ferreira & McMillan (2007)] have shown that the boundary fixed effects are not able to
fully control for unobservable amenities, as observable characteristics of neighborhoods,
particularly characteristics of neighbors, remain important in explaining prices even with
the inclusion of boundary fixed effects. Their identification strategy is to control for ob-
servable characteristics of the house and the neighborhood in addition to boundary fixed
effects.

There are two potential issues related to this augmented version of the BFE approach.
First, if there still exists unobservables that vary across the boundary, the estimates would
reflect not only the direct but also the indirect valuation through these unobserved ameni-
ties (as well as the contamination bias). A second and subtler issue is that adding observ-
able characteristics of neighbors as control variables may change the interpretation of the
resulting MWTP estimates. To the extent that attendance in a public school is largely
based on residential location, characteristics of neighbors might act as proxy variables to
the family background of the peers inside the school, implying MWTP estimates to be

---

4 This bias is sometimes referred as “spillover bias” and was originally discussed in Lewis (1973) in the context of unions.
interpreted instead as the valuation of the component of school quality that is orthogonal to peer effects.

In this paper, I propose an alternative method to control for contemporaneous unobservable amenities. An advantage of this method is that it controls for unobservables, including those affected by sorting, such as characteristics of neighbors. Another advantage is that, unlike the BFE approach, it controls for the “neighborhood amenities” (e.g., living close to highly educated neighbors) without controlling for the “school amenities” (e.g., having peers with highly educated parents). This feature is particularly interesting in the context of school quality, as it allows the MWTP estimates to be comparable to the tuition offered by private schools. Indeed, all parents living close to a private school are exposed to their neighbors’ characteristics, however they have access to the peer effects inside the school only if they pay for tuition. Finally, another potential advantage of the proposed method is that it does not rely on the geographical discontinuity in school quality, so it may in principle be feasible to apply this method to the estimation of the valuation of amenities that vary smoothly across boundaries, where the BFE approach is unfeasible.

My identification strategy has three steps, and explores variation in consumer surpluses of different groups of households within neighborhood and within time. The Long Form of the 2000 Census data is used to define 40 groups: 20 groups of parents, defined by the age of the oldest child (from ages 0 to 19) and 20 groups of non-parents, defined by the age of the head of household (from ages 31 to 50). Exact date of birth data is used to define the groups in such a way that children from ages 6 through 18 are observed to attend grades kindergarten through twelve, respectively. In the first step, I estimate the mean flow utility of each group for each neighborhood in a dynamic discrete choice model based on Rust (1987), Berry, Levinsohn & Pakes (1995) and a synthetic cohort assumption. In this step I suggest a novel approach to estimate a dynamic choice model with some advantages and some disadvantages over the standard approaches. A particular advantage that is crucial to this paper is that it enables one to estimate the flow utilities with cross-sectional data.

In the second step, I treat the endogeneity problem with a technique known as Proxy-

---

5 Some papers have used quasi-experimental identification strategies to estimate the valuation of other amenities, such as clean air (e.g., Chay & Greenstone (2005)), airport noise (e.g., Pope (2008)) and exposure to hazardous waste (e.g., Greenstone & Gallagher (2008)). Because these papers do not explicitly control for contemporaneous unobservables, they identify price capitalization, rather than MWTP.

6 Parents of group c are assumed to have similar preferences next year to parents in group c + 1 this year, and non-parents are allowed to expect to have similar preferences next year to either non-parents one year older or parents of group 0. The conditional state probabilities are observed in the data and assumed known.
The estimated mean flow utility at the group-neighborhood level is written linearly as a function of observed and unobserved neighborhood level amenities (i.e., non-excludable amenities), allowing for each group to value these neighborhood level amenities differently. I estimate panel data regressions of some groups with the mean flow utilities of other groups as proxy variables for the unobserved neighborhood level amenities. Specifically, the groups of non-parents and parents with children too young to attend school are used as proxy, since they do not value school quality in the flow utility sense. Any unobservable amenity that is valued by groups used as proxy are controlled in these regressions. Intuitively, although these groups do not value school quality in the flow utility sense, they do value the unobservable amenities in the neighborhood that are correlated with school quality.

Finally, in the third step, data on the number of children attending each grade in each group is used to decompose the MWTP estimates per year-family into the MWTP per year-child-grade.

Using data from Minnesota, I estimate that, for a one standard deviation (16%) increase in test scores, parents are willing to pay yearly for each child $2400 in primary (K-5) school, $1400 in middle (6-8) school and $2700 in secondary (9-12) school. These values correspond respectively to the bottom 56 percentile of the distribution of private tuition spending in the U.S. for primary schools, the bottom 26 percentile for middle schools and the bottom 23 percentile for secondary schools.

This paper is organized as follows. In section 2 I introduce a benchmark model with no moving costs in order to focus on the endogeneity problem and to present the method to control for contemporaneous unobservables proposed in this paper. Section 3 introduces a dynamic model of neighborhood choice with moving costs, and presents the identification strategy of the paper in three steps. In section 4 I show the data, the empirical results, and robustness checks. In section 5 I discuss some features of the

---

7 This method was first proposed by Chamberlain (1977) and Pudney (1982) and was used by Heckman & Scheinkman (1987) in the estimation of hedonic wage regressions. To the best of the author’s knowledge, this paper is the first to apply this technique to the estimation of preference parameters and also the first to suggest its implementation in a discrete choice framework.

8 The panel data are written essentially as low-dimensional factors, where the factors are the neighborhood level amenities and the loading coefficients are the preference parameters, which are allowed to be different across groups.

9 The exclusion restriction is that they do not value the school quality amenity that is provided in that particular period. However, they are allowed to value the school quality amenity in a cumulative utility sense, and, even in the flow utility sense, they are allowed to value amenities that are often proxied by school quality, such as the average neighbor’s level of education.

10 The median per year private school tuition expenditure in the U.S. is $2151 for elementary schools, $2473 for middle schools and $4624 for secondary schools. Source: Current Population Survey, October 1997, School Enrollment. All values are in 2000 dollars.
methodology and results in comparison to the previous literature, and finally in section I conclude.

2 Neighborhood Sorting and Endogeneity

This section describes the endogeneity issues related to the identification of the marginal willingness to pay (MWTP) for school quality, and illustrates the solution proposed in this paper in a benchmark model with no moving costs.

2.1 Basic Setup

Suppose there are two groups of households in the data: parents and non-parents, indexed respectively as \( p \) and \( np \). Each household chooses a neighborhood to live among \( J \) options. Each neighborhood is characterized by a set of amenities, with some of them unobserved to the researcher. Because there is no moving costs, households consider only their flow utility when deciding where to live.\(^{11}\)

Let the flow utility of household \( i \) in group \( c \) when choosing neighborhood \( j \) be

\[
U_{i,c,j} = SQ_j \theta_c + P_j \phi_c + \xi_{c,j} + \epsilon_{i,c,j}, \quad c = p, np, \quad j = 1, \ldots, J,
\]

(1)

where \( SQ \) and \( P \) are observable neighborhood-level measures of school quality and price, respectively, \( \xi \) is the unobserved variable allowed to be correlated to the observable amenities, and \( \epsilon_{i,c,j} \) is an iid extreme value 1 (EV1) type error.

In this context, the MWTP for school quality of group \( c \) is measured as

\[
MWTP_c = -\frac{\theta_c}{\phi_c},
\]

and \( \theta_c \) and \( \phi_c \) are the parameters of interest.

Collecting all the terms at the group-neighborhood level:

\[
\delta_{c,j} = SQ_j \theta_c + P_j \phi_c + \xi_{c,j}, \quad c = p, np, \quad j = 1, \ldots, J,
\]

(2)

where \( \delta_{c,j} \) can be described as the mean flow utility of group \( c \) for neighborhood \( j \).

Under the assumptions of this simplified model, it is easy to estimate \( \delta_{c,j} \) with information on households’ neighborhood choices as in Berry (1994).\(^{12}\)

\(^{11}\)In the absence of moving costs, only differences in the level of amenities available in the particular period can have an influence in household’s decisions.

\(^{12}\)Following Berry (1994), if \( n_{c,j} \) is the number of households of group \( c \) who live in neighborhood \( j \), and if \( \delta_{c,1} = 0 \) is assumed for each \( c \) (for normalization), then \( \delta_{c,j} = \log(n_{c,j}) - \log(n_{c,1}), \quad j = 2, \ldots, J. \) Under the
focus on the endogeneity issues that may arise in the identification of $\theta_c$ and $\phi_c$ in equation (2) once $\delta$, $SQ$ and $P$ are observed.\footnote{Section \ref{section:identification} addresses the concern that $\delta$ is estimated rather than observed directly from the data.}

2.2 The Endogeneity Problem

The variable $\xi$ in equation (2) accounts for the unobserved neighborhood amenities. Thus, $P$ is correlated to $\xi$ as any amenity that is observed to households should affect price, including those unobserved to the researcher. This source of endogeneity has long been discussed in the literature of demand estimation (e.g., Bresnahan (1987), Berry, Levinsohn & Pakes (1995) and Bayer, Ferreira & McMillan (2007)).

An additional endogeneity issue has been the main concern of the literature on valuation of school quality (e.g., Black & Machin (2011)). To the extent that $SQ$ is correlated to $\xi$, the valuation for school quality will not be identified, as it will reflect the preferences for other unobserved amenities correlated to school quality.

In this paper, I argue that researchers generally do not observe data in a point in time where on average only school quality changes, even in experimental approaches. This implies that the only way of solving the endogeneity problem to identify the MWTP parameter is by controlling for contemporaneous unobservables.

To illustrate the idea, suppose that in period $t$ we randomly assign neighborhoods to either the treatment or the control group, and in the treatment group we improve school quality somehow. In period $t$, we have an experiment: neighborhoods in both groups are on average the same except for the difference in school quality. In period $t + 1$, sorting happens as a function of the original change: households who value school quality the most tend to sort to the neighborhoods in the treatment group, and households who value school quality the least tend to sort to the ones in the control group. In period $t + 1$, we no longer have an experiment: neighborhoods in the treatment and control groups are on average different not only in terms of school quality, but also in terms of neighbors’ characteristics. These groups are likely to remain different or get even more different in periods $t + 2$ onwards, as people with different preferences for neighbors’ characteristics sort also.\footnote{The sorting may also occur because of heterogeneity of preferences for price. What is important for the argument is that it would be a coincidence if neighbor’s characteristics go back to the same levels of period $t$, so that we would have the conditions of an experiment again.} It follows that an experimental design identifies the valuation for school quality per se only in period $t$.\footnote{Simplified assumption of no moving costs there is a direct link between households’ neighborhood choices and their flow utility. In section \ref{section:identification} this assumption will be relaxed, which will make the identification of $\delta_{c,j}$ more cumbersome.}
In this example, a period is abstractly defined as the minimum unit of time for sorting to occur. Since sorting is likely to happen more frequently than researchers observe the data, it is generally unlikely to observe the data at exactly period $t$, so a change in school quality is generally observed by researchers in the same period as changes in amenities that were affected by that change, such as characteristics of neighbors. It follows that the only way to identify the valuation of school quality per se is by explicitly controlling for contemporaneous unobservables.\footnote{Actually, even if the data was observed in period $t$, as might be the case in high frequency repeated sales data, the direct valuation would be identified only under two strong assumptions: first, that the price fully adjusted to the increase in demand that was not yet realized; second, households expect the change in school quality, but do not expect the change in neighbor’s characteristics due to the original change in school quality.}

As a consequence, any attempt to measure the willingness to pay for school quality not explicitly controlling for $\xi$ will be measuring the total change in the valuation of the neighborhood due to the change in school quality, which is the sum of the change in the valuation of the neighborhood due to school quality per se (the direct valuation) and the change (up to that point) in the valuation of the neighborhood due to the change in neighbor’s characteristics induced by the original change in school quality (the indirect valuation). Figure 1 illustrates the idea in a path diagram.\footnote{More formally, because of sorting a case can be made for the addition of an equation $\xi = f(SQ) + error$ in structural form. This implies for instance that relevant (i.e., correlated to $SQ$) candidates for instrumental variables will be generally invalid (i.e., correlated to $\xi$). This argument holds in both discrete choice and hedonic price frameworks, and for any amenity. The first chapter of Caetano (2009) describes this endogeneity issue in more detail.}

In order to identify the MWTP for school quality, it is important to control fully for any indirect valuation for two reasons. First, because the indirect valuation is a function of valuations of amenities other than school quality. Second, because not controlling for it will provide MWTP estimates with contamination bias. As illustrated in the example, any positive indirect valuation in each particular neighborhood is associated with a negative indirect valuation in another neighborhood.\footnote{The indirect valuation due to sorting may not add to zero in the aggregate if non-linearities exist.} Households leave neighborhoods in the control group to go to neighborhoods in the treatment group and vice-versa. This bias is commonly studied in the program evaluation literature, and happens when the control group is indirectly affected by the program through spillovers generated by the effect in the treatment group (e.g., Heckman, Lalonde & Smith (1999) and Miguel & Kremer (2004)). It follows that the difference between treatment and control groups will not reflect the valuation of school quality, as it will include the indirect valuation not once, but twice, when it should not be included at all.\footnote{Formally, this can be understood as a change in $SQ_j$ not only causing a change in $\xi_j$ (as seen in the figure) but also a change in $\xi_k$ for some $k \neq j$.}
In the next subsection I propose a method to fully control for the indirect valuation. The method is a generalized version of a technique known as Proxy-IV, which was first proposed by Chamberlain (1977) and Pudney (1982) and was used by Heckman & Scheinkman (1987) in the estimation of wage regressions.

2.3 The Proxy-IV Method

The Proxy-IV method leverages substantially out of the linearity assumption in equation (2). $\xi_{c,j}$ in equation (2) is decomposed in two latent terms, where one is controlled and the other is assumed exogenous. More precisely, if $\xi_{c,j} := Q_{j} \lambda_{c} + \mu_{c,j}$, then equation (2) can be re-written as

$$\delta_{c,j} = SQ_{j} \theta_{c} + P_{j} \phi_{c} + Q_{j} \lambda_{c} + \mu_{c,j}, \quad c = p, np, \quad j = 1, \ldots, J,$$

(3)

where the following identification condition is assumed to hold:

**Assumption 2.1. Selection on Unobservables**

$$E[\mu_{c,j}|SQ_{j}, P_{j}, Q_{j}] = 0, \quad c = p, np, \quad j = 1, \ldots, J.$$

Intuitively, the idea behind the Proxy-IV method is to use the $\delta$s of one group (e.g., non-parents) as a proxy for $Q$ in the equation of another group (e.g., parents). Since both parents and non-parents are exposed to the same level of amenities in each particular neighborhood, including $Q$, under certain assumptions non-parental valuation of neighborhood $j$ (i.e., $\delta_{np,j}$) should provide all information about $Q_{j}$ necessary to absorb the
source of endogeneity in the equation of $\delta_{p,j}$.

Substituting $Q$ of the non-parental equation into $Q$ of the parental equation in (3):

$$
\delta_{p,j} = SQ_j, \tilde{\theta} + P_j, \tilde{\phi} + \delta_{np,j}, \tilde{\lambda} + \tilde{\mu}_j, \quad j = 1, \ldots, J,
$$

(4)

where

$$
\tilde{\theta} := \theta_p - \theta_{np}, \tilde{\lambda},
$$

(5)

$$
\tilde{\phi} := \phi_p - \phi_{np}, \tilde{\lambda},
$$

(6)

$\tilde{\lambda} := \frac{\delta_{np}}{\delta_{np}}$ and $\tilde{\mu}_j := \mu_p, j - \mu_{np}, j, \tilde{\lambda}$.

In contrast to equation (3), equation (4) includes only observed variables, but at a cost: equations (5) and (6) show that the use of $\delta_{np,j}$ as a proxy variable for $Q_j$ in equation (4) made $\theta_p$ and $\phi_p$ to be no longer directly identified out of the $SQ$ and $P$ coefficients.\(^{19}\)

The Proxy-IV method therefore requires additional assumptions about the preferences to identify the coefficients of interest. This section describes two particularly simple assumptions in the spirit of the ones that will be made in the next section to identify the MWTP for school quality, but other assumptions can be made depending on the prior knowledge of the researcher about the problem at hand.\(^{20}\)

An assumption that identifies $\theta_p$ in equation (5) is:

**Assumption 2.2. Exclusion Restriction on $SQ$:** $\theta_{np} = 0$

Assumption 2.2 asserts that non-parents do not value school quality in the flow utility sense. Note that this assumption only asserts that non-parents do not value the school quality services provided this year in the neighborhood, since at the moment they do not have any child attending school. In particular, it allows them to be forward looking agents and to value school quality this year in the cumulative utility sense, in a more general framework with moving costs. Additionally, in the flow utility sense non-parents are allowed to value neighborhood amenities that are often proxied by school quality. For instance, non-parents may value living close to well educated neighbors, but they do not value (in the flow utility sense) school peer effects due to children of well educated

---

\(^{19}\)This happened because $SQ$ and $P$ are neighborhood level variables. The coefficients of interest will be directly identified in equation (4) if $SQ$ and $P$ were observed instead at the neighborhood-group level. In that case, not only $\delta_{np,j}$, but also $SQ_{np,j}$ and $P_{np,j}$ would need to be used as proxy for $Q_j$.

\(^{20}\)See section 3.3 for a more detailed discussion about these assumptions.
neighbors, so the “school amenity” is included in $SQ$ and the “neighborhood amenity” is included in $Q$.

A similar assumption to 2.2 about the coefficient of $P$ is not plausible, as everyone is likely to care about price even in the flow utility sense. An alternative assumption that still identifies $\phi_p$ in equation (6) is that parents and non-parents value money with the same intensity.

**Assumption 2.3. Exclusion Restriction on $P$:** $\phi_{np} = \phi_p$

Under these assumptions, $\theta_p$ and $\phi_p$ are identified in equation (4), hence the MWTP for school quality is identified.

Note that not necessarily all relevant unobservables will be controlled with this method. In fact, the approach can control for only unobserved amenities valued by groups used as proxy. In the example of this section, unobservables such as neighbor’s characteristics will be proxied, but unobservables such as the presence of children’s park in the neighborhood will not be proxied, as non-parents do not value such an amenity in the flow utility sense. Moreover, to be included in $Q$ the unobserved amenity needs to be non-excludable. That is, by assumption, different groups are all exposed to the same level of the unobserved amenity $Q$ once they choose the same neighborhood, although they may value that amenity differently (i.e., $\lambda$ may vary with $c$).

In a more general setting, but still under the assumption of linearity in equation (3), the method offers substantial flexibility: (a) More than one group can be used as proxy variable in order to account for more than one unobserved amenity, hence weakening assumption 2.1; (b) Groups can be defined so as to weaken assumptions 2.1, 2.2 and 2.3; (c) Assumptions 2.2 and 2.3 can be relaxed; (d) Over-identification tests can be performed. All these extensions will be implemented in section 4. Section 3.3 presents the generalized version of the Proxy-IV method implemented in section 4.

### 3 Identification Strategy

This section explains the strategy to identify the MWTP for school quality per year-child-grade developed in this paper. I begin by presenting a standard dynamic model of neighborhood choice in the spirit of Rust (1987). I then show the identification strategy.

---

\[\text{See section 5 for a discussion on this topic.}\]

\[\text{Under the testable assumption that } \lambda \neq 1, \phi_p \text{ is identified since } 1 - \lambda \text{ is identified through the coefficient of } \delta_{np,j} \text{ in equation (4).}\]

\[\text{As } \text{Chamberlain (1977) noted, the use of } \delta \text{ as proxy for } Q \text{ introduces an omitted variables problem easily solvable under the same assumptions with the use of } \delta \text{ of a third group as IV. For this reason, Chamberlain (1977) referred to this method as Proxy-IV. Section 3.3 provides further discussion on this topic.}\]
in three steps. In the first step, the mean flow utility is identified. I suggest a non-standard estimation procedure that has some advantages and some disadvantages over standard techniques. For this paper, one particularly important advantage is that it offers a feasible way to estimate a dynamic model with a cross-section data. The second step uses a generalized version of the Proxy-IV method described in section 2 to estimate the MWTP for school quality per year-family. Since families may have more than one child, and these children may be attending different grade levels, the third step shows the identification of the MWTP per year-child-grade out of the MWTP per year-family.

3.1 A Dynamic Model of Neighborhood Choice

The setup of the model is similar to the standard dynamic models of the literature (e.g., Aguirregabiria & Mira (2007)). At each period, each household chooses a neighborhood with full information of the current period and only partial information of the future, with the unknown part of the future assumed to have a known distribution.

At each period $t$, each household chooses a neighborhood knowing that, with some cost, it may choose to move out of this neighborhood in the next period. Because there are moving costs, each household’s current decision also affects its decision in future periods, so households take into account the option value of their choice.

There are $J$ neighborhoods and $C$ groups, with $n_{c,t}$ households in group $c$ in period $t$. Each neighborhood is characterized as a set of amenities, such as public school quality and average rent. The future discount, denoted $\beta$, is the same for all households, and assumed known to the researcher. A household in group $c$ expects to live for $T_c$ more periods, but the researcher does not observe this value.

Specifically, in the beginning of each period $t$ household $i$ of group $c$ observes the state variable $S_{i,c,t}$ and decides on a sequence of neighborhoods from period $t$ onwards, denoted $(j_{i,c,t}) = \{j_{i,c,t}, j_{i,c,t+1}, \ldots, j_{i,c,T_c}\}$. Let $d_k(j)$ be equal to 1 if $k = j$, and zero otherwise. Thus, $d_j(j_{i,c,t}) = 1$ if and only if, in period $t$, household $i$ of group $c$ chooses $j_{i,c,t} = j$.

For each household, the value of the path $(j_{i,c})_t$ is the current valuation plus the discounted sum of the valuations of future periods. Since the value of $S_{i,c,s}$ for $s > t$ is not known as of period $t$, the household chooses the path which maximizes the expected cumulative utility, conditional on all information available in period $t$. Thus, the household’s

---

24 See Keane & Wolpin (1997) and Keane & Wolpin (2001) for important exceptions to this rule.
25 In section 4, $\beta$ is assumed equal to .95, although $\beta = .9$ and $\beta = .99$ are also shown to generate similar MWTP estimates.
26 This is a departure from standard models, which assume $T_c$ known (e.g., Aguirregabiria & Mira (2007)).
dynamic optimization problem can be expressed as the value function

\[
V(S_{i,c,t}) := \max_{(i,c,t)} \mathbb{E} \left[ \sum_{s=t}^{T} \beta^{s-t} \sum_{J} d_{j}(j_{i,c,s}) U_{j}(S_{i,c,s}) \mid S_{i,c,t} \right].
\]  

(7)

where \(U_{j}(S_{i,c,t})\) is the household’s flow indirect utility of neighborhood \(j\) in period \(t\).

The choice-specific value function is defined as

\[
V_{j}(S_{i,c,t}) := U_{j}(S_{i,c,t}) + \beta \mathbb{E} \left[ V(S_{i,c,t+1}) \mid S_{i,c,t}, j_{i,c,t} = j \right].
\]  

(8)

I decompose the state variable in two components: \(S_{i,c,t} := (W_{i,c,t}, \epsilon_{i,c,t})\). Each household, as of period \(t\), knows the distribution of each future realization of \(W\), conditional on its choice in period \(t\) and on \(W_{i,c,t}\). However, the household cannot predict \(\epsilon_{i,c,t+1}\) until it reaches period \(t+1\). Before then, it treats \(\epsilon_{i,c,t+1}\) as a random variable with a known distribution, also known to the researcher. The following assumption decomposes the flow indirect utility according to these two components of the state variable:

**Assumption 3.1. Additive Separability and Logit Error**

\[
U_{j}(S_{i,c,t}) := u_{j}(W_{i,c,t}) + \epsilon_{i,c,t,j}.
\]

where \(\epsilon_{i,c,t,j}\) is iid as Extreme Value 1 type (EV1).

The mean flow utility of group \(c\) for neighborhood \(j\) in period \(t\) is defined as the expected value of the flow utilities of neighborhood \(j\) across all households of group \(c\) in period \(t\):

\[
\delta_{c,t,j} := E(u_{j}(W_{i,c,t})).
\]  

(9)

The mean flow utility is the group’s current period valuation of choice \(j\), and, similarly to equation (4), is written linearly as a function of the neighborhood amenities:

\[
\delta_{c,t,j} = SQ_{i,j} \theta_{c,j} + P_{i,j} \phi_{c,j} + \xi_{c,t,j},
\]  

(10)

where \(SQ_{i,j}\) and \(P_{i,j}\) are respectively measures of the level of public school quality and the price of neighborhood \(j\) in period \(t\), \(\theta_{c,j}\) and \(\phi_{c,j}\) are preference parameters, and \(\xi_{c,t,j}\) is an unobserved term that varies at the group-period-neighborhood level.

In the estimation of this model in section 4, each period corresponds to one year, with the current year being 2000. The neighborhoods correspond to school attendance areas inside a school district. There are 40 groups of parents and non-parents in the data set.
For parents, there are 20 groups with index ranging from 0 to 19. The index refers to the age of the oldest child in the household, and was defined so as to have a direct connection with the grade level that child is attending in public school, so that children attending kindergarten through twelfth grades will be of age 6 through 18, respectively. For non-parents, there are also 20 groups with index ranging from 31 to 50, referring to the age of the head of the household.

The next three sections explain the three sequential steps to estimate the MWTP for school quality per year-child-grade. Section 3.2 presents the identification of the dependent variable in equation (10): the mean flow utility. In section 3.3, \( \theta_{c,t} \) and \( \phi_{c,t} \) in equation (10) are identified using the Proxy-IV method. These coefficients only identify the MWTP for school quality per year-family, so section 3.4 shows how the MWTP for school quality per year-grade-child is identified from that.

### 3.2 Step 1: Identification of the Mean Flow Utility

Because of the presence of moving costs, there is no direct connection between households' observed neighborhood choices and their mean flow utility. Proposition 1 establishes this connection under standard assumptions of the literature (e.g., Rust (1987); Aguirregabiria & Mira (2007)):

**Proposition 1.** Under assumption 3.1 and other standard assumptions (see appendix), I can write

\[
V_j(S_{i,c,t}) = v_j(W_{i,c,t}) + \epsilon_{i,c,t,j},
\]

where \( \epsilon_{i,c,t,j} \) is the same as in assumption 3.1, and

\[
\delta_{c,t,j} = E\left[ v_j(W_{i,c,t}) - \beta \sum_{W \in \mathcal{W}} \left( \gamma + \log \sum_{r=1}^{j} \exp(v_r(W)) \right) P(W|W_{i,c,t}, j_{i,c,t} = j) \right]
\]

where \( \gamma \approx 0.577 \) is the Euler’s constant\(^{27}\)

**Proof.** This result is a corollary of Rust (1987). Its proof as well as a detailed discussion of its assumptions are provided in the appendix.

Although this proposition makes only standard assumptions in the literature of dynamic demand estimation, it suggests a non-standard procedure to estimate the mean flow

\(^{27}\gamma \) is equal to the mean of the extreme value distribution.
utility $\delta_{t,j}$. Equation (11) shows a structure for $V_j(S_{i,c,t})$ analogous to the one in assumption 3.1 but for the choice-specific value functions. Because the errors are distributed as EV1, I can specify $v_j(W_{i,c,t})$ instead of $u_j(W_{i,c,t})$ and estimate $v_j(W_{i,c,t})$ using standard static discrete choice techniques (e.g., Train (2003)). The mean flow utility can then be recovered by the estimation of $v_j(W_{i,c,t})$ and of $v_j(W)$ for each future $W$ because of equation (12) as long as the conditional transition probabilities are known or estimated.

In comparing standard methods of dynamic demand estimation (e.g., Rust (1987); Aguirregabiria & Mira (2007)) with this non-standard method, some important differences emerge. On the one hand, this procedure may be disadvantageous compared to standard procedures, since the choice-specific cumulative utilities are much more complex to specify than the flow utilities, as they are a function not only of the amenities of the chosen neighborhood, but also a discounted weighted function of the amenities of all neighborhoods depending on each household’s expectations of moving in the future. On the other hand, practically, techniques in static discrete choice estimation such as the ones suggested by Berry (1994) and Berry, Levinsohn & Pakes (1995) allow for the feasible estimation of fixed effects at the group-neighborhood level, which provide the opportunity to estimate a flexible and yet feasible specification of the choice specific value functions which would not be achieved if the flow utilities were specified instead. Unlike the standard procedure, this allows for $W$ to contain unobservable components varying at the group-period-neighborhood level. Moreover, as discussed above, it allows $T_c$ to be unknown to the researcher. Finally, from the perspective of implementation, under additional assumptions this non-standard procedure allows for the feasible estimation of a dynamic model with only a cross-sectional data.

The current choice-specific value functions are specified as

$$v_j(W_{i,c,t}) := \Delta_{t,j} + \text{NEW}_{i,c,t,j} \Phi_{c,t} + f(W_{i,c,t}) \Gamma_{c,t}.$$
where \( W_{i,c,t} := \{ \Delta_{c,i,j}^{j=1}, \Phi_{c,t}, \{ NEW_{i,c,t,j} \}_{j=1}^{J} \} \). NEW_{i,c,t,j} is an indicator variable for whether neighborhood \( j \) is new for household \( i \) in period \( t \): \( NEW_{i,c,t,j} = 1 \) if \( j_{i,c,t-1} \neq j \), and zero otherwise, so that \( \Phi \) is the moving cost parameter. The fixed effect \( \Delta_{c,i,j} \) captures, for group \( c \) as a whole, the average flow value of living in neighborhood \( j \) in period \( t \) plus all the perceived continuation value conditional on living in that neighborhood in period \( t \). For instance, if households in group \( c \) on average believe that, conditional on living in neighborhood \( j \) now, next year they will be more likely to be in neighborhood \( l \neq j \) than households in other groups, then this will be reflected in \( \Delta_{c,i,j} \). It will also be reflected in \( \Delta_{c,i,j} \) if households of different groups have or expect in the future to have different preferences for amenities.

The following assumption constrains \( W_{i,c,t+1} \) so that it can be inferred using only data on period \( t \).

**Assumption 3.2. Synthetic Cohorts**

For \( X = \{ \Delta, \Phi \} \), \( t = 2000 \) and for each \( j \):

If \( c = 0, \ldots, 18 \):

\[
X_{c,t+1,j} = X_{c+1,t,j}
\]

If \( c = 31, \ldots, 49 \):

\[
X_{c,t+1,j} = X_{0,t,j} \quad \text{with probability} \quad \Pi_{c,j}^{0}
\]

\[
X_{c,t+1,j} = X_{c+1,t,j} \quad \text{with probability} \quad 1 - \Pi_{c,j}^{0}
\]

where \( \Pi_{c,j}^{0} := n_{c+1,j}/(n_{c+1,t} + n_{c+1,t}) \), \( n_{c+1,t} \) is the number of households of group \( c + 1 \), and \( n_{c+1,j} \) is the number of households whose head is \( c + 1 \) years of age and whose oldest child is zero years of age (i.e., a newborn).

For \( c = 0, \ldots, 18 \), assumption 3.2 implies that parents in group \( c \) expect to have next year the same mean cumulative valuation of each neighborhood (and moving costs) as groups one year older have this year. The assumption is similar for non-parents (i.e., \( c = 31, \ldots, 49 \)), but the transition is more complicated, since they may become parents next year. There is a likelihood \( \Pi_{c+1}^{0} \) of becoming a parent one year older, and a likelihood \( 1 - \Pi_{c+1}^{0} \) of becoming a non-parent one year older. Households by assumption perceive

---

33One can understand the dynamic problem as a nested optimization problem where each period is a nest. In this case, \( \Delta_{c,i,j} \) identifies the inclusive value of neighborhood \( j \) for group \( c \) in period \( t \). See [Train (2003)].
these probabilities to be the ones observed in period \( t \).

This assumption is closely related to the assumption of synthetic cohorts from the demography literature (e.g., Newell (1990)). One instance in which it would be invalid is if households believed that a specific neighborhood is gentrifying. In that case, \( \Delta_{c,i,t+1,j} \) is likely to be larger than \( \Delta_{c,i,t,j} \) or \( \Delta_{0,i,j} \) for that specific neighborhood. The Proxy-IV method developed in section 3.3 will mitigate the repercussions of deviations of assumption 3.2 from the perspective of the identification of the MWTP for school quality to the extent that the expectation of gentrification may be written as an unobserved non-excludable amenity.

Under assumption 3.2, the future choice-specific value functions are specified analogously to the current ones:

\[
v_j(W_{i,c,t+1}) := \Delta_{c,i,t+1,j} + \text{NEW}_{i,c,t+1,j} \Phi_{c+1,j}, \quad c = 0, ..., 18
\]

\[
v_j(W_{i,c,t+1}) := \Delta_{0,i,t,j} + \text{NEW}_{i,c,t+1,j} \Phi_{0,j} \quad \text{with probability } \Pi^0_{c+1,i,t}, \quad c = 31, ..., 49
\]

\[
v_j(W_{i,c,t+1}) := \Delta_{c+1,i,t,j} + \text{NEW}_{i,c,t+1,j} \Phi_{c+1,j} \quad \text{with probability } 1 - \Pi^0_{c+1,i,t}, \quad c = 31, ..., 49
\]

where \( W_{i,c,t+1} := \left(\{\Delta_{c,i,t+1,j}\}_{j=1}^4, \Phi_{c,t+1}, \{\text{NEW}_{i,c,t+1,j}\}_{j=1}^4\right) \) and \( \Pi^0_{c+1,i,t} \) is given in assumption 3.2.

Equations (13) and (14) show respectively the specifications of the choice-specific value functions of period \( t \) and \( t + 1 \) as function only of data on period \( t \) and estimable parameters. Equation (11) and assumption 3.1 guarantee the identification of \( v_j(W_{i,c,t}) \) in equation (13). Its parameters are estimated as in Berry, Levinsohn & Pakes (1995). Assumption 3.2 and the identification of the parameters of \( v_j(W_{i,c,t}) \) guarantee identification of \( v_j(W) \) for each possible future \( W \) in equation (14). From equation (12), and from the

---

34Note that this assumption imposes restrictions with regard to years \( t = 2000, \) and \( t + 1 = 2001 \) only, so the space of time for which the method requires households to not expect gentrification is relatively small. One could relax this assumption with longitudinal data by estimating the actual gentrification from 2000 to 2001 in each neighborhood and assuming households have the correct expectations about it. However, it is important to highlight that this is an assumption about expectations, and not about reality, hence the ad hoc nature of the assumption will continue even in the presence of longitudinal data.

35Assumption 3.2 is necessary to the identification of \( \delta_{c,j} \) through equation (12), however it may be of smaller importance in the identification of the MWTP for school quality. The Proxy-IV method developed in section 3.3 accounts for unobservable variables of the kind \( Q_{c,j}, \lambda_{c,j} \). The bias in the identification of \( \delta_{c,j} \) in equation (12) is of the form \( \nu_{c,j,t} \). Though \( \nu_{c,j,t} \) is not likely expressible in the form \( Q_{c,j}, \lambda_{c,j} \), it may be approximated by one such form, so that only the residual of such approximation has an effect in the final identification of the MWTP for school quality.

36Note that NEW\(_{i,c,t+1,j} \) is observed only with data at period \( t \), since \( j_{i,c,t} \) is observed.
transition probabilities assumed in 3.2 this completes the identification of $\delta_{c,t,j}$.

3.3 Step 2: Identification of the MWTP for School Quality Per Year-Family

Since the mean flow utility was identified in the last section, the next step is the treatment of the endogeneity problem in equation (10). In what follows, the year index $t$ will be suppressed for simplicity, as the data will be observed in year $t = 2000$ only. Re-writing equation (10):

$$
\delta_{c,j} = SQ_j \theta_c + P_j \phi_c + \xi_{c,j}, 
$$

(15)

As discussed in section 2, identification of $\theta_c$ and $\phi_c$ in (15) is unusually difficult in this context because of neighborhood sorting and because $\xi_{c,j}$ generally includes unobservable characteristics of neighbors.

To address this problem, this paper uses a generalized version of the Proxy-IV method discussed in section 2. The term $\xi_{c,j}$ is decomposed into a term including a finite number of unobserved neighborhood amenities, $Q_j$, plus a random term $\mu_{c,j}$. The mean flow utility can be rewritten as

$$
\delta_{c,j} = SQ_j \theta_c + P_j \phi_c + Q_j \lambda_c + \mu_{c,j}, 
$$

(16)

where $Q_j$ is a vector of fixed size $R$. $Q_j$ includes only unobserved amenities that are non-excludable; that is, all households are exposed to the same unobserved amenities once they choose that neighborhood. However, households of different groups are allowed to value these amenities differently.

The next assumption states that all endogeneity of equation (16) can be captured by $R$ non-excludable unobserved amenities, with allowance for different valuation weights across groups.

Assumption 3.3. Selection on Unobservables

For all groups $c$ and neighborhoods $j$, let $\mu_{c,j}$ be as defined in equation (16). Then,

$$
E [\mu_{c,j} | SQ_j, P_j, Q_j] = 0. 
$$

(17)

Intuitively, assumption 3.3 is weaker the larger is $R$. It implies that, if $Q$ was observed, then the parameters of equation (16) would be trivially identified.

See the appendix for details about the estimation.
Following Heckman & Scheinkman (1987)'s notation, I divide the groups in three subsets, each creating a system of equations of its own. The $\delta$'s of the groups in the first subset will be used to provide a proxy for $Q$ in the equations of the second subset. This procedure will solve the original endogeneity problem, but it will introduce a new source of endogeneity which is easily solved under the same assumptions with the inclusion of the $\delta$'s of the groups in the third subset as instrumental variables for the $\delta$'s of the groups in the first subset.

Formally, if $R_s$ is the number of groups in subset $G_s$, I write three systems of equations, one for each subset:

$$\delta_{j(1)} = SQ_j \theta_{(1)} + P_j \phi_{(1)} + Q_j \lambda_{(1)} + \mu_{j(1)},$$  \hspace{1cm} (18)
$$\delta_{j(2)} = SQ_j \theta_{(2)} + P_j \phi_{(2)} + Q_j \lambda_{(2)} + \mu_{j(2)},$$  \hspace{1cm} (19)
$$\delta_{j(3)} = SQ_j \theta_{(3)} + P_j \phi_{(3)} + Q_j \lambda_{(3)} + \mu_{j(3)},$$  \hspace{1cm} (20)

for $j = 1,...,J$

where $SQ_j$ and $P_j$ are scalars, $Q_j$ is a $1 \times R$ vector, $\delta_{j(s)}$, $\mu_{j(s)}$, $\theta_{(s)}$ and $\phi_{(s)}$ are each $1 \times R_s$ vectors, and $\lambda_{(s)}$ is a $R \times R_s$ matrix, $s = 1,2,3$.

**Assumption 3.4.** $\lambda_{(1)}$ has full rank $R$.

Intuitively, assumption 3.4 states that there is sufficient heterogeneity in the preferences of the unobserved amenities across groups in the first subset. It implies that any change in the unobserved amenities will be translated into a change in the mean utilities of the groups in $G_1$ as a whole, i.e., $\delta_{j(1)}$ will span the space generated by the columns of $Q$. Solving equation (18) for $Q$ and substituting it in equation (19):

$$\delta_{j(2)} = SQ_j \left(\theta_{(2)} - \theta_{(1)} \tilde{\lambda}_{(2)}\right) + P_j \left(\phi_{(2)} - \phi_{(1)} \tilde{\lambda}_{(2)}\right) + \delta_{j(1)} \tilde{\lambda}_{(2)} + \mu_{j(2)} - \mu_{j(1)} \tilde{\lambda}_{(2)},$$  \hspace{1cm} (21)

for $j = 1,...,J$

where $\tilde{\lambda}_{(2)} \equiv \lambda_{(1)}^{-1} \lambda_{(1)} \lambda_{(2)}^{-1}$ is an $R_1 \times R_2$ matrix measuring the relative preferences of the unobserved amenities for groups in subsets 1 and 2. In equation (21), the

---

38 For example, assume $R = R_1 = 2$, with $c = A,B$ indexing the groups from $G_1$, and let $\lambda_{c,1}$ and $\lambda_{c,2}$ denote the coefficients of the first and the second unobserved amenities, respectively. Then the assumption is that $\lambda_{A,1} \lambda_{B,2} - \lambda_{A,2} \lambda_{B,1} \neq 0$, i.e., both groups cannot value with the same intensity one amenity relative to the other.
mean utilities of groups in \( G_2 \) are written as a function of the observed amenities and the mean utilities of groups in \( G_1 \), and no longer as a function of \( Q \).

Note that, since \( \tilde{\mu}_{j(2)} \) is correlated with \( \delta_{j(1)} \) because of equation (18), equation (21) does not identify \( \tilde{\theta}_{(2)}, \tilde{\phi}_{(2)} \) and \( \tilde{\lambda}_{(2)} \) directly. However, as pointed out by Chamberlain (1977), under the previous assumptions \( \delta_{j(3)} \) can be used as instrumental variables (IV) for \( \delta_{j(1)} \), hence \( \tilde{\theta}_{(2)}, \tilde{\phi}_{(2)} \) and \( \tilde{\lambda}_{(2)} \) are nonetheless identified.

Still, the parameters of interest \( \theta_c \) and \( \phi_c \) for \( c \in G_2 \) are not directly identified out of \( \tilde{\theta}_{(2)}, \tilde{\phi}_{(2)} \) and \( \tilde{\lambda}_{(2)} \), hence additional assumptions need to be made. According to equation (21), the parameters of interest are implicitly given by the two systems of equations:

\[
\tilde{\theta}_c = \theta_c - \theta_{(1)} \tilde{\lambda}_c, \quad c \in G_2, \tag{22}
\]

\[
\tilde{\phi}_c = \phi_c - \phi_{(1)} \tilde{\lambda}_c, \quad c \in G_2. \tag{23}
\]

The systems of equations (22) and (23) have each \( R_2 \) equations and \( R_1 + R_2 \) unknowns. If at least \( R_1 \) restrictions in the parameters of each of these systems are made, then \( \theta_c \) and \( \phi_c \) are identified for each \( c \in G_2 \).

In section 4 I implement the method with \( G_2 = \{6, \ldots, 18\} \) since the oldest child of the households of these groups are observed to be attending respectively kindergarten through grade 12 in a public school. The rest of the groups are divided between \( G_1 \) and \( G_3 \) depending on the specification, provided that \( R_3 \geq R_1 \) holds. The assumptions that guarantee the identification of \( \theta_c, c \in G_2 \), are

**Assumption 3.5. Exclusion Restriction on SQ:** \( \theta_c = 0, \quad c \in G_1 \)

Groups \( c = 31, \ldots, 49 \) are non-parents, so they should not value public school quality in the flow utility sense, as discussed in section 2. Moreover, groups \( c = 0, \ldots, 5 \) are parents but do not have a child attending public school this year, so, following the same logic, they should also not value public school quality in the flow utility sense. This implies that any group except groups in \( G_2 = \{6, \ldots, 18\} \) is an appropriate candidate to be included in \( G_1 \).

For the identification of \( \phi_c, c \in G_2 \), the assumption is that groups used as proxies as well as group 6 have on average the same flow value for the amenity \( P \).

---

39 Assumption 3.3 guarantees validity of \( \delta_{j(3)} \). Its relevance is trivial, as both \( \delta_{j(1)} \) and \( \delta_{j(3)} \) share the same level of amenities.

40 The unknowns are \( \theta_c \) and \( \phi_c \) for \( c \in G_1 \cup G_2 \).

41 Formally, one needs to add at least \( R_1 \) linearly independent equations to each system.

42 Groups 19 and 50 are not included in this step because they are the last groups of non-parents and parents, respectively, so \( \delta_{c,j} \) is not identified for them. See equation (12).

43 For the equations to be linearly independent, testable assumptions on \( \lambda_s \) are implied by assumption 3.6. Intuitively, \( \phi \) are identified only to the extent that \( \lambda \) do not vary across groups similarly to \( \phi \). See footnote 22.
Assumption 3.6. Exclusion Restriction on \( P \):

\[
\phi_c = \phi_0 \quad c \in G_1
\]

Under these assumptions, \( \theta_c \) and \( \phi_c \) for \( c \in G_2 \) are identified, and the GMM estimator is consistent.

The choice of groups in \( G_1 \) is essential not only for assumptions 3.5 and 3.6 but also for assumption 3.3. Intuitively, the larger \( R \) is, the weaker is assumption 3.3, hence the weakness of assumption 3.3 depends on the choice of \( R_1 \), as \( R \) cannot be larger than \( R_1 \). The use of diverse groups in \( G_1 \) may be helpful to span a higher space of \( Q \) if existent. For instance, if average income of neighbors and existence of children’s park are both important unobservables, then using two nonparental groups as proxy will not help to span the space of the children’s park unobservable, as a nonparent does not value it in the flow utility sense. However, using as proxy groups one nonparental group and one group of parents whose oldest child is too young to attend school will span the space of both unobservables, as this latter group do value children’s park in the flow utility sense.

3.4 Step 3: Identification of the MWTP for School Quality Per Year-Child-Grade

Step 2 in section 3.3 estimated the MWTP for school quality per year-family for groups \( c \in G_2 = \{6, \ldots, 18\} \). In the final step, I show how to identify the MWTP for school quality per year-child-grade for grades kindergarten to twelve out of the MWTP estimates from step 2.

Families of group \( c \) may have more than one child, and may have children at different grades. If \( N_{i,c,g} \) is the number of children from family \( i \) of group \( c \) who are observed to be attending grade \( g \) in the data, then the expected number of children attending grade \( g \) across all households from group \( c \) is defined as \( N_{c,g} := E(N_{i,c,g}) \).

Because of the way the groups are defined, families at groups 6 through 18 have their oldest child attending grade kindergarten through grade 12, respectively. The MWTP that is estimated in step 2 is written as a weighted average of the MWTP per year-child-

\[ R = R_1 = 1. \]

\[ ^{44} \text{However, Pudney (1982) shows that the choice of which } R_1 \text{ groups to use as proxy is asymptotically irrelevant in terms of efficiency.} \]

\[ ^{45} \text{However, note that } R \text{ is unknown, so it is actually not known how many unobserved variables are controlled for in equation } (16). \text{ Heuristically, when a new group is included in } G_1 \text{ and the } R^2 \text{ rises, it suggests that one more dimension of } Q \text{ is being spanned. When groups are included in } G_1 \text{ and the } R^2 \text{ does not rise, it suggests that this additional group did not help span an additional dimension of } Q. \]

\[ ^{46} \text{See section } 4 \text{ for a discussion about the link between age of the child and the grade the child is attending.} \]
grade of all households, using information on the grade the children of each household are attending:

\[
MWTP^S_{c,g} = \sum_{g=0}^{c-6} MWTP^S_{c,g} N_{c,g}
\]

(24)

\[c = 6, ..., 18\]

where \(g = 0, ..., 12\) indexes grades kindergarten through grade 12, respectively, and \(MWTP^S_{c,g}\) is defined as the MWTP for school quality per year-child-grade \(g\) for families of group \(c\).

Assumption 3.7 below guarantees identification by stating that a household values the school quality for each child only as function of the grade that child is attending:

**Assumption 3.7.** For each \(g = 0, ..., 12\):

\[
MWTP^S_{c,g} = MWTP^S_{g} \quad \forall c \in G_2
\]

(25)

This assumption implies, for instance, that households’ valuation of school quality per child for a specific grade does not change as a function of birth order, or as a function of sibling spacing, or even as a function of the age of the oldest child:

\(MWTP^S_{g}\), implicitly defined in assumption 3.7, is referred to the *MWTP for School Quality per year-child-grade* \(g\), for grades \(g = 0, ..., 12\), and is the main parameter of interest in this paper.

Substituting equation (25) into equation (24):

\[
MWTP^S_{c,g} = \sum_{g=0}^{c-6} MWTP^S_{g} N_{c,g}
\]

(26)

\[c = 6, ..., 18\]

The system of equations (26) has 13 unknowns (\(MWTP^S_{g}\), \(g = 0, ..., 12\)) and 13 linearly independent equations (\(c = 6, ..., 18\)), which guarantees the identification of the MWTP for school quality per year-child-grade, from kindergarten to twelfth grade.\(^{47}\)

\(^{47}\)Formally, let \(N\) be an upper triangular matrix with \(N_{c,g}\) as its \((c-5,g+1)\)th element, \(c = 6, ..., 18\) and \(g = 0, ..., 12\). Then \(MWTP^S = N^{-1}.MWTP^S\) where \(MWTP^S\) is the \(13 \times 1\) vector with \(MWTP^S_{g}\) as elements, \(g = 0, ..., 12\), and \(MWTP^S\) is the \(13 \times 1\) vector with \(MWTP^S_{c,g}\) as elements, \(c = 6, ..., 18\). The matrix \(N\) is invertible since it is a triangular matrix and its diagonal elements are all strictly positive.
The estimation of steps 2 and 3 is carried out simultaneously via GMM, with unknown parameters $MWTP_{k}^{SQ}, g = 0, ..., 12, \tilde{\lambda}_c$ and $\phi_c, c = 6, ..., 18$.

4 Data and Empirical Results

In this section I present the data and the results of the paper. I also explain in detail how I implement the methodology presented in section 3.

The data set is the restricted-access, or long form, version of the 2000 decennial Census of Population and Housing for the state of Minnesota. It is a $(1:6)$ sample of all households in the U.S., containing detailed information on characteristics of houses, households and individuals within households. The restricted-access version of the Census data contains two additional pieces of information that are crucial to the analysis of this paper. First, it has information on where households are living down to a Census block, which is a geographical area similar to a street block. I use this information to identify neighborhoods as the elementary school attendance areas (SAA), which are each generally linked to one and only one school for each grade. I merge the Census data with average test score data at grade 5 for each school from the Minnesota’s Department of Education.

The second information is the exact date of birth of each individual. Most public schools in U.S. use a birth date cutoff to assign children to kindergarten. If the child will turn five before a given cutoff date, the child enters kindergarten in that year. Otherwise, the child waits one more year. In the case of MN, that cutoff is September first. I use the exact date of birth of the children to define the group to which each family belongs. The group is determined by the age of the oldest child as of September first 2000, so that the group matches the grade the oldest child is attending. The group of parents whose oldest child is 6 is identified as the cohort of parents whose oldest child turned 5 in the 12 months interval prior to September 1, 1999, since then this child would be observed in kindergarten in April first 2000, which is the date the Census is reported. Analogously, the group of parents whose oldest child is 5 is identified as the group of parents whose oldest child turned 4 in the 12 months interval prior to September first 1999, so that they are still not attending school in April 2000.

48 $\tilde{\theta}_c = \theta_c$ is identified by $\tilde{\theta}_c = MWTP_c^{SQ} \phi_c$, where $MWTP_c^{SQ}$ is given by (26).

49 The public-access version of the 2000 Census data has only geographical identifiers down to a Public Use Microdata Area (PUMA), which is an area with at least 100,000 individuals.

50 The test score data used in this paper come from the Minnesota Comprehensive Assessments (MCAs), which is a statewide standardized test in Minnesota covering reading and math at the school level. The data are publicly available at the Minnesota Department of Education (http://education.state.mn.us/MDE/index.html).
Table 1 shows the summary statistics of the observed characteristics of the SAAs as well as of households for MN. These represent the neighborhood observable characteristics in the sample. The sample has 345 SAAs, with on average over 800 houses each. The average rent is around $550, with a standard deviation of $220. Additionally, the average house value is just under $150,000, with standard deviation of over $50,000. The SAAs are also shown to be diverse. For instance, the average income of the neighborhoods is around $65,000 with standard deviation of $25,000, and the average proportion of neighbors with college degree or more is 28% with standard deviation of 17%.

The average test score, which is the measure of school quality used in the paper, is 1387, with standard deviation of 223. The amount of variation in this variable is very similar to the amount of variation of the test score used in Bayer, Ferreira & McMillan (2007).

I restrict the analysis to 40 groups, which are disjoint subsets of the total population of households. I define the groups in two different ways, depending on whether the household has a child (i.e., parents), or not (i.e., non-parents). For parents, I define the groups by the age of the oldest child, ranging from ages 0 to 19. For non-parents, I define the groups by the age of the head of household, ranging from ages 31 to 50.

Table 2 shows the summary statistics of the observed characteristics of the groups of parents used in the analysis. It is interesting to note that proportion of race and employment status do not vary much across groups. The average household income is between $56,000 and $71,000 for parents of each group, and the average income is very similar for consecutive groups. The average proportion of parents who are homeowners vary from 75% to 90%, with older parents being more likely to be homeowners. The moving rate also depends substantially on the group, with younger parents moving more than older parents. Moreover, the proportion of parents who have two children and three or more children naturally grows as groups get older. The likelihood of the oldest child attending a private school is very high for parents whose oldest child is not old enough to attend public schools, but becomes lower and remains around 10% for later groups. Finally, the proportion of parents who are employed is substantially high, around 95%.

Analogously, Table 3 shows the summary statistics of the observed characteristics of the groups of non-parents used in this analysis. The household income for groups of non-parents is similar to parents and across groups. In contrast to parents, proportion of non-parents who are employed are relatively low and varies around 55%. The propor-

---

31 Bayer, Ferreira & McMillan (2007) uses a measure of school quality with the standard deviation amounting to 14% of the mean of the variable, which is very similar to the 16% of this paper. Black (1999), however, has a school quality measure with standard deviation equals to 5% of the mean.

32 The average age of the parents whose oldest child is 0 year old is 31 years, so parents of the first group are on average exactly in the same generation of non-parents of the first group.
tion of non-parents who are homeowners is also little lower than for parents, with older non-parents being homeowners with a higher likelihood. The moving rate depends substantially on the group, as for parents, with older non-parents moving much less than younger ones.

Tables 2 and 3 show that there is substantial variation across groups in several important characteristics that may affect the preference for school quality as well as the preference for other amenities, some of which may be unobserved and correlated to school quality. This is crucial to the implementation of the approach developed in this paper, since it exploits variation across groups with respect to the preferences of observed and unobserved amenities.

Table 4 shows the main results of the paper. This table shows the MWTP per year, child and grade as of % of rent for an increase of 5% in school quality. The rows represent the grades, and the columns represent different specifications. Column I shows the MWTP estimates implied by a panel regression of the mean flow utilities of groups 6 through 18 on the average school quality per SAA and the average rent per SAA, without any other control variable, nor controlling for unobservables. The GMM estimates of this regression imply a MWTP of around 10% of rent for primary school, with 16% for kindergarten, very low and not significative values for middle school, and values for high school around 7%. Column II shows the MWTP estimates of the same panel regression, still not controlling for unobservables, but adding a list of control variables which account for observed amenities, such as neighborhood average income, proportion of blacks and proportion of highly educated neighbors. The estimates of this regression show a substantial increase on the MWTP for school quality in comparison to the baseline specification. This result is expected, since school quality and price are likely positively correlated to unobserved amenities, with price being likely the largest source of endogeneity. The MWTP estimates are very high for elementary and high school grades, with 38% for kindergarten and 23% for 12th grade, but around 6% for middle school.

Columns III through VI of table 4 show the results of the GMM estimation using the Proxy-IV method of controlling for unobservables, as described in steps 2 and 3 of section. I divided the groups in three sets: the proxy set, the set of interest and the set of instruments. The set of interest includes groups 6 through 18, which correspond to the households with school age children that attend a public school in the correctly

---

53 This table shows results for \( \beta = .95 \). The results do not change significantly for choices of different \( \beta \)s in the range .90 through .99.

54 The detailed list of controls can be seen in the notes at the end of Table 4.

55 Under the presence of endogeneity in the baseline regression, both \( \theta \)s and \( \phi \)s are likely overestimated, but \( \phi \)s are likely more so, making the MWTP coefficients likely underestimated.
pre-specified grade.  

In all specifications, I make the following assumptions: first, I assume that the groups used as proxy variables do not value school quality in the flow utility sense (assumption 3.5). Second, I assume that the group of age 6 values rent on average the same as each group used as proxy (assumption 3.6). As one can see in table 2, these groups have relatively similar distribution of household income, which is likely to be strongly correlated to the preference for average rent. With these assumptions, I can estimate the coefficients of school quality and average rent for each group, from 6 to 18, hence obtaining an estimate of the MWTP for school quality per year, child and grade, as described in section 3.

In each of columns III to VI, I change the number of proxies used. Intuitively, the more diverse proxies are added, the higher is the space of unobservables that is spanned. As I add more proxy variables, I assign more groups to the set used as instrumental variables. Column III shows the results using one proxy: the mean flow utilities of group 34. This regression does not include any observed amenity, except average school quality and average rent. The results are shown to be similar to the results of column II, even though the specification of column II adds a list of several observed control variables. Moreover, the \( R^2 \) of the regression of column III is substantially larger than the \( R^2 \) from column II, suggesting that allowing for one generic unobserved amenity controls for more variation of the dependent variable than the list of observed amenities used in column II.

It may be the case that there exists an unobserved amenity that is correlated to school quality that is valued (in the flow utility sense) by parents with children of school age, but not valued by non-parents. Availability of children’s parks may be an example of such amenities. Column IV adds another proxy variable to control for more unobserved amenities: the mean flow utilities for group 4. The MWTP estimates reduce a lot, especially for elementary school, as expected, since this group used as proxy is more likely to control for amenities that are valued by parents with children attending elementary schools. Column V adds one more proxy: the mean flow utilities from group 41. It may be the case that two unobservables are not enough to control for the endogeneity problem. Indeed, there can be an unobservable amenity that only relatively older people value or

---

56 The estimate of the mean flow utility was calculated across all households of group \( c \) that attended a public school at the correctly pre-specified grade (i.e., the grade that an age \( c \) student is supposed to be attending in MN). In total there are between 1900 and 4000 such households included in the estimation per group, depending on the group. It is important that this number is large enough so that the estimates are as close as possible to the true value. See Berry, Linton & Pakes (2004) for a useful discussion about the asymptotic properties of the Berry, Levinsohn & Pakes (1995) approach.

57 For each specification, the lists of groups used as proxies and of groups used as instrumental variables are included in the table.
are exposed. The results reduce even more for both primary and secondary school.

Finally, column VI adds to the specification of column V another proxy variable: group 2. The fit does not improve significatively, and the results are similar to the ones in column V. Columns V and VI together show that three unobservable amenities seem to be enough to take care of the endogeneity with regard to the variables school quality and average rent.\footnote{Instead of group 2, other groups were also added with similar results.}

As an additional robustness test, for each specification using the Proxy-IV method I provide the p-value of the over-identification test (i.e., J test). This test can be understood as a joint test for assumptions \ref{assumption3.3}, \ref{assumption3.5} and \ref{assumption3.6}. Rejecting the null hypothesis is evidence that at least one of these assumptions is not valid for that specification. Indeed, specification III is rejected at 1\% of significance, and specification IV is rejected at 5\% of significance. Specifications V and VI are not rejected even at 30\% of significance, providing further evidence that three proxy groups, two of non-parents and one of parents, seem to take care of the original endogeneity problem.

The results of our preferred specifications (specifications V and VI) show that parents are willing to pay 11\% more each year to send one of their children to attend a school that is 5\% better if the child is attending elementary school. The corresponding estimates for middle school and secondary school are respectively 6\% and 13\%. The value for the first grade seems to be larger than the remaining elementary school grades, and the value for grades 10 and 12 are much larger than the remaining secondary grades.

Table 5 shows the magnitude of the effects found in specification VI in table 4. These numbers are calculated over the average rent in MN, which is $544. Parents are willing to pay $2370 more per year to send one of their children to attend primary school in a school that is one standard deviation better. The corresponding valuations are $1374 for middle school and $2691 for secondary school.

5 Discussion

This section discusses important remarks with regard to the methodology developed in this paper as well as to the interpretation of the results in the previous section.

5.1 Comparison with private school tuition

The estimates of this paper can help shed some light on the level of competition of public and private schools across grades by creating a useful benchmark. An average family is willing to pay $2370 a year for its child to attend a one standard deviation better public
school in elementary grades. To the extent that a private school with quality similar to that public school costs more than $2370, that family is likely to stay in the public school system.

The estimates imply that the MWTP for a one standard deviation improvement of school quality inside the public school system is cost-equivalent compared to attending the 56 bottom (i.e., top 44) percentile of the private school tuition distribution for elementary grades, 26 bottom percentile for middle school and 23 bottom percentile for secondary school. To the extent that private school tuition is a strong indicator of school quality, these estimates suggest that private schools exert more competing pressure to public schools in elementary grades than in middle and secondary grades. It follows that policies with the goal of increasing competition between public and private schools may want to target elementary schools. Indeed, most applications to voucher programs in the U.S. are at the elementary level.

5.2 Comparison with the BFE Approach

There are mainly two reasons why the BFE approach and the proposed approach provide different estimates, and understanding these reasons help shed light in the contributions of this paper. First, these methods attempt to solve the endogeneity problem in different ways. While the BFE approach suggests a selection on observables approach conditional on boundary fixed effects, the Proxy-IV approach proposed in this paper suggests a selection on unobservables approach, by controlling for unobservables that other groups of households value. If there is an unobservable amenity varying across attendance boundaries which is valued by these groups of households, such unobservable can only be dealt with the Proxy-IV approach. Alternatively, if there is an unobservable amenity that is not valued by the households in the proxy groups and does not vary across the attendance boundary, such unobservable can only be dealt with the BFE approach. Moreover, controlling for observable characteristics of neighbors as done in the augmented version of the BFE approach may change the interpretation of the MWTP estimate. If characteristics of neighbors are a good proxy for the family background of the peers inside the school, then the estimate would be interpreted as the MWTP for the component of school quality that is orthogonal to peer effects. The Proxy-IV method controls for the characteristics of

59 A full analysis on this issue is more complex than suggested here. Families may consider other differences between a private and a public school besides school quality, such as religious curriculum and location. Moreover, the supply of available neighborhoods may be such that families are not able to be indifferent between the previous and the new neighborhood except for school quality and prices. This complex analysis is beyond the scope of this paper.

neighbors outside of the school (i.e., neighborhood amenities) without controlling for the characteristics of neighbors inside of the school (i.e., school amenities). This feature is particularly interesting in the context of school quality, as it enables the MWTP estimates to be comparable to the tuition offered by private schools. Indeed, all parents living close to a private school are exposed to their neighbors, however they have access to the peer effects inside the school only if they pay for tuition.

The second main reason why the BFE approach and the proposed approach may provide different estimates is that they identify the MWTP parameter at different units of measurement. The estimates of this paper are at the year-child-grade level, while the BFE estimates are at a different unit. The particular unit at which the MWTP parameter is estimated in the BFE literature depends on whether the parameter is estimated by the hedonic or the discrete choice approach. As [Bayer, Ferreira & McMillan (2007)] pointed out, the BFE hedonic approach identifies the MWTP parameter of the marginal household, while the discrete choice approach identifies the MWTP parameter of the average household. Because of moving costs, the unit of measurement of both the hedonic approach and the static discrete choice approach (as in [Bayer, Ferreira & McMillan (2007)]) will likely span more than one year and one grade for the marginal and the average household, respectively. Moreover, the marginal and the average household may not be a parent, let alone a parent with one child.

By allowing for heterogeneity in preferences across households in the discrete choice framework, [Bayer, Ferreira & McMillan (2007)] is able to go further than the hedonic literature and estimate the MWTP parameter for subgroups of the population. For instance, they find 35% higher MWTP estimates for parents with children under 18 compared to non-parents.\textsuperscript{61}

This paper follows [Bayer, Ferreira & McMillan (2007)] in allowing for heterogeneity of preferences in a discrete choice framework. However, the way in which the heterogeneity of preferences is explored is different.\textsuperscript{62} In [Bayer, Ferreira & McMillan (2007)],

\textsuperscript{61}The MWTP estimate for this subgroup of parents is also not at the year-child-grade level. Because it is a static discrete choice framework, this estimate spans more than one year. Moreover, this estimate is at the level of the typical household in this subgroup of parents, which is likely to have more than one child, and these children are likely to be attending different grades.

\textsuperscript{62}The two approaches also model heterogeneity of preferences differently. In [Bayer, Ferreira & McMillan (2007)], heterogeneity of preferences is controlled through the interaction of observed household characteristics and observed neighborhood characteristics. In contrast, in this paper the heterogeneity of preferences is controlled through the estimation of fixed effects at the group-neighborhood level instead of at the neighborhood level. It is unclear which approach is better in absorbing the heterogeneity in preferences. On the one hand, in order to control for fixed effects at the group-neighborhood level, groups must be defined using only a few observable household characteristics (e.g., age of oldest child for parents and age of head of household for non-parents), otherwise there will not be enough households in each group-neighborhood cell. On the other hand, controlling for fixed effects at the group-neighborhood level allows one to control for unobservable heterogeneity of preferences across groups.
the heterogeneity of preferences is used to present MWTP estimates for different subgroups of the population of households. In this paper the heterogeneity of preferences is used to explore variations in consumer surpluses in order to treat the endogeneity problem. The need to explicitly control for contemporaneous unobservables suggests the need to identify the parameter using variation within neighborhood and within time. I use variation in the share of each neighborhood across different groups to address this problem.

In contrast, the discrete choice BFE approach developed by Bayer, Ferreira & McMillan (2007) explores variation in the level of school quality across attendance areas but within a neighborhood.

The method provided in this paper is generally applicable to treat the endogeneity problem in the context of other amenities, as it explores a variation that is prevalent in the neighborhood context for any amenity, even if the level of that amenity varies smoothly across geographical boundaries. However, unlike the variation used in BFE approaches, the variation used in this paper is difficult to be explored in hedonic models, as that framework explores variations in the level of amenities and prices, not variation in consumer surpluses.

5.3 Endogeneity and Externality

In this paper I argue that, because any change in school quality causes a change in unobserved characteristics due to sorting, and because we likely observe the data only after at least some sorting has occurred, any method not explicitly controlling for contemporaneous unobservables will identify the total valuation of school quality, which is the sum of the valuation of school quality per se (i.e., the direct valuation) and the valuation of unobservable amenities that changed (up to that point) because of the change in school quality (i.e., the indirect valuation).

The indirect valuation can be understood as the externality caused by the change in school quality. For instance, an increase in school quality in a particular neighborhood may cause the quality of neighbors to increase due to sorting, generating a positive externality to anyone living in that neighborhood. An example of a more traditional type of externality would be the reduction in local crime because of the higher school quality. A policy maker may be interested in a MWTP parameter that encompasses such externalities. However, all methods available in the literature, including the one developed in this

---

63 Households of different types pay the same to live in a neighborhood even though they may value it differently. However, if a particular type values a neighborhood more than other types, then relatively more households of that type will be observed living in that neighborhood.

64 Lochner & Moretti (2003) finds that an increase in the level of education reduces crime, which suggests that an increase in school quality might have such a positive effect.
paper, are unable to identify the valuation due to externalities.

The approach developed in this paper provides estimates of the private valuation of school quality only. It does not include any spillover of the increase in school quality other than the spillovers to which only households with children at school age are exposed. By fully controlling for the indirect valuation, it does not include both the traditional externalities and the externalities due to sorting.

Any approach that does not fully control for the indirect valuation will include a component of the externality in the estimate. However, because of the contamination bias discussed in section 2, that component of the externality will be included more than once.\(^{65}\)

Future research is needed to measure the value of the externality due to school quality, and to disentangle the sorting externality and the more traditional type of externality.

6 Conclusion

In this paper I estimate the parental value of school quality. I argue the need to impose more structure in the analysis for two reasons: (1) to identify a parameter that is more useful to the policy maker; (2) to solve the endogeneity problem present in the neighborhood context. I estimate a dynamic neighborhood choice model in order to identify the valuation of school quality at a unit of measurement comparable to the tuition charged by private schools. I also provide a new perspective on the endogeneity problem present in the neighborhood context, and develop a novel identification strategy assuming selection on unobservables.

I find that parental valuation is higher for elementary and secondary grades compared to middle school grades. In comparison to private school tuition, the results suggest that private schools exert more competition pressure on public schools at elementary grades.

This paper develops two novel methodologies that may be useful in other applications. The non-standard approach to estimate a dynamic choice model that I developed in this paper has some advantages and some disadvantages with respect to the standard method. One advantage is that it enables one to estimate a dynamic model without information on the conditional choice probabilities. This method may be useful in the urban economics literature, where the available data with detailed geographic information is often a cross section.

The approach to treat the endogeneity problem developed in this paper may potentially be used to estimate the valuation of other neighborhood amenities, where the same

\(^{65}\)It is not clear whether the contamination bias will be positive. Because of sorting, a local positive externality implies a negative non-local externality, suggesting a positive bias. However, if both the local and non-local externalities are positive (as may be the case in the crime example), then the bias will be negative.
endogeneity issues due neighborhood sorting are present, yet alternative identification strategies such as the boundary fixed effects approach are unfeasible.

Future research is needed to access whether the methods developed in this paper can be successfully applied to other contexts.

References


Pope, J. (2008). Buyer information and the hedonic: the impact of a seller disclosure on

Pudney, S. (1982). Estimating latent variable systems when specification is uncertain:
Generalized component analysis and the eliminant method. *Journal of American Sta-
tistical Association, 77*(380), 883–889.

*Regional Science and Urban Economics, 15*(1), 41–63.

Reback, R. (2005). House prices and the provision of local public services: Capitalization


librium benefits of large changes in spatially delineated public goods*. *International

omy, 64*(5), 416–424.

A Details about Proposition 1

For each household $i$, group $c$, neighborhood $j$ and period $t$, assume:

**Assumption A.1.** (Conditional independence) Let $P(A|B)$ denote the conditional probability of $A$ given $B$, then

$$P(W_{i,c,t+1}|S_{i,c,t}, j_{i,c,t} = j) = P(W_{i,c,t+1}|W_{i,c,t}, j_{i,c,t} = j).$$

**Assumption A.2.** (Discrete support of $W$)

The support of $W_{i,c,t}$ is discrete and finite: $W_{i,c,t} \in \mathbb{W} = \{W^{(1)}, W^{(2)}, \ldots, W^{(|T|)}\}$, $|T| < \infty$.

Assumptions 3.1 and the assumptions above are similar to the assumptions used in Rust (1987) and also in the majority of the literature on estimation of dynamic models. However, there is one important difference: $W$ is assumed observed to the researcher in Rust (1987) but it is allowed to be unobserved (although estimable) in this paper.

Assumption A.2 is mostly technical. Assumptions A.1 together with 3.1 are crucial to the analysis, because they substantially reduce the information gap between the household and the researcher. In period $t$, the household observes $\epsilon_{i,c,t}$, while the researcher does not. Assumption 3.1 gives the researcher information about the distribution of $\epsilon_{i,c,t}$, and restricts the effect that this shock can have in future shocks (in fact, it states that it has no effect at all). Assumption A.1 also restricts the effect that $\epsilon_{i,c,t}$ has in the continuation value. Taken altogether, these assumptions transform $\epsilon_{i,c,t}$ into a simple random shock which has no dynamic effect beyond the direct effect in the household’s choice of where to live in period $t$, which the researcher observes. Since the future is unknown to both the household and the researcher in the same way, effectively the entire dynamic optimization problem faced by the household becomes available to the researcher.

**Proposition A.1.** Under assumptions 3.1, A.1, and A.2, $V_j(S_{i,c,t})$ can be written as

$$V_j(S_{i,c,t}) = v_j(W_{i,c,t}) + \epsilon_{i,c,t,j},$$

where $\epsilon_{i,c,t,j}$ is the same as in assumption 3.1 and

$$v_j(W_{i,c,t}) := u_j(W_{i,c,t}) + \beta \sum_{W \in \mathbb{W}} \left( \gamma + \log \sum_{r=1}^{j} \exp(v_r(W)) \right) P(W|W_{i,c,t}, j_{i,c,t} = j).$$

**Proof.** See Rust (1987).
The result of proposition 1 follows directly from proposition A.1 and equation (9).

**B Details about the estimation in Step 1**

Households choose the neighborhood that yields the highest utility among the neighborhoods available. Household \(i\) of group \(c\) chooses neighborhood \(j\) in period \(t\) if and only if the utility of choosing neighborhood \(j\) is at least as high as the utility of choosing any other neighborhood:

\[
j_{i,c,t} = j \iff \epsilon_{i,c,t,j} - \epsilon_{i,c,t,r} \geq v_r(W_{i,c,t}) - v_j(W_{i,c,t}), \quad \forall r.
\]

(29)

From assumption 3.1, \(\epsilon_{i,c,t,j} - \epsilon_{i,c,t,r}\) has a logit distribution, and the probability that household \(i\) of group \(c\) chooses neighborhood \(j\) in period \(t\) is:

\[
P_j(W_{i,c,t}) = \frac{\exp(v_j(W_{i,c,t}))}{\sum_{r=1}^{J} \exp(v_r(W_{i,c,t}))}.
\]

(30)

The maximum likelihood estimate of this problem is the value of the parameter that maximizes the sum across households of the log-likelihood that each household chooses the neighborhood as observed in the data. The log-likelihood function is written as

\[
LL_{c,t}(W_{i,c,t}, \{j_{i,c,t}^*\}_{i=1}^{I_c}) = \sum_{i=1}^{I_c} \sum_{j=1}^{J} d_j(j_{i,c,t}^*) \log(P_j(W_{i,c,t})), \quad c = 1, ..., C, \quad \forall t.
\]

(31)

where \(j_{i,c,t}^*\) is the choice observed in the data for each \(i, c, t\).

I estimate \(\Delta_{c,t}\) and \(\Phi_{c,t}\) by maximizing equation (31) independently for each group \(c \in \{1, ..., C\}\), and time \(t = 2000\).

It is necessary to estimate the choice-specific value functions both of period \(t\) and period \(t+1\). Estimation of \(v_j(W_{i,c,t})\) for \(t = 2000\) is trivial, and only requires to plug in the parameters estimated in (31) in equation (13). For year 2001, I plug in the estimates from the \(t = 2000\) into equation (14).

Assuming that the regularity conditions which guarantee that the Maximum Likelihood estimators \(\Delta_{c,t}\) and \(\Phi_{c,t}\) are consistent are satisfied, the following lemma guarantees the mean convergence of \(\hat{v}_j(W_{i,c,t})\) and \(\hat{v}_j(W_{i,c,t+1})\).

---

\[\text{Due to the large number of parameters, it is unfeasible to estimate all coefficients of this model using a standard numerical optimization algorithm, such as Newton-Raphson. Instead, I write the } \Delta_s \text{ as function of the other parameter } \Phi \text{ using a contraction, as in Berry, Levinsohn & Pakes (1995).}\]

\[\text{See McFadden (1973), McFadden (1977) and Berry, Levinsohn & Pakes (1995).}\]

40
The Law of Large numbers and lemma B.1 guarantee that $a$ be
\[ \hat{\theta}(W_{i,c,t}) := v_j(W_{i,c,t}). \]
\[ \hat{\theta}(W_{i,c,t+1}) := v_j(W_{i,c,t+1}). \]

Then, $E(|\hat{\theta}(W_{i,c,t}) - v_j(W_{i,c,t}|) \rightarrow 0$ and $E(|\hat{\theta}(W_{i,c,t+1}) - v_j(W_{i,c,t+1})| \rightarrow 0$ for each $i, c, j$ and $t = 2000$.

**Proof.** The $\hat{\theta}(W_{i,c,t})$ is a linear “plug-in” estimator of $W_{i,c,t}$. From the convergence in distribution of the MLE estimator, $E(||\sqrt{n}(\hat{W}_{i,c,t} - W_{i,c,t})||^2) = O_p(1), \forall c,t$. Assume that $W_{i,c,t}$ have finite second moments. Then, by Cauchy-Schwartz, the result follows.

Finally, I estimate $\delta_{i,c,t}$ using an empirical version of equation (12):

\[ \hat{\delta}_{i,c,t} = \frac{1}{n_{c,t}} \sum_{j=1}^{n_{c,t}} \left( \hat{\theta}(W_{i,c,t}) - \beta \left( \gamma + \log \sum_{r=1}^J \exp(\hat{\theta}(W_{i,c,t+1})) \right) \right), \quad c = 0, ..., 18 \]
\[ (32) \]
\[ \hat{\delta}_{i,c,t} = \frac{1}{n_{c,t}} \sum_{j=1}^{n_{c,t}} \left( \hat{\theta}(W_{i,c,t}) - \beta \left[ \sum_{j=1}^{n_{c,t}} \Pi_{c,t}^J \left( \gamma + \log \sum_{r=1}^J \exp(\hat{\theta}(W_{i,c,t+1})) \right) \right) \right), \quad c = 31, ..., 49 \]
\[ (33) \]

where $\gamma$ is the Euler’s constant, $\Pi_{c,t}^J := 1 - \Pi_{c,t}^0$, $W_{i,c,t}^0 := \left( \{\Delta_{c+1,t,j}\}_{j=0}^J, \Phi_{c+1,t}, \{NEW_{i,c,t+1, j}\}_{j=1}^J \right)$, and $W_{i,c,t+1}^0 := \left( \{\Delta_{c+1,t,j}\}_{j=1}^J, \Phi_{c+1,t}, \{NEW_{i,c,t+1, j}\}_{j=1}^J \right)$.

**Proposition B.1.** For all $i, c$ and $t = 2000$, let $\hat{W}_{i,c,t}$ be a consistent estimator of $W_{i,c,t}$. Let $\hat{\delta}_{i,c,t}$ be defined as in equations (32) and (33). Then $\hat{\delta}_{i,c,t}$ is a consistent estimator of $\delta_{i,c,t}$.

**Proof.** Let the term inside the sum in equations (32) and (33) be $a_{i,c,t,j}$, and its true value be $a_{i,c,t,j}$. The Continuous Mapping Theorem and lemma [B.1] imply that $E(|a_{i,c,t,j} - a_{i,c,t,j}^*|) \rightarrow 0$. Apply Markov’s theorem to $\frac{1}{n_{c,t}} \sum_{i=1}^{n_{c,t}} (a_{i,c,t,j} - a_{i,c,t,j})$ to show that it is $o_p(1)$. The Law of Large numbers and lemma [B.1] guarantee that $\frac{1}{n_{c,t}} \sum_{i=1}^{n_{c,t}} a_{i,c,t,j} \rightarrow \delta_{i,c,t}$. 

41
Table 1
Characteristics of the School Attendance Areas

<table>
<thead>
<tr>
<th>Variables</th>
<th>Full Sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>St. Dev.</td>
</tr>
<tr>
<td><strong>Neighborhood Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average test score</td>
<td>1387</td>
<td>223</td>
</tr>
<tr>
<td>Number of houses</td>
<td>862</td>
<td>763</td>
</tr>
<tr>
<td>Average number of rooms</td>
<td>6.4</td>
<td>0.8</td>
</tr>
<tr>
<td>Average number of bedrooms</td>
<td>3.0</td>
<td>0.4</td>
</tr>
<tr>
<td>Average rent ($)</td>
<td>544</td>
<td>212</td>
</tr>
<tr>
<td>Average house value ($)</td>
<td>144,215</td>
<td>59,219</td>
</tr>
<tr>
<td><strong>Neighbors’ Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household Income ($)</td>
<td>65,127</td>
<td>24,389</td>
</tr>
<tr>
<td>Household employment: proportion head is employed (%)</td>
<td>71</td>
<td>13</td>
</tr>
<tr>
<td>Household employment: proportion both head and spouse are employed (%)</td>
<td>46</td>
<td>13</td>
</tr>
<tr>
<td>Head of household’s education: proportion with high-school degree only (%)</td>
<td>28</td>
<td>11</td>
</tr>
<tr>
<td>Head of household’s education: proportion with college degree or more (%)</td>
<td>28</td>
<td>17</td>
</tr>
<tr>
<td>Age of head of household</td>
<td>46.0</td>
<td>3.7</td>
</tr>
<tr>
<td>Proportion of households who own their house (%)</td>
<td>80</td>
<td>15</td>
</tr>
<tr>
<td>Head of household’s race: proportion of black (%)</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Head of household’s race: proportion of non-black, non-white (%)</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Proportion of households whose head is a parent (%)</td>
<td>62</td>
<td>6</td>
</tr>
<tr>
<td>Proportion of households that have one child (%)</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>Proportion of households that have two children (%)</td>
<td>22</td>
<td>7</td>
</tr>
<tr>
<td>Proportion of households that have three or more children (%)</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>Number of people of 65 years of age or older in the household</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Number of people of 18 years of age or younger in the household</td>
<td>1.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Proportion of households whose oldest child is attending a private school (%)</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Number of school attendance areas</td>
<td>345</td>
<td></td>
</tr>
<tr>
<td>Number of households</td>
<td>153,102</td>
<td></td>
</tr>
</tbody>
</table>

42
Table 2

Summary Statistics
Groups of Parents (by the age of the oldest child)

<table>
<thead>
<tr>
<th>Group (obs)</th>
<th>Household Income ($)</th>
<th>Homeowner (%)</th>
<th>Black (%)</th>
<th>Other Race (%)</th>
<th>High School Degree (%)</th>
<th>College Degree (%)</th>
<th>Employed (%)</th>
<th>Age 2 Children (%)</th>
<th>3 or more Children (%)</th>
<th>Private School (%)</th>
<th>Moved Last Year (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>60,000</td>
<td>75</td>
<td>2</td>
<td>7</td>
<td>61</td>
<td>32</td>
<td>95</td>
<td>29</td>
<td>2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(1802)</td>
<td>(37,000)</td>
<td>(43)</td>
<td>(14)</td>
<td>(26)</td>
<td>(49)</td>
<td>(47)</td>
<td>(22)</td>
<td>(6)</td>
<td>(15)</td>
<td>(–)</td>
<td>(–)</td>
</tr>
<tr>
<td>1</td>
<td>56,000</td>
<td>75</td>
<td>2</td>
<td>4</td>
<td>61</td>
<td>33</td>
<td>96</td>
<td>30</td>
<td>8</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(3109)</td>
<td>(37,000)</td>
<td>(43)</td>
<td>(14)</td>
<td>(18)</td>
<td>(49)</td>
<td>(47)</td>
<td>(20)</td>
<td>(6)</td>
<td>(27)</td>
<td>(–)</td>
<td>(–)</td>
</tr>
<tr>
<td>2</td>
<td>58,000</td>
<td>75</td>
<td>3</td>
<td>6</td>
<td>61</td>
<td>32</td>
<td>96</td>
<td>31</td>
<td>28</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(3968)</td>
<td>(38,000)</td>
<td>(44)</td>
<td>(16)</td>
<td>(24)</td>
<td>(49)</td>
<td>(47)</td>
<td>(20)</td>
<td>(6)</td>
<td>(45)</td>
<td>(–)</td>
<td>(–)</td>
</tr>
<tr>
<td>3</td>
<td>59,000</td>
<td>80</td>
<td>2</td>
<td>6</td>
<td>62</td>
<td>32</td>
<td>96</td>
<td>32</td>
<td>50</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>(4477)</td>
<td>(38,000)</td>
<td>(40)</td>
<td>(12)</td>
<td>(23)</td>
<td>(48)</td>
<td>(47)</td>
<td>(19)</td>
<td>(6)</td>
<td>(50)</td>
<td>(19)</td>
<td>(22)</td>
</tr>
<tr>
<td>4</td>
<td>60,000</td>
<td>81</td>
<td>2</td>
<td>4</td>
<td>62</td>
<td>33</td>
<td>96</td>
<td>32</td>
<td>57</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>(4546)</td>
<td>(39,000)</td>
<td>(39)</td>
<td>(14)</td>
<td>(19)</td>
<td>(48)</td>
<td>(47)</td>
<td>(20)</td>
<td>(6)</td>
<td>(49)</td>
<td>(29)</td>
<td>(38)</td>
</tr>
<tr>
<td>5</td>
<td>59,000</td>
<td>82</td>
<td>3</td>
<td>4</td>
<td>64</td>
<td>30</td>
<td>96</td>
<td>33</td>
<td>56</td>
<td>17</td>
<td>26</td>
</tr>
<tr>
<td>(5015)</td>
<td>(39,000)</td>
<td>(38)</td>
<td>(17)</td>
<td>(19)</td>
<td>(48)</td>
<td>(46)</td>
<td>(20)</td>
<td>(6)</td>
<td>(50)</td>
<td>(37)</td>
<td>(44)</td>
</tr>
<tr>
<td>6</td>
<td>62,000</td>
<td>83</td>
<td>3</td>
<td>4</td>
<td>64</td>
<td>32</td>
<td>96</td>
<td>34</td>
<td>56</td>
<td>23</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>61,000</td>
<td>84</td>
<td>2</td>
<td>5</td>
<td>64</td>
<td>29</td>
<td>96</td>
<td>34</td>
<td>53</td>
<td>27</td>
<td>11</td>
</tr>
<tr>
<td>(5560)</td>
<td>(42,000)</td>
<td>(37)</td>
<td>(14)</td>
<td>(21)</td>
<td>(48)</td>
<td>(46)</td>
<td>(20)</td>
<td>(6)</td>
<td>(50)</td>
<td>(45)</td>
<td>(31)</td>
</tr>
<tr>
<td>8</td>
<td>63,000</td>
<td>85</td>
<td>2</td>
<td>4</td>
<td>66</td>
<td>28</td>
<td>96</td>
<td>35</td>
<td>50</td>
<td>33</td>
<td>11</td>
</tr>
<tr>
<td>(5825)</td>
<td>(42,000)</td>
<td>(36)</td>
<td>(14)</td>
<td>(21)</td>
<td>(47)</td>
<td>(45)</td>
<td>(20)</td>
<td>(6)</td>
<td>(50)</td>
<td>(47)</td>
<td>(31)</td>
</tr>
<tr>
<td>9</td>
<td>63,000</td>
<td>85</td>
<td>2</td>
<td>5</td>
<td>66</td>
<td>28</td>
<td>96</td>
<td>36</td>
<td>47</td>
<td>37</td>
<td>11</td>
</tr>
<tr>
<td>(5996)</td>
<td>(44,000)</td>
<td>(36)</td>
<td>(14)</td>
<td>(22)</td>
<td>(47)</td>
<td>(45)</td>
<td>(21)</td>
<td>(6)</td>
<td>(50)</td>
<td>(48)</td>
<td>(31)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are reported in parenthesis except in the first column where the number of observations is reported.
Table 2 (cont.)

Summary Statistics
Groups of Parents (by the age of the oldest child)

<table>
<thead>
<tr>
<th>Group (obs)</th>
<th>Household Income ($)</th>
<th>Home-owner (%)</th>
<th>Black (%)</th>
<th>Other Race (%)</th>
<th>High School Degree (%)</th>
<th>College Degree (%)</th>
<th>Employed (%)</th>
<th>Age</th>
<th>2 Children (%)</th>
<th>3 or more Children (%)</th>
<th>Private School (%)</th>
<th>Moved Last Year (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>65,000</td>
<td>86</td>
<td>2</td>
<td>5</td>
<td>67</td>
<td>27</td>
<td>96</td>
<td>37</td>
<td>46</td>
<td>38</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>(6582)</td>
<td>(46,000)</td>
<td>(35)</td>
<td>(15)</td>
<td>(21)</td>
<td>(47)</td>
<td>(44)</td>
<td>(21)</td>
<td>(6)</td>
<td>(50)</td>
<td>(48)</td>
<td>(33)</td>
<td>(37)</td>
</tr>
<tr>
<td>11</td>
<td>63,000</td>
<td>86</td>
<td>3</td>
<td>4</td>
<td>68</td>
<td>24</td>
<td>96</td>
<td>38</td>
<td>43</td>
<td>41</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>(6761)</td>
<td>(42,000)</td>
<td>(35)</td>
<td>(16)</td>
<td>(20)</td>
<td>(47)</td>
<td>(43)</td>
<td>(19)</td>
<td>(6)</td>
<td>(50)</td>
<td>(49)</td>
<td>(30)</td>
<td>(35)</td>
</tr>
<tr>
<td>12</td>
<td>65,000</td>
<td>87</td>
<td>2</td>
<td>4</td>
<td>69</td>
<td>23</td>
<td>95</td>
<td>39</td>
<td>41</td>
<td>42</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>(6925)</td>
<td>(46,000)</td>
<td>(33)</td>
<td>(14)</td>
<td>(20)</td>
<td>(46)</td>
<td>(42)</td>
<td>(21)</td>
<td>(6)</td>
<td>(49)</td>
<td>(49)</td>
<td>(28)</td>
<td>(36)</td>
</tr>
<tr>
<td>13</td>
<td>63,000</td>
<td>86</td>
<td>2</td>
<td>5</td>
<td>69</td>
<td>23</td>
<td>96</td>
<td>40</td>
<td>40</td>
<td>44</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>(7681)</td>
<td>(43,000)</td>
<td>(34)</td>
<td>(15)</td>
<td>(21)</td>
<td>(46)</td>
<td>(42)</td>
<td>(19)</td>
<td>(6)</td>
<td>(49)</td>
<td>(50)</td>
<td>(29)</td>
<td>(36)</td>
</tr>
<tr>
<td>14</td>
<td>65,000</td>
<td>89</td>
<td>2</td>
<td>4</td>
<td>69</td>
<td>25</td>
<td>95</td>
<td>41</td>
<td>41</td>
<td>40</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>(7527)</td>
<td>(43,000)</td>
<td>(31)</td>
<td>(14)</td>
<td>(21)</td>
<td>(46)</td>
<td>(43)</td>
<td>(21)</td>
<td>(6)</td>
<td>(49)</td>
<td>(49)</td>
<td>(28)</td>
<td>(31)</td>
</tr>
<tr>
<td>15</td>
<td>65,000</td>
<td>89</td>
<td>2</td>
<td>4</td>
<td>69</td>
<td>24</td>
<td>97</td>
<td>42</td>
<td>37</td>
<td>42</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>(8846)</td>
<td>(44,000)</td>
<td>(31)</td>
<td>(15)</td>
<td>(21)</td>
<td>(46)</td>
<td>(43)</td>
<td>(18)</td>
<td>(6)</td>
<td>(48)</td>
<td>(49)</td>
<td>(26)</td>
<td>(31)</td>
</tr>
<tr>
<td>16</td>
<td>68,000</td>
<td>91</td>
<td>1</td>
<td>4</td>
<td>69</td>
<td>25</td>
<td>97</td>
<td>43</td>
<td>38</td>
<td>39</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>(9242)</td>
<td>(44,000)</td>
<td>(29)</td>
<td>(12)</td>
<td>(19)</td>
<td>(46)</td>
<td>(43)</td>
<td>(18)</td>
<td>(6)</td>
<td>(49)</td>
<td>(49)</td>
<td>(23)</td>
<td>(30)</td>
</tr>
<tr>
<td>17</td>
<td>71,000</td>
<td>92</td>
<td>2</td>
<td>3</td>
<td>69</td>
<td>24</td>
<td>96</td>
<td>44</td>
<td>36</td>
<td>38</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>18</td>
<td>71,000</td>
<td>92</td>
<td>1</td>
<td>3</td>
<td>67</td>
<td>26</td>
<td>97</td>
<td>45</td>
<td>38</td>
<td>37</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>(11606)</td>
<td>(43,000)</td>
<td>(27)</td>
<td>(10)</td>
<td>(18)</td>
<td>(47)</td>
<td>(44)</td>
<td>(17)</td>
<td>(6)</td>
<td>(49)</td>
<td>(48)</td>
<td>(22)</td>
<td>(24)</td>
</tr>
<tr>
<td>19</td>
<td>70,000</td>
<td>92</td>
<td>2</td>
<td>4</td>
<td>73</td>
<td>19</td>
<td>96</td>
<td>45</td>
<td>38</td>
<td>36</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>(6570)</td>
<td>(39,000)</td>
<td>(27)</td>
<td>(14)</td>
<td>(21)</td>
<td>(45)</td>
<td>(39)</td>
<td>(20)</td>
<td>(6)</td>
<td>(49)</td>
<td>(48)</td>
<td>(25)</td>
<td>(26)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are reported in parenthesis except in the first column where the number of observations is reported.
Table 3
Summary Statistics
Groups of Non-Parents (by the age of the head of the household)

<table>
<thead>
<tr>
<th>Group (obs)</th>
<th>Household Income ($)</th>
<th>Home-owner (%)</th>
<th>Black (%)</th>
<th>Other Race (%)</th>
<th>High School Degree (%)</th>
<th>College Degree (%)</th>
<th>Employed (%)</th>
<th>Moved Last Year (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>60,000</td>
<td>67</td>
<td>3</td>
<td>6</td>
<td>55</td>
<td>42</td>
<td>51</td>
<td>35</td>
</tr>
<tr>
<td>(1319)</td>
<td>(35,000)</td>
<td>(47)</td>
<td>(16)</td>
<td>(23)</td>
<td>(50)</td>
<td>(49)</td>
<td>(50)</td>
<td>(48)</td>
</tr>
<tr>
<td>32</td>
<td>62,000</td>
<td>70</td>
<td>2</td>
<td>5</td>
<td>51</td>
<td>45</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>(1161)</td>
<td>(39,000)</td>
<td>(46)</td>
<td>(15)</td>
<td>(22)</td>
<td>(50)</td>
<td>(50)</td>
<td>(50)</td>
<td>(46)</td>
</tr>
<tr>
<td>33</td>
<td>60,000</td>
<td>73</td>
<td>2</td>
<td>5</td>
<td>58</td>
<td>36</td>
<td>50</td>
<td>23</td>
</tr>
<tr>
<td>(1103)</td>
<td>(37,000)</td>
<td>(45)</td>
<td>(15)</td>
<td>(21)</td>
<td>(49)</td>
<td>(48)</td>
<td>(50)</td>
<td>(42)</td>
</tr>
<tr>
<td>34</td>
<td>61,000</td>
<td>73</td>
<td>4</td>
<td>4</td>
<td>57</td>
<td>37</td>
<td>44</td>
<td>24</td>
</tr>
<tr>
<td>(1158)</td>
<td>(40,000)</td>
<td>(44)</td>
<td>(19)</td>
<td>(19)</td>
<td>(49)</td>
<td>(48)</td>
<td>(50)</td>
<td>(43)</td>
</tr>
<tr>
<td>35</td>
<td>61,000</td>
<td>73</td>
<td>2</td>
<td>3</td>
<td>62</td>
<td>35</td>
<td>42</td>
<td>23</td>
</tr>
<tr>
<td>(1083)</td>
<td>(41,000)</td>
<td>(44)</td>
<td>(15)</td>
<td>(17)</td>
<td>(49)</td>
<td>(48)</td>
<td>(49)</td>
<td>(42)</td>
</tr>
<tr>
<td>36</td>
<td>59,000</td>
<td>74</td>
<td>3</td>
<td>2</td>
<td>68</td>
<td>26</td>
<td>43</td>
<td>20</td>
</tr>
<tr>
<td>(1098)</td>
<td>(38,000)</td>
<td>(44)</td>
<td>(18)</td>
<td>(15)</td>
<td>(47)</td>
<td>(44)</td>
<td>(49)</td>
<td>(40)</td>
</tr>
<tr>
<td>37</td>
<td>57,000</td>
<td>73</td>
<td>2</td>
<td>4</td>
<td>64</td>
<td>30</td>
<td>44</td>
<td>21</td>
</tr>
<tr>
<td>(1031)</td>
<td>(39,000)</td>
<td>(44)</td>
<td>(12)</td>
<td>(19)</td>
<td>(48)</td>
<td>(46)</td>
<td>(50)</td>
<td>(40)</td>
</tr>
<tr>
<td>38</td>
<td>60,000</td>
<td>76</td>
<td>3</td>
<td>4</td>
<td>65</td>
<td>31</td>
<td>40</td>
<td>21</td>
</tr>
<tr>
<td>(1048)</td>
<td>(42,000)</td>
<td>(43)</td>
<td>(16)</td>
<td>(19)</td>
<td>(48)</td>
<td>(46)</td>
<td>(49)</td>
<td>(40)</td>
</tr>
<tr>
<td>39</td>
<td>58,000</td>
<td>79</td>
<td>3</td>
<td>3</td>
<td>67</td>
<td>26</td>
<td>43</td>
<td>19</td>
</tr>
<tr>
<td>(1073)</td>
<td>(37,000)</td>
<td>(41)</td>
<td>(17)</td>
<td>(18)</td>
<td>(47)</td>
<td>(44)</td>
<td>(49)</td>
<td>(40)</td>
</tr>
<tr>
<td>40</td>
<td>57,000</td>
<td>79</td>
<td>2</td>
<td>3</td>
<td>67</td>
<td>24</td>
<td>42</td>
<td>18</td>
</tr>
<tr>
<td>(1107)</td>
<td>(43,000)</td>
<td>(41)</td>
<td>(13)</td>
<td>(18)</td>
<td>(47)</td>
<td>(43)</td>
<td>(49)</td>
<td>(39)</td>
</tr>
<tr>
<td>41</td>
<td>58,000</td>
<td>79</td>
<td>2</td>
<td>3</td>
<td>71</td>
<td>23</td>
<td>44</td>
<td>14</td>
</tr>
<tr>
<td>(1176)</td>
<td>(38,000)</td>
<td>(41)</td>
<td>(15)</td>
<td>(17)</td>
<td>(46)</td>
<td>(42)</td>
<td>(50)</td>
<td>(35)</td>
</tr>
<tr>
<td>42</td>
<td>60,000</td>
<td>80</td>
<td>2</td>
<td>3</td>
<td>70</td>
<td>22</td>
<td>47</td>
<td>16</td>
</tr>
<tr>
<td>(1211)</td>
<td>(44,000)</td>
<td>(40)</td>
<td>(15)</td>
<td>(17)</td>
<td>(46)</td>
<td>(42)</td>
<td>(50)</td>
<td>(37)</td>
</tr>
<tr>
<td>43</td>
<td>59,000</td>
<td>78</td>
<td>1</td>
<td>4</td>
<td>72</td>
<td>22</td>
<td>44</td>
<td>15</td>
</tr>
<tr>
<td>(1119)</td>
<td>(42,000)</td>
<td>(41)</td>
<td>(12)</td>
<td>(19)</td>
<td>(45)</td>
<td>(41)</td>
<td>(50)</td>
<td>(36)</td>
</tr>
<tr>
<td>44</td>
<td>59,000</td>
<td>82</td>
<td>3</td>
<td>2</td>
<td>71</td>
<td>21</td>
<td>51</td>
<td>13</td>
</tr>
<tr>
<td>(1255)</td>
<td>(42,000)</td>
<td>(38)</td>
<td>(16)</td>
<td>(16)</td>
<td>(45)</td>
<td>(41)</td>
<td>(50)</td>
<td>(34)</td>
</tr>
<tr>
<td>45</td>
<td>59,000</td>
<td>83</td>
<td>2</td>
<td>3</td>
<td>71</td>
<td>23</td>
<td>54</td>
<td>11</td>
</tr>
<tr>
<td>(1280)</td>
<td>(42,000)</td>
<td>(38)</td>
<td>(14)</td>
<td>(16)</td>
<td>(45)</td>
<td>(42)</td>
<td>(50)</td>
<td>(31)</td>
</tr>
<tr>
<td>46</td>
<td>60,000</td>
<td>82</td>
<td>2</td>
<td>4</td>
<td>72</td>
<td>23</td>
<td>53</td>
<td>13</td>
</tr>
<tr>
<td>(1351)</td>
<td>(44,000)</td>
<td>(38)</td>
<td>(15)</td>
<td>(19)</td>
<td>(45)</td>
<td>(42)</td>
<td>(50)</td>
<td>(34)</td>
</tr>
<tr>
<td>47</td>
<td>62,000</td>
<td>86</td>
<td>1</td>
<td>3</td>
<td>69</td>
<td>23</td>
<td>56</td>
<td>12</td>
</tr>
<tr>
<td>(1484)</td>
<td>(43,000)</td>
<td>(34)</td>
<td>(12)</td>
<td>(16)</td>
<td>(46)</td>
<td>(42)</td>
<td>(50)</td>
<td>(32)</td>
</tr>
<tr>
<td>48</td>
<td>64,000</td>
<td>87</td>
<td>1</td>
<td>3</td>
<td>68</td>
<td>26</td>
<td>60</td>
<td>11</td>
</tr>
<tr>
<td>(1542)</td>
<td>(44,000)</td>
<td>(33)</td>
<td>(11)</td>
<td>(17)</td>
<td>(47)</td>
<td>(44)</td>
<td>(49)</td>
<td>(31)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are reported in parenthesis except in the first column where the number of observations is reported.
<table>
<thead>
<tr>
<th>Group (obs)</th>
<th>Household Income ($)</th>
<th>Homeowner (%)</th>
<th>Black (%)</th>
<th>Other Race (%)</th>
<th>High School Degree (%)</th>
<th>College Degree (%)</th>
<th>Employed (%)</th>
<th>Moved (%)</th>
<th>Last Year (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>49 (1776)</td>
<td>64,000</td>
<td>89</td>
<td>1</td>
<td>2</td>
<td>68</td>
<td>26</td>
<td>66</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>50 (1806)</td>
<td>68,000</td>
<td>88</td>
<td>1</td>
<td>2</td>
<td>64</td>
<td>30</td>
<td>66</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard deviations are reported in parenthesis except in the first column where the number of observations is reported.
Table 4
Main Results

MWTP (as % of rent) for an Increase in 5% in School Quality

<table>
<thead>
<tr>
<th>Grade</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>15.72**</td>
<td>38.10**</td>
<td>37.75**</td>
<td>21.09**</td>
<td>13.46*</td>
<td>10.58</td>
</tr>
<tr>
<td></td>
<td>(5.54)</td>
<td>(5.82)</td>
<td>(3.64)</td>
<td>(5.88)</td>
<td>(8.14)</td>
<td>(8.88)</td>
</tr>
<tr>
<td>1</td>
<td>9.97**</td>
<td>27.89**</td>
<td>28.52**</td>
<td>19.25**</td>
<td>14.96**</td>
<td>14.25**</td>
</tr>
<tr>
<td>2</td>
<td>4.02</td>
<td>19.28**</td>
<td>20.99**</td>
<td>16.03**</td>
<td>11.25</td>
<td>10.26</td>
</tr>
<tr>
<td></td>
<td>(3.89)</td>
<td>(4.16)</td>
<td>(2.48)</td>
<td>(5.57)</td>
<td>(7.52)</td>
<td>(7.95)</td>
</tr>
<tr>
<td>3</td>
<td>-0.45</td>
<td>8.19**</td>
<td>13.13**</td>
<td>14.07**</td>
<td>12.38**</td>
<td>12.80**</td>
</tr>
<tr>
<td></td>
<td>(3.16)</td>
<td>(3.89)</td>
<td>(2.11)</td>
<td>(3.17)</td>
<td>(5.62)</td>
<td>(5.54)</td>
</tr>
<tr>
<td>4</td>
<td>-3.53</td>
<td>4.16</td>
<td>7.97**</td>
<td>9.24**</td>
<td>10.70*</td>
<td>9.88*</td>
</tr>
<tr>
<td></td>
<td>(2.64)</td>
<td>(3.17)</td>
<td>(1.74)</td>
<td>(2.85)</td>
<td>(5.09)</td>
<td>(5.71)</td>
</tr>
<tr>
<td>5</td>
<td>-1.33</td>
<td>6.44*</td>
<td>9.58**</td>
<td>9.96**</td>
<td>10.74*</td>
<td>9.97</td>
</tr>
<tr>
<td></td>
<td>(3.19)</td>
<td>(3.37)</td>
<td>(2.14)</td>
<td>(4.14)</td>
<td>(5.73)</td>
<td>(6.43)</td>
</tr>
<tr>
<td>6</td>
<td>-0.17</td>
<td>5.84*</td>
<td>9.44**</td>
<td>8.61**</td>
<td>7.70</td>
<td>7.98</td>
</tr>
<tr>
<td></td>
<td>(3.21)</td>
<td>(3.12)</td>
<td>(1.96)</td>
<td>(3.65)</td>
<td>(5.90)</td>
<td>(6.23)</td>
</tr>
<tr>
<td>7</td>
<td>-1.74</td>
<td>6.95**</td>
<td>10.58**</td>
<td>7.67**</td>
<td>6.87</td>
<td>6.36</td>
</tr>
<tr>
<td></td>
<td>(3.11)</td>
<td>(3.21)</td>
<td>(1.83)</td>
<td>(6.12)</td>
<td>(6.87)</td>
<td>(7.79)</td>
</tr>
<tr>
<td>8</td>
<td>-0.88</td>
<td>8.03**</td>
<td>11.26**</td>
<td>8.73**</td>
<td>5.80</td>
<td>5.30</td>
</tr>
<tr>
<td></td>
<td>(3.08)</td>
<td>(3.57)</td>
<td>(2.15)</td>
<td>(4.11)</td>
<td>(6.21)</td>
<td>(6.63)</td>
</tr>
<tr>
<td>9</td>
<td>5.04</td>
<td>14.24**</td>
<td>17.43**</td>
<td>14.08**</td>
<td>12.46**</td>
<td>11.80*</td>
</tr>
<tr>
<td></td>
<td>(3.36)</td>
<td>(3.65)</td>
<td>(2.01)</td>
<td>(5.19)</td>
<td>(5.89)</td>
<td>(6.79)</td>
</tr>
<tr>
<td>10</td>
<td>3.73</td>
<td>15.74**</td>
<td>19.07**</td>
<td>18.68**</td>
<td>14.59**</td>
<td>13.93**</td>
</tr>
<tr>
<td></td>
<td>(3.02)</td>
<td>(3.93)</td>
<td>(1.93)</td>
<td>(3.57)</td>
<td>(4.89)</td>
<td>(5.61)</td>
</tr>
<tr>
<td>11</td>
<td>5.36</td>
<td>17.56**</td>
<td>20.96**</td>
<td>14.69**</td>
<td>13.05**</td>
<td>11.35</td>
</tr>
<tr>
<td></td>
<td>(3.56)</td>
<td>(4.21)</td>
<td>(1.92)</td>
<td>(5.64)</td>
<td>(6.55)</td>
<td>(7.67)</td>
</tr>
<tr>
<td>12</td>
<td>8.75**</td>
<td>22.95**</td>
<td>26.41**</td>
<td>18.72**</td>
<td>16.17**</td>
<td>14.20*</td>
</tr>
<tr>
<td></td>
<td>(4.37)</td>
<td>(4.76)</td>
<td>(2.25)</td>
<td>(6.11)</td>
<td>(6.75)</td>
<td>(7.92)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Controls?</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>No</th>
<th>No</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proxies (cohorts)</td>
<td>–</td>
<td>–</td>
<td>34</td>
<td>4, 34</td>
<td>4, 34, 41</td>
<td>2, 4, 34, 41</td>
</tr>
<tr>
<td>P-value for J test</td>
<td>–</td>
<td>–</td>
<td>.00**</td>
<td>.03**</td>
<td>.35</td>
<td>.32</td>
</tr>
<tr>
<td>R-squared</td>
<td>.0221</td>
<td>.2517</td>
<td>.3839</td>
<td>.3980</td>
<td>.4162</td>
<td>.4189</td>
</tr>
<tr>
<td>Observations</td>
<td>4475</td>
<td>4475</td>
<td>4475</td>
<td>4475</td>
<td>4475</td>
<td>4475</td>
</tr>
</tbody>
</table>

Note: Standard errors corrected for clustering by school attendance area and for heteroskedasticity. *: Statistically significant at the 10% level; **: Statistically significant at the 5% level. Dependent Variable: estimated mean flow utilities for cohorts 6 through 18. Proxy Variables: estimated mean flow utilities for cohorts referred in each column. Instruments: estimated mean flow utilities for cohorts referred in each column. Control Variables: Average Income of neighbors, proportion of black neighbors, proportion of neighbors with race other than white or black, proportion of neighbors with college degree or more, average number of rooms, proportion of homeowners in the neighborhood, average house price in the neighborhood. Underlying Assumptions: groups used as proxies do not value average school quality in the flow utility sense; groups used as proxies have on average the same flow value for the amenity average rent as group 6; \( \beta = 4/3 \). Underlying Sample: Parents whose oldest child is attending a public school and is not a defier, i.e., the child was assigned to the correct grade. The average monthly rent is $544.
Table 5
Magnitude of the results (in $)
(From a baseline $544 monthly rent)

<table>
<thead>
<tr>
<th></th>
<th>MWTP(5%)</th>
<th>MWTP(1 St. Dev)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Per month</td>
<td>Per year</td>
</tr>
<tr>
<td>Elementary School</td>
<td>63</td>
<td>737</td>
</tr>
<tr>
<td>Middle School</td>
<td>36</td>
<td>427</td>
</tr>
<tr>
<td>High School</td>
<td>70</td>
<td>837</td>
</tr>
</tbody>
</table>

*Note:* 1 St. Dev. $\approx$ 16%. Values derived from estimates from specification VI in table 4.