Estimation of Parental Valuation of School Quality in the U.S.*

Gregorio Caetano†

April 2009

*I am particularly thankful to David Card for many helpful discussions and invaluable advice throughout the duration of this project, and to Patrick Bayer, Ken Chay, Stefano DellaVigna, Fernando Ferreira, Patrick Kline, David Lee, Robert McMillan, Enrico Moretti, John Quigley and Steven Raphael for guidance. This work also benefited from discussions with Felipe Barros, Tiberio Caetano, Mario Cappizzani, Andres Donangelo, Constança Esteves-Sorenson, Ashley Langer, Nate Miller, Maria Carolina Nizarala-Martinez, Ram Rajagopal, Claudia Sitgraves, Melissa Tartari and participants in the U.C. Berkeley Labor Seminar. I would also like to thank Angela Andrus, Arnie Reznek and Jon Stiles of the Census Bureau for their competence and kindness. The research in this paper was conducted while I was Special Sworn Status researcher of the U.S. Census Bureau at the Berkeley Census Research Data Center. 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 insure that no confidential data are revealed. All remaining errors are my own.

†Department of Economics, University of California, Berkeley. E-mail: caetano@econ.berkeley.edu. Phone: +1(510)789-3324. Address: 508-1 Evans Hall #3880, Berkeley, CA 94720-3880 (USA). http://gregorio.caetano.googlepages.com/

Abstract

In this paper I present new estimates of the value of school quality, derived from differences in rental prices for homes in different school districts, using detailed geographic, household and individual information from the restricted-access 2000 U.S. Decennial Census data. The main contribution of the paper is to develop an approach to identify and estimate how much parents are willing to pay for an improvement in school quality per year, per child and for each grade, from kindergarten to twelfth grade, hence providing to the policy maker a measure that can be directly used to make cost-benefit analysis of investments to improve school quality. The previous literature has not focused on providing a specific unit for the estimate, either per year, per child, or per grade. I build on Berry, Levinsohn and Pakes (1995), Rust (1987) and Heckman and Scheinkman (1987) to extend existing discrete choice models of residential location to address three critical issues. First, I account for differences across families in the value of local school quality by allowing different values for families with children in different grades. Second, I use a dynamic framework to distinguish between the value of school quality services in a given year, and the option value of living in districts with different values of school quality. Third, by combining information on the location choices of families with no children, and those with children of different ages, I control for unobserved neighborhood amenities, including characteristics of the people who live in the neighborhood, that potentially confound the identification of school quality valuation due to sorting. I implement the approach using over 200 school districts in New Jersey. My estimates suggest that parents with a child in high school are willing to pay about 2.7% higher rent per year in order to send that child to a district with 5% higher average test scores. The corresponding valuations are 1.6% for a family with a child in middle school, and 1.4% for a family with a child in elementary school.

1 Introduction

The goal of this paper is to identify and estimate how much parents are willing to pay for a giving improvement of school quality (as measured by students' test scores), in order to inform the policy-maker about the benefit yielded by the outcome of a given investment in public school quality. This measure is referred to as the marginal willingness to pay (MWTP) for school quality.

In general, an investment to improve school quality can have a different effect across years and across grades. Moreover, the number of children who are benefited from this improvement during the period of one year and for each grade also matter for the estimation of the total benefit of the investment. Therefore, the MWTP for school quality has to be measured in the unit per year, per child and for each grade, so that the total benefit of a given investment can be compared to the total cost of the investment. The main contribution of this paper is to develop an approach to identify and estimate the MWTP for school quality exactly for this unit, i.e. per year, per child and per grade, from kindergarten to twelfth grade. The literature has not yet focused its attention on giving a specific unit to the MWTP estimate in any of the dimensions per year, per child, or per grade.[1]

Since there is no direct market to purchase public school quality, researchers have focused on interpreting public school quality as a neighborhood amenity, and measuring the valuation of school quality indirectly through the housing market. The housing market introduces at least three challenges to the estimation of the MWTP for school quality in the unit proposed. The first challenge is that the family is the agent who makes the residential decision, and not the child. Since families in general have more than one child, and the children can be attending different grades, one needs to disentangle the valuation per child and per grade from the valuation of each family. The second challenge is that, because of moving costs, when families make residential decisions, they take into account not only the current per year valuation of school quality, but also the option value of residing in each neighborhood, which is a function of the future levels of school quality as well as of the preferences for the grades that the children of the family are expected to attend in the future.[2] This makes it hard to disentangle the valuation of school quality per year and per grade

[1] See Black (1999) and Bayer, Ferreira and McMillan (2007) for the state of the art in this literature.

[2] Throughout the paper, the "current year" refers to the year 2000 and the "future year" refers to the year 2001.

from the cumulative valuation[3] of school quality, which spans more than one year and more than one grade.

The third challenge is the serious endogeneity problem which is inherent in the estimation of preferences for neighborhood amenities. Essentially, since amenities are bundled in a neighborhood, some of them may be correlated to school quality, and unobserved by the researcher. Any procedure which does not address this problem will provide estimates of the preference for school quality that partly reflect the preference for the unobservable amenities. Although this problem has been the focus of the literature, no method so far fully addresses the endogeneity generated by household sorting.[4] The default procedure of treatment of endogeneity with instrumental variables is inadequate in the case of neighborhoods, because the parameter of interest is the direct effect of a variation of school quality in the utility of a neighborhood, as opposed to the total effect.[5] If households sort because of neighborhood amenities, a variation in school quality, no matter how random is its source, will generate a change in the neighbors' characteristics, which are themselves amenities, some of which unobservable to the researcher. Random variations of school quality are therefore not useful in identifying the portion of the variation in the utility of the neighborhood due exclusively to school quality, and not for instance to changes in the racial composition of the neighborhood.

In this paper I use detailed geographic, household and individual information from the restricted-access 2000 U.S. Decennial Census data, and develop an unified approach to overcome these three challenges.[6] I develop and estimate a dynamic discrete choice model of neighborhood decisions

[3]The family's current cumulative valuation of a particular neighborhood is the present discounted flow of utility this family expects to enjoy from the consumption of this neighborhood in the current period. If a family decides on a path of neighborhoods to live in each period, starting in a given neighborhood for this period, and then staying or moving according to the expected best path from the next period on, then the current cumulative valuation of this neighborhood is the expected discounted utility of this path.

[4]Following Tiebout (1956)'s seminal paper, there is an important line of theoretical work in local public finance on the effects of sorting in the housing market. Important contributions include Epple and Zelenitz (1981), Epple, Filimon, and Romer (1984, 1993), Benabou (1993, 1996), Fernandez and Rogerson (1996), and Nechyba (1997). Also, interesting empirical contributions on sorting include Black (1999), Epple and Sieg (1999), Barrow (2002), Urquiola (2005), Chay and Greenstone (2005), Rothstein (2006) and Bayer, Ferreira and McMillan (2007).

[5]For simplicity throughout the paper I refer the indirect utility as just utility.

[6]More generally, note that these three challenges are also present when one wants to estimate the MWTP for any other amenity, such as crime rate, racial composition and clean air, so in principle the approach developed in this paper can be also applied to the estimation of MWTP for any other amenity. For policy reasons, the parameter that one would want to estimate is the MWTP for a given amenity per year and per person.

with moving costs as a method of estimating each family's per year value of a given neighborhood. Then, I relate the per year value of each neighborhood to observed neighborhood-level amenities, including school quality, and estimate the MWTP for school quality per year at the family level, controlling for unobservable amenities in order to account for the endogeneity problem, even when it is due to sorting. Finally, I disentangle the MWTP for school quality per year, per child and per grade from the MWTP for school quality per year at the family level, accounting for the fact that families may have more than one child attending school, and that these children may be attending different grades.

As previously stated, the literature on the estimation of the MWTP for school quality has focused exclusively on the problem of endogeneity. In a seminal paper, Black (1999) addressed the endogeneity problem in a hedonic price framework[7] by comparing the prices of similar houses close to each other, but on different sides of the school attendance boundaries. Her analysis assumes that, with the exception of school quality, all the other neighborhood amenities remain constant across the boundary. Hence, the difference in house prices across the boundary, controlling for observed house characteristics, is interpreted as entirely due to differences in school quality. More recently, Bayer, Ferreira and McMillan (2007) tested her assumption with a more detailed data set, and found that some observed neighborhood amenities, including characteristics of neighbors, can explain variation in house prices across boundaries, concluding that the boundary strategy alone cannot fully address the endogeneity problem.[8] They also showed convincing evidence of substantial sorting across boundaries, which brings further evidence of the necessity to deal with the endogeneity problem brought up by sorting, even with the use of boundary fixed effects.[9] Finally, they built on the standard literature on endogeneity in discrete choice models first introduced by Berry (1994) and Berry, Levinsohn and Pakes (1995), henceforth BLP,[10] and esti-

[7]See the seminal work of Rosen (1974) for a discussion on hedonic price regressions.

[8]Bayer, Ferreira and McMillan (2007)'s identification assumption is that, conditional on observed characteristics of the house and the neighborhood, any variation in price across the boundary is due to differences in school quality.

[9]Bayer, Ferreira and McMillan (2007) warn that sorting may be also causing unobservable differences in the neighborhoods' house characteristics, which could in part be accounting for the differences in prices across the boundary. For example, wealthier households may sort to neighborhoods with better schools, and may also be more likely to hire a landscaper or paint their house. These unobservable improvements in housing quality increase houses' values, and this increase would be mistakenly attributed to school quality. Note that this type of endogeneity due to sorting prevails even if the neighborhoods were indeed the same across boundaries, except for school quality.

[10]BLP develop a discrete choice estimation approach in two steps in order to deal with the endogeneity problem. In

mated a static discrete choice model of residential location decisions, both incorporating Black's boundary identification idea, and allowing for heterogeneity of preferences for amenities.

This paper builds on BLP, Rust (1987), and Heckman and Scheinkman (1987) to create a framework to identify and estimate the MWTP for school quality per year, per child and per grade. The empirical strategy proposed in this paper has three steps. In the first step, I divide the population of families in several cohorts of parents and non-parents, defined by the age of the oldest child for parents, and by the age of the head of the family for non-parents, and estimate a dynamic discrete choice model of neighborhood decision with moving costs, building on BLP, Rust (1987) and a synthetic cohort assumption.[11] The goal of this step is to estimate the mean per year valuation at the cohort-neighborhood level.[12] I use the insight from Rust (1987) that each family's current per year utility of living in a given neighborhood can be written as the family's current cumulative utility of that neighborhood plus a known continuous function of the family's future cumulative utilities of all neighborhoods. I then estimate each family's current per year utilities for each neighborhood by estimating the family's current and future cumulative utilities for all neighborhoods, hence estimating the per year mean utility at the cohort-neighborhood level by averaging the per year utilities of the families of each cohort, for each neighborhood. In order to estimate the current cumulative valuations of each family for each neighborhood, I estimate *separately for each*

the first step they estimate a discrete choice model using fixed effects at the choice level. In the second step, they use the estimated fixed effect for each choice as a dependent variable in a regression on observed choice characteristics, including price. The key insights relevant for this paper are twofold: first, the fixed effects in the first step carry the source of endogeneity to the second step. In the second step, the dependent variable is then evaluated in a linear model, so standard instrumental variable methods to solve the endogeneity problem become feasible. The second insight is that the estimated fixed effect has an economic interpretation: it is the mean utility of each choice across all individuals of the population. This idea is of great assistance in thinking of instrumental variables to solve the endogeneity problem in the second step. In particular, BLP build on Bresnahan (1987) and suggest a set of instrumental variables for price, but assuming exogeneity of the products' observable characteristics (except price). Similar instruments were used in Bayer, Ferreira and McMillan (2007) in their application to residential choice, with the assumption of exogeneity of the amenities conditional on the boundary fixed effects. The approach proposed in this paper does not use such instruments, and deals with the endogeneity of price as well as school quality using a different approach that also accounts for the endogeneity due to sorting.

[11] See Newell (1990) for a discussion about synthetic cohorts.

[12] The first step of my approach can be seen as a modified version of the first step of the standard (demand side) BLP approach, with basically two differences. First, the goal is to estimate the mean valuation at the cohort-neighborhood level, as opposed to the neighborhood, in order to create a panel data for the second step at a lower level than the neighborhood level, which will allow me to deal with the endogeneity problem with a novel approach. Second, the goal is to estimate the mean per year valuation, as opposed to the mean cumulative valuation, which will be crucial not only for being able to disentangle the valuation of school quality per year and per grade, but also for the endogeneity treatment, because it makes the exclusion restrictions relatively weaker.

cohort a *static* discrete choice model with moving costs, using as utility the cumulative utility, with a simple modification of the standard BLP specification to add moving costs: a term with fixed effects at the cohort-neighborhood level, a term with the interaction of observed amenities with observed family's characteristics, and a term capturing moving costs.[13] Since the data set used is only cross-sectional, I cannot estimate the future cumulative utilities directly. I make a synthetic cohort assumption to estimate each family's future cumulative utility of each neighborhood.[14] More specifically, I use the estimated parameters of the current cumulative valuations of cohort $c+1$ as the parameters of the expected future cumulative valuations of cohort c. Additionally, I use the current observed neighborhood amenities interacted with the current family's attributes in the place of the expected future neighborhood amenities interacted with the expected future family's attributes. These two assumptions imply, for instance, that in 2000, a family whose oldest child is 4 years old expects to value (cumulatively) each neighborhood in 2001 the same as a similar family whose oldest child is 5 years old values the same neighborhood in 2000.[15] With these assumptions, I am able to estimate the current and future cumulative valuations of each family for each neighborhood, which in turn allow me to estimate the mean per year valuation of each cohort for each neighborhood.

The mean per year valuations estimated in the first step are a panel at the cohort-neighborhood level. In the second step, I build on Heckman and Scheinkman (1987)[16] and develop an approach

[13] My specification of the cumulative utility is similar to the specification of the utility in Bayer, Ferreira and McMillan (2007), but with the addition of a term that accounts for moving costs, and the fact that I estimate the discrete choice model separately for each cohort, so in reality I am estimating fixed effects at the cohort-choice level, rather than at the choice level, and I am allowing the coefficients of the interaction parameters to be different across cohorts. Bayer, Keohane and Timmins (2006) estimated a static discrete choice model of residential choice using the BLP approach with a similar term to account for moving costs in order to estimate the MWTP for clean air. See also Bayer, McMillan, Murphy and Timmins (2007).

[14] Hotz and Miller (1993) and Hotz, Miller, Sanders and Smith (1994) also suggest the use of synthetic cohorts to estimate a dynamic model when the researcher is able to estimate the conditional choice probabilities, i.e., the likelihood of choosing each neighborhood conditional on living previously in a given neighborhood. I do not observe this information in the decennial U.S. Census. However, I observe whether or not each family moved last year, and I use this information to identify the moving cost parameter.

[15] "Similar" in this context refers to the family's observed characteristics, as well as whether the family moved or not last year. It is important to mention that the approach proposed in this paper is robust to deviations of the assumption of synthetic cohorts because of the strategy of step two. See section 4.

[16] Their approach builds on earlier work by Chamberlain (1977) and Pudney (1982). Chamberlain (1977) originally referred to the technique used in this second step as Proxy-IV method. Heckman and Scheinkman's (1987) original application is a test for the equality across sectors of the returns to skills in the labor market. In that paper, they do not go as far as to identify and estimate the coefficients of interest, as I do in this paper, since they were only interested in testing for the equality of the coefficients across sectors.

to recover the estimates of the family's MWTP for school quality per year *by cohort*, controlling for unobservable neighborhood-level amenities which may be correlated with school quality. I specify the per year valuations as linear functions of observable and unobservable neighborhood-level amenities plus a random term, where the parameters of preference for both the observed and unobserved amenities can vary with cohort. This panel system of linear equations has a particular structure that I explore to account for endogeneity. The insight is to impose restrictions in the parameters of the system in order to identify the MWTP for school quality for the cohorts of interest. In this context, the cohorts of interest are the cohorts of parents whose oldest child is at school age.[17]

The structure of the panel system of equation implies that cohorts are exposed to the same amenities (observed and unobserved) if they choose the same neighborhood, though they may value them differently. I make the assumption that the cohorts of families without children of school age (i.e., either non-parents or parents whose oldest child is too young to attend school) do not value school quality in the current period.[18] Although these cohorts do not value school quality in the instantaneous sense,[19] they value amenities that are correlated to school quality, so these cohorts' instantaneous valuations bring information that can be used to control for the unobserved amenities to which the cohorts that value school quality in the instantaneous sense are exposed.[20]

To understand the intuition behind this procedure, suppose, as an example, that average neighborhood income is an unobserved amenity. Richer households place higher value on school quality, so the average neighborhood income is a source of endogeneity in the panel system of linear equa-

[17]One drawback of the Census data is that the grade that the child attends is not perfectly observed in the data. Instead, one only observes in which category of grades each child is attending (i.e., kindergarten, grades 1-4, grades 5-8 and grades 9-12). I use information on each child's exact date of birth available in the restricted-access Census data, as well as the kindergarten age entry rules of each state, in order to improve the match between cohort assignment and grade assignment. I am also able to control for a "fuzzy design" with observed dummy variables at the family's level for whether the predicted grade of the oldest child of that family is indeed inside the reported category and whether the oldest child is indeed attending a public school.

[18]I allow these cohorts to have a preference for school quality in terms of expected cumulative valuation, since they may expect to have a child who will attend a public school in the future. However, I assume that they do not value the school quality services provided in the current period, since they do not yet have children attending school.

[19]Throughout the paper, I am going to refer to per year utility as the instantaneous utility.

[20]This proxy adds a known source of endogeneity, which arises from the correlation between the mean per year valuation of the proxy group and the random error term of the proxy group. I follow Chamberlain (1977)'s suggestion and use the mean per year valuations of a third group of cohorts as instrument for the mean per year valuation of the proxy group of cohorts, since the valuations of the group used as instrument are assumed not correlated with the random terms of the proxy group.

tions described above, since neighborhoods with higher school quality will tend to attract higher income residents. Hence there will be correlation between school quality and average income of the neighborhood that cannot be disentangled with a random source of variation of school quality. In what follows, I am going to explain how the proposed approach deals with this endogeneity problem. Suppose there is an increase in school quality in a given neighborhood. This increase in school quality will likely generate an increase in the average income of the neighborhood, because richer people will sort to that neighborhood. As a result, an increase in school quality will affect the mean per year valuation of that neighborhood for families with children of school age, both directly because they value school quality in the current period, and indirectly through the increase in the neighborhood average income. However, the increase in school quality will affect the mean per year valuation of childless families and families with children too young to attend school only indirectly, through the increase in the average income of the neighborhood, since they do not enjoy the amenity school quality in the current period, but they do enjoy the amenity average income of the neighborhood. For this reason, the mean per year valuation of families that do not value school quality in the current period can be used as proxy variable for the average income of the neighborhood, in order to control for the indirect effect of a change in school quality in the valuation of parents with children at school age.[21]

Finally, in the third step I have the MWTP for school quality per year for each cohort of families whose oldest child is attending school. I then need to estimate the MWTP for school quality per year, per child and per grade, by decomposing the valuation of the families, since they can have more than one child attending school, and they could also be attending different grades. For that, I use individual detailed information on each child of each family available in the restricted-access version of the 2000 Census data to know, for each family, the grade that each child is attending. I make the assumption that, conditional on observed family attributes, such as household income,

[21]This identification strategy is related to the strategy from Barrow (2002), who specified the utility function with both school quality and another term with the interaction of school quality and a *dummy* variable of presence of children under age 18 in the household. My approach differs from hers mainly for three reasons: first, I make assumptions with respect to the per-period valuation, rather than the cumulative valuation; Second, I allow for parents with children at different grades and non-parents of different ages to value observed and unobserved amenities differently. Third, rather than using standard multinomial logit estimates, I build on BLP and Heckman and Scheinkman (1987) to allow for more flexible forms of the unobserved term.

number of children and marital status, only the grade that the child is attending matters for the valuation of school quality. In particular, birth order, sibling spacing and the age of the parents do not matter for the valuation of school quality, conditional on observed attributes of families. For instance, take two similar (in terms of observed family attributes) families with two children each, one family whose oldest child is attending grade 5, and the other family whose oldest child is attending grade 8 and the second oldest child is attending grade 5. My assumption is that the value that the two families place on school quality for the children who are attending grade 5 is the same.

I find that, for the state of New Jersey, parents are willing to pay, for a 5% increase in test scores, an exceeding percentage over rent value of 1.4% in elementary school, 1.6% in middle school, and 2.7% in high school, with relatively high levels in grades 11 and 12. The average monthly rent is $755, so this represents a yearly average MWTP of $780 for a one standard deviation (i.e., 16%) increase in test scores in secondary school, $460 for middle school and $400 for primary school. To compare with the estimates encountered in the previous literature, Black (1999) finds a MWTP for school quality of 2.5% of house prices, and Bayer, Ferreira and McMillan (2007) find a MWTP for school quality of under 1% of house prices. Both values correspond to a 5% increase in school quality. Besides the differences in the samples,[22], three main reasons account for the differences in the estimates. First, the estimates from this paper measure the per year, per child and per grade valuation of school quality, while previous estimates have an unknown unit (i.e., not at the year-child-grade level). Second, the estimates from this paper correspond to the average MWTP for school quality across families whose oldest child is attending a public school, while the previous estimates correspond to the average[23] MWTP for school quality across the whole population.[24] Finally, the approach proposed in this paper extends the treatment of endogeneity to account for sorting.

[22]Black (1999) uses a sample of school districts in Massachusetts, and Bayer, Ferreira and McMillan (2007) use a sample of school districts in the San Francisco Bay Area.

[23]See Epple (1987) and Ekeland, Heckman and Nesheim (2004) for a discussion of what exactly the hedonic coefficients identify.

[24]Bayer, Ferreira and McMillan (2007) interact observed household characteristics with observed house and neighborhood amenities, allowing for families with children younger than 18 years of age to have a higher preference for school quality. However, their approach does not allow families with children at different grades to sort differently across neighborhoods, and do not allow families with children younger than 18 years of age to sort differently than families with children older than 18 years of age other than because of the observed amenities.

From a methodological point of view, this paper makes two additional contributions. I develop two independent modifications of the BLP approach, one to deal with dynamics, and other to deal with endogeneity. In the first contribution, I substitute the first step of the standard BLP approach, that typically estimates the mean utility for each choice, for a method that estimates the mean instantaneous utility instead. I use the specification of BLP with a term allowing for moving costs as well as synthetic cohort assumptions to estimate the current and future cumulative utilities, and then I use an "integrated" version of a Bellman equation from Rust (1987) to link them and estimate the mean instantaneous utility. The second step of BLP can then be used as in the standard case, but with the mean instantaneous utility as dependent variable, therefore identifying the coefficients of the mean instantaneous utility instead of the coefficients of the mean utility.[25] The contribution is in the fact that this method is feasible even when the researcher has access only to one wave of cross sectional data and a very large set of parameters.[26] Moreover, this method allows for unobserved heterogeneity and an unknown time horizon even with such constricted data.[27] This method can have direct application in the fields of local public finance and urban economics, as it is very difficult to find data with crucial features for these fields (i.e., geographical identifiers at a low level and household data) and at the same time features that allow the estimation of dynamic models.

The second contribution builds on BLP and Heckman and Scheinkman (1987) to create an alternative approach to deal with endogeneity in discrete choice models when disaggregated data

[25] Also related to the approach developed in this paper, there is a very recent literature on the dynamic demand for durable goods, of which houses and neighborhoods are examples. Key contributions include Melnikov (2001), Carranza (2007), Erdem, Imai and Keane (2003), Hendel and Nevo (2006), Gowrisankaran and Rysman (2007), Schiraldi (2006) and Bayer, McMillan, Murphy and Timmins (2007). In particular, Bayer, McMillan, Murphy and Timmins (2007)'s application is on residential location choices, as in this paper.

[26] In this paper, I estimated over 400 parameters for each of the 38 cohorts. Because of the size of the parameter set, Rust's (1987) approach is unfeasible. Moreover, because we do not observe the neighborhood each family was living last year, we cannot estimate the conditional choice probabilities and use a method in the spirit of Hotz and Miller (1993). My approach can be seen as a modified version of Hotz and Miller (1993), where instead of estimating the conditional choice probabilities with a flexible specification and using the representation theorem to relate them with the cumulative valuations, one estimates the current cumulative valuation directly as if in a static model, with a very flexible specification, and use synthetic cohorts assumptions to get an estimate of the expected future cumulative valuations as well, later using Rust's (1987) insight to estimate the mean instantaneous valuation. Adding fixed effects at the cohort-neighborhood level is crucial in this approach, because it allows each cohort to have different option value for each neighborhood, hence not needing to rely on the interaction parameters to explain variation in the cumulative valuations across cohorts within the same neighborhood.

[27] For an excellent recent survey on methods of estimation for dynamic discrete choice models, see Aguirregabiria and Mira (2007).

is available. I modify the first step of the standard BLP approach to create a panel data set in the second step at the group-choice level, rather than a cross-sectional data set at the choice level. This panel has a structure that can be used to deal with the endogeneity problem. The requirement is twofold: first, one needs to assume that all the endogeneity can be controlled by choice-level unobserved characteristics, allowed to be valued differently across groups of individuals. Second, that it is possible to impose conditions in the relationship of the parameters of preferences of some groups, in order to identify the parameters of preferences of other groups.[28] This method can be applied in situations where Instrumental Variables seem to be inadequate or unavailable, as in this paper. This is clearly the case for choices with network effects (i.e., choices for which people care about functions of characteristics of who else makes the same choice). Examples are choice of occupation, college choice, school choice, fashion goods, websites, software and, as shown in this paper, neighborhood choice, since people care about who are their neighbors. Moreover, the method can be applied in situations when the researcher does not want to make assumptions with regard to the specification of control variables. The idea is to control for unobservables, and assume that whatever the unobservable is, it is the same across all people (although people are allowed to value it differently). For instance, in the case of this paper, we as researchers do not know which are the characteristics that change because of sorting due to variation of any of the observed amenities. Moreover, we do not know which characteristics of neighbors people care about (e.g. racial composition, income, wealth, level of education, how "nice" are the neighbors), and even what is the function of these characteristics (e.g. average, median, standard deviation, etc.). Any misspecification on this account will bias our estimates. The insight is to assume that different groups are exposed similarly to these unobserved amenities, without the need to specify their nature.

This paper is organized as follows. In section 2, I introduce the dynamic model of neighborhood choice, and derive the MWTP for school quality as a function of the components of the model. In section 3, I present the identification strategy. In section 4, I build the estimation approach in three steps. In section 5, I show the empirical implementation, the results and also some robustness

[28]This last requirement is not necessary if one is interested in measuring the MWTP of a given characteristic that changes at the group-choice level, instead of at the choice level.

checks. I conclude in section 6 with a summary and discussion of future research.

2 A Dynamic Model of Neighborhood Choice

In this section, I introduce a dynamic model of neighborhood choice based on Rust (1987), and arrive at the definition of MWTP for school quality for each grade in this model. The setup is similar to the standard models in the literature.[29] The basic idea is that, at each period, the household chooses a neighborhood, with full information of the current period, and only partial information of the future. The unknown part of the future is assumed random, and specific to the period, household and choice.

Suppose that there are C cohorts, indexed by c. The cohorts in this model refer to the age of the head of household, in childless households, and to the age of the oldest child, in households with children. Each cohort has I_c households, indexed by $i = 1, ..., I_c$.

There are J neighborhoods, indexed by $j = 1, ..., J$. The neighborhoods in this model are school districts. Each neighborhood has a set of characteristics, such as school quality and average rent, which I will refer as "amenities." The household takes the neighborhood's amenities in consideration when deciding where to live.

Each period "t" corresponds to one year. The amenities are measured per neighborhood per period, and the preferences are measured per household per period. Amenities and preferences are divided according to three categories: they can be observed to the researcher ($\boldsymbol{W}$), they can be unobserved but estimable to the researcher ($\boldsymbol{\theta}$), or they can be unobserved and not estimable to the researcher ($\boldsymbol{\epsilon}$). Let $\boldsymbol{S}_{i,c,t} := (\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}, \boldsymbol{\epsilon}_{i,c,t})$. I will refer to $\boldsymbol{S}_{i,c,t}$ as "state variable." At each period t, the household observes $\boldsymbol{S}_{i,c,t}$ before making the decision. Also, the household is able, as of period t, to perfectly predict the future distribution of $\boldsymbol{W}$ and $\boldsymbol{\theta}$, conditional on its choice in period t and the present $\boldsymbol{W}$ and $\boldsymbol{\theta}$. However, the household cannot predict $\boldsymbol{\epsilon}_{i,c,t+1}$ until it reaches period $t+1$, and thus treats it as random variable as of period t with a known distribution, also known to the researcher.

A household in cohort c expects it will live for T_c periods, but the researcher does not observe

[29] See Aguirregabiria and Mira (2007).

this value. Also, $T_{c+1} = T_c - 1$, so that childless households where the head is 30 years old expect to survive one year longer than childless households where the head is 31 years old. Analogously, a household whose oldest child is 5 years old will survive one year longer than a household whose oldest child is 6 years old.

At each period t, the household observes $\boldsymbol{S}_{i,c,t}$ and chooses the neighborhood it is going to live, based on the fact that, though it may be costly, it can move out of this neighborhood in the next period. Because there are moving costs, the household's present decision also affects its decisions in future periods as well, so the household takes into account the option value of each choice. The future discount, denoted β, is the same for all households, and known to the researcher. When the next period arrives, the household observes the realization of $\boldsymbol{S}_{i,c,t+1}$, and makes a new choice of neighborhood using the same procedure as in the last period.

Specifically, household i of cohort c decides in period t on a sequence of neighborhoods from period t onwards, denoted $(\boldsymbol{j}_{i,c})_{\boldsymbol{t}} = \{j_{i,c,t}, j_{i,c,t+1}, ..., j_{i,c,T_c}\}$. 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, individual i of cohort c chooses $j_{i,c,t} = j$. The household's valuation of neighborhood j in period t is given by $U_j(\boldsymbol{S}_{i,c,t})$. Therefore, the household's valuation in period s can be written as $\sum_{j=1}^{J} d_j(j_{i,c,s}) U_j(\boldsymbol{S}_{i,c,s})$.

The value of the path $(\boldsymbol{j}_{i,c})_{\boldsymbol{t}}$ is the current valuation, plus the discounted sum of the valuations of future periods. However, since the value of $\boldsymbol{S}_{i,c,s}$ for $s > t$ is not known as of period t, the household chooses the path which maximizes de *expected* cumulative value, conditional on all information available in period t, given in $\boldsymbol{S}_{i,c,t}$. This function is known as "value function", and is denoted $\boldsymbol{V}(\boldsymbol{S}_{i,c,t})$. The household's dynamic optimization problem can be expressed as

$$\boldsymbol{V}(\boldsymbol{S}_{i,c,t}) := \max_{(\boldsymbol{j}_{i,c})_{\boldsymbol{t}}} E\left[\sum_{s=t}^{T_c} \beta^{s-t} \sum_{j=1}^{J} d_j(j_{i,c,s}) U_j(\boldsymbol{S}_{i,c,s}) \middle| \boldsymbol{S}_{i,c,t}\right]. \tag{1}$$

Define:[30]

$$V_j(\boldsymbol{S}_{i,c,t}) := U_j(\boldsymbol{S}_{i,c,t}) + \beta E\left[\boldsymbol{V}(\boldsymbol{S}_{i,c,t+1}) \middle| \boldsymbol{S}_{i,c,t}\,,\, j_{i,c,t} = j\right]. \tag{2}$$

[30]The expectation in equation (2) is taken over the distribution of the random components of $\boldsymbol{S}_{i,c,t+1}$, conditional on $\boldsymbol{S}_{i,c,t}$ and $j_{i,c,t} = j$.

Then, it is possible to write the value function as a Bellman equation[31]

$$\boldsymbol{V}(\boldsymbol{S}_{i,c,t}) = \max_j V_j(\boldsymbol{S}_{i,c,t}). \tag{3}$$

Assumption 2.1. *(Additive separability) The current period valuation of choice j can be decomposed in two separable parts,*

$$U_j(\boldsymbol{S}_{i,c,t}) = u_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) + \varepsilon_{i,c,t,j}.$$

Assumption 2.1 entirely parameterizes part of the current period valuation of choice j. The first component, $u_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ will be referred to as "instantaneous utility."[32] Define the "mean instantaneous utility" for neighborhood j of cohort c in period t as the expected value of the instantaneous utilities across households of cohort c in period t:[33]

$$\delta_{c,t,j} = \boldsymbol{E}(u_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})). \tag{4}$$

The mean instantaneous utility is the cohort's current period valuation of choice j, and satisfies

$$\delta_{c,t,j} = X_{t,j}\pi_{c,t} + \xi_{c,t,j}, \tag{5}$$

where the $X_{t,j}$ are neighborhood j's observable amenities in period t, $\pi_{c,t}$ is the parameter vector that measures the preferences for each observed neighborhood amenity in period t, and $\xi_{c,t,j}$ is an unobserved term that varies at the cohort-period-neighborhood level.

Let $\pi_{c,t}^{\text{SQ}}$ be the coefficient of the amenity "school quality", which in this context is defined as average test score,[34] and let $\pi_{c,t}^{\text{Rent}}$ be the coefficient of the average neighborhood rent. Then, the

[31] Bellman (1957).

[32] For notational purposes, I refer to "instantaneous utility" as the estimable part of the actual instantaneous utility, since this part is the one that is going to be used frequently from now on.

[33] The expectation in equation (4) is taken over the population of households of cohort c in period t, but the results shown in sections 3 and 4 are also valid if the expectations are taken over any subset of this population, as long as the underlying sample of this subset is large enough, so that the asymptotic arguments remain valid. In section 5, I implement the approach with the expectation taken over the population of households of cohort c in period t whose oldest child is attending a public school and was assigned to the correct grade.

[34] More details in section 5.

value of school quality services in period t, or *MWTP for school quality for cohort c in period t*, is defined as[35]

$$\text{MWTP}_{c,t}^{SQ} \equiv -\frac{\pi_{c,t}^{\text{SQ}}}{\pi_{c,t}^{\text{Rent}}}. \tag{6}$$

As stated earlier, the cohorts refer to the age of the oldest child in families with children. More specifically, the cohorts correspond the age of the oldest child in the household as of the cutoff date for admission into kindergarten in 2000. Hence, if the cutoff is October 1, cohort 6 is composed of the households whose oldest child completed 6 years of age between October 2 of 1999, and October 1 of 2000, which is the year the data was collected.[36] This way, cohort 6 corresponds to the households whose oldest child is in kindergarten when the data was collected, cohort 7 corresponds to the households whose oldest child is in first grade when the data was collected, and so on. Therefore, the MWTP for school quality for cohorts between 6 and 18 gives the MWTP for school quality for families whose oldest child is attending grades between kindergarten and twelfth grade.[37]

The MWTP defined in equation (6) is not yet the MWTP for school quality per year, per child and per grade, which is the goal of this paper. Instead, it is the MWTP for school quality per year at the family level, and since a family can have two children at different grades, they are not in general the same. In section 3.3 I will show how I am able to identify the desired measure from $\text{MWTP}_{c,t}^{SQ}$.

Section 3 develops an approach for identification of the MWTP for school quality per year, per child and per grade. Section 4 shows explicitly how this approach can be used for estimation in three steps. Section 5 implements the three steps with the 2000 decennial Census data, and shows

[35] In principle, any other variable measured in dollars could be used in the denominator, such as average house value. I use the average rent, rather than the average house value, for two reasons: first, rent is more accurately measured than house value in the Census data (See Bayer, Ferreira and McMillan (2007) for a discussion about potential misreporting of house values.) Second, rent is a more robust measure than house value for the context of this paper, since it is more tightly connected to the actual per-period value of the house. House value is also a function of the expected future re-selling value of the house, so it can depend significatively on potential bubbles or bursts in the housing market.

[36] The Census data is reported as of April first, 2000.

[37] As one can see in section 5, since I observe the grade each child is attending as well as whether the child is attending a public school, I can control for any "fuzzy design" due to the fact that not all children who are at an age to enter in a public school at a given grade are indeed attending a public school at exactly that grade.

the results of the paper.

3 Identification of the MWTP for School Quality

The goal of this paper is to identify and estimate the MWTP for school quality per year, per child and per grade. For that, I am going to first identify $\text{MWTP}^{SQ}_{c,2000}$, which by equation (6) is a function of the coefficients $\pi^{\text{SQ}}_{c,2000}$ and $\pi^{\text{Rent}}_{c,2000}$. Equation (5) relates these coefficients to the mean instantaneous utility. Identification through equation (5) requires first that the mean instantaneous utilities $\delta_{c,j,2000}$ be identified, and second, that the issue of the likely correlation between $X_{j,2000}$ and $\xi_{c,2000,j}$ be addressed. Section 3.1 will show a strategy for identification of the $\delta_{c,2000,j}$, and section 3.2 will discuss endogeneity in equation (5), and present a strategy for identification of the coefficients $\pi^{\text{SQ}}_{c,2000}$ and $\pi^{\text{Rent}}_{c,2000}$. Finally, section 3.3 shows how to identify the MWTP for school quality per year, per child and per grade from $\text{MWTP}^{SQ}_{c,2000}$.

3.1 Identification of the mean instantaneous utility

I develop a strategy for identification of the $\delta_{c,t,j}$, which will be presented in the next three subsections. Here, I summarize the approach: in subsection 3.1.1, I arrive at the identification of $\delta_{c,t,j}$ as a function of the present and future choice-specific value functions, which will be defined later, but consist on the period's cumulative utility of a choice at a given period. This function can be seen in equation (9). In subsection 3.1.2, I arrive at an estimator of the choice-specific value functions using a BLP-type specification approach. The choice-specific value functions will be expressed as functions of observable neighborhood and household characteristics in each period, and a cohort-period-neighborhood specific fixed effect, as can be seen in equation (10). The estimation of this equation for future periods requires data on future characteristics of neighborhoods and households, which are not observable when one has access to only one cross-sectional data set, as in this paper. In subsection 3.1.3, I use the idea of synthetic cohorts, and express future choice-specific valuation functions as functions of present neighborhood and household characteristics. This can be seen in equation (11), and completes the identification of $\delta_{c,t,j}$.

3.1.1 Expressing the mean instantaneous utility as a function of the present and future choice-specific value functions.

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

Assumption 3.1. *(Conditional independence) Let $\boldsymbol{P}(A)$ denote the probability of A, then*

$$\boldsymbol{P}(\boldsymbol{W}_{i,c,t+1}|\boldsymbol{S}_{i,c,t}, j_{i,c,t} = j) = \boldsymbol{P}(\boldsymbol{W}_{i,c,t+1}|\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}, j_{i,c,t} = j).$$

Assumption 3.2. *(Extreme value i.i.d. unobservables)*

$\varepsilon_{i,c,t,j}$ is distributed iid Extreme Value 1.

Assumption 3.3. *(Discrete support of $\boldsymbol{W}$)*

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

Assumptions 2.1 and the three assumptions above are equivalent to the assumptions in Rust (1987), and are also used in the majority of the literature on estimation of dynamic models.[38]

Assumption 3.3 is mostly technical. Assumptions 3.1 and 3.2 are crucial to the analysis, because they substantially reduce the information gap between the household and the researcher. In period t, the household observes $\boldsymbol{\epsilon}_{i,c,t}$, while the researcher does not. Assumption 3.2 gives the researcher information about the distribution of $\boldsymbol{\epsilon}_{i,c,t}$, and restricts the effect that this shock can have in future shocks (in fact, it states that it has no effect at all). Assumption 3.1 restricts the effect that $\boldsymbol{\epsilon}_{i,c,t}$ has in the future value of the observable variables. Taken together, these assumptions transform $\boldsymbol{\epsilon}_{i,c,t}$ in 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, the entire dynamic optimization problem faced by the household is now available to the researcher, except for the exact value of $\boldsymbol{\epsilon}_{i,c,t}$, and the values of the parameters.

[38] See Keane and Wolpin (1997, 2001) for interesting exceptions to this rule.

Proposition 1. *Under assumptions 2.1, 3.1, 3.2, and 3.3, the $V_j(\boldsymbol{S}_{i,c,t})$ defined in equation* (2) *can be written as*

$$V_j(\boldsymbol{S}_{i,c,t}) = v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) + \varepsilon_{i,c,t,j}, \tag{7}$$

where $\varepsilon_{i,c,t,j}$ is the same as in assumption 2.1, and

$$\begin{aligned} v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) :=& u_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) + \\ &+\beta \sum_{\boldsymbol{W}\in\mathbb{W}} \left(\gamma + \log\left(\sum_{r=1}^{J} \exp(v_r(\boldsymbol{W}, \boldsymbol{\theta}_{c,t+1}))\right)\right) \boldsymbol{P}(\boldsymbol{W}|\boldsymbol{W}_{i,c,t}\,,\, j_{i,c,t} = j), \end{aligned} \tag{8}$$

where $\gamma \approx 0.577$ is the Euler's constant.[39]

Proof. See Rust (1987). □

I will refer to the $v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ as the "choice-specific value functions."[40] These functions represent the non-random part of the cumulative utility of choice j. Equation (7) shows a structure for $V_j(\boldsymbol{S}_{i,c,t})$ which is identical to the one in assumption 2.1 in everything but the fact that the estimable part is expressed in terms of the choice-specific value functions, instead of the instantaneous utility functions. The advantage of the use of the choice-specific value functions is that these functions can be directly estimated in a static discrete choice framework. This is due to the fact that there exists a direct correspondence between the choice-specific value function and the actual choices of the household, which is not the case with the instantaneous utility.

From proposition 1, I can write, for each i, c, t and j,

$$\begin{aligned} \delta_{c,t,j} = E\Bigg(& v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) - \\ & -\beta \sum_{\boldsymbol{W}\in\mathbb{W}} \left(\gamma + \log \sum_{r=1}^{J} \exp v_r(\boldsymbol{W}, \boldsymbol{\theta}_{c,t+1})\right) \boldsymbol{P}(\boldsymbol{W}_{i,c,t+1} = \boldsymbol{W}|\boldsymbol{W}_{i,c,t}\,,\, j_{i,c,t} = j)\Bigg). \end{aligned} \tag{9}$$

[39] γ is equal to the mean of the extreme value distribution.

[40] Again, as in the case of instantaneous utilities, for notational reasons I am defining the "choice-specific value functions" as only the estimable part of the actual choice-specific value functions, since this is the part that is going to appear frequently in the rest of the paper.

Intuitively, the assumptions of proposition 1 guarantee that the future decisions planned by the household in t for each possible state ($\boldsymbol{W} \in \mathbb{W}$) are similar to the actual decisions it will make in the future, up to a random error. From equation (9), in order to identify $\delta_{c,t,j}$ for all j, it is enough to identify $v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ and $v_j(\boldsymbol{W}, \boldsymbol{\theta}_{c,t+1})$, for all j and $\boldsymbol{W} \in \mathbb{W}$.

3.1.2 Specification of the present choice-specific value functions.

In this section I specify $v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ as a linear function of a set of K neighborhood observable amenities $X_{t,j}$, M observable household characteristics $Z_{i,c,t}$, an indicator variable for whether neighborhood j is new for household i in period t: $\text{NEW}_{i,c,t,j} = 1$ if $j_{i,c,t-1} \neq j$, and zero otherwise, and a fixed effect at the cohort-period-neighborhood level $\Delta_{c,t,j}$. The parameters of this specification are expressed as $\boldsymbol{\theta}_{c,t} \equiv \left(\{\Delta_{c,t,j}\}_{j=1}^{J}, \Gamma_{c,t}, \Phi_{c,t}\right)$, and the vector $\boldsymbol{W}_{i,c,t} = \left(Z_{i,c,t}, \{X_{t,j}\}_{j=1}^{J}, \{\text{NEW}_{i,c,t,j}\}_{j=1}^{J}\right)$. I specify $v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ as

$$v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) := \Delta_{c,t,j} + X_{t,j} Z_{i,c,t} \Gamma_{c,t} + \text{NEW}_{i,c,t,j} \Phi_{c,t}. \tag{10}$$

The fixed effect $\Delta_{c,t,j}$ captures unobservables at the cohort-period-neighborhood level. The other terms of the RHS of the equation highlight the heterogeneity of the choice-specific value function across households within cohort c, period t and choice j. Finally, $\Phi_{c,t}$ represents moving costs for each c and t.

Equation (10) is a sharp departure from standard methods of estimation of dynamic models.[41] Here, I am parametrically specifying the choice-specific value functions, rather than the instantaneous utility functions. As discussed in section 3.1.1, this is a key feature of the proposed approach.

It is important to notice that the specification itself is closely related to the first step equation of the BLP approach. In the context of housing, Bayer, Ferreira and McMillan (2007) apply the BLP approach with a very similar specification. This approach is different from theirs mainly for three reasons. First, equation (10) has one additional term to account for moving costs.[42] Second, equation (10) identifies the mean utilities at the cohort-period-neighborhood level, rather than at

[41] See Aguirregabiria and Mira (2007).

[42] See also Bayer, Keohane and Timmins (2006) and Bayer, McMillan, Murphy and Timmins (2007).

the neighborhood level. Third, the unit of choice in equation (10) is the neighborhood, rather than the house.

Equations (7), (10), and assumption 3.2 guarantee the identification of the parameters of $v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ (provided some regularity conditions are satisfied) inside of a maximum likelihood, discrete-choice framework (see BLP). As a result, $v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ is itself identified.

3.1.3 Identification of the future choice-specific value functions with cross-sectional data

The following assumption constrains $\boldsymbol{\theta}_{c,t+1}$ and $\boldsymbol{W}_{i,c,t+1}$, so that they can be inferred using only data on period t. For each i, c, j and t, assume:

Assumption 3.4. *(Synthetic cohort)*[43]

$$(i) \quad \boldsymbol{\theta}_{c,t+1} = \boldsymbol{\theta}_{c+1,t}$$

$$(ii) \quad \boldsymbol{P}(Z_{i,c,t+1} = Z_{i,c,t}\,,\, X_{t+1,j} = X_{t,j} | \boldsymbol{W}_{i,c,t}\,,\, j_{i,c,t} = j) = 1$$

Assumption 3.4 implies that a household in cohort c expects to have in the next period the same choice-specific valuation as another household (with the same observed characteristics Z) in cohort $c+1$ has now. Also, it implies that household i expects to have the same characteristics Z, and that the amenities X in all neighborhoods will remain the same. There are some arguments for why this assumption may not be strong in the case of this paper. First, observe that the stationarity is required with respect to the expectations of the households, not with respect to reality. Second, since this paper deals with periods $t = 2000$, and $t+1 = 2001$ only, the space of time for which stationarity is required is rather small. Finally, because of the strategy developed in section 3.2, this assumption may not be of importance in the identification of the MWTP coefficients.[44]

[43] Assumption 3.4 can be trivially generalized to expected trends both in the observable variables and in the parameters, as long as the expected trend is known by the researcher. In this case, all the results hold, but with the modified version of $v_j(\boldsymbol{W}_{i,c,t+1}, \boldsymbol{\theta}_{c,t+1})$. The assumption is

Assumption 3.4(b): *(Generalized version of assumption 3.4)*

$$v_j(\boldsymbol{W}_{i,c,t+1}, \boldsymbol{\theta}_{c,t+1}) = f_t(v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c+1,t})).$$

where f_t is known.

[44] See footnote 47. Moreover, assumption 3.4 is actually stronger than the necessary. The weaker version of this assumption is: $v_j(\boldsymbol{W}_{i,c,t+1}, \boldsymbol{\theta}_{c,t+1}) = v_j(\boldsymbol{W}^p_{i,c,t+1}, \boldsymbol{\theta}_{c+1,t})$ where $\boldsymbol{W}^p_{i,c,t+1} := (Z_{i,c,t}, \{X_{t,j}\}_{j=1}^J, \{NEW_{i,c,t+1,j}\}_{j=1}^J)$. See equation (11).

This assumption is related to the idea of synthetic cohorts[45] from the demography literature: for cohort c, I create the imaginary group of all households in cohort c one period later, and then "plug in" the estimated parameters from cohort $c+1$. For instance, consider the cohort of parents whose oldest child is of age 5. I create the imaginary group, which contains all these parents one year later when their oldest child would be of age 6. In order to create the choice-specific value functions of the parents of the synthetic cohort, I use the $\boldsymbol{W}$'s from the parents of cohort 5 and the $\boldsymbol{\theta}$'s from the cohort 6. Since both these cohorts are observed in period t, under assumption 3.4 all the information necessary to estimate the choice-specific value function of cohort 5 one year later is available in the data.

Under assumption 3.4, for each household i of cohort c, and for each neighborhood j, I can write the future choice-specific value function as:

$$v_j(\boldsymbol{W}_{i,c,t+1}, \boldsymbol{\theta}_{c,t+1}) = \Delta_{c+1,t,j} + X_{t,j} Z_{i,c,t} \Gamma_{c+1,t} + \mathrm{NEW}_{i,c,t+1,j} \Phi_{c+1,t}. \tag{11}$$

Equations (10) and (11) show respectively the choice-specific value functions of period t and $t+1$ as function only of data on period t and parameters.[46] Since the parameters of $v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ are identified (as seen in the last section), $v_j(\boldsymbol{W}_{i,c,t+1}, \boldsymbol{\theta}_{c+1,t})$ is identified. From equation 9, this completes the identification of $\delta_{c,t,j}$.[47]

[45] See Newell (1990).

[46] Note that $\mathrm{NEW}_{i,c,t+1,j}$ is only a function of $d_j(j^*_{i,c,t})$, where $j^*_{i,c,t}$, the actual choice of neighborhood of household i in period t, is part of the observed data in period t for each i and each c.

[47] Assumption 3.4 is necessary for identification of $\delta_{c,j}$ through equation (9). However, in the identification of the MWTP for school quality, this assumption may be of smaller importance. The reason is because the method for dealing with endogeneity developed in section 3.2 accounts for unobservable variables of the kind $Q_{t,j}\lambda_{c,t}$. The bias of misspecification in equation (11) can be expressed as a term $B_{i,c,t,j}$. This term will generate a bias of identification in equation (9) of the form

$$\beta \boldsymbol{E}\left(\sum_{\boldsymbol{W} \in \mathbb{W}} \log \left(\frac{\sum_{r=1}^{J} e^{B_{i,c,t,j}} \exp v_r(\boldsymbol{W}, \boldsymbol{\theta}_{c,t+1})}{\sum_{r=1}^{J} \exp v_r(\boldsymbol{W}, \boldsymbol{\theta}_{c,t+1})} \right) \boldsymbol{P}(\boldsymbol{W}_{i,c,t+1} = \boldsymbol{W} | \boldsymbol{W}_{i,c,t}\,,\, j_{i,c,t} = j) \right) = \nu_{c,t,j}.$$

Though $\nu_{c,t,j}$ is not likely expressible in the form $Q_{t,j}\lambda_{c,t}$, it can probably 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.

3.2 Identification of the parameters of the MWTP for school quality per year at the family level

Since the mean instantaneous utility is identified, the next step is the treatment of the problem of endogeneity in equation (5). In what follows, I will assume that the mean instantaneous utility $\delta_{c,j,t}$ is known. I will suppress the time index t, in order to ease the notation burden. Hence, all variables and parameters in this section have a subscript t which was omitted. Re-writing equation (5):

$$\delta_{c,j} = X_j \pi_c + \xi_{c,j}, \tag{12}$$

Identification in (12) depends on the correlation between $\xi_{c,j}$ and X_j. Concisely, these terms are likely correlated because neighbors sort according to X, and ξ generally includes characteristics of neighbors.

To address this problem, I assume that the term $\xi_{c,j}$ can be decomposed into a finite number of unobservable neighborhood amenities, Q_j, plus a random term $\mu_{c,j}$. The unobservable neighborhood amenities Q_j enter the mean instantaneous utility with different weights depending on the cohort, λ_c, in order to allow cohorts to value the unobservable amenity differently. The mean instantaneous utility can be rewritten as

$$\delta_{c,j} = X_j \pi_c + \overbrace{Q_j \lambda_c + \mu_{c,j}}^{\xi_{c,j}}, \tag{13}$$

where Q_j is a vector of fixed size R. Q_j includes all the unobserved amenities that are non-excludable, in the sense that every household is exposed to the same amenity once it chooses that neighborhood.[48]

Note that the specification of equation (13) is flexible enough to capture some unobservable amenities that are not non-excludable. For instance, in equation (13) a situation where a given cohort is excluded from an unobservable amenity is equivalent to a situation where the amenity is

[48]The standard definition of a public good is a good that is non-rival and non-excludable. This means, respectively, that consumption of the good by one individual does not reduce availability of the good for consumption by others, and that no one can be effectively excluded from using the good. Q_j is assumed to be the set of all unobserved non-excludable amenities, including those that may be rival, as long as households expect a similar exposition to the amenity when deciding where to live. See Samuelson (1954) and Mas-Colell, Whinston and Green (1995).

non-excludable, but the cohort does not value that unobservable amenity. An example of where this may be relevant is the consideration of unobservable amenities which are the result of interaction with the public school. Childless households are excluded from the enjoyment of these amenities. The error structure of equation (13) would capture this effect, but would interpret it as if the childless cohorts do not value the part of the amenity average income of neighbors that is related to peer effects.[49]

Additionally, the specification of equation (13) captures situations where a cohort is exposed to two times more of a given unobservable amenity than another cohort. This is equivalent in equation (13) to both cohorts being exposed the same way to the amenity but the first cohort valuing that amenity twice as much as it truly does. An example of where this may be relevant is the consideration of unobservable amenities which are the result of interaction with different grades of the public school. The amenity may be different depending on the grade. The error structure of equation (13) would capture this effect, but would interpret it as if cohorts had the same exposition to the amenity, but different valuations.[50]

The next assumption states that all of the endogeneity of equation (13) can be captured by R non-excludable unobserved amenities, with allowance for different valuation weights across cohorts.

Assumption 3.5. *(Non-excludable amenities decomposition) For all cohorts c and neighborhoods j, let* $\mu_{c,j}$ *be as defined in equation* (13)*. Then,*

$$E\left[\mu_{c,j}|X_j, Q_j\right] = 0. \tag{14}$$

Assumption 3.5 implies that if Q was observed, the parameters of equation (13) would be trivially identified. Intuitively, this assumption is stronger the smaller is R.

From this point forward, I follow Heckman and Scheinkman (1987)'s approach. The basic

[49]For instance, parents may care about the average income of neighbors because of two reasons: first, parents may think that their children will benefit from peer effects within the school. Second, parents may like to live close to higher income neighbors because of a reason not related to the school. In contrast, non-parents have only the second reason to care about the average income of neighbors, since they do not have a child attending school.

[50]For instance, the infra-structure of the school is likely to be correlated to school quality, and is likely to be used in different intensity depending on the grade. As an example, the school library or the football field can be very important if one is in grade 12, but not as important if one is in the first grade.

idea is to divide the cohorts in three groups. In equation (13), the unobservables Q_j do not vary per cohort. The first group will therefore be used to provide a proxy for the unobservables Q_j in the second group. This will solve part of the endogeneity problem, but introduce a new kind of endogeneity. The third group will then be used as instrument.

Write equation (13) in matrix form as:

$$\boldsymbol{\delta}_j = X_j\boldsymbol{\pi} + Q_j\boldsymbol{\lambda} + \boldsymbol{\mu}_j, \tag{15}$$
$$\text{for } j = 1, ..., J$$

where $\boldsymbol{\delta}_j$ is a $1 \times (C-1)$ vector, $\boldsymbol{\pi}$ is a $K \times (C-1)$ matrix and $\boldsymbol{\lambda}$ is a $R \times (C-1)$ matrix.[51]

Divide the cohorts in three groups, G_1, G_2 and G_3, and let N_1, N_2 and N_3 be the respective number of cohorts in each group. The system of equations (15) can then be broken by groups:

$$\boldsymbol{\delta}_{j(1)} = X_j\boldsymbol{\pi}_{(1)} + Q_j\boldsymbol{\lambda}_{(1)} + \boldsymbol{\mu}_{j(1)}, \tag{16}$$
$$\boldsymbol{\delta}_{j(2)} = X_j\boldsymbol{\pi}_{(2)} + Q_j\boldsymbol{\lambda}_{(2)} + \boldsymbol{\mu}_{j(2)}, \tag{17}$$
$$\boldsymbol{\delta}_{j(3)} = X_j\boldsymbol{\pi}_{(3)} + Q_j\boldsymbol{\lambda}_{(3)} + \boldsymbol{\mu}_{j(3)}, \tag{18}$$
$$j = 1, ..., J$$

where $\boldsymbol{\delta}_{j(n)}$ and $\boldsymbol{\mu}_{j(n)}$ are $1 \times N_n$ vectors, $\boldsymbol{\pi}_{(n)}$ is a $K \times N_n$ matrix and $\boldsymbol{\lambda}_{(n)}$ is a $R \times N_n$ matrix, $n = 1, 2, 3$. I make the following assumption:

Assumption 3.6. $\boldsymbol{\lambda}_{(1)}$ *has full rank R.*

Assumption 3.6 states that there is sufficient heterogeneity in the preferences of the unobserved amenities across cohorts in group 1. This implies that any change in the unobserved amenities will necessarily be translated into a change in the mean utility of group 1 as a whole. With this

[51] Cohort C was dropped because I do not identify $\delta_{C,j}$ for each j, since I do not use information on choices of cohort $C+1$. See equation (9).

assumption, group 1 will generate a proxy variable for Q_j: re-write equation (16) as

$$Q_j = \left(\boldsymbol{\delta}_{j(1)} - X_j \boldsymbol{\pi}_{(1)} - \boldsymbol{\mu}_{j(1)}\right) \boldsymbol{\lambda}'_{(1)} \left(\boldsymbol{\lambda}_{(1)} \boldsymbol{\lambda}'_{(1)}\right)^{-1}. \tag{19}$$
$$j = 1, \ldots, J$$

Equation (19) shows that the unobservable amenities Q_j of neighborhood j can be written as a function of $\boldsymbol{\delta}_{j(1)}$ and X_j plus an error, provided that $\left(\boldsymbol{\lambda}_{(1)} \boldsymbol{\lambda}'_{(1)}\right)$ can be inverted (i.e., assumption 3.6 holds). This implies that one can use $\boldsymbol{\delta}_{j(1)}$ and X_j as proxy for Q_j. Substituting equation (19) in equation (17):

$$\boldsymbol{\delta}_{j(2)} = X_j \overbrace{\left(\boldsymbol{\pi}_{(2)} - \boldsymbol{\pi}_{(1)} \tilde{\boldsymbol{\lambda}}_{(2)}\right)}^{\tilde{\boldsymbol{\pi}}_{(2)}} + \boldsymbol{\delta}_{j(1)} \tilde{\boldsymbol{\lambda}}_{(2)} + \overbrace{\boldsymbol{\mu}_{j(2)} - \boldsymbol{\mu}_{j(1)} \tilde{\boldsymbol{\lambda}}_{(2)}}^{\tilde{\boldsymbol{\mu}}_{j(2)}}, \tag{20}$$
$$j = 1, \ldots, J$$

where $\tilde{\boldsymbol{\lambda}}_{(2)} \equiv \left(\boldsymbol{\lambda}'_{(1)} \left(\boldsymbol{\lambda}_{(1)} \boldsymbol{\lambda}'_{(1)}\right)^{-1} \boldsymbol{\lambda}_{(2)}\right)$ is a $N_1 \times N_2$ matrix measuring the relative preferences of the unobserved amenities for groups 1 and 2. Equation (20) can be written more clearly as:

$$\boldsymbol{\delta}_{j(2)} = X_j \tilde{\boldsymbol{\pi}}_{(2)} + \boldsymbol{\delta}_{j(1)} \tilde{\boldsymbol{\lambda}}_{(2)} + \tilde{\boldsymbol{\mu}}_{j(2)}. \tag{21}$$
$$j = 1, \ldots, J$$

In equation (21), the mean utility of group 2 for neighborhood j is written as a function of the observed amenities and the mean utility of group 1 for the same neighborhood j. Since both groups 1 and 2 are exposed to the same unobserved amenities, by controlling for $\boldsymbol{\delta}_{j(1)}$, one is controlling for changes in the unobserved amenities that are correlated to changes in the observed amenities.

Since $\tilde{\boldsymbol{\mu}}_{j(2)}$ is obviously correlated with $\boldsymbol{\delta}_{j(1)}$ because of equation (16), this equation does not identify the parameters $\tilde{\boldsymbol{\pi}}_{(2)}$ and $\tilde{\boldsymbol{\lambda}}_{(2)}$. However, one can use $\boldsymbol{\delta}_{j(3)}$ as instrumental variables (IV) for $\boldsymbol{\delta}_{j(1)}$. The following assumption gives the conditions for the consistency of the IV estimator of the regression of $\boldsymbol{\delta}_{j(2)}$ on X_j and $\boldsymbol{\delta}_{j(1)}$, using $\boldsymbol{\delta}_{j(3)}$ as IV for $\boldsymbol{\delta}_{j(1)}$.

Assumption 3.7. *(Identification of the parameters of preference) For each j:*

$$E\left[\tilde{\boldsymbol{\mu}}_{j(2)}|X_j, \boldsymbol{\delta}_{j(3)}\right] = \mathbf{0},$$

$$COV\left(\boldsymbol{\delta}_{j(1)}, \boldsymbol{\delta}_{j(3)}\right) \text{ is non-singular.}$$

Though $\tilde{\boldsymbol{\pi}}_{(2)}$ and $\tilde{\boldsymbol{\lambda}}_{(2)}$ are identified, the parameters of interest are $\boldsymbol{\pi}_{(n)}$, or more specifically $\boldsymbol{\pi}^{\text{SQ}}_{(n)}$ and $\boldsymbol{\pi}^{\text{Rent}}_{(n)}$, which are each $1 \times N_n$ vectors of $\boldsymbol{\pi}_{(n)}$, with the coefficients of respectively average school quality and average house rent for each cohort of group n, $n = 1, 2, 3$.[52] I now show how to use $\tilde{\boldsymbol{\pi}}_{(2)}$ and $\tilde{\boldsymbol{\lambda}}_{(2)}$ to identify the coefficients of interest.

Let $\tilde{\pi}^{\text{SQ}}_c$, $\tilde{\pi}^{\text{Rent}}_c$ and $\tilde{\lambda}_c$ be defined as elements of $\tilde{\boldsymbol{\pi}}^{\text{SQ}}_{(n)}$, $\tilde{\boldsymbol{\pi}}^{\text{Rent}}_{(n)}$ and $\tilde{\boldsymbol{\lambda}}_{(n)}$ respectively, for each c in G_n.[53] Equation (20) states the parameters of interest implicitly, through the two equations:

$$\tilde{\pi}^{\text{SQ}}_c = \pi^{\text{SQ}}_c - \boldsymbol{\pi}^{\text{SQ}}_{(1)}\tilde{\lambda}_c \qquad c \in G_2, \tag{22}$$

$$\tilde{\pi}^{\text{Rent}}_c = \pi^{\text{Rent}}_c - \boldsymbol{\pi}^{\text{Rent}}_{(1)}\tilde{\lambda}_c \qquad c \in G_2. \tag{23}$$

From equation (21) and assumption 3.7, $\tilde{\pi}^{\text{SQ}}_c$, $\tilde{\pi}^{\text{Rent}}_c$ and $\tilde{\lambda}_c$ are identified for each c in G_2, so the systems of equations (22) and (23) have each N_2 equations and $N_1 + N_2$ unknowns.[54] Therefore, if I can assume N_1 linearly independent restrictions in the parameters of each of the systems, I am able to identify π^{SQ}_c and π^{Rent}_c for each c in G_2.

In section 5, I explain how I generate N_1 linearly independent restrictions in the system of equations (22) and (23). It suffices, for now, to point that this will be achieved by assuming that $\pi^{\text{SQ}}_c = 0$ for the cohorts of childless households, and the cohorts of households whose oldest child is up to five years old. Also, I will assume that the coefficients π^{Rent}_c are the same across some cohorts.

[52] The reason that one cannot directly identify the estimates of interest is that the Xs are also used as proxy variables together with the δs. This issue was not present in Heckman and Scheinkman (1987), since they did not want to estimate the coefficients of interest. Instead, they only wanted to test for whether the coefficients of interest were the same across sectors.

[53] $\tilde{\pi}^{\text{SQ}}_c$ and $\tilde{\pi}^{\text{Rent}}_c$ are scalars, and $\tilde{\lambda}_c$ is a $N_n \times 1$ vector.

[54] The unknowns are π^{SQ}_c and π^{Rent}_c for $c \in G_1 \cup G_2$.

3.3 Identification of the MWTP for school quality per year, per child and per grade

Let $\text{MWTP}^{SQ}_{c,g,t}$, $g = 0, 1, ..., 12$, where $g = 0, ..., 12$ represents kindergarten through grade 12, respectively, be defined as the MWTP for school quality per year, per child and per grade for families of cohort c. Families of cohort c can have more than one child, and can have children at different grades. Let $N_{i,c,g,t}$ be the number of children from family i of cohort c who are attending grade g in period t. Let the expected number of children attending grade g in period t, among every family i from cohort c be defined as $N_{c,g,t} := E_c(N_{i,c,g,t})$. Note that $N_{c,g,t}$ is identified using the restricted 2000 Census data set, since we know the exact date of birth of each child inside a household, hence I can relate the age of each child with the grade that child is attending.[55]

I am going to make the following assumption:

Assumption 3.8. *For each c and t:*

$$MWTP^{SQ}_{c,t} = \sum_{g=0}^{c} MWTP^{SQ}_{c,g,t} N_{c,g,t} \tag{24}$$

$$MWTP^{SQ}_{c,g,t} = MWTP^{SQ^{\star}}_{g,t} \quad \forall c \tag{25}$$

where $\text{MWTP}^{SQ^{\star}}_{g,t}$ is the MWTP for school quality per year, per child and per grade, for grades $g = 0, ..., 12$.

Assumption 3.8 implies that, for a given period t, a family values the school quality for each child only as function of the grade that child is attending. In particular, parents do no vary their valuation for school quality per child 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 (hence the age of the head of the family). Note that all these assumptions are with regard to the valuation of school quality controlling for observed family attributes, such income and number of children. In particular, I allow for two families with different number of children to value school quality for one of their children who are attending the same grade differently.[56]

[55] Again, I deal with a fuzzy design by restricting the analysis only to those families such that all their children's age implied a grade that was inside the category that the child actually reported.

[56] This assumption therefore does not necessarily contradicts the debate on the quantity vs. quality of children. See Becker and Lewis (1973).

Substituting equation (25) into equation (24):

$$\text{MWTP}_{c,t}^{SQ} = \sum_{g=0}^{c} \text{MWTP}_{g,t}^{SQ^{\star}} N_{c,g,t} \tag{26}$$

$$c = 6, 7, ..., 18$$

For each period t, the system of equations (26) has 13 unknowns ($\text{MWTP}_{g,t}^{SQ^{\star}}$, $g = 0, ..., 12$) and 13 linearly independent equations ($c = 6, ..., 18$), which guarantees the identification of the MWTP for school quality per year, per child and per grade, from kindergarten to twelfth grades.

4 Estimation of the MWTP for School Quality

This section explains the empirical strategy that will be implemented in this paper, which is a direct consequence of the identification effort. The identification relied first on the identification of the parameters of both the present and future choice-specific value functions. These will be estimated through a BLP-type approach using data only on period $t = 2000$, and will be plugged into the equation of the choice-specific value functions using the assumptions of section 3, generating the estimated choice-specific value functions. These are then used in the integrated version of a Bellman equation, which enables one to estimate the mean instantaneous utilities. In the next step, the estimated mean instantaneous utilities are plugged into the equation that gives the coefficients of the preference for school quality and rent per year for each cohort. This generates a system of equations which is estimated through a Three-Stage Least Squares (3SLS) regression using the technique of the division in groups discussed in the last section. The estimated coefficients are then used to calculate the estimated MWTP for school quality per year for each cohort, and with assumption 3.8 we can then easily estimate the MWTP for school quality per year, per child and per grade.

The empirical strategy to estimate the MWTP for school quality for each grade is in three steps. The following subsections explain these steps in more detail.

4.1 Step 1: Estimation of the Mean Instantaneous Valuation at the Cohort-Neighborhood Level

The first step uses a similar approach to BLP in order to consistently estimate the parameters of the choice-specific value functions under relatively weak assumptions. The idea is to control for an unobserved term at the cohort-period-neighborhood level. This term allows the household to have a much more complex information set than of the researcher.[57] Examples of the types of effects for which this term controls are different preferences for unobservable amenities across different cohorts, and different expectations about the cohort's own future preferences and constraints.

The idea is to use the specification (10), equation (7) of proposition 1, and assumption 3.2 to estimate $\boldsymbol{\theta}_{c,t}$ for each c in a Multinomial Logit framework using the cross-sectional data at period t.

Households choose the neighborhood that yields the highest utility among the neighborhoods available. Household i of cohort 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:

$$j_{i,c,t} = j \Longleftrightarrow \varepsilon_{i,c,t,j} - \varepsilon_{i,c,t,r} \geq v_r(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) - v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}), \quad \forall r. \tag{27}$$

From assumption 3.2, the $\varepsilon_{i,c,t,j}$ are randomly distributed extreme value 1 for each j. Hence, $\varepsilon_{i,c,t,j} - \varepsilon_{i,c,t,r}$ has a logit distribution, and the probability that household i of cohort c chooses neighborhood j in period t is:

$$\boldsymbol{P}_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) = \frac{\exp(v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}))}{\sum_{r=1}^{J} \exp(v_r(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}))}. \tag{28}$$

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

[57] Intuitively, although I allow each household's expectation with respect to future preferences to be unobserved to the researcher, I am able to estimate the parameters of preferences out of the assumption that its expectation is similar to the current preferences for households one year older, which are estimable using current data.

$$LL_{c,t}(\boldsymbol{W}_{i,c,t}, \{j^*_{i,c,t}\}_{i=1}^{I_c}, \boldsymbol{\theta}_{c,t}) = \sum_{i=1}^{I_c} \sum_{j=1}^{J} d_j(j^*_{i,c,t}) \log(\boldsymbol{P}_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})), \quad c = 1, ..., C, \forall t. \tag{29}$$

where $j^*_{i,c,t}$ is the choice observed in the data for each i, c, t.

I observe $\boldsymbol{W}_{i,c,t}$ and $j^*_{i,c,t}$ for each i and j, and estimate $\boldsymbol{\theta}_{c,t}$ by maximizing equation (29) independently for each cohort $c \in \{1, ..., C\}$, and time $t = 2000$.[58]

It is necessary to estimate the choice-specific value functions, both of period t and period $t+1$. Estimation of $v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ for $t = 2000$ is trivial, and only requires to plug in the estimates estimated in (29) in the equation for $v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ as defined in (10). The estimated choice-value function for $t = 2000$ is

$$\hat{v}_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) = \hat{\Delta}_{c,t,j} + X_{t,j} Z_{i,c,t} \hat{\Gamma}_{c,t} + \text{NEW}_{i,c,t,j} \hat{\Phi}_{c,t}. \tag{30}$$

For the estimation of $v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ for period 2001, I plug in the estimates from the last step into the equation of $v_j(\boldsymbol{W}_{i,c,t+1}, \boldsymbol{\theta}_{c,t+1})$ as defined in equation (11):

$$\hat{v}_j(\boldsymbol{W}_{i,c,t+1}, \boldsymbol{\theta}_{c,t+1}) = \hat{\Delta}_{c+1,t,j} + X_{t,j} Z_{i,c,t} \hat{\Gamma}_{c+1,t} + \text{NEW}_{i,c,t+1,j} \hat{\Phi}_{c+1,t}. \tag{31}$$

Observe that t in equation (31) refers to year 2000, for which all the components of the equation are either available or estimable.

Assuming that the regularity conditions which guarantee that the Maximum Likelihood estimator $\boldsymbol{\theta}_{c,t}$ is consistent are satisfied[59], the following lemma guarantees the mean convergence of $\hat{v}_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ and $\hat{v}_j(\boldsymbol{W}_{i,c,t+1}, \boldsymbol{\theta}_{c,t+1})$, which is necessary for proposition 2.[60]

Lemma 1. *For all c and t, let* $\boldsymbol{\theta}_{c,t} \equiv \left(\{\Delta_{c,t,j}\}_{j=1}^{J}, \Gamma_{c,t}, \Phi_{c,t}\right)$*, and* $\hat{\boldsymbol{\theta}}_{c,t}$ *be a consistent estimator of*

[58] 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. I write the Δs as function of the other parameters Γ and Φ using a contraction, as first suggested by BLP, and optimize the log-likelihood of (29) writing it as a concentrated maximum likelihood.

[59] See McFadden (1973), McFadden (1977) and BLP.

[60] Proof of Lemma 1: The $\hat{v}_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ is a linear "plug-in" estimator of $\hat{\boldsymbol{\theta}}_{c,t}$. From the convergence in distribution of the MLE estimator, $E(||\sqrt{n}(\hat{\boldsymbol{\theta}}_{c,t} - \boldsymbol{\theta}_{c,t})||^2) = O_p(1), \forall c, t$. Assume the $\boldsymbol{W}_{i,c,t}$ have finite second moments. Then, by Cauchy-Schwartz, the result follows.

$\boldsymbol{\theta}_{c,t}$. *Let* $v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})$ *be defined as in either equation* (10) *or* (11). *Define*

$$\hat{v}_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) = v_j(\boldsymbol{W}_{i,c,t}, \hat{\boldsymbol{\theta}}_{c,t}).$$

Then, $\boldsymbol{E}(|\hat{v}_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) - v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})|) \to 0$.

Finally, I estimate $\delta_{c,t,j}$ using an empirical version of equation (9).[61]

$$\hat{\delta}_{c,t,j} = \frac{1}{I_c} \sum_{i=1}^{I_c} \left(\hat{v}_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) - \beta \left(\gamma + \log \sum_{r=1}^{J} \exp \hat{v}_r(\boldsymbol{W}_{i,c,t+1}, \boldsymbol{\theta}_{c,t+1}) \right) \right), \tag{32}$$

where γ is the Euler's constant.

Proposition 2. *For each i, c, t and j, let* $E(|\hat{v}_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t}) - v_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\theta}_{c,t})|) \to 0$. *Let* $\hat{\delta}_{c,t,j}$ *be defined as in equation* (32). *Then* $\hat{\delta}_{c,t,j}$ *is a consistent estimator of* $\delta_{c,t,j}$.

Proof. Let the term inside the sum in equation (32) be $\hat{a}_{i,j}$, and its true value be $a_{i,j}$. The Continuous Mapping Theorem assures that $E(|\hat{a}_{i,j} - a_{i,j}|) \to 0$. Apply Markov's theorem to $\frac{1}{I_c}\sum_{i=1}^{I_c}(\hat{a}_{i,j} - a_{i,j})$ to show that it is $o_p(1)$. The Law of Large numbers and proposition 1 guarantee that $\frac{1}{I_c}\sum_{i=1}^{I_c} a_{i,j} \xrightarrow{p} \delta_{c,t,j}$. □

4.2 Step 2: Estimation of MWTP for school quality per year at the family level

In the second step, the mean instantaneous utilities at the cohort-neighborhood level for period $t = 2000$ calculated in the previous step are plugged into equation (13). Estimation follows Heckman and Scheinkman (1987). For notational clarity, I will drop the subscript t, as in the identification part (section 3.2). All variables and coefficients must be interpreted as having the subscript $t = 2000$.

Equation (13) is now rewritten with the estimated mean instantaneous utilities

$$\hat{\delta}_{c,j} = X_j \pi_c + Q_j \lambda_c + \mu_{c,j} + \hat{e}_{c,j}, \tag{33}$$

[61] Note that because of assumption 3.4, the conditional distribution of $\boldsymbol{W}_{i,c,t+1}$ has only one point in the support. This implies that the sum over $\boldsymbol{W} \in \mathbb{W}$ from equation (8) has only one term.

where the $\hat{e}_{c,j}$ are residuals of the estimation of $\delta_{c,j}$. Equation (33) shows that in practice the error term contains one further component.[62] Under the assumptions of section 3 and some regularity conditions, this will not cause consistency problems in the estimation of the MWTP for school quality using the proposed method.[63]

A very important feature of this method was first pointed out by Pudney (1982). He demonstrates that any choice of assignment of cohorts to group 1 generates coefficient estimates that are asymptotically equivalent in efficiency, provided that assumption 3.6 holds. Another important feature is that if assumption 3.6 is valid for R, then the estimator of interest is robust to specifications with any number of unobservable amenities smaller or equal to R.

Although Pudney's (1982) result imply that the choice of assignment of cohorts to groups are irrelevant with respect to asymptotic efficiency, in practice the choice of assignment is important. The coefficients of group 3 are not identified, and therefore it is preferable to assign to group 3 those cohorts whose MWTP for school quality is not important to be estimated. At the same time, the additional moment conditions that are required by the method are only related to the coefficients of groups 1 and 2. These additional conditions are obtained by making assumptions about the values of some coefficients of these groups. This means that if part of the coefficients of one cohort can be set to zero, this cohort would be well suited to be assigned to group 1 or 2. However, it is likely that this cohort is also not interesting, so it would be well suited for group 3. This is a practical aspect of the implementation that I will address when dividing the cohorts in section 5.

4.3 Step 3: Estimation of MWTP for school quality per year, per child and per grade

Finally, in the last step I estimate $N_{c,g}$ (period $t = 2000$) and use the estimates of MWTP_c^{SQ} as well as of $N_{c,g}$ to estimate $\text{MWTP}_g^{SQ^\star}$, for $g = 0, ..., 12$, using equation (26).

[62] A similar error is also present in the second step of the BLP approach, due to the fact that the dependent variable is estimated rather than observed.

[63] The argument why the presence of $\hat{e}_{c,j}$ is not an issue in this estimation relies on the fact that the three-stage least squares estimator, which is the estimator used in this approach, is a linear estimator. If $E(\hat{\delta}_{c,j} - \delta_{c,j})^2 \to 0$, the result follows from Cauchy-Schwartz, as long as some regularity conditions are satisfied. To see this, let the coefficients estimated in equation (33) be denoted α, then $\hat{\alpha}$ takes the form $\hat{\alpha} = \alpha + A_n B_n(\mu + \hat{e}) = \alpha + o_p(1) + \frac{1}{n}\sum_{i=1}^{n} f_{i,n}\hat{e}_{i,n}$. If, for all i, $E(\hat{e}_{i,n})^2 \to 0$, the result follows trivially from Cauchy-Schwartz, as long as $E(f_{i,n})^2 < a$, for some $a < \infty$.

5 Empirical Implementation and Results

In this section I present the data and the results of the paper, and explain how I implement the estimation methodology presented in section 4.

The data set is the restricted-access, or long form, version of the 2000 decennial Census of Population and Housing. It is a $(1:7)$ 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.[64] I use this information to identify the neighborhoods as the elementary school districts (SD). I used the geographical information to merge the Census data with school quality data from the New Jersey (NJ)'s Department of Education, which provides information on the proportion of students who were proficient in the ESPA[65] test for elementary grades for each school, as well as on total enrollment by grade.[66] My measure of school quality is a weighted average proportion of proficient students for elementary schools, across all schools in the district. The weights are total school enrollment in elementary grades.[67]

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 most schools in NJ, that cutoff is October first. I use the exact date of birth of the children to define the cohorts to which the family belongs. The cohort is determined by the age of the oldest child in October first 2000, so that the cohort matches the grade the oldest child is attending. The cohort of parents whose oldest child is 6 is identified as the cohort of parents

[64] 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. In this paper I am estimating a model of local choices of households, so I need an identifier at a lower geographical level than the PUMA.

[65] ESPA is the May 1999 Elementary School Proficiency Test.

[66] In this version of the paper, I only show results for the state of NJ.

[67] The results shown in the next section seem very robust to different measures of school quality. I checked the robustness of the results with two different types of measures. First, I used analogous measures for either middle school or high school, rather than elementary school. Second, I used the maximum of the proportion of proficient students in elementary school across schools in the district, rather than the average.

whose oldest child turned 5 in the 12 months interval prior to October 1, 1999, since then this child would supposedly be observed in kindergarten in April first 2000, which is the date the Census is reported. Analogously, the cohort of parents whose oldest child is 5 is identified as the cohort of parents whose oldest child turned 4 in the 12 months interval prior to October first 1999, so that they are still not supposed to be attending school in April 2000.

Column I of table 1 shows the summary statistics of the observed characteristics of the SDs for the full sample of NJ. These represent the neighborhood observable characteristics in the estimation (or X in equation (10)). There are 501 SDs in the full sample of NJ, with on average of over 800 houses in each of the sample SD's. The average test score, which is the measure of school quality used in the paper, is in average 59, with standard deviation of 8. The amount of variation in this variable is very similar to the amount of variation of the test score used in other papers.[68] The average rent is around \$750, with a standard deviation of \$240. Additionally, the average house value is just over \$200,000, with standard deviation of over \$100,000. With regard to the characteristics of the neighbors, SDs are also shown to be very diverse. For instance, the average income of the neighborhoods is around \$75,000 with standard deviation of \$35,000, and the average proportion of neighbors with college degree or more is 33% with standard deviation of 18%.

In this analysis, I restrict attention to 38 cohorts, which are disjoint subsets of the total population of households. I define the cohorts 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 cohorts by the age of the oldest child, ranging from ages 1 to 19. For non-parents, I define the cohorts by the age of the head of household, ranging from ages 32 to 50.[69]

Table 2 shows the summary statistics of the observed characteristics of the cohorts of parents used in this analysis (i.e., the Z in equation (10)). It is interesting to notice that proportion of race and employment status do not vary much across cohorts. The average household income is between \$78,000 and \$87,000 for parents of each cohort, and the average income is very similar for

[68] Bayer, Ferreira and McMillan (2007) use a measure of school quality with the standard deviation amounting to 14% of the mean of the variable, as in this paper.

[69] The average age of the parents whose oldest child is 1 year old is 32 years, so parents of the first cohort are on average exactly in the same generation of non-parents of the first cohort.

consecutive cohorts. In contrast, the average proportion of households with children whose head has a college degree or more depends significatively on the age of the oldest child, with parents of younger children being more likely to be highly educated. The average proportion of parents who are homeowners vary from 60% to 75%, with older parents being homeowners with a higher likelihood. The moving rate also depends substantially on the cohort, 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 cohorts 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, and is fairly similar for older cohorts, around 14%. Finally, the proportion of "defiers", defined as students who are attending a different grade from the one they were supposed to be attending according to their date of birth, depends substantially on the age of the oldest child.[70]

Analogously, Table 3 shows the summary statistics of the observed characteristics of the cohorts of non-parents used in this analysis. The household income for cohorts of non-parents is around $\$70,000$, which is around $\$10,000$ lower than the average for the cohorts of parents. Similarly to parents, the average proportion of households without children whose head has a college degree or more depends significatively on the age of the head, with younger non-parents being more likely to be highly educated. The average proportion of non-parents who are homeowners is much lower than for parents, varying from 40% to 65%, with older non-parents being homeowners with a higher likelihood. The moving rate depends substantially on the cohort, as for parents, with older non-parents moving much less than younger ones.

Tables 2 and 3 show that there is substantial variation across cohorts 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 important to the approach of this paper, since it exploits variation across cohorts with respect to the preferences of observed and unobserved amenities.

In order to implement the method developed in section 4, I restricted the sample of school

[70]Part of this variation across grades is likely to be due to measurement error in this variable. The grade each child is attending is reported in the Census as categories. I defined the variable defier using information only on these categories. For instance, a child who is supposed to be attending third grade but is in fact attending fourth grade cannot be observed as a defier, since the category of third and fourth grades is the same.

districts only to those that contained at least one household for each of the 38 cohorts. Column II of Table 1 shows the characteristics of each school district in this new sample. In total there are 224 SDs in this new sample, so there could be sample selection, and the results of the paper would not be generalizable to the results for the full sample of SDs in NJ. As can be seen in table 1, with the exception of average size of the SD, the full sample and the sample used in the analysis of the paper do not seem to differ much.

Table 4 shows the results of hedonic price regressions using house rent as dependent variable and several observed characteristics of the house and neighborhood amenities, in order to provide the reader with a benchmark to compare with the main results of the paper. The first row of the table shows the MWTP for a 5% increase in school quality for increasingly more complex specifications. Overall, the results show very similar estimates to the ones available in the literature, even after controlling for a wide list of characteristics of the house and the neighborhood. As can be seen in the table, once the proportion of highly educated neighbors is included in the regressions (column III), the coefficient on test score reduces substantially, and then remains fairly constant no matter how many additional control variables are included.[71] The estimates are around 1.5% of the rent value for a 5% increase in school quality, which are similar to the estimates found in the literature.[72]

Table 5 shows the results of the paper from step 2. This table shows the MWTP per year at the family level for an increase of 5% in school quality for each grade of the oldest child of the family. The rows represent the grades, and the columns represent different specifications. Column I shows the estimates of a simple GLS panel regression of the mean instantaneous utilities of cohorts 6 through 18 on the average school quality per neighborhood and the average rents per neighborhood, without any other control variable, nor controlling for unobservables. The estimates of this regression imply a MWTP per year at the family level of around 3% of rent for primary school, 3.5% for middle school and around 4.5% for secondary school. Column II shows the

[71] Hedonic price regression results for households who own their house in the sample are also similar to the ones reported in table 4. The summary statistics of the full sample of houses in NJ, as well as rented and owned houses are included in Table A1.

[72] Black (1999) found that parents are willing to pay around 2.5% more on house values for a 5% increase in school quality for a subset of SDs in Massachusetts, and Bayer, Ferreira and McMillan (2007) found that parents are willing to pay under 1% more on house values for a 5% increase in school quality in the San Francisco Bay Area. They both control for boundary fixed effects in their results, which table 4 does not.

estimates of the same GLS panel regression of the mean instantaneous utilities of cohorts 6 through 18 on the average school quality per neighborhood and the average rents per neighborhood, still not controlling for unobservables, but adding a list of control variables which measure observed amenities, such as neighborhood average income, and proportion of highly educated people.[73] The estimates of this regression represent a substantial reduction on the MWTP for school quality per year at the family level. This result is expected, since school quality is supposedly correlated to amenities which are positively valued. The MWTP estimates are around 2% for primary school and 3.5% for secondary school.

Columns III through VI of table 5 show the results of step 2 in section 4.[74] I divided the cohorts in three groups: the proxy group, the group of interest and the group of instruments. The group of interest includes cohorts 6 to 18. These cohorts correspond to the households with school age children, so the MWTP per year for each grade of the oldest child is identified from the parameters of preference of these cohorts. In step 1 of section 4, I estimated the mean instantaneous utilities for each neighborhood for a subgroup of each of cohorts 6 through 18. The subgroup is defined as the group of households of cohort c, ranging from 6 to 18, whose oldest child is attending a public school in the correctly pre-specified grade.[75]

In all specifications, I make the following assumptions: first, I assume that the cohorts used as proxy variables do not value school quality in the instantaneous sense, since I only use as proxy variables the cohorts of non-parents and parents whose oldest child is not yet of school age. Second, I assume that the cohort of age 6 values average rent on average the same as cohorts of age 1, 5 and 15. Note that in step 1 I am controlling for interactions of observed amenities, including average rent, and all observed household characteristics included in tables 2 and 3, including income. Moreover, as one can see in table 2, these cohorts have relatively similar distribution of household income, which is likely to be strongly correlated to the preference for average rent. There is also

[73]The detailed list of controls can be seen in the notes at the end of Table 5.

[74]This table shows results for $\beta = .95$. The results do not change significatively for choices of different βs in the range .85 through .99.

[75]The estimate of the mean instantaneous utility was calculated across all households of cohort c with the dummy variables "private school" and "defier" equal to zero. In total there are between 1900 and 4000 households included in the estimation per cohort, depending on the cohort. It is important that this number be large enough so that the estimates are as close as possible to the true value. See Berry, Linton and Pakes (2004) for an interesting discussion about the asymptotic properties of the BLP approach.

some formal evidence in favor of the assumption. I assigned cohort 1, which is assumed not to value school quality, to the group of interest. I estimated its coefficient of school quality and it is not significatively different from zero. Moreover, I estimated the coefficients of price with the assumption that the coefficients of price for cohorts from 1 to 15 behave linearly as a function of the age age of the oldest child, and test the coefficient of the linear trend, and it was not significatively from zero.[76] These two assumptions provide three additional restrictions each. With these, I can estimate the coefficients of school quality and average rent for each cohort, from 6 to 18, hence obtaining an estimate of the MWTP for school quality per year for the group at the family level, for different cohorts.

In each of columns III to V, I vary the number of unobservables for which I allow, by varying the number of cohorts assigned to the group of proxy variables. As I add more proxy variables, I assign more cohorts to the group used as instrumental variables.[77] Column III shows the results using a proxy for one unobservable. The cohort used as proxy is cohort 49. 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 specification of column II adds a list of 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. The MWTP per year at the family level for an increase in school quality of 5% is around 2% for primary schools, 2.7% for middle school and 4% for secondary school.

It could be the case that there exists an unobserved amenity that is correlated to school quality that is valued by parents with children of school age, but not valued by non-parents. Availability of children parks may be an example of such amenities. Column IV adds another proxy variable to control for more unobserved amenities: the mean instantaneous utilities for cohort 1. The results are fairly similar to column III, even though the R^2 increased substantially. Column V adds one more proxy: the mean instantaneous utilities from cohort 2. The results are again very similar to column IV, but there is no substantial improvement in the fit ($F = 1.09$ is lower than the critical

[76] Analogously, I also tested for quadratic and cubic trends, and also could not reject the null hypothesis.

[77] The list of cohorts used as instrumental variables for each specification is included in the notes of the table.

value $F^* = 1.72$ at 5% of significance).

Finally, column VI adds to the specification of column V the same list of control variables as in column II. The fit does not improve significatively ($F = .58$ is lower than the critical value $F^* = 1.28$ at 5% of significance), and the results are the same. Columns V and VI together show that two unobservable amenities seem to be enough to take care of the endogeneity with regard to the variables school quality and average rents.[78]

It is important to compare the results between columns I and II and columns III through VI. The specifications of column I and II use the proposed approach only up to step two, and run a simple GLS regression, column I with only average rent as control variable, column II with other control variables as well. Columns I and II is not directly related to previous estimates in the literature because the specifications of table 5 use the mean instantaneous utilities as dependent variables, rather than the mean cumulative utilities. The other columns use all the proposed method, including step two, running a Three Stage Least Square controlling for unobservable amenities. One can see that there is not much difference between the results of column II and the results of columns III through VI, even though columns III through VI have a much better fit. This suggests that, once I use step one to measure the instantaneous valuation of each neighborhood, rather than the cumulative valuation of each neighborhood, most of the endogeneity can be controlled using observed amenities only.

The upper panel of table 6 shows the results from step 3 for the specification of column VI., with the estimation of the MWTP for school quality per year, per child and per grade. The results show that parents are willing to pay to send one of their child to attend one year of a school that is 5% better an exceeding percentage over rent value of 1.4%, if the child is attending elementary school. The corresponding estimates for middle school and secondary school are respectively 1.6% and 2.7%. The value for the kindergarten seems to be larger than the rest for elementary school, and the value for grades 11 and 12 are much larger than the rest. One interpretation is that parents perceive an investment of improvement in school quality for the last years of high school to be the most rewarding in terms of human capital. These results suggest the policy maker, everything

[78] I added up to three more proxy variables, with very similar results.

else constant, invest more in policies that benefit secondary school versus elementary and middle school.

The lower panel of Table 6 shows the magnitude of the effects reported in the upper panel. These numbers are calculated over the average rent in NJ, which is $755. Parents whose oldest child is attending a public school are willing to pay $35 more per month over rent to send one of their children to attend a school for one year in a school that is one standard deviation better for primary school. The corresponding valuations are $40 for middle school and $65 for secondary school.

6 Conclusion

In this paper, I develop an approach to identify and estimate the MWTP for school quality per year, per child and per grade through the neighborhood decision. I extend the literature by addressing three challenges. First, because of moving costs people consider not only the per year valuation of each neighborhood, but also the option value of living in that neighborhood, which makes it hard to disentangle the MWTP for school quality for each family per year and per grade. Second, people decide which neighborhood to live taking into account not only the amenity school quality but also other amenities, some of them unobserved to the researcher and intrinsically correlated to any observed amenity due to sorting, which generates an endogeneity problem that cannot be solved with instrumental variables. Third, the family, rather than the child is the one who decides where to live, and families in general have more than one child attending school, possibly in different grades, which makes it difficult to disentangle the MWTP for school quality per year, per child and per grade, from the MWTP for school quality per year at the family level.

Essentially, I use the same insight to address the three issues: the fact that the choices of some groups of families can be used to infer the preferences of other groups of families. In the dynamic issue, I use the choices of older families to infer the expected future cumulative valuations of younger families. In the endogeneity issue, I use the choices of childless families and families with children too young to attend school to infer the unobservable neighborhood amenities. Finally, for the family-child issue, I relate the preferences for school quality of different families with children

at same grade.

Methodologically, this paper makes two additional contributions by suggesting two modifications of the BLP approach. The first is an extension of BLP to a dynamic framework with one wave of cross-sectional data (hence no way of estimating conditional choice probabilities as in Hotz and Miller (1993)), building on Rust (1987). This is achieved with the transformation of the family's dynamic optimization problem into an "integrated" Bellman equation which involves cumulative valuations, instead of instantaneous utilities, because the former can be naturally inferred from the families' location decisions, hence the BLP fixed effects are feasible to be estimated with their contraction method. Moreover, because the data available in this paper is cross-sectional, expected future cumulative valuations of younger families are inferred using the current cumulative valuations of the older families through a synthetic cohort assumption.

The second contribution is an alternative method of treatment of endogeneity in discrete choice methods, building on BLP and Heckman and Scheinkman (1987). This method is capable of controlling for unobserved neighborhood-level amenities, without the need of pre-specifying them other than assuming that all groups of families are exposed to the same unobservable amenities (however, they may value them differently). With this assumption, the choice of some groups can be used to construct a proxy of the unobservable amenities which can serve to control for endogeneity for other groups. This approach can be implemented together with the instrumental variables for price proposed in BLP, or without them, as in this paper. In either case, the assumptions will be the exogeneity of the characteristics of the choice (except price in case of the BLP instrument) *conditional on unobserved characteristics*.

Although these two modifications can be used separately, I implemented an unified version using the restricted-access 2000 decennial Census, and obtained values in New Jersey that increase from primary to secondary school. Families whose oldest child is in elementary school are willing to pay 1.4% over rental prices for a 5% increase in school quality, increasing to 1.6% when the child is in middle school, and to 2.7% when the child is in high school.

In the future, I intend to open the education production function and estimate the willingness to pay for an improvement of other variables that are correlated to school quality, such as class size,

spending, teacher quality, and most interestingly, characteristics of peers, such as level of education, racial composition and income of parents. With the proposed approach, one can disentangle the family's valuation of the amenity "characteristics of neighbors" and the amenity "characteristics of the children's peers". For instance, non-parents may value (in the instantaneous sense) to live close to highly educated neighbors, but this valuation is not with respect to student's peers. However, parents with children at school value (in the instantaneous sense) both amenities, so by using the instantaneous valuation of non-parents as proxy variable, we can disentangle these two amenities, hence estimating parental valuation of peer characteristics within the school.

Moreover, I intend to use the proposed method to estimate the MWTP for other amenities. An important feature of the proposed method is that it does not require a discontinuity of the level of the amenities, as in Black's (1999) strategy, so it can in principle be applied for instance to estimate MWTP for amenities such as crime rate, clean air and racial or income composition. For these coefficients to be identified, one would need to make *a priori* assumptions for some groups with regard to the parameters of preference of the specific amenity for which we want to estimate the valuation. For instance, one could assume that the preference with respect to racial composition, conditional on the race and the income category of the household, does not vary as a function of whether the household has a child or not. So we can use non-parents' *cumulative* valuation for school quality for each group defined by race and income category as proxy for unobserved amenities that are correlated to racial composition, so as to estimate parental racial preference.

References

Aguirregabiria, V. & Mira, P. (2007). Dynamic discrete choice structural models: a survey. *Working Paper, University of Toronto Department of Economics*, (297).

Bajari, P. & Benkard, C. L. (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. *Journal of Political Economy*, *113*(6), 1239–1276.

Bajari, P. & Kahn, M. E. (2003). Estimating housing demand with an application to explaining

racial segregation in cities. NBER Working Papers 9891, National Bureau of Economic Research, Inc.

Barrow, L. (2002). School choice through relocation: Evidence from the Washington, D.C. area. *Journal of Public Economics*, *86*(2), 155–189.

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

Bayer, P., Keohane, N., & Timmins, C. (2006). Migration and hedonic valuation: The case of air quality. *NBER Working Paper*, (W12106).

Bayer, P., McMillan, R., Murphy, A., & Timmins, C. (2007). A dynamic model of demand for houses and neighborhoods. Mimeo.

Becker, G. & Lewis, H. (1973). On the interaction between the quantity and quality of children. *Journal of Political Economy*, *81*(2), 2:S279–S288.

Bellman, R. (1957). *Dynamic Programming*. Princeton University Press, NJ.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *Rand Journal of Economics*, *25*(2), 242–262.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890.

Berry, S., Linton, O., & Pakes, A. (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *Review of Economic Studies*, *71*, 613–654.

Black, S. (1999). Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics*, *114*, 577–599.

Bresnahan, T. (1981). Departures from marginal-cost pricing in the american automobile industry. *Journal of Econometrics*, *17*, 201–227.

Bresnahan, T. (1987). Competition and collusion in the american automobile oligopoly: The 1955 price war. *Journal of Industrial Economics*, *35*, 457–482.

Carranza, J. (2007). Estimation of demand for dierentiated durable goods. Mimeo.

Chamberlain, G. (1977). *Kinometrics: Determinants of Socioeconomic Success Within and Between Families*, chapter 7: An Instrumental Variable Interpretation of Identification in Variance-Components and Mimic Models, (pp. 235–254). Amsterdam: North Holland.

Chay, K. & Greenstone, M. (2005). Does air quality matter? Evidence from the housing market. *Journal of Political Economy*, *113*(2), 376–424.

Cropper, M., Deck, L., Kishor, N., & McConnell, K. (1993). Valuing product attributes using single market data: A comparison of hedonic and discrete choice approaches. *The Review of Economics and Statistics*, *75*(2), 225–232.

Ekeland, I., Heckman, J. J., & Nesheim, L. (2004). Identification and estimation of hedonic models. *Journal of Political Economy*, *112*(1), S60–S109.

Epple, D. (1987). Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products. *Journal of Political Economy*, *95*(1), 59–80.

Epple, D., Filimon, R., & Romer, T. (1984). Equilibrium among local jurisdictions: toward an integrated treatment of voting and residential choice. *Journal of Public Economics*, *24*(3), 281–308.

Epple, D., Filimon, R., & Romer, T. (1993). Existence of voting and housing equilibrium in a system of communities with property taxes. *Regional Science and Urban Economics*, *23*(5), 585–610.

Epple, D. & Romano, R. (1998). Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, *88*(1), 33–62.

Epple, D. & Sieg, H. (2006). Estimating equilibrium models of local jurisdictions. *Journal of Political Economy*, *107*(1), 645–681.

Epple, D. & Zelenitz, A. (1981). The implications of competition among jurisdictions: Does Tiebout need politics? *Journal of Political Economy*, *89*(6), 1197–1217.

Erdem, T., Imai, S., & Keane, M. (2003). Consumer price and promotion expectations: Capturing consumer brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and Economics*, (1), 5–64.

Fernandez, R. & Rogerson, R. (1996). Income distribution, communities and the quality of public education: A policy analysis. *The Quarterly Journal of Economics*, *111*(1), 135–164.

Gowrisankaran, G. & Rysman, M. (2007). Dynamics of consumer demand for new durable goods. Mimeo.

Heckman, J. & Scheinkman, J. (1987). The importance of bundling in a Gorman-Lancaster model of earnings. *Review of Economic Studies*, *54*(2), 243–255.

Hendel, I. & Nevo, A. (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrics*, *74*(6), 1637–1673.

Hotz, J., Miller, R. A., Sanders, S., & Smith, J. (1994). A simulation estimator for dynamic models of discrete choice. *Review of Economic Studies*, *61*(2), 265–289.

Keane, M. & Wolpin, K. (1997). The career decisions of young men. *Journal of Political Economy*, *105*, 473–522.

Keane, M. & Wolpin, K. (2001). The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, *42*, 1051–1103.

Lemieux, T. (1998). Estimating the effects of unions on wage inequality in a panel data model with comparative advantage and nonrandom selection. *Journal of Labor Economics*, *16*(2), 261–291.

Mas-Colell, A., Whinston, M., & Green, J. (1995). *Microeconomic Theory*. New York: Oxford University Press.

McFadden, D. (1973). *Conditional Logit Analysis of Qualitative Choice Behavior*. Frontiers in Econometrics. New York: Academic Press.

McFadden, D. (1977). Modelling the choice of residential location. Cowles Foundation Discussion Papers 477, Cowles Foundation, Yale University.

Melnikov, O. (2001). Demand for differentiated durable products: The case of the U.S. computer printer market. Mimeo.

Nechyba, T. J. & Strauss, R. P. (1998). Community choice and local public services: A discrete choice approach. *Regional Science and Urban Economics*, *28*(1), 51–73.

Newell, C. (1990). *Methods and Models in Demography*. Guilford Press.

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

Quigley, J. M. (1985). Consumer choice of dwelling, neighborhood and public services. *Regional Science and Urban Economics*, *15*(1), 41–63.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55.

Rothstein, J. (2006). Good principals or good peers? Parental valuation of school characteristics, Tiebout equilibrium, and the effects of inter-district competition. *The American Economic Review*, *96*(4), 1333–1350.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, *55*(5), 999–1033.

Samuelson, P. (1954). The pure theory of public expenditure. *Review of Economics and Statistics*, *36*(4).

Schiraldi, P. (2006). Automobile replacement: A dynamic structural approach. Mimeo, Boston University.

Tiebout, C. (1956). A pure theory of local expenditures. *The Journal of Political Economy*, *64*(5), 416–424.

Urquiola, M. (2005). Does school choice lead to sorting? evidence from tiebout variation. *The American Economic Review*, *95*(4), 1310–1326.

Table 1
Characteristics of the School Districts

	FULL SAMPLE		LOGIT SAMPLE	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std.Dev.**
Neighborhood characteristics				
Average test score	59.43	8.51	57.41	9.12
Number of houses	848.54	712.35	1638.29	1533.78
Average number of rooms	6.1	0.8	5.8	0.8
Average number of bedrooms	2.8	0.4	2.7	0.4
Proportion of houses built 1960 or later	51%	21%	52%	23%
Proportion of business properties	2%	1%	2%	1%
Average rent	$768	$240	$755	$184
Average house value	$218,529	$118,807	$197,864	$83,863
Proportion of vacant houses for rent	1%	2%	1%	1%
Proportion of vacant houses for sale	1%	1%	1%	1%
Neighbor's characteristics				
Average income per household	$75,058	$34,904	$71,465	$26,861
Household employment: proportion head is employed	56%	13%	55%	10%
proportion both head and spouse are employed	30%	9%	29%	8%
Head of household's education: proportion with high-school or more	85%	9%	84%	10%
proportion with college or more	33%	18%	33%	17%
Head of household's race - proportion of black	7%	13%	10%	15%
proportion of non-black, non-white	6%	6%	8%	7%
Average head of household's age	51.4	3.2	50.6	2.8
Proportion of married households	55%	14%	53%	12%
Average number of children per household	0.7	0.2	0.7	0.1
Average number of cars per household	1.8	0.3	1.7	0.3
Proportion of household who own their house	76%	17%	70%	18%
Proportion of households who moved last year	14%	5%	15%	4%
Number of observations	501		224	

Table 2
Summary Statistics
Cohorts of Parents (by the age of the oldest child)

Cohort (Obs)		Home Owner	Head Employed	Head Black	Other Race	College or more	Household Income	Moved last year	Two children	3 or more children	Private School	Defier
Age 1	**Mean**	0.59	0.91	0.12	0.16	0.43	80,369.73	0.32	0.04	--	--	--
(3693)	**St. Dev.**	(0.49)	(0.28)	(0.33)	(0.37)	(0.50)	(85,120.17)	(0.47)	(0.21)	--	--	--
Age 2	**Mean**	0.59	0.92	0.13	0.18	0.42	78,876.97	0.29	0.20	--	--	--
(3926)	**St. Dev.**	(0.49)	(0.27)	(0.33)	(0.38)	(0.49)	(73,467.56)	(0.46)	(0.40)	--	--	--
Age 3	**Mean**	0.62	0.92	0.13	0.17	0.40	80,503.03	0.27	0.37	0.02	--	--
(4192)	**St. Dev.**	(0.49)	(0.27)	(0.33)	(0.38)	(0.49)	(83,166.04)	(0.44)	(0.48)	(0.14)	--	--
Age 4	**Mean**	0.64	0.92	0.14	0.17	0.38	81,221.37	0.25	0.48	0.06	0.46	0.01
(4236)	**St. Dev.**	(0.48))	(0.27)	(0.35)	(0.37)	(0.49)	(85,769.84)	(0.43)	(0.50)	(0.23)	(0.50)	(0.12)
Age 5	**Mean**	0.63	0.92	0.14	0.18	0.38	78,834.97	0.23	0.50	0.09	0.51	0.12
(4573)	**St. Dev.**	(0.48)	(0.27)	(0.34)	(0.38)	(0.48)	(76,360.46)	(0.42)	(0.50)	(0.29)	(0.50)	(0.33)
Age 6	**Mean**	0.64	0.91	0.15	0.15	0.37	81,585.27	0.23	0.51	0.14	0.24	0.17
(4597)	**St. Dev.**	(0.48)	(0.29)	(0.36)	(0.36)	(0.48)	(92,241.78)	(0.42)	(0.50)	(0.34)	(0.42)	(0.38)
Age 7	**Mean**	0.65	0.91	0.16	0.16	0.37	79,143.44	0.20	0.50	0.17	0.18	0.08
(4789)	**St. Dev.**	(0.48)	(0.29)	(0.36)	(0.37)	(0.48)	(83,881.51)	(0.40)	(0.50)	(0.37)	(0.38)	(0.28)
Age 8	**Mean**	0.66	0.91	0.15	0.16	0.35	80,709.53	0.19	0.50	0.20	0.16	0.02
(4915)	**St. Dev.**	(0.47)	(0.29)	(0.36)	(0.36)	(0.48)	(83,615.84)	(0.39)	(0.50)	(0.40)	(0.37)	(0.13)
Age 9	**Mean**	0.66	0.91	0.16	0.16	0.34	82,261.63	(0.17	0.48	0.22	0.15	0.02
(5065)	**St. Dev.**	(0.47)	(0.29)	(0.37)	(0.37)	(0.48)	(90,809.27)	0.38)	(0.50)	(0.42)	(0.36)	(0.14)
Age 10	**Mean**	0.67	0.91	0.15	0.14	0.34	80,419.80	0.17	0.48	0.24	0.14	0.11
(5439)	**St. Dev.**	(0.47)	(0.29)	(0.36)	(0.34)	(0.47)	(84,351.92)	(0.37)	(0.50)	(0.43)	(0.35)	(0.31)
Age 11	**Mean**	0.68	0.90	0.15	0.14	0.34	81,730.99	0.16	0.46	0.26	0.14	0.14
(5187)	**St. Dev.**	(0.47)	(0.30)	(0.36)	(0.35)	(0.47)	(90,918.76)	(0.36)	(0.50)	(0.44)	(0.35)	(0.35)

Table 2
Summary Statistics (cont.)
Cohorts of Parents (by the age of the oldest child)

Cohort (Obs)		Home Owner	Head Employed	Head Black	Other Race	College or more	Household Income	Moved last year	Two children	3 or more children	Private School	Defier*
Age 12	**Mean**	0.70	0.91	0.16	0.15	0.35	82,634.07	0.15	0.45	0.27	0.15	0.02
(5163)	**St. Dev.**	(0.46)	(0.29)	(0.37)	(0.35)	(0.48)	(91,100.75)	(0.36)	(0.50)	(0.44)	(0.35)	(0.16)
Age 13	**Mean**	0.71	0.92	0.16	0.14	0.35	83,321.73	0.15	0.46	0.27	0.14	0.02
(5093)	**St. Dev.**	(0.45)	(0.27)	(0.37)	(0.35)	(0.48)	(89,328.51)	(0.35)	(0.50)	(0.45)	(0.35)	(0.15)
Age 14	**Mean**	0.71	0.91	0.16	0.14	0.34	81,146.10	0.13	0.45	0.27	0.13	0.12
(5346)	**St. Dev.**	(0.46)	(0.28)	(0.37)	(0.35)	(0.47)	(80,492.34)	(0.34)	(0.50)	(0.45)	(0.34)	(0.33)
Age 15	**Mean**	0.73	0.91	0.16	0.15	0.35	81,598.33	0.13	0.42	0.27	0.14	0.17
(5424)	**St. Dev.**	(0.45)	(0.28)	(0.36)	(0.36)	(0.48)	(84,245.40)	(0.34)	(0.49)	(0.44)	(0.34)	(0.38)
Age 16	**Mean**	0.72	0.92	0.16	0.13	0.34	85,449.01	0.13	0.40	0.27	0.13	0.04
(5728)	**St. Dev.**	(0.45)	(0.28)	(0.36)	(0.34)	(0.47)	(93,311.82)	(0.33)	(0.49)	(0.44)	(0.34)	(0.20)
Age 17	**Mean**	0.73	0.92	0.15	0.12	0.35	87,582.91	0.11	0.40	0.26	0.12	0.04
(6150)	**St. Dev.**	(0.45)	(0.27)	(0.36)	(0.33)	(0.48)	(95,704.83)	(0.32)	(0.49)	(0.44)	(0.33)	(0.19)
Age 18	**Mean**	0.73	0.91	0.15	0.14	0.34	86,692.48	0.10	0.39	0.26	0.13	0.12
(5983)	**St. Dev.**	(0.44)	(0.28)	(0.36)	(0.34)	(0.47)	(88,001.75)	(0.31)	(0.49)	(0.44)	(0.34)	(0.33)
Age 19	**Mean**	0.68	0.90	0.18	0.15	0.25	75,692.26	0.11	0.38	0.28	0.15	0.28
(4088)	**St. Dev.**	(0.47)	(0.31)	(0.39)	(0.36)	(0.43)	(72,733.65)	(0.31)	(0.49)	(0.45)	(0.36)	(0.45)

* Defier is defined as equal to one if the child is attending the same grade as it is expected according to his or her birth date, and zero otherwise.

Table 3
Summary Statistics
Cohorts of Non-Parents (by the age of the head of household)

Cohort (Obs)		Home Owner	Head Employed	Head Black	Other Race	College or more	Household Income	Moved last year
Age 32	**Mean**	0.42	0.42	0.13	0.15	0.50	70,683.91	0.32
(2521)	**St. Dev.**	(0.49)	(0.49)	(0.34)	(0.35)	(0.50)	(59,521.55)	(0.47)
Age 33	**Mean**	0.46	0.41	0.14	0.14	0.46	72,052.67	0.31
(2436)	**St. Dev.**	(0.50)	(0.49)	(0.35)	(0.34)	(0.50)	(72,138.42)	(0.46)
Age 34	**Mean**	0.45	0.39	0.15	0.11	0.44	69,388.98	0.29
(2383)	**St. Dev.**	(0.50)	(0.49)	(0.36)	(0.32)	(0.50)	(54,285.57)	(0.46)
Age 35	**Mean**	0.47	0.38	0.16	0.12	0.42	71,944.31	0.27
(2449)	**St. Dev.**	(0.50)	(0.49)	(0.36)	(0.33)	(0.49)	(67,365.34)	(0.45)
Age 36	**Mean**	0.52	0.39	0.15	0.10	0.43	74,034.31	0.24
(2369)	**St. Dev.**	(0.50)	(0.49)	(0.36)	(0.30)	(0.49)	(70,205.73)	(0.42)
Age 37	**Mean**	0.50	0.38	0.17	0.12	0.42	72,186.81	0.24
(2251)	**St. Dev.**	(0.50)	(0.49)	(0.37)	(0.33)	(0.49)	(73,063.94)	(0.42)
Age 38	**Mean**	0.52	0.36	0.16	0.08	0.39	69,496.10	0.23
(2255)	**St. Dev.**	(0.50)	(0.48)	(0.37)	(0.28)	(0.49)	(68,718.15)	(0.42)
Age 39	**Mean**	0.53	0.36	0.16	0.09	0.38	70,212.76	0.21
(2154)	**St. Dev.**	(0.50)	(0.48)	(0.36)	(0.28)	(0.49)	(64,325.71)	(0.41)
Age 40	**Mean**	0.53	0.35	0.17	0.09	0.39	69,389.31	0.21
(2176)	**St. Dev.**	(0.50)	(0.48)	(0.37)	(0.28)	(0.49)	(65,849.15)	(0.40)
Age 41	**Mean**	0.54	0.36	0.16	0.08	0.36	67,830.01	0.21
(2120)	**St. Dev.**	(0.50)	(0.48)	(0.37)	(0.28)	(0.48)	(67,546.03)	(0.41)
Age 42	**Mean**	0.54	0.35	0.18	0.09	0.35	67,361.92	0.18
(2278)	**St. Dev.**	(0.50)	(0.48)	(0.38)	(0.28)	(0.48)	(69,972.43)	(0.39)
Age 43	**Mean**	0.56	0.36	0.16	0.08	0.38	66,961.68	0.17
(2140)	**St. Dev.**	(0.50)	(0.48)	(0.37)	(0.27)	(0.49)	(63,895.50)	(0.38)
Age 44	**Mean**	0.56	0.37	0.17	0.07	0.36	69,312.88	0.16
(2229)	**St. Dev.**	(0.50)	(0.48)	(0.38)	(0.25)	(0.48)	(70,550.42)	(0.36)
Age 45	**Mean**	0.58	0.38	0.16	0.07	0.35	73,017.45	0.16
(2273)	**St. Dev.**	(0.49)	(0.48)	(0.36)	(0.25)	(0.48)	(90,527.92)	(0.36)
Age 46	**Mean**	0.59	0.39	0.15	0.07	0.36	69,803.82	0.15
(2252)	**St. Dev.**	(0.49)	(0.49)	(0.36)	(0.26)	(0.48)	(66,429.75)	(0.36)

Table 3
Summary Statistics (cont.)
Cohorts of Non-Parents (by the age of the head of household)

Cohort (Obs)		Home Owner	Head Employed	Head Black	Other Race	College or more	Household Income	Moved last year
Age 47	**Mean**	0.62	0.40	0.16	0.07	0.36	70,091.73	0.15
(2328)	**St. Dev.**	(0.49)	(0.49)	(0.37)	(0.25)	(0.48)	(72,214.65)	(0.36)
Age 48	**Mean**	0.63	0.42	0.18	0.07	0.37	70,782.97	0.13
(2390)	**St. Dev.**	(0.48)	(0.49)	(0.38)	(0.25)	(0.48)	(64,579.28)	(0.34)
Age 49	**Mean**	0.64	0.43	0.17	0.07	0.39	75,145.41	0.13
(2522)	**St. Dev.**	(0.48)	(0.50)	(0.38)	(0.25)	(0.49)	(81,951.39)	(0.34)
Age 50	**Mean**	0.65	0.44	0.17	0.07	0.39	77,819.51	0.12
(2613)	**St. Dev.**	(0.48)	(0.50)	(0.37)	(0.26)	(0.49)	(89,657.59)	(0.33)

Table 4
Hedonic Price Regressions

	I	II	III	IV	V	VI	VI	VII	VIII
MWTP for 5% increase in Test Score	5.62%	5.23%	1.62%	1.55%	1.94%	1.03%	1.42%	1.62%	1.49%
Test Score	0.087**	0.081**	0.025**	0.024**	0.030**	0.016*	0.022**	0.025**	0.023**
	(0.007)	(0.009)	(0.006)	(0.006)	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)
House built between 1990-1994		-1.209*	-1.216*	-1.220*	-1.212*	-1.140*	-1.11	-1.12	-1.121*
		(0.599)	(0.574)	(0.575)	(0.572)	(0.573)	(0.568)	(0.569)	(0.569)
House built between 1980-1989		-2.346**	-2.240**	-2.238**	-2.222**	-2.197**	-2.179**	-2.166**	-2.174**
		(0.565)	(0.523)	(0.527)	(0.518)	(0.517)	(0.515)	(0.516)	(0.518)
House built between 1970-1979		-2.900**	-2.573**	-2.570**	-2.539**	-2.557**	-2.546**	-2.534**	-2.539**
		(0.632)	(0.600)	(0.602)	(0.596)	(0.597)	(0.596)	(0.596)	(0.597)
House built between 1960-1969		-2.803**	-2.422**	-2.422**	-2.401**	-2.477**	-2.459**	-2.452**	-2.461**
		(0.626)	(0.592)	(0.594)	(0.588)	(0.590)	(0.588)	(0.589)	(0.592)
House built between 1950-1959		-2.784**	-2.410**	-2.412**	-2.398**	-2.514**	-2.499**	-2.499**	-2.508**
		(0.647)	(0.618)	(0.620)	(0.618)	(0.618)	(0.617)	(0.619)	(0.619)
House built between 1940-1949		-3.072**	-2.651**	-2.654**	-2.645**	-2.736**	-2.720**	-2.716**	-2.720**
		(0.649)	(0.624)	(0.624)	(0.625)	(0.625)	(0.624)	(0.624)	(0.625)
House built on or before 1939		-3.103**	-2.791**	-2.794**	-2.798**	-2.819**	-2.801**	-2.796**	-2.799**
		(0.593)	(0.591)	(0.591)	(0.597)	(0.589)	(0.586)	(0.586)	(0.587)
Number of rooms		0.287**	0.241**	0.236**	0.259**	0.298**	0.296**	0.298**	0.298**
		(0.043)	(0.030)	(0.029)	(0.028)	(0.026)	(0.026)	(0.026)	(0.026)
Number of Bedrooms		0.526**	0.639**	0.646**	0.623**	0.594**	0.592**	0.588**	0.588**
		(0.087)	(0.085)	(0.084)	(0.081)	(0.076)	(0.075)	(0.075)	(0.075)
House is a business property		-0.007	0.055	0.048	0.070	0.124	0.121	0.120	0.120
		(0.140)	(0.128)	(0.126)	(0.125)	(0.118)	(0.119)	(0.120)	(0.119)
House between 1 and 10 acres		-0.058	-0.084	-0.125	-0.067	0.336**	0.348**	0.357**	0.358**
		(0.166)	(0.129)	(0.120)	(0.120)	(0.119)	(0.120)	(0.119)	(0.118)
House is over 10 acres		-0.917**	-0.946**	-0.987**	-0.929**	-0.286	-0.293	-0.302	-0.302
		(0.257)	(0.206)	(0.206)	(0.204)	(0.200)	(0.199)	(0.197)	(0.196)
Proportion of highly educated neighbors			7.682**	7.009**	6.495**	6.076**	5.322**	4.668**	6.107**
			(0.470)	(0.910)	(1.093)	(1.138)	(1.077)	(1.111)	(1.221)
Average Income				0.006	0.008	0.025**	0.031**	0.036**	0.032**
				(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Controls 1	no	no	no	no	yes	yes	yes	yes	yes
Controls 2	no	no	no	no	no	yes	yes	yes	yes
Controls 3	no	no	no	no	no	no	yes	yes	yes
Controls 4	no	no	no	no	no	no	no	yes	yes
Controls 5	no	no	no	no	no	no	no	no	yes
Constant	2.419**	3.414**	3.923**	3.839**	2.778*	8.355**	5.522*	-0.726	2.743
	(0.389)	(0.930)	(0.669)	(0.689)	(1.336)	(2.354)	(2.411)	2.689	(2.918)
Observations	137,239	137,239	137,239	137,239	137,239	137,239	137,239	137,239	137,239
R-squared	0.08	0.16	0.24	0.24	0.25	0.26	0.27	0.27	0.27

Notes: Dependent Variable: house rent value.
Robust standard errors in parenthesis. *significant at 5%; ** significant at 1%.
Controls 1: (1) Percentage of head of household with high-school degree or more, (2) Percentage of black head of household, (3) Percentage of non-black, non-white head of household
Controls 2: Percentage of houses built between (1) 1990-1994, (2) 1980-1989, (3) 1970-1979, (4) 1960-1969, (5) 1950-1959, (6) 1940-1949, (7) Percentage of houses built on or before 1939, (8) Average number of rooms, (9) Average number of bedrooms, (10) Percentage of business properties, (11) Percentage of properties between 1 and 10 acres, (12) Percentage of properties larger than 10 acres, (13) Percentage of households who own their house
Controls 3: (1) Percentage of vacant houses for rent
Controls 4: (1) Percentage of married households
Controls 5: (1) Average number of cars per household, (2) Percentage of head of household employed, (3) Percentage of households where both the head and the spouse are employed.

Table 5
Results from Step 2
MWTP by Year for an Increase in 5% in School Quality at the Family Level

Age of the oldest child	I	II	III	IV	V	VI
6	3.25*	2.12*	2.21*	2.29*	2.28*	2.27
	(1.50)	(0.99)	(1.00)	(1.07)	(1.13)	(1.17)
7	3.19*	1.95*	2.20*	2.32*	2.20*	2.21*
	(1.33)	(0.85)	(0.88)	(0.96)	(1.00)	(1.03)
8	2.89*	1.69*	1.90	2.02**	2.18**	2.14**
	(1.33)	(0.75)	(1.00)	(0.74)	(0.76)	(0.76)
9	2.98*	2.02	1.80*	1.80*	1.60	1.54
	(1.51)	(1.15)	(0.78)	(0.86)	(0.84)	(0.83)
10	2.59**	2.27*	2.10*	2.18*	2.30*	2.27*
	(1.01)	(0.97)	(0.84)	(0.99)	(1.11)	(1.13)
11	3.12*	2.21*	2.60**	2.27**	2.10*	2.03*
	(1.42)	(0.89)	(0.96)	(0.84)	(0.85)	(0.84)
12	3.54	2.54	2.70*	2.83*	2.80*	2.69
	(2.11)	(1.46)	(1.37)	(1.34)	(1.42)	(1.38)
13	3.12*	2.92*	2.60**	3.35*	3.10*	2.92*
	(1.47)	(1.22)	(1.00)	(1.32)	(1.35)	(1.30)
14	3.89**	2.83**	3.20**	2.91**	2.67**	2.57**
	(1.35)	(0.97)	(1.19)	(0.87)	(0.80)	(0.79)
15	4.12**	3.58**	3.80**	3.86**	3.69**	3.67**
	(1.32)	(1.30)	(1.23)	(1.36)	(1.38)	(1.38)
16	3.87**	3.28**	3.70**	3.91**	3.70**	3.61*
	(1.11)	(1.12)	(1.12)	(1.27)	(1.43)	(1.43)
17	5.12**	4.11**	3.80**	4.92**	4.60**	4.57**
	(1.91)	(1.28)	(1.12)	(1.27)	(1.13)	(1.14)
18	5.59**	3.89**	4.40**	4.44**	4.56**	4.52**
	(2.18)	(1.43)	(1.38)	(1.68)	(1.23)	(1.23)
Controls?	No	Yes	No	No	No	Yes
Proxies (cohorts)	-	-	49	1, 49	1-2, 49	1-2, 49
Instruments (cohorts)	-	-	45-48	3-5, 45-48	3-5, 40-48	3-5, 40-48
R-squared	0.186	0.283	0.454	0.642	0.643	0.654
Observations	3136	3136	3136	3136	3136	3136

Notes:
Standard errors calculated using the Delta method, and corrected for clustering by SD and for heteroskedascity.
In all columns, average rent is included as control variable.
Dependent Variable: estimated mean instantaneous utilities for cohorts 6 through 18.
Proxy Variables: estimated mean instantaneous utilities for cohorts referred in each column.
Instruments: estimated mean instantaneous utilities for cohorts referred in each column.
Control Variables: Average Income, proportion of black neighbors, proportion of neighbors with race other than white or black, proportion of highly educated neighbors, average number of rooms, proportion of homeowners in the neighborhood.
Underlying Assumptions: preference for average school quality of cohort of proxies equal to zero; preference for average rent for cohorts of age 1, 5, 6 and 15 are the same on average.
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.

Table 6

MWTP^{SQ} *per year, per child, per grade*												
Elementary School						**Middle School**			**High School**			
K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
2.2	2.0	1.5	0.6	1.3	1.0	1.6	1.8	1.4	2.5	2.3	3.2	2.9
(1.1)	(1.0)	(0.7)	(0.7)	(1.0)	(0.7)	(1.3)	(1.2)	(0.7)	(1.3)	(1.4)	(1.1)	(1.2)
		1.4					**1.6**			**2.7**		
		(0.4)					**(0.6)**			**(0.6)**		

Magnitude of the results

(From a baseline $755 monthly rent)

	MWTP^{SQ}(5%)		**MWTP^{SQ}(1 St.Dev)**	
	Per month	**Per year**	**Per month**	**Per year**
Elementary School	11	130	35	405
Middle School	12	145	40	460
High School	21	245	65	780

1 St. Dev. $\approx$ 16%

Values derived from estimates from specification VI in table 5.

Table A1
Summary Statistics
House Characteristics

	FULL SAMPLE		RENTERS		HOMEOWNERS	
House variables	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
House built between 1999-2000	1%	11%	1%	9%	1%	12%
1995-1998	4%	20%	3%	16%	5%	21%
1990-1994	5%	21%	3%	18%	5%	23%
1980-1989	12%	33%	10%	30%	13%	34%
1970-1979	14%	34%	16%	36%	13%	33%
1960-1969	16%	37%	17%	38%	15%	36%
1950-1959	17%	38%	16%	36%	18%	38%
1940-1949	10%	30%	12%	33%	9%	29%
on or before 1939	21%	41%	22%	42%	20%	40%
Number of rooms	5.7	2.0	3.9	1.5	6.6	1.7
Number of bedrooms	2.7	1.1	1.7	0.9	3.1	0.9
Property area: 1-10 acres	9%	28%	2%	13%	12%	33%
Property area: > 10 acres	1%	8%	0.3%	6%	1%	9%
Rent	$693	$372	$693	$372		
House for rent (vacant)	2%	13%	5%	22%		
House value	$209,068	$153,554			$209,068	$153,554
House for sale (vacant)	1%	10%			2%	12%
Business property	2%	13%	1%	9%	2%	15%
Number of observations	425,119		137,239		287,880	