Neighborhood Sorting and the Value of Public School Quality

Gregorio Caetano*

August 2016

Abstract

This paper develops a method to estimate the parental value of public school quality with two novel features. First, it estimates the value of public school quality in the same unit in which public schools' costs are measured and private school tuition is charged: per year, per child at each grade level. Second, it develops a novel approach to control for unobservables correlated to school quality, including those generated by sorting. People without school-age children enjoy neighborhood-level amenities but do not enjoy school-level amenities, so data about their residential choice can be used to control for neighborhood unobservables, isolating the value of school quality per se. I embed this idea into a dynamic model of neighborhood choice, building on previously unconnected literatures. Using the 2000 U.S. Census data, I find that parents tend to value school quality more in elementary and high school grades relative to middle school grades. However, improving public school quality currently costs more than is worth to parents even at the most valued grades, so externalities in education are necessary to justify such investments. These findings highlight the importance of improving the efficiency with which school resources are spent. *Keywords: School Quality, Willingness to Pay, Neighborhood Sorting, Neighborhood Amenities, Dynamic Demand Estimation, Public Schools, Housing Market. JEL Code: C5, H4, I2, R2, R3.*

*Department of Economics, University of Rochester, 238 Harkness Hall Box 270156, Rochester NY 14627. E-mail: gregorio.caetano@rochester.edu. I thank David Card for many helpful discussions and invaluable advice throughout the duration of this project. I also thank Carolina Caetano, Fernando Ferreira, Josh Kinsler, Patrick Kline, Ashley Langer, Vikram Maheshri, Robert McMillan, Enrico Moretti, Romans Pancs, Ronni Pavan, Claudia Sitgraves, Melissa Tartari and seminar participants for their comments. Yusuke Jinnai, Alexis Orellana and Daniel Ringo provided excellent research assistance. I am also grateful to the help provided by Angela Andrus, Arnie Reznek, Liliana Souza and Jon Stiles of the Census Bureau. The research results and conclusions expressed in this paper are mine and do not necessarily reflect the views of the Census Bureau. This paper has been screened to ensure that no confidential data are revealed. All remaining errors are my own.

1 Introduction

In recent years, there has been a growing interest in evaluating the efficiency with which public money is spent within the educational system. Information on the costs of running different public schools are readily available and so are the schools' quality measurements. However, any cost-benefit evaluation needs to incorporate a third, crucial ingredient: the parental value of public school quality. This paper estimates how much parents are willing to pay for public school quality per year, per child at each grade level. This unit of measurement is important for a comparison with the costs of improving public school quality and with the tuition charged by competing private schools.

Using the 2000 U.S. Census data for Minnesota, I find that parents are, on average, willing to pay \$2,400, \$1,400 and \$2,700 yearly per child for an investment that improves test scores by one standard deviation in elementary, middle and high school grades, respectively. Such investments, however, cost at least \$2,850 per pupil per year.[1] These findings highlight the importance of improving the efficiency with which public school resources are spent; otherwise externalities in education are necessary to justify such investments.

Because there is no direct market for purchasing public schooling, researchers have attempted to estimate the valuation of public school quality by comparing the prices of similar houses in neighborhoods that are also similar except for their level of school quality (Black and Machin (2011) provides a recent survey of this prolific literature). An important concern in this literature is to fully control for the demographic composition of neighbors, since otherwise they will bias the valuation estimates. Unobservable demographics of neighbors are post-determined, as they are affected by school quality through neighborhood choice. Thus, any bias due to omitted demographics of neighbors cannot be fully avoided with the use of instrumental variables; instead, these need to be explicitly controlled for.[2] The conventional way of handling this issue is to explicitly add observed demographics as controls, but this may not be enough to absorb unobservables. Moreover, it might also partially absorb the demographics of school peers, thus identifying only the value for the component of school quality that is orthogonal to school peers. Of course, such measure of the value of school quality is important for many policies aimed at changing school inputs, particularly in the short-run, but it is difficult to use it to infer the value parents give to an open enrollment policy, to a private school voucher system, or to other programs that enable

[1] See Greenwald, Hedges, and Laine (1996) and Hanushek (1997) for widely-known meta-analyses on the topic.

[2] Indeed, any initial change in school quality, no matter how random its source, will lead to the sorting of people across observed and unobserved characteristics. Thus, at any point in time neighborhoods are observed not only with different levels of school quality and prices, but also with systematically different neighbors.

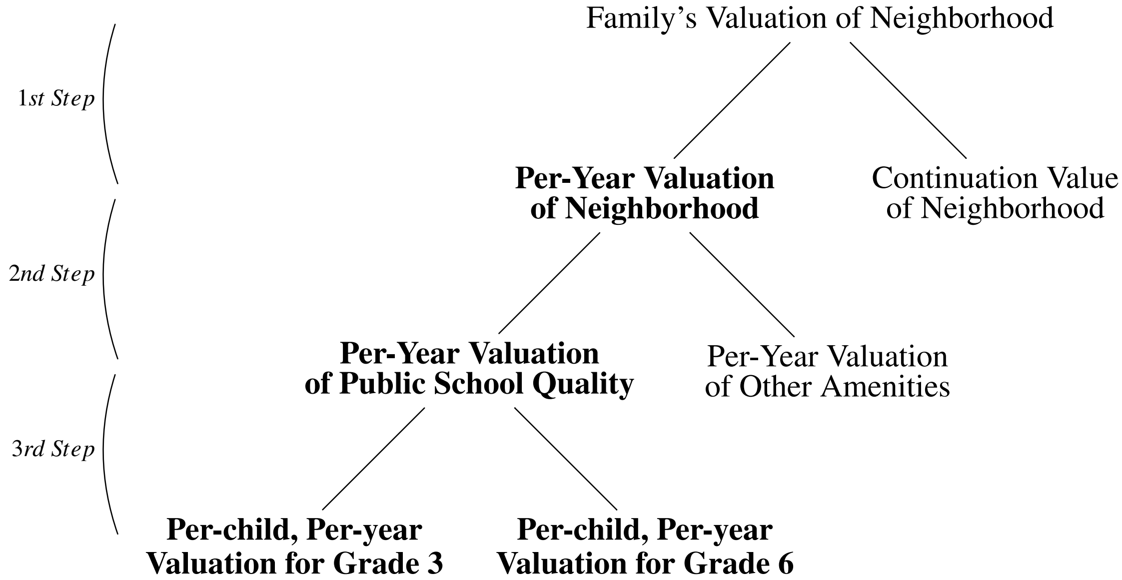
them to opt among schools with different peers.[3]

This paper complements this literature in two ways. First, it develops a new way of controlling for omitted variables. It controls for unobservables, including the demographics of neighbors, and it does so without controlling for the demographics of school peers. This strategy can be implemented on its own or together with other approaches to control for unobservables such as the use of boundary fixed effects (Black (1999)), which has been successful at controlling for much of the endogeneity in this context. Second, it provides estimates at a specific unit of measurement that facilitates cost-benefit analyses: per year, per child at each grade level. Because families may intend to stay in the same house for longer than one year, and because children of the same family may attend public schools in different grades, the estimate obtained directly from housing choices spans more than one year, one child and one grade.

The proposed approach has three steps. The first step estimates a dynamic model of residential sorting. Because of moving costs, families are aware that their residential choice today will influence their future residential decisions. Thus, this step extracts the family's per-year valuation of the neighborhood from the family's total valuation of the neighborhood (which also includes the continuation value of that neighborhood). The second step extracts the family's per-year valuation of public school quality from the family's per-year valuation of the neighborhood (which also includes the per-year valuation of other amenities). Here, neighborhood unobserved amenities are proxied by neighborhood *per-year* valuations of families without school-age children (including non-parents). These proxies control for the characteristics of neighbors and other confounders without controlling for the characteristics of school peers, since in that year these families do not enjoy the school amenities. The third step distributes the families' per-year valuations of public school quality among their children according to the grade each child attends, aggregating the information across all families to obtain an estimate of the parental value of public school quality per year, per child at each grade level. Figure 1 illustrates these three steps for a family with two children, one attending grade 3 and the other attending grade 6.

[3]For instance, estimates of the value of school quality incorporating peer characteristics can be helpful as an input for models studying the impact of educational policies in the spirit of Epple and Romano (1998) and Ferreyra (2007).

Figure 1: Three-Step Approach



Notes: This figure illustrates the three-step approach for a family with two children, one attending grade 3 and the other attending grade 6. Step 1 extracts the family's per-year valuation of the neighborhood from the family's total valuation of the neighborhood (which also includes the continuation value of that neighborhood). Step 2 extracts the family's per-year valuation of public school quality from the family's per-year valuation of the neighborhood (which also includes the per-year valuation of other amenities). Step 3 divides the family's per-year valuation of public school quality based on each child's grade of attendance.

To obtain estimates at the desired unit of measurement, the first step requires an estimation of a dynamic model with Decennial Census data - a cross-sectional dataset. To accommodate this data restriction, I make a synthetic cohort assumption to connect the contemporaneous residential choices of different families observed at different times in their life cycle, as if these choices were from the same family along its own life cycle.[4] For example, I assume that families with an 8-year-old child expect to have the same preferences next year as families with a 9-year-old child have this year. I provide evidence suggesting that this assumption is reasonable in the context of this paper. Moreover, I argue that in practice small violations of this assumption should not bias the estimate of the value of school quality, which is the main parameter of interest. Intuitively, violations of this assumption might bias the intermediate estimator of the per-year value of the neighborhood, but this bias should be absorbed by controls in the next step of the estimation procedure.

Methodologically, this paper relates to two separate literatures. The first literature estimates the valuation of non-marketed goods with horizontal neighborhood sorting models using a discrete choice framework. Three papers are particularly relevant to this study.

[4] Epple, Romano, and Sieg (2012) make a similar synthetic cohort assumption in an overlapping generations model to study the intergeneration conflict of provision of public education.

Bayer, Ferreira, and McMillan (2007) embeds the boundary fixed effects approach into a static discrete choice framework, building on insights from the Industrial Organization (Berry (1994), Berry, Levinsohn, and Pakes (1995)), and estimate the valuations of public school quality and neighbors' demographic composition. More recently, Bayer, McMillan, Murphy, and Timmins (2016) uses longitudinal data on homeowners to estimate the valuation of violent crime, air pollution and racial composition with a dynamic discrete choice model, building on insights from Rust (1987) and Hotz and Miller (1993). Finally, Mastromonaco (2014) builds on Bajari, Fruehwirth, Kim, and Timmins (2012) and Bayer et al. (2016) to estimate the general equilibrium impact of school quality on wealth accumulation, prices and the demographic composition of neighborhoods.[5] Besides the differences in methodology and underlying assumptions, Mastromonaco (2014) aims at explicitly studying the sorting due to school quality, while this paper aims at controlling for it. As in Bayer et al. (2007), this paper estimates the valuation of public school quality with a discrete choice framework, but with two key differences. First, it uses a different approach to treat the endogeneity problem. Second, it estimates a dynamic rather than a static model. The dynamic estimation approach I develop departs from standard approaches in a similar way as Bayer et al. (2016), but imposes assumptions to accommodate estimation with the Decennial Census data.

The second literature involves the use of a specific panel data method to control for unobservables. This technique, discussed in Chamberlain (1977) and Pudney (1982), has been applied to wage regressions (e.g., Heckman and Scheinkman (1987)), but so far has not been applied to choice models. The main methodological contribution of the paper lies in embedding this method in a discrete choice framework, thus connecting these two literatures. The approach that arises from this connection is potentially useful for a wide range of applications in demand estimation, because it provides a new way of controlling for unobservables, including those inherently endogenous such as the neighbors' demographic compositions. The key idea is to construct control variables from observed *choices* of others, not from observed *characteristics*. This is helpful because choices should reflect all characteristics, including the ones unobserved to econometricians.

The rest of the paper is organized as follows: Section 2 presents the novel approach to control for pre- and post-determined unobservables. Section 3 describes the data. Section 4 describes the identification strategy of the paper in three steps. Section 5 presents the empirical results and robustness checks. Section 6 provides a more detailed discussion of the results, and, finally, Section 7 offers some concluding remarks. Appendix A presents

[5] Bajari et al. (2012) develops a new approach to absorb unobservables in an hedonic framework by controlling for previous house prices under a rational expectations assumption.

some technical details, and Appendix B provides a Monte Carlo study of the approach to control unobservables developed in the paper.

2 Controlling for Unobservables

This section presents the novel idea of controlling for pre- and post-determined unobserved neighborhood amenities, including neighborhood demographics. For exposition purposes, I first present the idea in a simple benchmark model with no moving costs.

2.1 Basic Setup

Suppose there are two groups of households in the data: parents and non-parents, indexed respectively as p and np. Each household chooses a neighborhood to live among J options. Each neighborhood is characterized by a set of amenities, with some of them unobserved to the researcher. Because there is no moving costs, households consider only their flow utility when deciding where to live.[6] Let the flow utility of household i in group c when choosing neighborhood j be

$$U_{i,c,j} = SQ_j.\theta_c + P_j.\phi_c + \xi_{c,j} + \varepsilon_{i,c,j}, \qquad c = p, np, \quad j = 1, ..., J, \tag{1}$$

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

In this context, the marginal willingness to pay (MWTP) for school quality of group c is measured as the marginal rate of substitution between SQ and P, or $MWTP_c = -\theta_c/\phi_c$, hence θ_c and ϕ_c are the parameters of interest.

Collecting all the terms at the group-neighborhood level:

$$\delta_{c,j} = SQ_j.\theta_c + P_j.\phi_c + \xi_{c,j}, \qquad c = p, np, \quad j = 1, ..., J, \tag{2}$$

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

Under the assumptions of this simplified model, it is easy to estimate $\delta_{c,j}$ with information on households' neighborhood choices as in Berry (1994),[7] so this section focuses

[6] In the absence of moving costs, only differences in the level of amenities available in the particular period can have an influence in household's decisions.

[7] Following Berry (1994), if $n_{c,j}$ is the number of households of group c who live in neighborhood j, and if $\delta_{c,1} = 0$ is normalized for each c, then $\hat{\delta}_{c,j} = \log(n_{c,j}) - \log(n_{c,1})$, $j = 2, ..., J$. Under the simplified assumption of no moving costs there is a direct link between households' neighborhood choices and their flow utility. In Section

only on the endogeneity issues that may arise in the identification of θ_c and ϕ_c in equation (2) once δ, SQ and P are observed.[8]

The variable ξ in equation (2) accounts for unobserved neighborhood amenities. These include not only pre-determined amenities but also post-determined ones, such as amenities affected by sorting (e.g., composition of neighbors). P and SQ are in general correlated to ξ, rendering any naive estimation of θ and ϕ biased. Next, I introduce a novel way of controlling for unobservables in this context. The method is a generalized version of a technique that was first proposed by Chamberlain (1977) and Pudney (1982) and was used by Heckman and Scheinkman (1987) in the context of the estimation of wage regressions.

2.2 Controlling for Unobservables

The key idea that I exploit is that unobserved amenities correlated to school quality often affect the neighborhood choice of people that do not care about school quality *per se*. This allows me to use the choice of these people to proxy for these unobservables. To see this, let $\xi_{c,j}$ in equation (2) be decomposed in two latent terms, where one is controlled and the other is assumed exogenous. If $\xi_{c,j} := Q_j.\lambda_c + \mu_{c,j}$, then equation (2) is re-written as

$$\delta_{c,j} = SQ_j.\theta_c + P_j.\phi_c + Q_j.\lambda_c + \mu_{c,j}, \qquad c = p, np, \quad j = 1, ..., J, \tag{3}$$

where the following identification condition is assumed to hold:

$$E[\mu_{c,j}|SQ_j, P_j, Q_j] = 0, \quad c = p, np, \ j = 1, ..., J. \tag{4}$$

Intuitively, the idea behind the method is to use the δs of one group (e.g., non-parents) as a proxy for Q in the equation of another group (e.g., parents). Since both parents and non-parents, once they live in the same neighborhood, are exposed to the same level of amenities Q, under certain conditions non-parental valuation of neighborhood j (i.e., $\delta_{np,j}$) should provide all information about Q_j necessary to absorb the source of endogeneity in the equation of $\delta_{p,j}$.

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

4 this assumption will be relaxed, which will make the identification of $\delta_{c,j}$ more cumbersome.

[8] Section 4 also discusses the implications of $\delta_{c,j}$ being estimated rather than directly observed.

$$\delta_{p,j} = SQ_j.\tilde{\theta} + P_j.\tilde{\phi} + \delta_{np,j}.\tilde{\lambda} + \tilde{\mu}_j \qquad j = 1, ..., J, \tag{5}$$

where

$$\tilde{\theta} := \theta_p - \theta_{np}.\tilde{\lambda}, \tag{6}$$

$$\tilde{\phi} := \phi_p - \phi_{np}.\tilde{\lambda}, \tag{7}$$

$\tilde{\lambda} := \frac{\lambda_p}{\lambda_{np}}$ and $\tilde{\mu}_j := \mu_{p,j} - \mu_{np,j}.\tilde{\lambda}$.

Under the selection on unobservables assumption (4), $\tilde{\theta}$, $\tilde{\phi}$ and $\tilde{\lambda}$ are identified as the coefficients of SQ, P and δ_{np} in equation (5), respectively. To identify θ_p and ϕ_p, the actual coefficients of interest, two auxiliary assumptions need to be made. Here, I discuss broadly the auxiliary assumptions made in the paper, but other assumptions can be made in different contexts, depending on the prior information from the researcher about the amenities of interest and the groups.

To identify θ_p via equation (6), I assume $\theta_{np} = 0$. Non-parents are assumed to not value school quality in the *flow* utility sense. Note that the assumption concerns the school quality services provided only during this year in the neighborhood, since at the moment they do not have any child attending school.[9] Additionally, in the flow utility sense non-parents are allowed to value neighborhood amenities that are often proxied by school quality. For instance, non-parents may value living close to well educated neighbors, but they do not value (in the flow utility sense) school peer effects due to children of well educated neighbors. Thus, the "school amenity" is included in SQ and the "neighborhood amenity" is included in Q.

A similar assumption about the coefficient of P is not plausible, as everyone is likely to care about price even in the flow utility sense. Thus, to identify ϕ_p, I assume $\phi_{np} = \phi_p$ instead.[10] Under the testable assumption that $\tilde{\lambda} \neq 1$, ϕ_p is identified via equation (7) since $1 - \tilde{\lambda}$ is identified through the coefficient of $\delta_{np,j}$ in equation (5).

Under these assumptions, the MWTP for school quality is identified.[11] In a more general setting, but still under the assumption of linearity in equation (3), this approach

[9] As discussed in Section 4, they are allowed to be forward looking agents and to value school quality this year in the cumulative utility sense, in a more general framework with moving costs.

[10] The actual assumption made in the paper, Assumption 2 described in Section 4, is similar but weaker than this one.

[11] As Chamberlain (1977) noted, the use of δ as proxy for Q introduces an omitted variables problem easily solvable under the same assumptions with the use of δ of a third group as IV. For this reason, Chamberlain (1977) referred to this method as Proxy-IV. Section 4.2 provides further discussion on this topic.

offers substantial flexibility: (a) More than one group can be used as proxy variable in order to account for more than one unobserved amenity, hence weakening the selection on unobservables assumption; (b) Groups can be defined so as to weaken the auxiliary assumptions; (c) The auxiliary assumptions can be relaxed; (d) Tests that have the power to detect violations of each of these assumptions can be performed. All these extensions are implemented in Section 5. Section 4.2 presents the generalized version of the approach implemented in Section 5.

Remark 1. *Not necessarily all relevant unobservables will be controlled with this method. The approach can control for only unobserved amenities valued by groups used as proxy. In this example, unobservables such as neighbor's characteristics will be proxied, but unobservables such as the presence of children's parks in the neighborhood will not be proxied, as non-parents do not value such amenity in the flow utility sense. Moreover, to be included in Q the unobserved amenity needs to be non-excludable. That is, by assumption, different groups are all exposed to the same level of the unobserved amenity Q once they choose the same neighborhood, although they may value that amenity differently (i.e., λ may vary with c). In Appendix B, I present a Monte Carlo study of this method, arguing that in practice endogenous amenities that cannot be written as $Q_j.\lambda_c$ are still controlled for with the method, as long as enough groups are added as proxy.*

3 Data

The analysis is performed with the restricted-access, or long form, version of the 2000 Decennial Census of Population and Housing for the state of Minnesota. It is a $1/6$ sample of all families in the state, containing detailed information on the characteristics of houses, families and individuals within families. This data is uniquely suited to the analysis in this paper for two reasons. First, it identifies where families are living down to a census block, which is a geographical area similar to a street block. This allows for linking each house to one and only one public school for each grade using electronic boundary maps of all elementary, middle, and secondary school attendance areas.[12] The second unique piece of information available in the data is the exact date of birth of each person. This allows for linking each child in the family to one and only one grade using the kindergarten entry rule of the state of Minnesota, which states that children turning five before September 1st will enter kindergarten that year, while children turning five after that date will enter kindergarten the following year.

[12]These electronic boundary maps are difficult to obtain for a single school district, let alone for the whole state for all grades. This is one of the primary reasons why I restrict the analysis to the state of Minnesota only.

I merge this data with average student achievement data from the Minnesota Comprehensive Assessments (MCA), which is a statewide standardized test covering reading and math at the school level.[13] I define neighborhoods as elementary school attendance areas, as middle and secondary school attendance areas are larger and contain one or more elementary attendance areas.

Table 1 shows the summary statistics of the observed characteristics of the neighborhoods as well as of families in our sample. The sample has 345 neighborhoods, with, on average, over 800 houses each. The average monthly rent is \$544, with a standard deviation of \$212. Additionally, the average house value is \$144,215, with standard deviation of over \$59,219. The neighborhoods are quite diverse. For instance, the average income of the neighborhoods is around \$65,000 with standard deviation of \$24,000, and the average proportion of neighbors with college degree or more is 28% with standard deviation of 17%. The average test score, which is the measure of school quality used in the paper, is 1,387, with standard deviation of 223. The amount of variation in this variable is similar to the amount of variation of the test score used in previous studies.[14]

The analysis is restricted to 40 groups, which are disjoint subsets of the population of families. Groups are defined in two different ways, depending on whether the family has children (parents) or not (non-parents). For parents, groups are defined by the age of the oldest child, ranging from ages 0 to 19. For non-parents, groups are defined by the age of the head of the family, ranging from ages 31 to 50.[15] Specifically for the groups with children at school (i.e., groups 6 through 19), the sample is further restricted to only those who attend a public school[16] and who are attending the correct grade as prescribed by their exact birthdate.[17]

[13] This data is publicly available at the Minnesota Department of Education (http://education.state.mn.us/MDE/index.html).

[14] For example, Bayer et al. (2007) use a measure of school quality with the standard deviation amounting to 14% of the mean of the variable, comparable to 16% in this paper.

[15] The average age of the parent whose oldest child is 0 years old in our sample is equal to 31.

[16] The likelihood of each child attending a private school is relatively constant around 10% for all groups.

[17] Data on the grade each child attends are available in the census only at aggregated categories (kindergarten, grades 1-4, grades 5-8, and grades 9-12), so exact date of birth is necessary to predict which exact grade within each category the child is attending. If the predicted grade does not fall in the appropriate category (e.g., if the child was held back a year), then this observation is excluded from the analysis. Less than 1% of the observations are dropped because of this restriction.

Table 1

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

Table 2 shows the summary statistics of the observed characteristics of the groups of parents used in the analysis. The race distribution and employment status do not vary much across groups. The average family income is between $\$56,000$ and $\$71,000$ for parents in each group, and is similar for consecutive groups. The average proportion of parents who are homeowners varies from 75% to 90%, with older parents being more likely to be homeowners. Parents with younger children tend to be more educated; the proportion of families with a college degree varies from 20% to 30% for these groups. The moving rate also depends substantially on the age group, with younger parents moving more than older parents. Finally, the proportions of parents who have two children and three or more children understandably grow as parents get older.

Table 2

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

Group (obs)	**Household Income ($)**	**Home-owner (%)**	**Black (%)**	**Other Race (%)**	**High School Degree (%)**	**College Degree (%)**	**Employed (%)**	**Moved Last Year (%)**	**2 Children (%)**	**3 or More Children (%)**
0	60,000	75	2	7	61	32	95	37	2	–
(1,802)	(37,000)	(43)	(14)	(26)	(49)	(47)	(22)	(48)	(15)	(–)
1	56,000	75	2	4	61	33	96	33	8	–
(3,109)	(37,000)	(43)	(14)	(18)	(49)	(47)	(20)	(47)	(27)	(–)
2	58,000	75	3	6	61	32	96	31	28	–
(3,968)	(38,000)	(44)	(16)	(24)	(49)	(47)	(20)	(46)	(45)	(–)
3	59,000	80	2	6	62	32	96	27	50	4
(4,477)	(38,000)	(40)	(12)	(23)	(48)	(47)	(19)	(44)	(50)	(19)
4	60,000	81	2	4	62	33	96	25	57	9
(4,546)	(39,000)	(39)	(14)	(19)	(48)	(47)	(20)	(43)	(49)	(29)
5	59,000	82	3	4	64	30	96	24	56	17
(5,015)	(39,000)	(38)	(17)	(19)	(48)	(46)	(20)	(43)	(50)	(37)
6	62,000	83	3	4	64	32	96	21	56	23
(4,516)	(42,000)	(38)	(16)	(20)	(48)	(46)	(19)	(40)	(50)	(42)
7	61,000	84	2	5	64	29	96	19	53	27
(5,078)	(42,000)	(37)	(14)	(21)	(48)	(46)	(20)	(39)	(50)	(45)
8	63,000	85	2	4	66	28	96	16	50	33
(5,258)	(42,000)	(36)	(14)	(21)	(47)	(45)	(20)	(37)	(50)	(47)
9	63,000	85	2	5	66	28	96	16	47	37
(5,344)	(44,000)	(36)	(14)	(22)	(47)	(45)	(21)	(37)	(50)	(48)
10	65,000	86	2	5	67	27	96	16	46	38
(6,098)	(46,000)	(35)	(15)	(21)	(47)	(44)	(21)	(37)	(50)	(48)
11	63,000	86	3	4	68	24	96	14	43	41
(6,127)	(42,000)	(35)	(16)	(20)	(47)	(43)	(19)	(35)	(50)	(49)
12	65,000	87	2	4	69	23	95	15	41	42
(6,386)	(46,000)	(33)	(14)	(20)	(46)	(42)	(21)	(36)	(49)	(49)
13	63,000	86	2	5	69	23	96	15	40	44
(7,285)	(43,000)	(34)	(15)	(21)	(46)	(42)	(19)	(36)	(49)	(50)
14	65,000	89	2	4	69	25	95	11	41	40
(7,018)	(43,000)	(31)	(14)	(21)	(46)	(43)	(21)	(31)	(49)	(49)
15	65,000	89	2	4	69	24	97	11	37	42
(8,179)	(44,000)	(31)	(15)	(21)	(46)	(43)	(18)	(31)	(48)	(49)
16	68,000	91	1	4	69	25	97	10	38	39
(8,719)	(44,000)	(29)	(12)	(19)	(46)	(43)	(18)	(30)	(49)	(49)
17	71,000	92	2	3	69	24	96	8	36	38
(9,853)	(47,000)	(26)	(12)	(17)	(46)	(43)	(19)	(27)	(48)	(49)
18	71,000	92	1	3	67	26	97	6	38	37
(10,481)	(43,000)	(27)	(10)	(18)	(47)	(44)	(17)	(24)	(49)	(48)
19	70,000	92	2	4	73	19	96	7	38	36
(6,039)	(39,000)	(27)	(14)	(21)	(45)	(39)	(20)	(26)	(49)	(48)

Notes: Standard deviations are reported in parenthesis except in the first column where the number of observations is reported. Groups are defined by the age of the oldest child.

Table 3

Summary Statistics
Groups of non-parents (by the age of the head of the household)

Group (obs)	Household Income ($)	Home-owner (%)	Black (%)	Other Race (%)	High School Degree (%)	College Degree (%)	Employed (%)	Moved Last Year (%)
31	60,000	67	3	6	55	42	51	35
(1,319)	(35,000)	(47)	(16)	(23)	(50)	(49)	(50)	(48)
32	62,000	70	2	5	51	45	50	30
(1,161)	(39,000)	(46)	(15)	(22)	(50)	(50)	(50)	(46)
33	60,000	73	2	5	58	36	50	23
(1,103)	(37,000)	(45)	(15)	(21)	(49)	(48)	(50)	(42)
34	61,000	73	4	4	57	37	44	24
(1,158)	(40,000)	(44)	(19)	(19)	(49)	(48)	(50)	(43)
35	61,000	73	2	3	62	35	42	23
(1,083)	(41,000)	(44)	(15)	(17)	(49)	(48)	(49)	(42)
36	59,000	74	3	2	68	26	43	20
(1,098)	(38,000)	(44)	(18)	(15)	(47)	(44)	(49)	(40)
37	57,000	73	2	4	64	30	44	21
(1,031)	(39,000)	(44)	(12)	(19)	(48)	(46)	(50)	(40)
38	60,000	76	3	4	65	31	40	21
(1,048)	(42,000)	(43)	(16)	(19)	(48)	(46)	(49)	(40)
39	58,000	79	3	3	67	26	43	19
(1,073)	(37,000)	(41)	(17)	(18)	(47)	(44)	(49)	(40)
40	57,000	79	2	3	67	24	42	18
(1,107)	(43,000)	(41)	(13)	(18)	(47)	(43)	(49)	(39)
41	58,000	79	2	3	71	23	44	14
(1,176)	(38,000)	(41)	(15)	(17)	(46)	(42)	(50)	(35)
42	60,000	80	2	3	70	22	47	16
(1,211)	(44,000)	(40)	(15)	(17)	(46)	(42)	(50)	(37)
43	59,000	78	1	4	72	22	44	15
(1,119)	(42,000)	(41)	(12)	(19)	(45)	(41)	(50)	(36)
44	59,000	82	3	2	71	21	51	13
(1,255)	(42,000)	(38)	(16)	(16)	(45)	(41)	(50)	(34)
45	59,000	83	2	3	71	23	54	11
(1,280)	(42,000)	(38)	(14)	(16)	(45)	(42)	(50)	(31)
46	60,000	82	2	4	72	23	53	13
(1,351)	(44,000)	(38)	(15)	(19)	(45)	(42)	(50)	(34)
47	62,000	86	1	3	69	23	56	12
(1,484)	(43,000)	(34)	(12)	(16)	(46)	(42)	(50)	(32)
48	64,000	87	1	3	68	26	60	11
(1,542)	(44,000)	(33)	(11)	(17)	(47)	(44)	(49)	(31)
49	64,000	89	1	2	68	26	66	11
(1,776)	(44,000)	(32)	(10)	(15)	(47)	(44)	(48)	(31)
50	68,000	88	1	2	64	30	66	10
(1,806)	(49,000)	(32)	(11)	(13)	(48)	(46)	(47)	(30)

Notes: Standard deviations are reported in parenthesis except in the first column where the number of observations is reported. Groups are defined by the age of the head of household.

Analogously, Table 3 shows the summary statistics of the observed characteristics of the groups of non-parents used in this analysis. The family income for groups of non-parents is more similar across groups in comparison to parents, but the difference between parents and non-parents of the same age is small. The proportion of non-parents who are employed is around 55%, in contrast to the 95% found in parents. The proportion of non-parents who are homeowners is also a little lower than that of parents, with older non-parents having a higher likelihood of being homeowners. Similarly to parents, the moving rate of non-parents depends substantially on the age group, with older non-parents overall moving much less than younger ones.

Tables 2 and 3 show that there is substantial variation across groups in several important characteristics, which suggests that there is enough heterogeneity of preferences for public school quality, rents and other amenities. This cross-sectional heterogeneity is crucial to the implementation of the approach developed in this paper, since I exploit differential variation across groups with respect to the preferences of observed and unobserved amenities.

4 Identification Strategy

This section explains the proposed strategy to identify the parental valuation for school quality per year, per child at each grade level in three steps, as illustrated in Figure 1.

Basic Setup

The setup of the model is similar to the standard dynamic models in the literature (Arcidiacono and Ellickson (2011) provides a recent survey of this literature). Each year, families choose a neighborhood with full information of the current year and only partial information of the future. Families incur moving costs if they change neighborhoods, and they anticipate they will have to incur moving costs in the future if they decide to move.

Specifically, there are J neighborhoods and $C+1$ groups of families, with each group c containing $n_{c,t}$ families in year t. In the beginning of each year t, family i of group c observes the state variable $\boldsymbol{S}_{i,c,t}$ and decides the neighborhood they will live in t by maximizing the choice-specific value function $V_j(\boldsymbol{S}_{i,c,t})$ defined as

$$V_j(\boldsymbol{S}_{i,c,t}) := U_j(\boldsymbol{S}_{i,c,t}) + \beta \mathbb{E}\left[\boldsymbol{V}(\boldsymbol{S}_{i,c,t+1}) \middle| \boldsymbol{S}_{i,c,t}\,, d_{i,c,t} = j\right], \tag{8}$$

where $d_{i,c,t} = j$ if individual i chooses neighborhood j in period t. $U_j(\boldsymbol{S}_{i,c,t})$ is the family's flow (i.e., per year) indirect utility, β is the future discount (assumed known to the researcher), and $\boldsymbol{V}(\boldsymbol{S}_{i,c,t+1}) := \max_r V_r(\boldsymbol{S}_{i,c,t+1})$.

I start from standard assumptions in the literature (e.g., Rust (1987)):

Dynamic Assumptions. ***(Rust (1987))*** *Let* $\boldsymbol{S}_{i,c,t} := (\boldsymbol{W}_{i,c,t}, \boldsymbol{\epsilon}_{i,c,t})$.

1. ***Additive Separability:*** $U_j(\boldsymbol{W}_{i,c,t}, \boldsymbol{\epsilon}_{i,c,t}) := u_j(\boldsymbol{W}_{i,c,t}) + \varepsilon_{i,c,t,j}$, *where* $\varepsilon_{i,c,t,j}$ *is i.i.d. as Extreme Value 1 type (EV1).*
2. ***Conditional Independence:*** $\boldsymbol{P}(\boldsymbol{W}_{i,c,t+1}|\boldsymbol{W}_{i,c,t}, \boldsymbol{\epsilon}_{i,c,t}, d_{i,c,t} = j) = \boldsymbol{P}(\boldsymbol{W}_{i,c,t+1}|\boldsymbol{W}_{i,c,t}, d_{i,c,t} = j)$, *where* $\boldsymbol{P}(A|B)$ *denotes the conditional probability of A given B.*
3. ***Finite Support:*** *The support of* $\boldsymbol{W}_{i,c,t}$ *is discrete and finite:* $\boldsymbol{W}_{i,c,t} \in \mathbb{W}_{i,c,t} = \{\boldsymbol{w}^1_{i,c,t}, \boldsymbol{w}^2_{i,c,t}, ..., \boldsymbol{w}^{|\mathbb{T}|}_{i,c,t}\}$, $|\mathbb{T}| < \infty$.

These three assumptions are standard in most papers estimating dynamic discrete choice models (Arcidiacono and Ellickson (2011)). In words, the state variable is decomposed in two components: one component, $\boldsymbol{\epsilon}_{i,c,t}$, is idiosyncratic and has a trivial transition over time, while the other component, $\boldsymbol{W}_{i,c,t}$, has a more complex transition over time. Importantly, $\boldsymbol{\epsilon}_{i,c,t}$ is assumed irrelevant to predict $\boldsymbol{W}_{i,c,t+1}$, given $\boldsymbol{W}_{i,c,t}$ and $d_{i,c,t}$.

As discussed in Rust (1987), these assumptions imply that $V_j(\boldsymbol{S}_{i,c,t})$ can be written as

$$V_j(\boldsymbol{S}_{i,c,t}) = v_j(\boldsymbol{W}_{i,c,t}) + \varepsilon_{i,c,t,j}, \tag{9}$$

with

$$\begin{aligned} v_j(\boldsymbol{W}_{i,c,t}) :=& u_j(\boldsymbol{W}_{i,c,t}) + \\ &+\beta \sum_{\boldsymbol{w}_{i,c,t+1} \in \mathbb{W}_{i,c,t+1}} \Big(\gamma + \log \sum_{r=1}^{J} \exp(v_r(\boldsymbol{w}_{i,c,t+1}))\Big) \boldsymbol{P}(\boldsymbol{W}_{i,c,t+1} = \boldsymbol{w}_{i,c,t+1} | \boldsymbol{W}_{i,c,t}, d_{i,c,t} = j), \end{aligned} \tag{10}$$

where $\gamma \approx 0.577$ is the mean of the Extreme Value distribution. Next, I describe in detail the three sequential steps to estimate the valuation of school quality per year, per child at each grade level.

4.1 Step 1: Obtaining the Per-Year Valuation of the Neighborhood

Obtaining the per-year valuation of each neighborhood requires estimating a dynamic choice model. My estimation method departs from standard approaches in order to circumvent two major restrictions in the data. First, I do not directly observe $d_{i,c,t-1}$ in the data. I only observe $d_{i,c,t}$ and whether i just moved from $t-1$ to t. If the family does not move from $t-1$ to t, then $d_{i,c,t-1} = d_{i,c,t}$; otherwise, I have no information about $d_{i,c,t-1}$, as the family may have moved within neighborhood.[18] I circumvent this problem by assuming that families first decide whether to move (thus incurring a moving cost), and then, conditional on moving, they decide their destination. The moving cost parameter is assumed the same no matter the origin or destination neighborhood. This allows me to identify the moving cost parameter only off families who do not move from $t-1$ to t. The second restriction is that the data is cross-sectional. As discussed below, this restricts the types of assumptions I can make with respect to families' expectations of their own valuation of each neighborhood in the future.

To summarize my approach: first, I estimate the unobserved components of the state variable $\boldsymbol{W}_{i,c,t}$ directly, for all c; next, I plug in these estimates to calculate the conditional transition probabilities $\boldsymbol{P}(\boldsymbol{W}_{i,c,t+1} = \boldsymbol{W} | \boldsymbol{W}_{i,c,t}\,, d_{i,c,t} = j)$ under a synthetic cohort assumption. Finally, I calculate $u_j(\boldsymbol{W}_{i,c,t})$ via equation (10). I describe this approach in more detail below.

The choice-specific value functions for year t are written as

$$v_j(\boldsymbol{W}_{i,c,t}) := \Delta_{c,t,j} + \text{Move}_{i,c,t,j}.\Phi_{c,t} \tag{11}$$

where $\boldsymbol{W}_{i,c,t} := \left(\{\Delta_{c,t,r}, \text{Move}_{i,c,t,j}\}_{r=1}^{J}, \Phi_{c,t}\right)$. $\Delta_{c,t,j}$ can be interpreted as the mean cumulative value of neighborhood j among all families of group c as of period t. $\Phi_{c,t}$ is the moving cost parameter. $\text{Move}_{i,c,t,j} := \Big(\text{Stayer}_{i,c,t}.1_{\{j \neq d_{i,c,t-1}\}} + (1 - \text{Stayer}_{i,c,t})\Big)$, where $\text{Stayer}_{i,c,t}$ is an indicator for whether family i stayed in the same house from $t-1$ to t, and $1_{\{j \neq d_{i,c,t-1}\}}$ is an indicator variable for whether the expression in brackets is true. Intuitively, the moving cost parameter $\Phi_{c,t}$ will be included in the utility of a family of group c who moved for all j (so $\Phi_{c,t}$ is not identified off movers, as it is a constant across all neighborhood options). In contrast, $\Phi_{c,t}$ will be included in the utility of a family of group c who stayed in the same house from $t-1$ to t only if $j \neq d_{i,c,t-1}$. For stayer families, we observe $d_{i,c,t-1}$ because in that case $d_{i,c,t-1} = d_{i,c,t}$.

This specification reflects a departure from standard methods (e.g., Rust (1987); Hotz and Miller (1993)). The key difference is what represents a "state variable" in both approaches. In standard approaches, state variables are defined to be primitives of the model,

[18] Decennial Census contains data on where the family was located 5 years ago, allowing me to verify that indeed some families move within neighborhood.

such as the observed amenities of the neighborhood. Thus, choice specific value functions are written as a complicated function of these state variables.[19] An alternative approach followed here is to specify state variables not as primitives. This way, choice specific value functions can be written as a simple function of these more complex state variables. This allows for the estimation of a dynamic model using static discrete choice techniques (Berry et al. (1995)). A similar departure from standard methods is exploited in Bayer et al. (2016) (see Remark 2).

$\Delta_{c,t,j}$, the mean cumulative value of neighborhood j from the perspective of group c as of period t, is a particularly complex "reduced-form" state variable. It encompasses much of the variables that are typically defined as state variables in standard approaches. For instance, observed and unobserved amenities varying at the neighborhood level (including the ones valued differently across groups), or even amenities varying at the neighborhood-group level, are included in $\Delta_{c,t,j}$. Importantly, the continuation value from $t+1$ onwards of neighborhood j from the perspective of group c[20] is included in $\Delta_{c,t,j}$ too.

Using this approach instead of a standard approach involves a practical trade-off. On the one hand, in this approach it is difficult to model heterogeneity in $v_j(\boldsymbol{W}_{i,c,t})$ across families within group c. Any heterogeneity term should be a complicated function of all neighborhoods (not only neighborhood j) because of the continuation value. For instance, in order to allow for families to have heterogeneous preferences for a given observed amenity within group c, $v_j(\boldsymbol{W}_{i,c,t})$ should be written not only as a function of the level of that amenity in neighborhood j but also as a function of the levels of that amenity for all neighborhoods $k \neq j$.[21] On the other hand, this approach allows the econometrician to avoid making some assumptions about expectations that are difficult to test in the data.[22] The

[19]As pointed out by Arcidiacono and Ellickson (2011) (p. 369), "the main difference between static and dynamic discrete choice is that, in the former, the payoffs will generally be expressed as linear functions of the state variables (because they are primitives), whereas in the latter, the expressions are more complicated (...) How much more complicated will depend on the number of choices, the number of possible states and the distribution of the structural errors."

[20]This is the expected valuation of each neighborhood k next year from the perspective of a family of group c, weighted by the corresponding expectation to live next year in k given that the family lives in neighborhood j in year t.

[21]In the context of this paper, I allay concerns about this issue in three ways. First, I define groups along a dimension that should reflect much of the heterogeneity in my context. For instance, I define the groups of interest ($c = 6, ..., 18$) to be families with children attending public schools at different grades. Second, I assume that the moving cost parameter, the only parameter of the state variable that I allow to be heterogeneous within groups, is the same irrespective of the neighborhood of origin or destination. Third, I re-estimate the first step with an additional term in equation (11) reflecting the interaction of observed amenities in j with family observed characteristics, and find virtually no change in the main school quality valuation estimates obtained in the second and third steps. This is not surprising, as the method to handle the endogeneity problem in the second step is helpful to mitigate any bias from the first step, as I discuss in Remark 4.

[22]Indeed, these are assumptions about *expectations* as of period t. These do not necessarily relate to how the level

approach makes assumptions about how families of group c expect $\Delta_{c,t+1,j}$ to be as of period t. As discussed above, $\Delta_{c,t+1,j}$ is an implicit function of the levels of all amenities of all neighborhoods valued at the corresponding preferences in t', with $t' = t+1,...,T$. By making assumptions about how families expect this "reduced-form" state variable to evolve from t to $t+1$, the method avoids spelling out this implicit function, thus avoiding making specific assumptions about expectations. Specifically, it avoids assuming the value of the time horizon T; similarly, it avoids assuming families' specific expected transitions on the level of and preference for each amenity (observed or unobserved to the econometrician) from t to t', with $t' = t+1,...,T$.

In order to estimate $u_j(\boldsymbol{W}_{i,c,t})$ with cross-sectional data from year t only, I make the following assumption about how families expect their state variables to evolve from t to $t+1$:

Synthetic Cohort Assumption. *Let* $\boldsymbol{W}_{i,c,t+1} := \left(\{\Delta_{c,t+1,r}, Move_{i,c,t,r}\}_{r=1}^{J}, \Phi_{c,t+1}\right)$. *Families have the following expectation in year* $t = 2000$*:*

If c are parents $(c = 0,...,18)$:

$$\boldsymbol{P}\left(\underbrace{\left(\{\Delta_{c,t+1,r}, Move_{i,c,t+1,r}\}_{r=1}^{J}, \Phi_{c,t+1}\right)}_{\boldsymbol{W}_{i,c,t+1}} = \underbrace{\left(\{\Delta_{c+1,t,r}, 1_{\{r\neq j\}}\}_{r=1}^{J}, \Phi_{c+1,t}\right)}_{\boldsymbol{W}} | \boldsymbol{W}_{i,c,t}\, , d_{i,c,t} = j\right) = 1$$

If c are non-parents $(c = 31,...,49)$:

$$\boldsymbol{P}\left(\underbrace{\left(\{\Delta_{c,t+1,r}, Move_{i,c,t+1,r}\}_{r=1}^{J}, \Phi_{c,t+1}\right)}_{\boldsymbol{W}_{i,c,t+1}} = \underbrace{\left(\{\Delta_{0,t,r}, 1_{\{r\neq j\}}\}_{r=1}^{J}, \Phi_{0,t}\right)}_{\boldsymbol{W}^1} | \boldsymbol{W}_{i,c,t}\, , d_{i,c,t} = j\right) = \Pi_{c,t}$$

$$\boldsymbol{P}\left(\underbrace{\left(\{\Delta_{c,t+1,r}, Move_{i,c,t+1,r}\}_{r=1}^{J}, \Phi_{c,t+1}\right)}_{\boldsymbol{W}_{i,c,t+1}} = \underbrace{\left(\{\Delta_{c+1,t,r}, 1_{\{r\neq j\}}\}_{r=1}^{J}, \Phi_{c+1,t}\right)}_{\boldsymbol{W}^2} | \boldsymbol{W}_{i,c,t}\, , d_{i,c,t} = j\right) = 1 - \Pi_{c,t}$$

where $\Pi_{c,t} := n^0_{c+1,t}/(n^0_{c+1,t} + n_{c+1,t})$*,* $n_{c+1,t}$ *is the number of families of group* $c+1$*,* $\Delta_{0,t,r}$ *and* $\Phi_{0,t}$ *refer to the parameters from the group of parents whose oldest child is zero years of age (i.e., a newborn), and* $n^0_{c+1,t}$ *is the number of families whose head is* $c+1$ *years of age and whose oldest child is a newborn.*

This assumption states that parents expect to have the same mean cumulative valuation of each neighborhood (and moving costs) next year as parents one cohort ahead have

of amenities and their preferences *actually* evolve over time. Thus, there are some limits on how in practice one could weaken these assumptions with access to longitudinal data, since what is actually needed are data eliciting families' expectations. See Remark 2.

this year. The assumption is similar for non-parents, but the transition is more complicated, since they may become parents next year. There is a likelihood Π_{c+1} of becoming a parent (of a newborn), and a likelihood $1-\Pi_{c+1}$ of keep being a non-parent. By assumption, families perceive these probabilities to be the ones observed in period t.

This assumption is likely to be a good approximation in the context of Minnesota around the year 2000. The majority of schools in the state experienced only small changes in enrollment and demographic composition in the years leading up to 2000.[23] In Remark 4 and Appendix B, I argue that small violations of this assumption should not affect the estimates of interest.

Let $\delta_{c,t,j}$ denote the average of $u_j(\boldsymbol{W}_{i,c,t})$ across all families i belonging to group c: $\delta_{c,t,j} := \frac{1}{n_{c,t}}\sum_{i\in I_c}[u_j(\boldsymbol{W}_{i,c,t})]$, where I_c denotes the set of families in group c. Then the synthetic cohort assumption together with equation (10) imply:

If c are parents $(c=0,...,18)$:

$$\delta_{c,t,j} := \frac{1}{n_{c,t}}\sum_{i\in I_c}\Big[\underbrace{\Delta_{c,t,j}+\text{Move}_{i,c,t,j}.\Phi_{c,t}}_{v_j(\boldsymbol{W}_{i,c,t})} - \beta\Big(\gamma+\log\sum_{r=1}^{J}\exp\big(\underbrace{\Delta_{c+1,t,r}+1_{\{r\neq j\}}.\Phi_{c+1,t}}_{v_r(\boldsymbol{W}_{i,c,t+1})}\big)\Big)\Big]. \tag{12}$$

If c are non-parents $(c=31,...,49)$:

$$\delta_{c,t,j} := \frac{1}{n_{c,t}}\sum_{i\in I_c}\Big[\underbrace{\Delta_{c,t,j}+\text{Move}_{i,c,t,j}.\Phi_{c,t}}_{v_j(\boldsymbol{W}_{i,c,t})} -$$

$$-\beta\Big(\gamma+\Pi_{c,t}.\log\sum_{r=1}^{J}\exp\Big(\Big(\underbrace{\Delta_{0,t,r}+1_{\{r\neq j\}}.\Phi_{0,t}}_{v_j(\boldsymbol{W}^1_{i,c,t+1})}\Big)\Big)+(1-\Pi_{c,t}).\log\sum_{r=1}^{J}\exp\Big(\underbrace{\Delta_{c+1,t,r}+1_{\{r\neq j\}}.\Phi_{c+1,t}}_{v_j(\boldsymbol{W}^2_{i,c,t+1})}\Big)\Big)\Big]. \tag{13}$$

As shown in equations (12) and (13), $\delta_{c,t,j}$ is written only as a function of $\mathbb{D}_t := \Big(\Big\{\{d_{i,c,t},\{\text{Move}_{i,c,t,r}\}_{r=1}^{J}\}_{i\in I_{c,t}}, n^0_{c,t}, n_{c,t}\Big\}_{c=0}^{C}\Big)$ and $\Theta := \Big(\{\{\Delta_{c,t,r}\}_{r=1}^{J},\Phi_{c,t}\}_{c=0}^{C},\beta\Big)$. $\mathbb{D}_t$ comes directly from the 2000 Decennial Census data, and Θ can be estimated with $\mathbb{D}_t$ as in Berry et al. (1995), with β assumed equal to 0.95. Appendix A describes the details of the estimation process.

Remark 2. *It is useful to see how these assumptions could be relaxed if we had access to longitudinal data instead. This helps relate the dynamic estimation approach used here*

[23]The main valuation estimates barely change when the analysis is restricted to only schools where this assumption is most likely to hold, namely those below the median in terms of changes in enrollment or demographic composition.

with the approach developed in Bayer et al. (2016). With longitudinal data, I could model the expected transition of the choice specific value functions more flexibly. For instance, Bayer et al. (2016) estimate the expected choice specific value function in $t+1$ with a rational expectations assumption implied by their equation (25) (p. 915), reproduced here for $t+1$:

$$v^{\tau}_{j,t+1} = \rho^{\tau}_{0,j} + \sum_{l=0}^{L-1} \rho^{\tau}_{1,l} v^{\tau}_{j,t-l} + \sum_{l=0}^{L-1} X'_{j,t-l}\rho^{\tau}_{2,l} + \rho^{\tau}_{3,j}(t+1) + \omega^{\tau}_{j,t+1}$$

$v^{\tau}_{j,t+1}$ in their context is analogous to $\Delta_{c,t+1,j}$ in the context of this paper, where τ is analogous to c (although the definitions of groups are completely different).[24] *Individuals of type τ by assumption project the true value of $v^{\tau}_{j,t+1}$ onto $(v^{\tau}_{j,t-l}, X'_{j,t-l})$ for $l = 0,...,L-1$ to estimate their predicted value of $v^{\tau}_{j,t+1}$ as of period t. Thus, individuals of type τ use both current (t) and past ($t-1,...,t-(L-1)$) information to make this prediction, where "information" in their context corresponds to both mean cumulative valuations of the same type and neighborhood observables. In contrast, cross-sectional data prevents me from assuming that individuals project the true value of $\Delta_{c,t+1,j}$ onto their information set (as the dependent variable in the equation above, $\Delta_{c,t+1,j}$, cannot be estimated with data from t). Moreover, cross-sectional data restricts me to assume families use only current information to predict $\Delta_{c,t+1,j}$ (as $\Delta_{c,t-l,j}$ for $l = 1,...,L-1$ cannot be estimated with data from t).*

4.2 Step 2: Obtaining the Per Year Valuation of Public School Quality

In what follows, the year index $t = 2000$ is suppressed for simplicity. $\hat{\delta}_{c,j}$, the estimated per year valuation of neighborhood j for group c obtained from step 1, is written linearly as a function of the neighborhood amenities:

$$\hat{\delta}_{c,j} = SQ_j\theta_c + P_j\phi_c + \xi_{c,j}, \tag{14}$$

where SQ_j and P_j are respectively measures of the level of public school quality and the price of neighborhood j, θ_c and ϕ_c are preference parameters, and $\xi_{c,j}$ is an unobserved term that varies at the group-neighborhood level. ξ includes not only unobserved amenities of neighborhood j. It also includes any error due to violations of the dynamic and the

[24] Bayer et al. (2016) focus their analysis on homeowners, so they also rightly consider important concerns related to homeowners' wealth that are outside of the scope of this paper.

synthetic cohort assumptions made in the previous section. Thus, the discussion below allows for $\hat{\delta}_{c,j}$ to be a potentially biased estimator of $\delta_{c,j}$ (see Remark 4).

Identification of θ_c and ϕ_c in (14) is difficult to obtain because $\xi_{c,j}$ is generally correlated to SQ_j and to P_j. To address this endogeneity problem, I use a novel approach. The term $\xi_{c,j}$ is decomposed into a term including a finite number of unobserved neighborhood-level amenities, Q_j, plus a random term $\mu_{c,j}$. The mean flow utility can be rewritten as

$$\hat{\delta}_{c,j} = SQ_j\theta_c + P_j\phi_c + \overbrace{Q_j\lambda_c + \mu_{c,j}}^{\xi_{c,j}}, \tag{15}$$

where Q_j is a vector of fixed size R.

The next assumption states that all endogeneity of equation (15) can be captured by R such unobserved amenities, with allowance for different valuation weights across groups. Intuitively, Assumption 1 weakens as R becomes larger.

Assumption 1. ***Selection on Unobservables***

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

where $\mu_{c,j}$ is defined in equation (15).

If Q was observed, then the parameters of equation (15) would be trivially identified. Instead, I will proxy for Q. The procedure is as follows. First the groups are divided in three subsets. The δs of the groups in the first subset will be used as proxy for Q in the equations of the second subset. This procedure will solve the original endogeneity problem, but it will introduce a new source of endogeneity that is easily solved under the same assumptions using the δs of the groups in the third subset as IVs for the δs of the groups in the first subset.

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

$$\hat{\boldsymbol{\delta}}_{j(1)} = SQ_j\boldsymbol{\theta}_{(1)} + P_j\boldsymbol{\phi}_{(1)} + Q_j\boldsymbol{\lambda}_{(1)} + \boldsymbol{\mu}_{j(1)}, \tag{16}$$

$$\hat{\boldsymbol{\delta}}_{j(2)} = SQ_j\boldsymbol{\theta}_{(2)} + P_j\boldsymbol{\phi}_{(2)} + Q_j\boldsymbol{\lambda}_{(2)} + \boldsymbol{\mu}_{j(2)}, \tag{17}$$

$$\hat{\boldsymbol{\delta}}_{j(3)} = SQ_j\boldsymbol{\theta}_{(3)} + P_j\boldsymbol{\phi}_{(3)} + Q_j\boldsymbol{\lambda}_{(3)} + \boldsymbol{\mu}_{j(3)}, \qquad \text{for} \quad j = 1, \ldots, J \tag{18}$$

where SQ_j and P_j are scalars; Q_j is a $1 \times R$ vector; $\boldsymbol{\delta}_{j(s)}$, $\boldsymbol{\mu}_{j(s)}$, $\boldsymbol{\theta}_{(s)}$ and $\boldsymbol{\phi}_{(s)}$ are each $1 \times R_s$ vectors; and $\boldsymbol{\lambda}_{(s)}$ is a $R \times R_s$ matrix, $s = 1, 2, 3$.

Solving equation (16) for Q and substituting it into equation (17)[25]:

$$\hat{\boldsymbol{\delta}}_{j(2)} = SQ_j \overbrace{\left(\boldsymbol{\theta}_{(2)} - \boldsymbol{\theta}_{(1)}\tilde{\boldsymbol{\lambda}}_{(2)}\right)}^{\tilde{\boldsymbol{\theta}}_{(2)}} + P_j \overbrace{\left(\boldsymbol{\phi}_{(2)} - \boldsymbol{\phi}_{(1)}\tilde{\boldsymbol{\lambda}}_{(2)}\right)}^{\tilde{\boldsymbol{\phi}}_{(2)}} + \hat{\boldsymbol{\delta}}_{j(1)}\tilde{\boldsymbol{\lambda}}_{(2)} + \overbrace{\boldsymbol{\mu}_{j(2)} - \boldsymbol{\mu}_{j(1)}\tilde{\boldsymbol{\lambda}}_{(2)}}^{\tilde{\boldsymbol{\mu}}_{j(2)}},\ j = 1, ..., J \tag{19}$$

where $\tilde{\boldsymbol{\lambda}}_{(2)} \equiv \boldsymbol{\lambda}'_{(1)} \left(\boldsymbol{\lambda}_{(1)}\boldsymbol{\lambda}'_{(1)}\right)^{-1} \boldsymbol{\lambda}_{(2)}$ is an $R_1 \times R_2$ matrix measuring the relative preferences of the unobserved amenities for groups in subsets 1 and 2. In equation (19), the mean utilities of the groups in G_2 are written as a function of the observed amenities and the mean utilities of the groups in G_1, and no longer as a function of Q.

Note that, since $\tilde{\boldsymbol{\mu}}_{j(2)}$ is correlated with $\boldsymbol{\delta}_{j(1)}$ because of equation (16), equation (19) does not identify $\tilde{\boldsymbol{\theta}}_{(2)}$, $\tilde{\boldsymbol{\phi}}_{(2)}$ and $\tilde{\boldsymbol{\lambda}}_{(2)}$ directly via OLS. However, as pointed out by Chamberlain (1977), under the previous assumptions $\boldsymbol{\delta}_{j(3)}$ can be used as IV for $\boldsymbol{\delta}_{j(1)}$, hence $\tilde{\boldsymbol{\theta}}_{(2)}$, $\tilde{\boldsymbol{\phi}}_{(2)}$ and $\tilde{\boldsymbol{\lambda}}_{(2)}$ are nonetheless identified via 2SLS.

Still, the parameters of interest θ_c and ϕ_c for $c \in G_2$ are not directly identified from $\tilde{\boldsymbol{\theta}}_{(2)}$, $\tilde{\boldsymbol{\phi}}_{(2)}$ and $\tilde{\boldsymbol{\lambda}}_{(2)}$, hence additional assumptions need to be made. According to equation (19), the parameters of interest are implicitly given by the two systems of equations:

$$\tilde{\theta}_c = \theta_c - \boldsymbol{\theta}_{(1)}\tilde{\lambda}_c, \qquad c \in G_2, \tag{20}$$
$$\tilde{\phi}_c = \phi_c - \boldsymbol{\phi}_{(1)}\tilde{\lambda}_c, \qquad c \in G_2. \tag{21}$$

The systems of equations (20) and (21) have each R_2 equations and $R_1 + R_2$ unknowns (θ_c and ϕ_c for $c \in G_1 \cup G_2$). If at least R_1 restrictions in the parameters of each of these systems are made, then θ_c and ϕ_c are identified for each c in G_2.

In Section 5, this method is implemented with $G_2 = \{6, ..., 18\}$ since the oldest child of the families of these groups are observed to be attending respectively kindergarten through grade 12 in a public school. The rest of the groups are divided between G_1 and G_3 depending on the specification, provided that $R_3 \geq R_1$ holds. Assumption 2 makes exclusion restrictions that guarantee the identification of θ_c and ϕ_c, $c \in G_2$.

Assumption 2. ***Exclusion Restrictions:***

1. *$\theta_c = 0$, $\quad c \in G_1$,*
2. *$\phi_c = \phi_6 \quad c \in G_1$.*

[25] $R = rank(\boldsymbol{\lambda}_{(1)})$, as discussed in Remark 3.

To identify θ_c, $c \in G_2$, I assume $\theta_c = 0$ for the groups used as proxy. Groups $c = 31, ..., 49$ are non-parents, so they should not value public school quality in the flow utility sense, because they do not enjoy school amenities in that year. Moreover, groups $c = 0, ..., 5$ are parents but do not have a child attending public school this year. So, following the same logic, they should also not value public school quality in the flow utility sense. This implies that any group except groups in $G_2 = \{6, ..., 18\}$ is an appropriate candidate to be included in G_1.

To identify ϕ_c, $c \in G_2$, I assume that groups used as proxies as well as group 6 have on average the same flow value for the amenity P. This assumption can be relaxed if needed, but in practice doing so does not change the main results (see Section 5). This is not surprising, given how income tends to change little among those groups in Tables 2 and 3.

Under Assumptions 1 and 2, θ_c and ϕ_c for $c \in G_2$ are identified. The average per-year valuation of school quality for families in group c, also known as the average marginal willingness to pay (MWTP) for school quality, represents how much these families are willing to trade-off school quality for rents:

$$MWTP_c^{SQ} := -\frac{\theta_c}{\phi_c}. \tag{22}$$

Remark 3. ***Intuition:*** *Because $R \leq R_1$, the more groups are included as proxy, the larger is the dimension of Q that can be controlled for, and thus the weaker is Assumption 1. Also, the heterogeneity of groups included as proxy is important. Technically, $\hat{\boldsymbol{\delta}}_{j(1)}$ needs to span the space generated by the columns of Q so that all endogeneity is controlled for. This requires different groups to value Q differently. For instance, if the average income of neighbors and the existence of a children's park are both important unobservables, then using two non-parental groups as proxy will not help to span the space of the children's park unobservable, as non-parents do not enjoy it that year. However, using as proxy one non-parental group and one group of parents without school-age children will span the space of both unobservables, as this latter group does enjoy children's park in that year. Heuristically, when a new group is included in G_1 and the adjusted R^2 rises significantly, it suggests that one more dimension of Q is being spanned.*

Remark 4. ***What if $\hat{\delta}_{c,j}$ is Biased?*** *Many assumptions were made to identify $\delta_{c,j}$ in step 1, so it is important to discuss the implications of violations of those assumptions. Any violation of the assumptions made in step 1 can bias the main estimates only to the extent that $\hat{\delta}_{c,j}$ is biased. Let $\eta_{c,j} := \hat{\delta}_{c,j} - \delta_{c,j}$ denote this bias. $\eta_{c,j}$ should be a component of $\xi_{c,j}$ defined in equation* (14). *Thus, violations of any assumption made in step 1 constitute* *one more potential source of endogeneity that the method described in this section aims to control for. In particular, Assumptions 1 and 2 are sufficient to guarantee that* $MWTP_c^{SQ}$, $c \in G_2$ *are identified, even if the assumptions made in step 1 do not hold. For instance, consider a situation where group c expects neighborhood j to be gentrifying (a violation of the synthetic cohort assumption) as the only potential source of endogeneity. Then* $\delta_{c,j}$ *would be overestimated in the first step, since* $\mathbb{E}(\Delta_{c,t+1,j}) > \Delta_{c+1,t,j}, \Delta_{0,t,j}$ *(this can be seen in equations* (12) *and* (13)*). If* $\eta_{c,j}$ *can be absorbed by* $Q_j.\lambda_c$*, then it does not bias our estimates of* $MWTP_c^{SQ}$, $c \in G_2$*. In Appendix B, I present a Monte Carlo study that considers the possibility of* $\hat{\delta}_{c,j}$ *being biased in ways that cannot be absorbed by* $Q_j.\lambda_c$*. I show that its impact on the bias of* $MWTP_c^{SQ}$, $c \in G_2$ *is minimal as long as* R_1 *is sufficiently large relative to R.*

Remark 5. *At first glance it may appear to be inconsistent to specify both the choice-specific value functions in* (11) *and the flow utilities in* (14)*, but this is not the case. Equation* (11) *does not specify how the choice-specific value functions vary with each amenity of the neighborhood. The proposed approach only makes such decomposition at the flow utility level, according to equation* (14).

4.3 Step 3: Obtaining the Valuation of Public School Quality Per Year, Per Child at Each Grade Level

Step 2 in Section 4.2 estimated the MWTP for school quality per year, per family for groups $c \in G_2 = \{6, ..., 18\}$. In the final step, these estimates are used to calculate the MWTP for school quality per year, per child, and per grade from kindergarten to grade 12.

Families of group c may have more than one child, and may have children attending different grades. Let $N_{i,c,g}$ be the number of children from family i of group c who are observed to be attending grade g in the data, and let the average number of children attending grade g across all families from group c be defined as $N_{c,g} := \frac{1}{n_c}\sum_{i \in I_c} N_{i,c,g}$. Families of groups 6 through 18 have their oldest child attending grade kindergarten through grade 12, respectively. The MWTP that is estimated in step 2 is written as a weighted sum of the MWTP per year-child-grade of all families, using information on the grade that the children in each family are attending:

$$\text{MWTP}_c^{SQ} = \sum_{g=0}^{c-6} \text{MWTP}_{c,g}^{SQ}.N_{c,g}, \qquad c = 6, ..., 18 \tag{23}$$

where $g = 0, ..., 12$ indexes grades kindergarten through grade 12, respectively, and $\text{MWTP}_{c,g}^{SQ}$

is defined as the MWTP for school quality per year-child-grade g for families of group c.

Assumption 3 below guarantees identification by stating that a family values the school quality for each child only as function of the grade that child is attending:

Assumption 3.

$$MWTP^{SQ}_{c,g} = MWTP^{SQ}_{g} \qquad \forall c \in G_2,\, g = 0,...,12 \tag{24}$$

This assumption implies, for instance, that a family's valuation of school quality per child for a specific grade does not change as a function of birth order, or as a function of sibling spacing. MWTP^{SQ}_{g}, implicitly defined in Assumption 3, is referred to as the *MWTP for school quality per year-child-grade g*, for grades $g = 0,...,12$, and is the main parameter of interest in this paper.

Substituting equation (24) into equation (23):

$$\text{MWTP}^{SQ}_{c} = \sum_{g=0}^{c-6} \text{MWTP}^{SQ}_{g}.N_{c,g}, \qquad c = 6,...,18 \tag{25}$$

The system of equations (25) has 13 unknowns (MWTP^{SQ}_{g}, $g = 0,...,12$) and 13 linearly independent equations ($c = 6,...,18$), which guarantee the identification of the MWTP for school quality per year-child-grade, from kindergarten to grade 12. The estimation of steps 2 and 3 is carried out simultaneously via GMM, with unknown parameters $\text{MWTP}^{\text{SQ}}_{g}$, $g = 0,...,12$, $\tilde{\lambda}_c$ and ϕ_c, $c = 6,...,18$.[26]

Remark 6. ***Robustness Checks:*** *Under Assumptions 1, 2 and 3, this three-step approach identifies the valuation of school quality per year, per child at each grade level. This approach can be tailored to perform a variety of robustness checks: (a) Assumptions 2 and 3 can be each directly tested; (b) Groups can be defined so as to weaken Assumptions 1, 2 and 3. (c) The model is over-identified, so these assumptions can be tested jointly. All such checks are implemented in Section 5.*

5 Empirical Results

This section presents the main results and robustness checks of the paper, explaining along the way how the methodology presented in Section 4 is implemented. Table 4 shows the main results and robustness checks of the paper.[27] This table shows the MWTP per year, per child, and per grade as a percentage of rent for an increase of 5% in school quality.

[26] $\tilde{\theta}_c$ is identified by $\tilde{\theta}_c = \text{MWTP}^{\text{SQ}}_{c}.\phi_c$, where $\text{MWTP}^{\text{SQ}}_{c}$ is given by (25).

[27] This table shows results for $\beta = .95$. The results do not change significantly for choices of different βs in the range .90 through .99.

The rows represent the grades, and the columns represent different specifications. Column I refers to a baseline panel regression of the mean flow utilities of groups 6 to 18 on the average school quality and the average rent of each neighborhood, without any control variable. The GMM estimates of this regression imply a MWTP of around 10% of rent for primary school, with 16% for kindergarten, very low and not significant values for middle school, and around 7% for high school.

Column II shows the MWTP estimates of the same panel regression, still not controlling for unobservables, but adding a list of control variables that account for observed amenities: average income of neighbors, proportion of Black neighbors, proportion of neighbors with race other than White or Black, proportion of neighbors with college degree or more, average number of rooms and proportion of homeowners in the neighborhood. The estimates of this regression show a substantial increase on the MWTP for school quality in comparison to the baseline specification. This result is expected, since school quality and price are likely both positively correlated to unobserved amenities, with price being likely the largest source of endogeneity.[28] However, as described above, explicitly adding socio-demographic control variables may change the interpretation of the MWTP estimates, so columns I and II are not necessarily directly comparable. The MWTP estimates are very high for elementary and high school grades, with 38% for kindergarten and 23% for 12th grade, but around 6% for middle school.

Columns III through VI of Table 4 show the results of the GMM estimation using the proposed method of controlling for unobservables, as described in Sections 4.2 and 4.3. The groups are divided into three sets: the proxy set, the set of interest and the set of instruments. The set of interest includes groups 6 through 18, which correspond to the families with school age children that attend a public school in the correctly pre-specified grade. Other than that, the decision of the groups included in each set follows the general guidelines provided in Appendix B. Columns III through VI differ from each other with respect to the number of proxies used. Intuitively, the more proxies are added, the higher is the space of unobservables that can be spanned, provided the added proxy group values the unobservables differently from the other proxy groups. As more proxy variables are added, I assign more groups to the set used as instrumental variables. For each specification, the lists of groups used as proxies and of groups used as instrumental variables are noted in the table. Column III shows the results using one proxy: the mean flow utilities of group 34. The results are shown to be similar to the results of column II, even though the

[28]Heuristically, under the presence of endogeneity in the baseline regression, both θs and ϕs are likely overestimated, but ϕs are likely more so, as prices are likely more endogenous than school quality. Because ϕ is included in the denominator, naive OLS estimates would tend to be biased downward.

specification of column II adds a wide 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 adding one proxy variable controls for more variation of the dependent variable than adding the list of observed amenities used in column II. This is particularly desirable, given that the interpretation of the MWTP coefficient does not change by adding this generic control variable.

There may exist confounding amenities valued by parents with children of school age, but not valued by non-parents, such as the availability of children's parks. Column IV adds another proxy variable to control for such amenities: the mean flow utilities for group 4. The MWTP estimates reduce a lot, especially for elementary school. This is expected, since group 4 is likely to enjoy the same amenities that parents with children of similar ages tend to enjoy. Also, the standard errors become larger because the multicollinearity increases as more proxies are added (see discussion in Appendix B). Likewise, it is possible that two unobservables are not enough to control for the endogeneity problem. There could be an unobservable amenity, for example, that only relatively older people value or are exposed to. For this reason, column V adds one more proxy: the mean flow utilities from group 41. The estimates reduce even more for both primary and secondary school.

Finally, column VI adds to the specification of column V another proxy variable: group 2. The fit does not improve significantly, and the results are similar to the ones in column V. Columns V and VI together show that three unobservable amenities seem to be enough to take care of the endogeneity with regard to the variables school quality and average rent.[29]

[29] As robustness check, I changed the groups included in G_1 and G_3 and find the same results, provided that one group of parent with $c \leq 5$ and two groups of non-parents c' and c'' with c' and c'' sufficiently different from each other are included in G_1.

Table 4

Main Results						
MWTP (as % of rent) per year-child-grade for an increase in 5% in school quality						
Grade	**I**	**II**	**III**	**IV**	**V**	**VI**
K	15.72**	38.10**	37.75**	21.09**	13.46*	10.58
	(5.54)	(5.82)	(3.64)	(5.88)	(8.14)	(8.88)
1	9.97**	27.89**	28.52**	19.25**	14.99**	14.25**
	(4.24)	(4.57)	(2.71)	(3.53)	(6.19)	(6.24)
2	4.02	19.28**	20.99**	16.03**	11.25	10.26
	(3.89)	(4.16)	(2.48)	(5.57)	(7.52)	(7.95)
3	-0.45	8.19**	13.13**	14.07**	12.38**	12.80**
	(3.16)	(3.89)	(2.11)	(3.17)	(5.62)	(5.54)
4	-3.53	4.16	7.97**	9.24**	10.70*	9.88*
	(2.64)	(3.17)	(1.74)	(2.85)	(5.09)	(5.71)
5	-1.33	6.44*	9.58**	9.96**	10.74*	9.97
	(3.19)	(3.37)	(2.14)	(4.14)	(5.73)	(6.43)
6	-0.17	5.84*	9.44**	8.61**	7.70	7.98
	(3.21)	(3.12)	(1.96)	(3.65)	(5.90)	(6.23)
7	-1.74	6.95**	10.58**	7.67**	6.87	6.36
	(3.11)	(3.21)	(1.83)	(6.12)	(6.87)	(7.79)
8	-0.88	8.03**	11.26**	8.73**	5.80	5.30
	(3.08)	(3.57)	(2.15)	(4.11)	(6.21)	(6.63)
9	5.04	14.24**	17.43**	14.08**	12.46**	11.80*
	(3.36)	(3.65)	(2.01)	(5.19)	(5.89)	(6.79)
10	3.73	15.74**	19.07**	18.68**	14.59**	13.93**
	(3.02)	(3.93)	(1.93)	(3.57)	(4.89)	(5.61)
11	5.36	17.56**	20.96**	14.69**	13.05**	11.35
	(3.56)	(4.21)	(1.92)	(5.64)	(6.55)	(7.67)
12	8.75**	22.95**	26.41**	18.72**	16.17**	14.20*
	(4.37)	(4.76)	(2.25)	(6.11)	(6.75)	(7.92)
Controls?	**No**	**Yes**	**No**	**No**	**No**	**No**
Proxies (groups)	–	–	34	34 4	34, 41 4	34, 41 2, 4
Instruments (groups)	–	–	36–38	36–38 1, 3	36–38, 42–44 1, 3	36–38, 42-44 0, 1, 3, 5
P-value for J test	–	–	.00**	.03**	.35	.32
Adjusted R-squared	.0221	.2517	.3839	.3980	.4162	.4189
Observations	4,475	4,475	4,475	4,475	4,475	4,475

Notes: Dependent Variable: estimated mean flow utilities for groups 6 through 18. Proxy Variables: estimated mean flow utilities for groups referred in each column. Instruments: estimated mean flow utilities for groups referred in each column. Controls used in column II: average income of neighbors, proportion of Black neighbors, proportion of neighbors with race other than White or Black, proportion of neighbors with college degree or more, average number of rooms and proportion of homeowners in the neighborhood. The average monthly rent is $544. Robust standard errors clustered by school attendance area are in parenthesis. *: Statistically significant at the 10% level; **: Statistically significant at the 5% level.

As an additional robustness test, for each specification I provide the p-value of the over-identification test (i.e., J test). This test can be understood as a joint test for Assumptions 1, 2 and 3. Rejecting the null hypothesis is evidence that at least one of these assumptions is not valid for that specification. Indeed, specification III is rejected at 1% of significance, and specification IV is rejected at 5% of significance, suggesting that the test is powerful in this context. Specifications V and VI are not rejected even at 30% of significance, providing further evidence that the three proxy groups, two of non-parents and one of parents, seem to take care of the original endogeneity problem. As discussed in Appendix B, this over-identification test seems to have power to detect violations of Assumpion 1. I also perform other tests aiming at detecting violations of either Assumption 2[30] or Assumption 3[31], and find no evidence that these assumptions play an important role in the main results.

The results of the preferred specifications (specifications V and VI) show that parents are willing to pay 11% more a year to send each of their children to a 5% better elementary public school. The corresponding estimates for middle school and secondary school are respectively 6% and 13%.

Table 5

Magnitude of the Results

	MWTP(5%)		**MWTP(1 St.Dev)**	
	Per month	**Per year**	**Per month**	**Per year**
Elementary School	63**	757**	202**	2,422**
	(16)	(192)	(51)	(614)
Middle School	36**	433**	115**	1,386**
	(18)	(216)	(58)	(691)
High School	70**	840**	224**	2,688**
	(17)	(204)	(54)	(653)

Notes: 1 St. Dev. $\approx$ 16%. Coefficient values and standard errors derived from estimates of specification VI in Table 4. *: Statistically significant at the 10% level; **: Statistically significant at the 5% level. Values are in 2000 dollars.

Table 5 shows the magnitude of the effects found in specification VI in Table 4. These numbers are calculated over the average rent in Minnesota valued in 2000 dollars, which

[30]When I include $c = 4,5 \in G_2$, I find that $\theta_4 = \theta_5 = 0$, as expected. Moreover, when I relax the assumption by writing $\phi_c = \alpha_0^p + \alpha_1^p.c$, where $p = \{\text{parents}, \text{nonparents}\}$, for all $c \leq 9$ and $c \geq 31$, I find that $\alpha_1^p = 0$ for each p. Finally, when I restrict the data to families with levels of household income above the median of the sample I find similar results for each grade.

[31]When I restrict the data to only families with one child I find similar results for each grade.

is \$544. As the coefficients are averaged across grades, the standard errors become substantially smaller, and all coefficients become statistically significant at the 95% confidence level. Parents are willing to pay \$2,422 more per year to send one of their children to attend primary school in a school that is one standard deviation better. The corresponding values are \$1,386 for middle school and \$2,688 for secondary school.

6 Interpretation of Estimates

Although Table 5 offers insight into the statistical significance of the results, it is difficult to gather the economic significance of these valuation estimates. To provide this context, this section takes advantage of the unit of measurement of the estimates in Table 5 to discuss the implications of these results in more detail.

A policymaker can use these estimates to perform a cost-benefit analysis of a statewide reform that improves the achievement of students in all grades in one standard deviation (16%). The results presented in Table 5 suggest that parents will be willing to pay \$2,422 per child attending elementary school, \$1,386 per child attending middle school and \$2,688 per child attending high school. Meta-analyses on this topic suggest that the cost per student of increasing the test scores in one standard deviation is of at least \$2,850.[32] This value is directly comparable to the MWTP estimates from Table 5, suggesting that at the level of efficiency suggested by previous studies[33] this reform is not worth implementing unless it generates enough positive externality at the state level. In a back-of-the-envelope analysis, I calculate that this reform if performed in 2014 will generate a benefit of \$2.6 billion and a cost of \$3.2 billion.[34]

These estimates can also be used to study the level of competition among public schools by calculating what a family would be willing to additionally pay over its life cycle to live in a house located in a neighborhood that offers a one standard deviation better school for all grades, all else constant.[35] The willingness to pay will depend on how

[32] Greenwald et al. (1996) find that a one standard deviation increase in test score is generally achieved with a per student expenditure of about \$3,800. For investments targeted at teacher education and teacher experience, these values are of about \$2,850 and \$3,350, respectively. However, Hanushek (1997) suggests that these estimates may be too optimistic. All values are in 2000 dollars, inflated by the CPI index.

[33] In 1999-2000, the total expenditure per student in public schools in Minnesota was \$8,916 (Source: Common Core of Data.). This suggests that at least 32% of the average school expenditure per student needs to be additionally spent to achieve a 16% increase in test scores.

[34] In this calculation I use the total state enrollment in each grade in 2014 from the Common Core of Data as well as the estimates for each grade from specification VI of Table 4. Amounts are inflated to 2014 dollars by the CPI index.

[35] For simplicity, in this exercise the family is indifferent between the two houses except for school quality and prices.

many children are included in the family, on the stage of the children in their education careers and on how long the family expects to stay in the new house. Table 6 presents these results. For example, a family with one child about to enter kindergarten will be willing to pay in 2000 an additional $15,445 in order to send its child to a one standard deviation better school until 8th grade, all else constant.[36] Similarly, an average family with a child three years from attending kindergarten and another child about to attend 3rd grade will be willing to pay an additional $35,768 (=$18,599 + $17,169) in 2000 in order to send its children to a one standard deviation better school until both children finish high school, all else constant.[37]

Finally, the estimates in this paper can also help shed some light on the level of competition between public and private schools across grades. For instance, consider a family with one child about to attend kindergarten contemplating whether to send its child to a private school instead of the local public school, since the quality of the private school is one standard deviation higher. In a state such as Minnesota, which does not have a private school voucher program, that family will decide to keep its child in the local public school unless the private school costs less than $2,422 a year per child. To provide some rough benchmark, the average tuition charged by a private elementary school in the U.S. is $5,400, and the average tuition charged by a private secondary school in the U.S. is $8,470.[38] About 78% of the private schools in the U.S. cost more than $2,422 at elementary grades, and about 90% of the private schools in the U.S. cost more than $2,688 at secondary grades. Comparable data on private and public school quality are not available; however, to the extent that private school tuition is a strong indicator of private school quality, these estimates suggest that private schools exert little competing pressure to public schools unless a voucher system is in place, particularly for secondary grades.[39]

[36]For perspective, the average house price in Minnesota in 2000 is of $144,000, as can be seen in Table 1.

[37]In these present value calculations, I use an annual percentage rate (APR) of about 5.25% (i.e., $\beta = .95 \approx \frac{1}{1+APR}$).

[38]Source: Data on private school tuitions of all states from the 2007-2008 School and Staffing Survey (SASS), provided by the U.S. Department of Education. Data for Minnesota alone is not available in the public version. All figures are in 2000 dollars, deflated by the CPI.

[39]For comparison, by 1999-2000 the maximum voucher payment allowed in the Milwaukee, WI private school voucher program was for $5,106.

Table 6

WTP Per Child

1 St. Dev. increase in school quality

From	Until: 2nd Grade	5th Grade	8th Grade	12th Grade
3 Years Before Kindergarten	5,978	10,766	13,242	18,599
Kindergarten	6,972	12,557	15,445	21,692
3rd Grade	–	6,514	9,882	17,169
6th Grade	–	–	3,928	12,427
9th Grade	–	–	–	9,912

Notes: 1 St. Dev. $\approx 16\%$. Coefficient values derived from estimates of specification VI in Table 4. For example, the table shows that, all else constant, a family with one child about to enter 3rd grade will be willing to pay an additional $9,882 for a house in order to send its child to a one standard deviation better school until the child finishes 8th grade. Similarly, all else constant a family with one child about to enter kindergarten and another child about to enter 9th grade will be willing to pay an additional $31,604 (=$21,692+$9,912) for a house in order to send its children to a one standard deviation better school until both children finish high school. $\beta = .95$, so the annual percentage rate (APR) that is used for the present value calculations of this table is around 5.25%. Values are in 2000 dollars.

Naturally, a full analysis of school choice is much more complex than suggested here. For instance, it should include comparisons among public schools in different neighborhoods and comparisons between public and private schools, accounting for the fact that families have heterogeneous preferences not only across grades, but also across many other demographic characteristics. Moreover, it should acknowledge that different families might consider school "quality" to be a different amenity than the one perceived by policy makers. For instance, families may consider other differences between a private and a public school that might not be easily projected in measures of school quality, such as religious curriculum.[40] The intention of the admittedly speculative discussion in this section is to highlight the types of policies that can be informed with willingness to pay estimates for public school quality at this unit of measurement. In this sense, the results of this paper provide a preliminary step towards a more detailed empirical analysis about the competition among public schools and between public and private schools. For instance, estimates like the ones provided in this paper can be helpful as an input for models studying the impact of educational policies in the spirit of Epple and Romano (1998), Nechyba (1999), Nechyba (2000), Epple, Figlio, and Romano (2004) and Ferreyra (2007).

[40] A more complete analysis will also require knowledge of the whole distribution of the valuation estimates.

7 Conclusion

This paper estimates the parental willingness to pay for an improvement in school quality. It estimates a dynamic neighborhood choice model in order to identify the valuation of school quality per year, per child at each grade level, a unit of measurement comparable to the tuition charged by private schools and to available public education costs, facilitating cost-benefit analyses.

Parental valuation is found to be higher for elementary and secondary grades compared to middle school grades. One potential explanation is that some parents seek shorter-term investments on their children's career because of budget constraints. This will lead some parents to postpone sending their children to a better school until later grades. More broadly, the results suggest that parental valuation does not seem to outweigh costs, even at the most valued grades. These findings highlight the importance of improving the efficiency of the way school resources are typically spent, otherwise externalities are needed to justify such investments.

Methodologically, I develop a novel strategy to control for confounding unobservable amenities, including those that are post-determined, such as amenities affected by sorting. Being able to fully control for the composition of neighbors is crucial to isolate the valuation of any neighborhood amenity *per se*. Going forward, this method might be particularly useful to estimate the preference for neighborhood amenities other than school quality (such as neighborhood price), where the same endogeneity issues are present, and yet complementary identification strategies such as the boundary fixed effects approach might be unfeasible. A practical solution is also developed to estimate a dynamic choice model with a high dimensional state space using only cross-sectional data, provided some synthetic cohort assumptions can be made. This method can be useful in the labor, public and urban economics literatures, that often require detailed geographical and family data available mostly in cross-sectional data sets such as the Decennial Census. Future research is needed to determine whether these methods can be successfully applied to other contexts.

References

P. Arcidiacono and P. B. Ellickson. Practical methods for estimation of dynamic discrete choice models. *Annu. Rev. Econ.*, 3(1):363–394, 2011.

P. Arcidiacono and R. Miller. Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 7(6):1823–1868, 2011.

P. Bajari, J. C. Fruehwirth, K. I. Kim, and C. Timmins. A rational expectations approach to hedonic price regressions with time-varying unobserved product attributes: The price of pollution. *The American Economic Review*, 102(5):1898–1926, 2012.

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

P. Bayer, R. McMillan, A. Murphy, and C. Timmins. A dynamic model of demand for houses and neighborhoods. *Econometrica*, 84(3):893–942, 2016.

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

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

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

S. Black and S. Machin. Housing valuations of school performance. *Handbook of the Economics of Education*, 3:485–519, 2011.

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

J. Clapp, A. Nanda, and S. Ross. Which school attributes matter? The influence of school district performance and demographic composition on property values. *Journal of Urban Economics*, 63(2):451–466, 2008.

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

D. Epple and H. Sieg. Estimating equilibrium models of local jurisdictions. *Journal of Political Economy*, 107(1):645–681, 1999.

D. Epple, D. Figlio, and R. Romano. Competition between private and public schools: Testing stratification and pricing predictions. *Journal of Public Economics*, 88(7):1215–1245, 2004.

D. Epple, R. Romano, and H. Sieg. The intergenerational conflict over the provision of public education. *Journal of Public Economics*, 96(3):255–268, 2012.

M. Ferreyra. Estimating the effects of private school vouchers in multidistrict economies. *The American Economic Review*, pages 789–817, 2007.

R. Greenwald, L. Hedges, and R. Laine. The effect of school resources on student achievement. *Review of Educational Research*, 66(3):361–396, 1996.

E. Hanushek. Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, 19(2):141–164, 1997.

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

J. Hotz and R. Miller. Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies*, 60(3):497–529, July 1993.

R. Mastromonaco. A dynamic general equilibrium analysis of school quality improvements. 2014.

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

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

T. Nechyba. School finance induced migration and stratification patterns: The impact of private school vouchers. *Journal of Public Economic Theory*, 1(1):5–50, 1999.

T. Nechyba. Mobility, targeting, and private-school vouchers. *American Economic Review*, pages 130–146, 2000.

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

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

A Details About the Estimation in Step 1

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

$$d_{i,c,t} = j \Longleftrightarrow \varepsilon_{i,c,t,j} - \varepsilon_{i,c,t,r} \geq v_r(\boldsymbol{W}_{i,c,t}) - v_j(\boldsymbol{W}_{i,c,t}), \quad \forall r. \tag{26}$$

Because $\varepsilon_{i,c,t,j}$ is assumed distributed as extreme value, $\varepsilon_{i,c,t,j} - \varepsilon_{i,c,t,r}$ has a logit distribution, and the probability that family i of group c chooses neighborhood j in period t is:

$$\boldsymbol{P}_j(\boldsymbol{W}_{i,c,t}) = \frac{\exp(v_j(\boldsymbol{W}_{i,c,t}))}{\sum_{r=1}^{J} \exp(v_r(\boldsymbol{W}_{i,c,t}))}. \tag{27}$$

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

$$LL_{c,t}(\{\boldsymbol{W}_{i,c,t}, d_{i,c,t}\}_{i=1}^{I_c}) = \sum_{i=1}^{I_{c,t}} \sum_{j=1}^{J} 1_{\{d_{i,c,t}=j\}} \log(\boldsymbol{P}_j(\boldsymbol{W}_{i,c,t})), \quad c = 0, ..., C, \quad \forall t. \tag{28}$$

where $1_{\{d_{i,c,t}=j\}}$ is an indicator for whether the expression in brackets is true.

I estimate $\boldsymbol{\Delta}_{c,t}$ and $\boldsymbol{\Phi}_{c,t}$ by maximizing equation (28) independently for each group $c \in \{0, ..., C\}$, and time $t = 2000$.[41] $\delta_{c,t,j}$ is estimated via equations (12) and (13) by plugging in estimates of $\boldsymbol{\Delta}_{c,t}$ and of $\boldsymbol{\Phi}_{c,t}$ along with data on $\Pi_{c,t}$ for all c. The technical results below guarantee the consistent estimation of $\delta_{c,t,j}$.

A.1 Technical Results

Assuming that the regularity conditions which guarantee that the Maximum Likelihood Estimators $\boldsymbol{\Delta}_{c,t}$ and $\boldsymbol{\Phi}_{c,t}$ are consistent are satisfied,[42] the following lemma guarantees the mean convergence of $\hat{v}_j(\boldsymbol{W}_{i,c,t})$ and $\hat{v}_j(\boldsymbol{W}_{i,c,t+1})$.

[41] Due to the large number of parameters, it is unfeasible to estimate all coefficients of this model using a standard numerical optimization algorithm, such as Newton-Raphson. Instead, I write the Δs as function of the other parameter Φ using a contraction, as in Berry et al. (1995).

[42] See McFadden (1973), McFadden (1977) and Berry et al. (1995).

Lemma A.1. *For all i, c and* $t = 2000$*, let* $\hat{\boldsymbol{W}}_{i,c,t'}$ *be a consistent estimator of* $\boldsymbol{W}_{i,c,t'}$ *for* $t' = t, t+1$*. Define*

$$\hat{v}_j(\boldsymbol{W}_{i,c,t}) := v_j(\hat{\boldsymbol{W}}_{i,c,t}).$$
$$\hat{v}_j(\boldsymbol{W}_{i,c,t+1}) := v_j(\hat{\boldsymbol{W}}_{i,c,t+1}).$$

Then, $\boldsymbol{E}(|\hat{v}_j(\boldsymbol{W}_{i,c,t}) - v_j(\boldsymbol{W}_{i,c,t})|) \to 0$ *and* $\boldsymbol{E}(|\hat{v}_j(\boldsymbol{W}_{i,c,t+1}) - v_j(\boldsymbol{W}_{i,c,t+1})|) \to 0$ *for each i, c, j and* $t = 2000$*.*

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

Finally, I estimate $\delta_{c,t,j}$ using an empirical version of equation (12) and (13):

For $c = 0, \ldots, 18$:

$$\hat{\delta}_{c,t,j} := \frac{1}{n_{c,t}} \sum_{i \in I_c} \Big[\underbrace{\hat{\Delta}_{c,t,j} + 1_{\{j \neq d_{i,c,t-1}\}}.\hat{\Phi}_{c,t}}_{\hat{v}_j(\boldsymbol{W}_{i,c,t})} - \beta \Big(\gamma + \log \sum_{r=1}^{J} \exp \big(\underbrace{\hat{\Delta}_{c+1,t,r} + 1_{\{r \neq j\}}.\hat{\Phi}_{c+1,t}}_{\hat{v}_r(\boldsymbol{W}_{i,c,t+1})} \big) \Big) \Big]. \tag{29}$$

For $c = 31, \ldots, 49$:

$$\hat{\delta}_{c,t,j} := \frac{1}{n_{c,t}} \sum_{i \in I_c} \Big[\underbrace{\hat{\Delta}_{c,t,j} + 1_{\{j \neq d_{i,c,t-1}\}}.\hat{\Phi}_{c,t}}_{\hat{v}_j(\boldsymbol{W}_{i,c,t})} -$$
$$- \beta \Big(\gamma + \Pi_{c,t}.\log \sum_{r=1}^{J} \exp \Big(\Big(\underbrace{\hat{\Delta}_{0,t,r} + 1_{\{r \neq j\}}.\hat{\Phi}_{0,t}}_{\hat{v}_j(\boldsymbol{W}_{i,c,t+1})} \Big) \Big) + (1 - \Pi_{c,t}).\log \sum_{r=1}^{J} \exp \Big(\underbrace{\hat{\Delta}_{c+1,t,r} + 1_{\{r \neq j\}}.\hat{\Phi}_{c+1,t}}_{\hat{v}_j(\boldsymbol{W}_{i,c,t+1})} \Big) \Big) \Big]. \tag{30}$$

Proposition A.1. *For all i, c and* $t = 2000$*, let* $\hat{\boldsymbol{W}}_{i,c,t}$ *be a consistent estimator of* $\boldsymbol{W}_{i,c,t}$*. Let* $\hat{\delta}_{c,t,j}$ *be defined as in equations* (29) *and* (30)*. Then* $\hat{\delta}_{c,t,j}$ *is a consistent estimator of* $\delta_{c,t,j}$*, as defined by equations* (12) *and* (13)*, respectively.*

Proof. Let the term inside the sum in equations (29) and (30) be $\hat{a}_{i,c,j,t}$, and its true value be $a_{i,c,j,t}$. The continuous mapping theorem and Lemma A.1 imply that $E(|\hat{a}_{i,c,j,t} - a_{i,c,j,t}|) \to 0$. Apply Markov's theorem to $\frac{1}{n_{c,t}} \sum_{i \in I_{c,t}} (\hat{a}_{i,c,j,t} - a_{i,c,j,t})$ to show that it is $o_p(1)$. The Law of Large Numbers and Lemma A.1 guarantee that $\frac{1}{n_{c,t}} \sum_{i \in I_{c,t}} a_{i,c,j,t} \xrightarrow{p} \delta_{c,t,j}$. □

B Monte Carlo Study

In this Section, I present a Monte Carlo study of the approach to control for unobservables developed in this paper. To keep this study tractable, I focus on the second step of the estimation procedure. Specifically, I study how violations of Assumptions 1 and 2 affect the bias of the Proxy-IV estimators of $\theta_{(2)}$ and $\phi_{(2)}$ proposed in this paper. Re-writing equations (16), (17) and (18):

$$\hat{\boldsymbol{\delta}}_{j(s)} = SQ_j \boldsymbol{\theta}_{(s)} + P_j \boldsymbol{\phi}_{(s)} + \underbrace{Q_j \boldsymbol{\lambda}_{(s)} + \boldsymbol{\mu}_{j(s)}}_{\xi_{j,(s)}}, \quad G_s = \{1, ..., R_s\}, \quad s = 1,2,3, \quad j = 1, ..., J$$

where $\xi_{j,(s)}$ is the error term, and Q_j and $\hat{\delta}_{j(s)}$ are row vectors of dimensions R and R_s, respectively, with $s = 1,2,3$. The coefficients of interest are $\theta_{(2)}$ and $\phi_{(2)}$. In words, the Proxy-IV approach exploited in the paper involves using SQ_j, P_j and $\hat{\delta}_{j(3)}$ as Instrumental Variables for SQ_j, P_j and $\hat{\delta}_{j(1)}$, where the dependent variable is $\hat{\delta}_{j(2)}$.

Basic Setup

I start by setting $R = 1$ (i.e., Q_j has one dimension), $R_2 = 1$ (i.e., one group belonging to G_2, the set of groups of interest), $r := R_3 - R_1 = 3$ (i.e., r more groups used as IVs than groups used as proxies), $\boldsymbol{\theta}_{(2)} = 1$, $\boldsymbol{\phi}_{(2)} = -1$, $\boldsymbol{\lambda}_{(2)} = 1$ and[43]

$$(SQ_j, P_j, Q_j, \bar{\mu}_j) \sim N\left(mean = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{pmatrix} \right). \tag{31}$$

$\mu_{c,j}$, an element of $\boldsymbol{\mu}_{j,(s)}$ for $s \in \{1,2,3\}$, is defined so as to allow for the study of violations of Assumption 1. I set $\mu_{c,j} \sim N(\bar{\mu}_{c,j}, \sigma^2 = 1)$ for all c and j, where $\bar{\mu}_{c,j} = \bar{\mu}_j.\alpha^{\mu}_{c,j}$, with $\bar{\mu}_j$ defined in equation (31) and $\alpha^{\mu}_{c,j}$ being drawn from $uniform[0, \bar{\alpha}^{\mu}]$. For $\bar{\alpha}^{\mu} \neq 0$, SQ_j and P_j are endogenous ($\mu_{c,j}$ is correlated to SQ_j and P_j conditional on Q_j). Notice that $\mu_{c,j}$ is defined so as to not be absorbed by the term $Q_j.\lambda_c$.

Finally, the preference parameters change across groups within G_s for $s = 1,3$ as follows:

[43] Qualitative results are robust to other values of these and the other parameters that are kept constant throughout this Monte Carlo study.

$$\mathbb{X}_c = \alpha_s^{\mathbb{X}} + v_s^{\mathbb{X}}.\frac{(c-1)}{R_s}, \quad c \in G_s := \{1,2,...,R_s\}, \quad s = 1,3 \tag{32}$$

where $\mathbb{X} = \{\theta, \phi, \lambda\}$. For each $\mathbb{X} = \{\theta, \phi, \lambda\}$, $v_c^{\mathbb{X}}$ represents how different the corresponding preference parameter of each group is relative to other groups within the same G_s. For $s = 1$, I set $\alpha_1^{\lambda} = 10$, $v_1^{\lambda} = 1$, $v_1^{\theta} = 0$, $v_1^{\phi} = 0$, and consider variations of α_1^{θ} and α_1^{ϕ}. For $s = 3$, I set $\alpha_3^{\mathbb{X}} = 1$ and $v_3^{\mathbb{X}} = 1$, with $\mathbb{X} = \{\theta, \phi, \lambda\}$. These baseline values are maintained in all results below, unless otherwise noted.

The goal of this Section is to study the biases of $\hat{\theta}_{(2)}$ and $\hat{\phi}_{(2)}$ as Assumptions 1 or 2 are violated. Assumptions 1 and 2 in the text imply $\left(\bar{\alpha}^{\mu} = 0, \alpha_1^{\theta} = 0, \alpha_1^{\phi} = -1, v_1^{\theta} = 0, v_1^{\phi} = 0\right)$. Specifically, deviations from $\bar{\alpha}^{\mu} = 0$ imply violations of Assumption 1, and deviations from $\left(\alpha_1^{\theta} = 0, \alpha_1^{\phi} = -1, v_1^{\theta} = 0, v_1^{\phi} = 0\right)$ imply violations of Assumption 2. I consider variations in $\bar{\alpha}^{\mu}$, α_1^{θ} and α_1^{ϕ} for different combinations of values of R_1 and α_1^{λ}. I present results in terms of $\hat{\theta}_2 - \theta_2$ and $\hat{\phi}_2 - \phi_2$ for different values of these parameters. For each Monte Carlo iteration $m = 1,...,M$, with $M = 1,000$, I re-draw the data generation process defined above and calculate the biases $\hat{\theta}_2 - \theta_2$ and $\hat{\phi}_2 - \phi_2$, with $\hat{\theta}_2$ and $\hat{\phi}_2$ estimated using the Proxy-IV approach described in Step 2 (Section 4.2). I present the average of these biases across all M iterations, along with its 95% confidence interval.

B.1 Violations of Assumption 1 ($\bar{\alpha}^{\mu} = 0$):

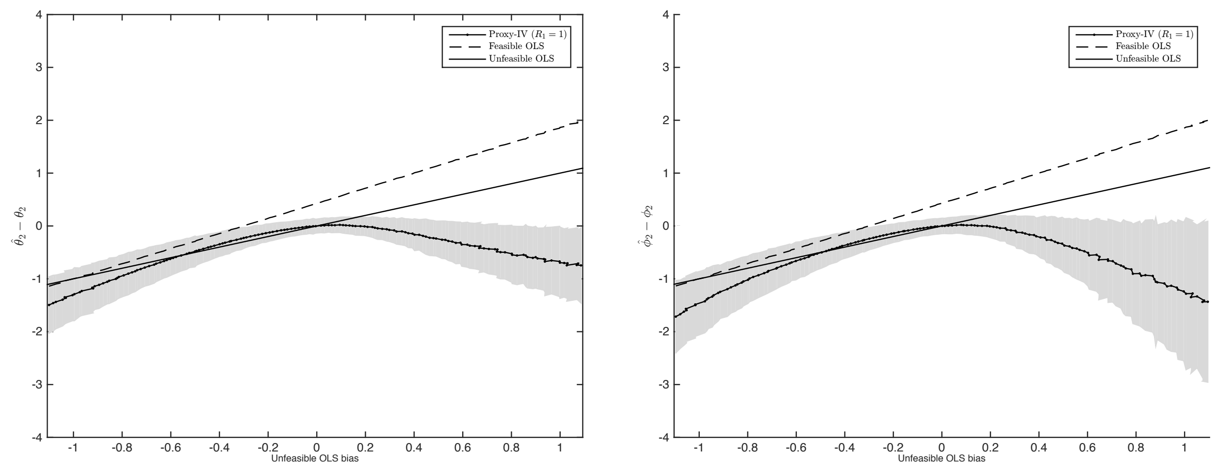
Figures 2-4 present what happens with $\hat{\theta}_2 - \theta_2$ and $\hat{\phi}_2 - \phi_2$ for violations of Assumption 1.[44] For each value of $\bar{\alpha}^{\mu}$, I estimate $\hat{\theta}_{OLS,2}^{u} - \theta_2$ and $\hat{\phi}_{OLS,2}^{u} - \phi_2$, the corresponding biases of the OLS estimators of a linear regression of $\delta_{j,2}$ onto SQ_j, P_j and Q_j. The superscript u stands for "unfeasible", since Q_j is not observed by the econometrician; nonetheless, the biases of these unfeasible OLS estimators offer an intuitive metric of the size of the violation of Assumption 1, as Assumption 1 is valid (and thus $\hat{\theta}_{OLS,2}^{u}$ and $\hat{\phi}_{OLS,2}^{u}$ are unbiased) if and only if $\bar{\alpha}^{\mu} = 0$.

For clarity in the exposition, first I present results for $R_1 = 1$. For different values of $\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\theta}_{OLS,2}^{u} - \theta_2\right)$ implied by different values of $\bar{\alpha}^{\mu}$, Figure 2(a) presents $\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\theta}_2 - \theta_2\right)$ along with its corresponding 95% confidence interval across all M iterations. As a benchmark, I also present $\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\theta}_{OLS,2}^{f} - \theta_2\right)$, the average bias of the standard, "feasible" OLS estimator of the linear regression of $\delta_{j,2}$ onto SQ_j and P_j (without including Q_j as control).

[44] In order to isolate the impact of violations of Assumption 1, I set $\alpha_1^{\theta} = 1$ and $\alpha_1^{\phi} = -1$ to maintain Assumption 2 as valid in this case.

When the unfeasible OLS estimator is unbiased (at the center of the horizontal axis), the Proxy-IV estimator is also unbiased, but the feasible OLS estimator is biased upward. This happens because, although Q_j is endogenous (leading to a bias feasible OLS estimator), Q_j is fully absorbed by $\delta_{j,(1)}$, thus making the Proxy-IV estimator unbiased. For violations of Assumption 1, the Proxy-IV estimator behaves better than even the unfeasible OLS estimator, since $\delta_{j,(1)}$ partially absorbs the endogeneity in $\mu_{j,(2)}$ while Q_j does not.[45] Figure 2(b) shows analogous results for ϕ.

Figure 2: Proxy-IV Bias for Violations of Assumption 1 ($R_1 = 1$)



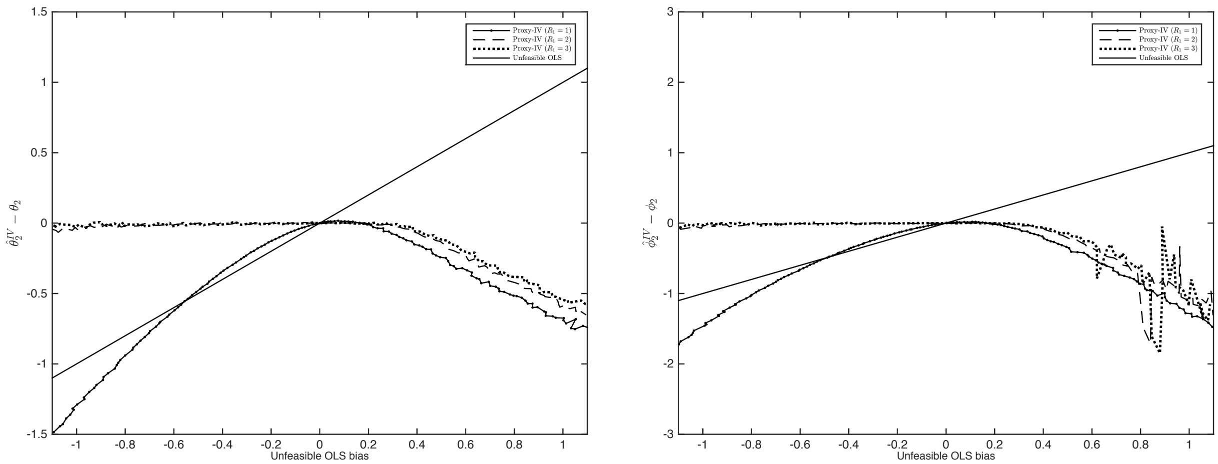
(a) $\hat{\theta} - \theta$ (b) $\hat{\phi} - \phi$

Notes: For each value of $\tilde{\alpha}^{\mu}$, each implying a different value of $\mathbb{E}\left[\mu_{c,j} | SQ_j, P_j, Q_j\right]$, Panel (a) (Panel (b)) presents the bias of three estimators of θ (ϕ): Unfeasible OLS, Feasible OLS and Proxy-IV (along with its 95% confidence interval). Feasible (Unfeasible) OLS refers to the OLS estimator in a regression of $\delta_{j,(2)}$ on SQ_j and P_j (on SQ_j, P_j and Q_j). Proxy-IV refers to the IV estimator in a regression of $\delta_{j,(2)}$ on SQ_j, P_j and $\delta_{j,(1)}$ with SQ_j, P_j and $\delta_{j,(3)}$ as IVs, where $R_1 = 1$ and $R_3 = 4$. The bias of the unfeasible OLS estimator, shown in the horizontal axis, is a more intuitive metric of $\mathbb{E}\left[\mu_{c,j} | SQ_j, P_j, Q_j\right]$, i.e., the degree of violation of Assumption 1. This assumption is valid when the bias of the Proxy-IV estimator of θ and ϕ is equal to zero, represented at the center of the corresponding horizontal axis.

Figure 3 presents the same results (without confidence interval, for clarity) for different values of R_1, but holding constant $R = 1$. It is clear that violations of Assumption 1 lead to less bias in the Proxy-IV estimator the larger is R_1. The reason, as discussed in Remark 4 in the text, is that $\delta_{j,c}$ for the additional groups c in G_1 end up further controlling (partially) for the endogenous $\mu_{j,(2)}$ over and above controlling for $Q_j.\lambda_{(2)}$. In Figure 3(b), the Proxy-IV estimator of ϕ is noisier with higher values of R_1 for larger violations of Assumption 1. This happens because $1 - \tilde{\lambda}_{(2)}$ approaches zero as $\mu_{j,(2)}$ becomes a larger component of $\xi_{j,(2)} := Q_j.\lambda_{(2)} + \mu_{j,(2)}$, making it non-invertible (see footnote 25).

[45]To see this, note that the plot of $\hat{\theta}_2 - \theta_2$ tends to lie on top of the 45 degree line to the left of the origin (of the horizontal axis) and below the 45 degree line to the right of the origin.

Figure 3: Proxy-IV Bias for Violations of Assumption 1 by Values of R_1



(a) $\hat{\theta} - \theta$ (b) $\hat{\phi} - \phi$

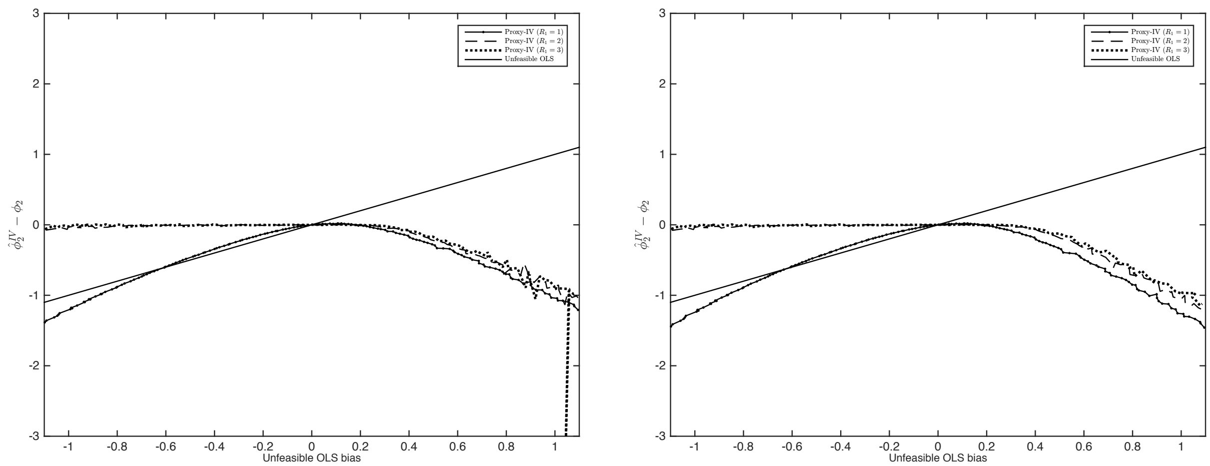
Notes: For each value of $\bar{\alpha}^{\mu}$, each implying a different value of $\mathbb{E}\left[\mu_{c,j}|SQ_j, P_j, Q_j\right]$, Panel (a) (Panel (b)) presents the bias of two estimators of θ (ϕ): unfeasible OLS and Proxy-IV. Unfeasible OLS refers to the OLS estimator in a regression of $\delta_{j,(2)}$ on SQ_j, P_j and Q_j. Proxy-IV refers to the IV estimator in a regression of $\delta_{j,(2)}$ on SQ_j, P_j and $\delta_{j,(1)}$ with SQ_j, P_j and $\delta_{j,(3)}$ as IVs, where R_1 is described in the Figure, and $R_3 = R_1 + 3$. The bias of the unfeasible OLS estimator, shown in the horizontal axis, is a more intuitive metric of $\mathbb{E}\left[\mu_{c,j}|SQ_j, P_j, Q_j\right]$, i.e., the degree of violation of Assumption 1. This assumption is valid when the bias of the Proxy-IV estimator of θ and ϕ is equal to zero, represented at the center of the corresponding horizontal axis.

This potential issue of noise is only of concern if elements of $\lambda_{(1)}$ are close enough to each other (low value of ν_1^{λ}) *and* close enough to $\lambda_{(2)}$ (value of α_1^{λ} too close to $\lambda_{(2)} = 1$). In Figure 4(a), we show analogous results to Figure 3(b) for $\nu_1^{\lambda} = \nu_3^{\lambda} = 5$ instead of $\nu_1^{\lambda} = \nu_3^{\lambda} = 1$,[46] and in Figure 4(b), we show analogous results to Figure 3(b) for $\alpha_1^{\lambda} = 20$ instead of $\alpha_1^{\lambda} = 10$. In both cases, $\hat{\phi}$ is much less noisy even for moderate violations of Assumption 1. Of course, such noise, if existent, should yield wider confidence intervals of the Proxy-IV estimators. Since the standard errors reported in Table 4 do not suggest the existence of such noise, in practice this does not seem to have been an issue in the context of this application.[47]

[46] Naturally, this heterogeneity within groups must happen in both the IV set (G_3) and the proxy set (G_1) in order for it to propagate to the second stage equation in the Proxy-IV approach.

[47] If elements of $\delta_{j,(1)}$ are almost equal to each other, then one might experience multicollinearity issues, contributing to even wider confidence intervals. This concern, together with the concern discussed in Figure 4(a), guided me to choose groups distant from each other in age when selecting the groups to include in G_1 (Section 5).

Figure 4: Proxy-IV Bias of $\hat{\phi}$ for Violations of Assumption 1 by Values of R_1 (Higher Values of ν_1^λ and α_1^λ)



(a) $\hat{\phi} - \phi$, Higher Values of $\nu_1^\lambda, \nu_3^\lambda$ (b) $\hat{\phi} - \phi$, Higher Value of α_1^λ

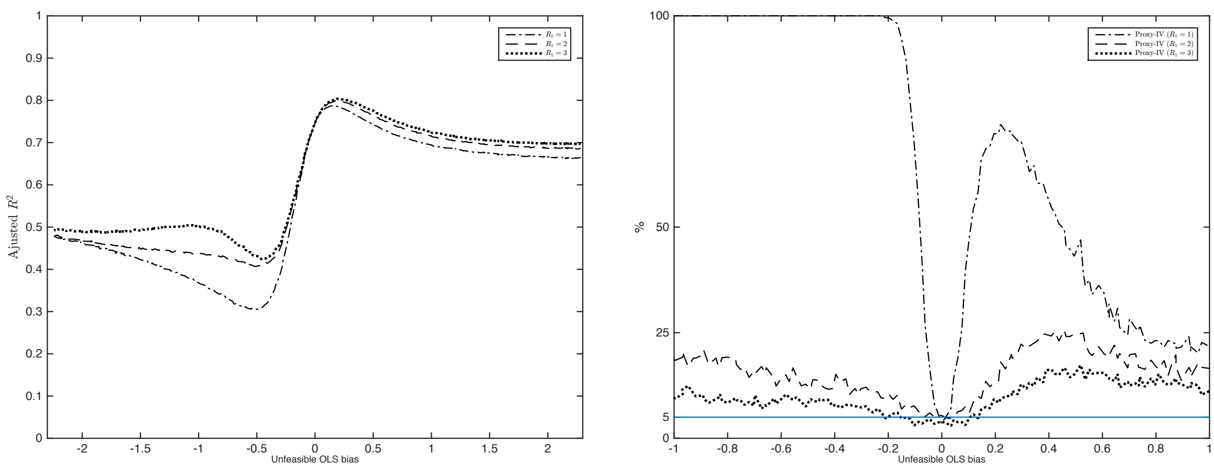
Notes: Both panels should be compared to Figure 3(b). Panel (a) (Panel (b)) has all parameter values as Figure 3(b), except for a higher value of ν_1^λ (α_1^λ): $\nu_1^\lambda = \nu_3^\lambda = 5$ instead of $\nu_1^\lambda = \nu_3^\lambda = 1$ ($\alpha_1^\lambda = 5$ instead of $\alpha_1^\lambda = 1$). For each value of $\bar{\alpha}^\mu$, each implying a different value of $\mathbb{E}\left[\mu_{c,j}|SQ_j, P_j, Q_j\right]$, both panels present the bias of two estimators of ϕ: Unfeasible OLS and Proxy-IV. Unfeasible OLS refers to the OLS estimator in a regression of $\delta_{j,(2)}$ on SQ_j, P_j and Q_j. Proxy-IV refers to the IV estimator in a regression of $\delta_{j,(2)}$ on SQ_j, P_j and $\delta_{j,(1)}$ with SQ_j, P_j and $\delta_{j,(3)}$ as IVs, where R_1 is described in the plot, and $R_3 = R_1 + 3$. The bias of the unfeasible OLS estimator, shown in the horizontal axis, is a more intuitive metric of $\mathbb{E}\left[\mu_{c,j}|SQ_j, P_j, Q_j\right]$, i.e., the degree of violation of Assumption 1. This assumption is valid when the bias of the Proxy-IV estimator of θ and ϕ is equal to zero, represented at the center of the corresponding horizontal axis.

Figure 5 provides some diagnostics that can be obtained by the Proxy-IV, depending on R_1. It shows an important trade-off involved in the decision of whether to add more groups as proxy (more elements in $\delta_{j,(1)}$). On the one hand, a higher value of R_1 leads to a lower bias, as discussed above. Figure 5(a) shows why. The adjusted R^2 of the Proxy-IV approach increases with a higher value of R_1. Moreover, this increase is higher the larger is the unfeasible OLS bias. Intuitively, when Assumption 1 is valid then additional elements of $\delta_{j,(1)}$ do not help absorb the endogenous component of $\xi_{j,(2)} := Q_j.\lambda_{(2)} + \mu_{j,(2)}$, but when it is invalid then these additional elements of $\delta_{j,(1)}$ partially absorb $\mu_{j,(2)}$, reducing the Proxy-IV bias. Figure 5(b) shows the other side of the trade-off. It presents how the proportion (among all M iterations) of rejections of the null hypothesis in the over-identification test (J test) changes with violations of Assumption 1 in terms of the bias of the unfeasible OLS estimator of θ.[48] The level of significance is set to 5%, so for the test to have power it needs to reject the null more than 5% of the times when the null is incorrect. It is clear that this test has non-trivial power in this context, as the proportion of rejections increases as Assumption 1 is violated. However, for higher values of R_1, this increase is less steep, reflecting the reduction in the power of the test to detect violations of Assumption 1.

[48]The qualitative results for the bias of the unfeasible OLS estimator of ϕ are similar.

Nonetheless, even for $R_1 = 3$ the test has some power to detect violations of Assumption 1. The results in Table 4 show that the over-identification test is rejected when $R_1 = 2$ (column IV of Table 4), suggesting that in practice the test is still powerful at least for $R_1 = 2$.

Figure 5: Diagnostics Obtained From Proxy-IV



(a) Adjusted R^2 from Proxy-IV

(b) Over-identification Test from Proxy-IV

Notes: For each value of $\bar{\alpha}^{\mu}$, each implying a different value of $\mathbb{E}\left[\mu_{c,j}|SQ_j, P_j, Q_j\right]$, and for different values of R_1, Panel (a) presents the adjusted R^2 and Panel (b) presents the over-identification (J) test of the Proxy-IV method. The Unfeasible OLS refers to the OLS estimator in a regression of $\delta_{j,(2)}$ on SQ_j, P_j and Q_j. Proxy-IV refers to the IV estimator in a regression of $\delta_{j,(2)}$ on SQ_j, P_j and $\delta_{j,(1)}$ with SQ_j, P_j and $\delta_{j,(3)}$ as IVs, where R_1 is described in the Figure and $R_3 = R_1 + 3$. The bias of the unfeasible OLS estimator of θ, shown in the horizontal axis, is a more intuitive metric of $\mathbb{E}\left[\mu_{c,j}|SQ_j, P_j, Q_j\right]$, i.e., the extent to which Assumption 1 fails. This Assumption is valid when the bias of the Proxy-IV estimator of θ and ϕ is equal to zero, represented at the center of the corresponding horizontal axis.

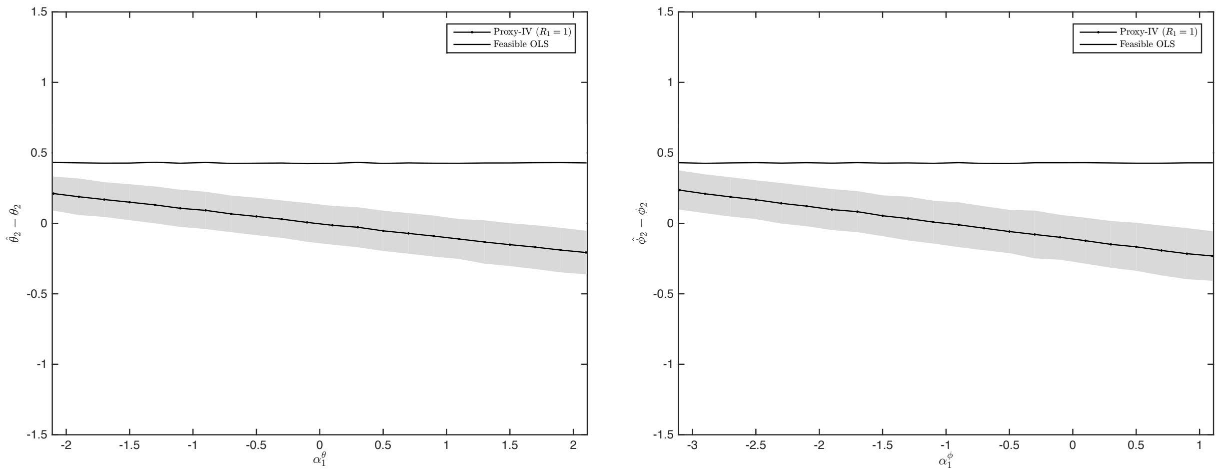
B.2 Violations of Assumption 2 $\left(\alpha_1^{\theta} = 0, \alpha_1^{\phi} = -1\right)$:

Figures 6(a)-7(a) present what happens with $\hat{\theta}_{(2)} - \theta_{(2)}$ for violations of Assumption 2.1.[49] For clarity in the exposition, first I present results for $R_1 = 1$. For different values of α_1^{θ}, Figure 6(a) presents $\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\theta}_{(2)} - \theta_{(2)}\right)$ along with its corresponding 95% confidence interval across all M iterations. Clearly, the estimator using the Proxy-IV method proposed in this paper behaves better than the feasible OLS estimator for moderate violations of Assumption 2.1 ($\alpha_1^{\theta} \approx 0$). Figure 7(a) shows that these results do not change for different levels of R_1. Figures 6(b)-7(b) present analogous results for $\hat{\phi}_{(2)} - \phi_{(2)}$ with respect to violations of Assumption 2.2.[50]

[49] In order to isolate the impact of violations of Assumption 2.1, I set $\bar{\alpha}^{\mu} = 0$ and $\alpha_1^{\phi} = -1$ to maintain Assumption 1 and Assumption 2.2, respectively.

[50] In order to isolate the impact of violations of Assumption 2.2, I set $\bar{\alpha}^{\mu} = 0$ and $\alpha_1^{\theta} = 0$ to maintain Assumption 1 and Assumption 2.1, respectively.

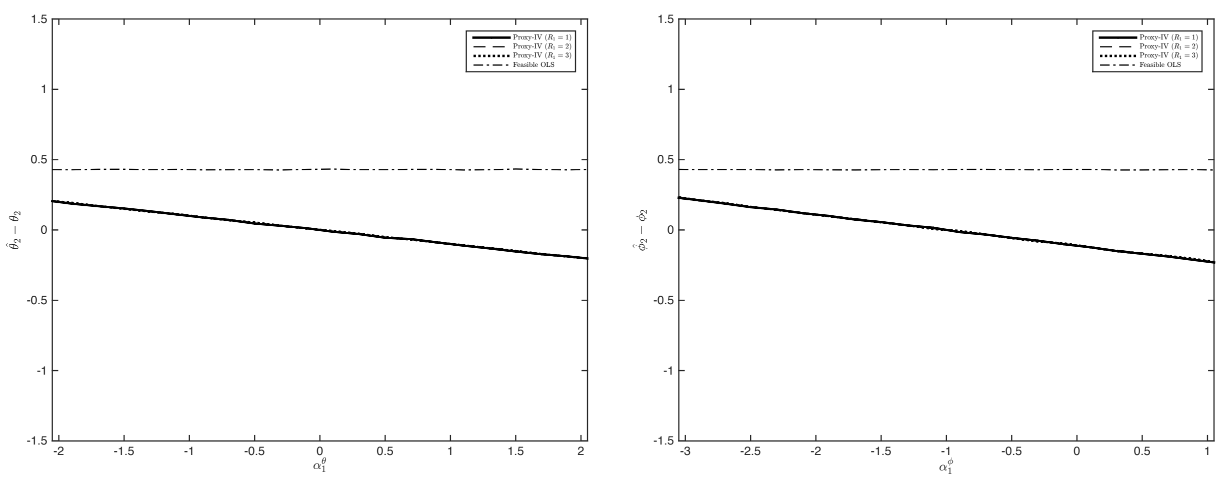
Figure 6: Proxy-IV Bias for Violations of Assumption 1 ($R_1 = 1$)



(a) $\hat{\theta} - \theta$ (b) $\hat{\phi} - \phi$

Notes: For each value of α_1^{θ} (α_1^{ϕ}), each implying a different degree of violation of Assumption 2.1 (Assumption 2.2), Panel (a) (Panel (b)) presents the bias of two estimators of θ (ϕ): Feasible OLS and Proxy-IV (along with its 95% confidence interval). Feasible OLS refers to the OLS estimator in a regression of $\delta_{j,(2)}$ on SQ_j and P_j. Proxy-IV refers to the IV estimator in a regression of $\delta_{j,(2)}$ on SQ_j, P_j and $\delta_{j,(1)}$ with SQ_j, P_j and $\delta_{j,(3)}$ as IVs, where $R_1 = 1$ and $R_3 = 4$. Assumption 2.1 is valid when $\alpha_1^{\theta} = 0$ and Assumption 2.2 is valid when $\alpha_1^{\phi} = -1$, both represented at the center of the corresponding horizontal axis.

Figure 7: Proxy-IV Bias for Violations of Assumption 1 by Values of R_1

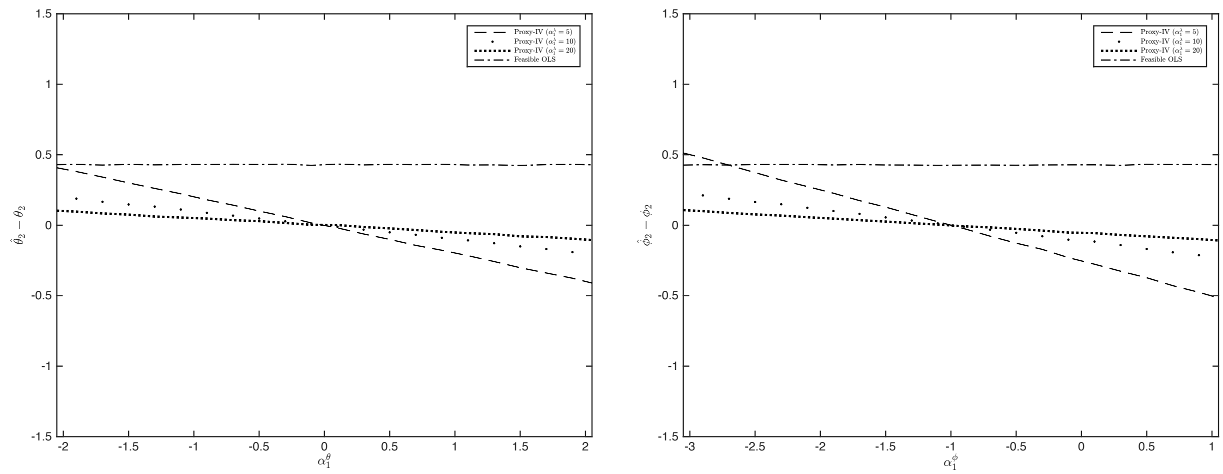


(a) $\hat{\theta} - \theta$ (b) $\hat{\phi} - \phi$

Notes: For each value of α_1^{θ} (α_1^{ϕ}), each implying a different degree of violation of Assumption 2.1 (Assumption 2.2), Panel (a) (Panel (b)) presents the bias of two estimators of θ (ϕ): Feasible OLS and Proxy-IV (along with its 95% confidence interval). Feasible OLS refers to the OLS estimator in a regression of $\delta_{j,(2)}$ on SQ_j and P_j. Proxy-IV refers to the IV estimator in a regression of $\delta_{j,(2)}$ on SQ_j, P_j and $\delta_{j,(1)}$ with SQ_j, P_j and $\delta_{j,(3)}$ as IVs, where R_1 is described in the Figure and $R_3 = R_1 + 3$. Assumption 2.1 is valid when $\alpha_1^{\theta} = 0$ and Assumption 2.2 is valid when $\alpha_1^{\phi} = -1$, both represented at the center of the corresponding horizontal axis.

Figure 8 presents results for different values of α_1^{λ}. The higher is α_1^{λ} relative to $\lambda_{(2)} = 1$, the flatter is the curve, thus the lower is the bias in the Proxy-IV method due to violations of Assumption 2.1 or 2.2.[51] Intuitively, the larger is the difference between elements of $\lambda_{(1)}$ and $\lambda_{(2)}$, the smaller is the repercussion of violations of Assumption 2 on the bias of the Proxy-IV estimators.

Figure 8: Proxy-IV Bias for Violations of Assumption 1 by Values of α_1^{λ}



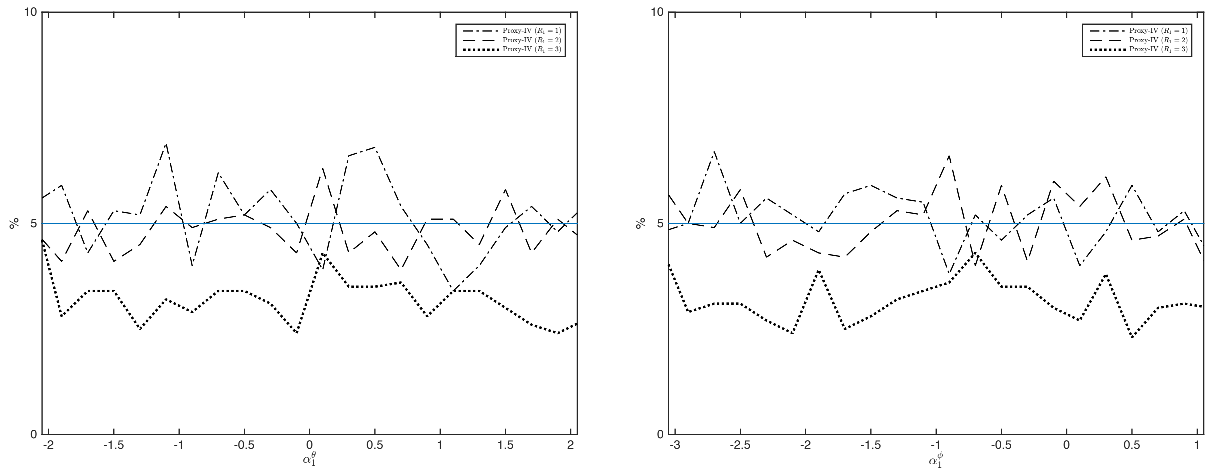
(a) $\hat{\theta} - \theta$ (b) $\hat{\phi} - \phi$

Notes: For each value of α_1^{θ} (α_1^{ϕ}), each implying a different degree of violation of Assumption 2.1 (Assumption 2.2), Panel (a) (Panel (b)) presents the bias of two estimators of θ (ϕ): Feasible OLS and Proxy-IV (along with its 95% confidence interval). Feasible OLS refers to the OLS estimator in a regression of $\delta_{j,(2)}$ on SQ_j and P_j. Proxy-IV refers to the IV estimator in a regression of $\delta_{j,(2)}$ on SQ_j, P_j and $\delta_{j,(1)}$ with SQ_j, P_j and $\delta_{j,(3)}$ as IVs, where $R_1 = 1$ and $R_3 = 4$ and α_1^{λ} is described in the Figure. Assumption 2.1 is valid when $\alpha_1^{\theta} = 0$ and Assumption 2.2 is valid when $\alpha_1^{\phi} = -1$, both represented at the center of the corresponding horizontal axis.

Figure 9(a) presents how the proportion (among all M iterations) of rejections of the null hypothesis in the over-identification test (J test) changes with violations of Assumption 2.1. The level of significance is set as 5%. When the null hypothesis of Assumption 2.1 (as well as Assumptions 1 and 2.2) is valid, the null is rejected around 5% of the times. For violations of Assumption 2.1 (but still maintaining Assumptions 1 and 2.2) the proportion of iterations where the null is rejected does not increase, suggesting that the over-identification test does not have power to detect violations of Assumption 2.1. As R_1 increases, the test does not increase its power to detect such violations either. These results are similar for violations of Assumption 2.2, as can be seen in Figure 9(b). In Section 5, to mitigate this concern I perform a more direct, powerful test to detect violations of Assumptions 2.1 and 2.2. For instance, to test for violations of Assumption 2.1 I incorporate in G_2 a group c that is similar to groups included in G_1, and test for whether $\theta_c = 0$. See footnote 30 in the text for more details.

[51] Similarly, I find a flatter curve for higher values of $\nu_1^{\lambda} = \nu_3^{\lambda}$.

Figure 9: Diagnostics Obtained From Proxy-IV



(a) Over-identification Test from Proxy-IV (b) Over-identification Test from Proxy-IV

Notes: For each value of α_1^θ (α_1^ϕ), each implying a different degree of violation of Assumption 2.1 (Assumption 2.2), Panel (a) (Panel (b)) presents the over-identification (J) test of the Proxy-IV method. Proxy-IV refers to the IV estimator in a regression of $\delta_{j,(2)}$ on SQ_j, P_j and $\delta_{j,(1)}$ with SQ_j, P_j and $\delta_{j,(3)}$ as IVs, where R_1 is described in the Figure and $R_3 = R_1 + 3$. Assumption 2.1 is valid when $\alpha_1^\theta = 0$ and Assumption 2.2 is valid when $\alpha_1^\phi = -1$, both represented at the center of the corresponding horizontal axis.

As discussed above, the findings of the Monte Carlo study suggest some guidelines for implementing the Proxy-IV approach. First, in order to mitigate any bias due to violations of Assumptions 1 or 2, one should attempt to increase R_1 as much as possible, provided that standard errors of $\hat{\phi}$ do not increase too much. It is also useful to make sure that the groups included in G_1 are sufficiently different from each other and from groups included in G_2. Further, over-identification tests are powerful to detect violations of Assumption 1, although they are more powerful the lower is R_1. Importantly, a different test must be performed if one wants to detect violations of Assumption 2, since such over-identification tests cannot detect them. I follow these guidelines when implementing the Proxy-IV approach in Section 5.