Pontificia Universidad Católica de Chile
Instituto de Economía
Magíster en Economía

TESIS DE GRADO
MAGÍSTER EN ECONOMÍA

Dominga Selman

Agosto, 2021

Pontificia Universidad Católica de Chile
Instituto de Economía
Magíster en Economía

WHAT YOU SEE IS WHAT YOU GET: A PARTIAL IDENTIFICATION APPROACH TO SCHOOL CHOICE

Dominga Selman

August, 2021

Advisors:

Pablo Celhay

Nicolás Figueroa

Abstract

Using a robust method of discrete choice analysis proposed by Barseghyan (2021), I estimate parents' preferences in the context of school choice. To account for information costs, I develop a model where parents only know about a subset of all available schools. Choice sets are unobservable and can have different sizes. I partially identify my model using Pre-K applications from Chile's new centralized school admissions system. My results suggest that current assumptions on the observability of agents' choice sets are too strong. However, the estimation method I use lacks computational tractability, so the challenge of finding an alternative approach to estimate parents' preferences in contexts of incomplete information remains.

1 Introduction

The concept of "rational" consumer is what lies at the core of classic economic theory. Although this concept has been given a much more specific definition over the years, it can be broadly summarized as the idea that the consumer is a maximizer. As Daniel McFadden explains in his Nobel Prize Lecture (McFadden, 2001), in the 1960s, advances in technology increased the amount of microeconomic data on individual behavior and the capacity with which researchers could process it. With this newly available data, economists re-examined the way they had, until then, modeled individual agent behavior in the context of consumer theory. In turn, this new framework led to the development of the now prominent discrete choice literature.

The pioneering work of Daniel McFadden lay the theoretical basis for discrete choice by leveraging the principle of revealed preference to learn about agents' (unobserved) preferences using data on their observed choices. In the tradition of McFadden (1973), most discrete choice models rely on the assumption that the econometrician observes the choice set of the consumer. Using this assumption, together with some restrictions on the structure of agents' preferences, the model's parameters can be point identified by the econometrician.

However, the assumption on the observability of agents' choice sets is unlikely at best. Sometimes choice sets are not observed by the econometrician because data on choice sets is not available but could be in principle. In other words, it is a matter of having the necessary resources to collect the data. Nevertheless, in some cases, choice sets can never be observed because they are complex mental constructs that are subject to each agents' individual characteristics.

Consumers may face information costs, cognitive limitations, or idiosyncratic preferences, leading them to choose from a random (unobservable) subset of the feasible set (see Matějka and McKay (2015)). For example, Goeree (2008) studies the U.S. personal computer market. He uses a discrete-choice model of limited consumer information arguing that when information costs are high, one cannot assume perfect information in markets where prices constantly change.[1]

I argue that in the context of the Chilean School Admission System, parents' choice sets are unobservable because parents' choice sets are mental constructs because such data cannot be collected. In a city with many schools, information costs can be high, especially if a centralized database containing each school's relevant information is not available. Parents may be unaware of every school in their city (let alone their city), but are likely familiar with the most emblematic schools or those near their homes. Parents' knowledge of schools could also depend on each family's network since they may know a school either because a neighbor or relative has attended it.

Furthermore, although a family may be aware of the existence of a school, it does not mean they possess all the relevant information to consider applying to it.[2] Acquiring information about relevant school attributes is no easy task for parents; they could search online, but information on specific school attributes is not always readily available. Therefore, parents may have to go to each school, which is a costly alternative, especially if they want information about more than one school. Further,

[1] For a more comprehensive review of the literature on preference estimation with unobserved choice sets see Crawford et al. (2021).

[2] Information on prices, standardized test score results, or religious orientation, amongst others.

families could also have location preferences (other than wanting a school near their homes) and only consider schools within specific areas.[3]

A common way to circumvent the unobservability of the agents' choice set is to assume that it perfectly coincides or is a subset of the feasible set. Econometricians can also use auxiliary information on the composition, distribution, or formation process of choice sets to work around this limitation. Be that as it may, these types of discrete choice models use revealed preferences at their core. Therefore, making erroneous assumptions about consumers' choice sets fundamentally conflicts with the theoretical basis of discrete choice models. Naturally, making wrong assumptions about their composition becomes more likely as its formation processes become more and more complex.

In all, correctly specifying each family's choice set seems nearly impossible in practice. Thus, the econometrician must make additional (and usually heroic) assumptions to achieve the classic discrete choice point identification. These approaches may appear very compelling due to their straightforward implementation. However, I argue that these assumptions significantly bias the econometrician's conclusions because they do not accurately represent agents' real choice sets.

A common approach is to define each family's' choice set as a function of their geographic location. For example, the econometrician could define a family's' choice set as the union of every school in an X-mile radius of their home. However, parents may be willing to commute a long distance for a school with specific attributes that they value. Therefore, when specifying each family's choice set, the econometrician may be leaving out families that commute to far schools because they are willing to trade-off distance for quality. For example, Chumacero et al. (2011) find that parents consider quality and location when choosing schools and they quantify the relevant trade-offs.

To challenge the current literature on school choice, I rely heavily on a recent paper by Barseghyan (2021), hereafter, Barseghyan et al. They propose a robust method of discrete choice analysis when agents' choice sets are unobserved using partial identification. Using their method, I show how in the context of Chile's new Centralized School Admissions System, mainstream assumptions on unobserved choice sets lead to results that are not robust.

In 2016 the Chilean Government implemented a new school admission system called "Sistema de Admisión Escolar" (SAE), radically changing how the public school admissions system works in Chile. This new policy introduced a deferred acceptance mechanism that reduced parents' application costs, created special quotas for students from lower-income families, and ended schools' ability to choose their students, among other things.

The reform publicized that this new system would reduce socio-economic segregation across schools. The logic behind it was that children from lower-income families would not be discriminated against and rejected by the "good quality" schools.[4] Additionally, they would have access to priority admission and, with the voucher system, could now afford to go to better schools, which typically have higher tuition costs.

[3] Parents could have trouble getting their kids to and from school, so they may only choose from a subset of schools near their workplace to reduce commuting costs, which is not observable by the econometrician.

[4] When referring to school quality, I am using standardized test scores as a proxy.

This policy choice suggests that the Chilean Government is working under the assumption that school supply is the primary driver of school segregation. Focusing exclusively on school supply to fight segregation leads policymakers to overlook a crucial aspect of the problem: parents' demand for schools is key because choice sets result from complex socio-economic, demographic, and idiosyncratic interactions unique to each family. I explore how assumptions about choice sets constrain estimation results of parents' preferences in practice.

2 Literature Review

School choice is a heavily discussed topic in education, debates around it have been around for more than 60 years and will most likely not die out soon. In the United States, school choice originated with the history of school desegregation; the historic Brown v. Board of Education decision in 1954 ruled that separating children in public schools based on race was unconstitutional. It signaled the end of legalized racial segregation of schools in the United States and got massive media attention. This case is arguably one of the most significant milestones regarding school choice and schooling systems.

Parties for and against racial school segregation would argue in their favor by making welfare, efficiency, and moral arguments. Many focused on how this ruling would affect students' future earnings because of possible (good or bad) spillover effects. For example, Billings et al. (2012) studied the impact of school Segregation on Educational Attainment and Crime and found that segregation widened racial gaps in middle school and high school math scores.

Modeling how parents choose and rank schools using empirical tools that leverage the principle of revealed preferences and data from matching markets, has been a central topic in this literature. These empirical methods build on the existing literature that estimates random utility models of consumer preferences (see Berry et al. (1995)).

McFadden's pioneering work developed the theoretical basis of discrete choice models, where he outlines a general procedure to formulate econometric models of aggregate choice behavior from individual choices. He develops a conditional logit analysis that has useful empirical properties (McFadden, 1973). These models are used to study which and how parents value different school attributes [5] (see Holme (2002), Hastings et al. (2005) and Bosetti (2004)).

In analyzing how parents choose schools, Hastings et al. (2009) investigate the importance of socio-economic heterogeneity in preferences to explain inequality in enrollment at good schools. Along the same lines, using a random utility model that allows for choice-specific unobservable characteristics and deals with potential endogeneity, Gallego and Hernando (2010) study how parents choose schools. They find that the two crucial school attributes are test scores and distance to school. Their results suggest there is much heterogeneity in preferences and that the valuation of most school attributes depends on household characteristics.[6]

Classic discrete choice models, such as Logit, Probit, Multinomial Logit, and Mixed Logit, among many others, require the econometrician specify the agent's choice set to achieve point identification.

[5] Such as test scores, tuition, commuting time, number of students per class, number of teachers per student, curriculum, and religious orientation, among many others.

[6] This and similar literature motivates my utility function specification and why I argue that accounting for heterogeneity in preferences is necessary.

However, assumptions must be made because a common challenge in the school choice literature is that agents often have imperfect information about the available schools and their characteristics. Nevertheless, assumptions about what information the agent has or does not have can bias results because agents knowing all relevant alternatives is vital when using a revealed preference approach.

For example, Hastings et al. (2007) use a field experiment to examine the degree to which information costs impact parental choices and their revealed preferences for academic achievement. They find that providing families with simplified information significantly increases the average test score of their chosen school.

Along the same lines, Neilson et al. (2019) study how a personalized information provision intervention that targets families of public schools Pre-K students entering elementary schools in Chile changes parents' choices. This intervention resulted in parents' choices shifting toward schools with higher average test scores, higher value-added, higher prices, and ones further from their homes.

Nevertheless, when analyzing parents' decision-making process, school attributes are not the only variables at play because there can also be a strategic component. Depending on each cities' school admission system, some students may have strong incentives to game the system, which can be enormously prejudicial to non-strategic students (see Abdulkadiroğlu et al. (2009)). Thus, designing these systems is a challenge for policymakers because efficiency is not their only desirable attribute.[7]

In one of their seminal papers, Abdulkadiroğlu and Sönmez (2003) were the first to formulate the school choice problem as a mechanism design problem. They analyzed existing school choice plans (Boston, Columbus, Minneapolis, and Seattle) and exposed profound shortcomings such as their vulnerability to preference manipulation and inefficiency.[8] They proposed a practical solution for these critical issues in the form of two alternative mechanisms; the Gale-Shapley Student Optimal Stable Mechanism and Top Trading Cycles Mechanism, both of which are strategy-proof and Pareto efficient.

Although mechanisms being strategy-proof is important from a social welfare point of view, it can also affect research studies because strategic games are not always truth-telling. In other words, parents have incentives to lie about their preferences when ranking schools in the application process to secure a place in more popular schools (see Abdulkadiroğlu and Sönmez (2003)).

Parents lying about their true preferences is problematic because most discrete choice models use revealed preferences at their core. The Chilean Centralized School Admissions System uses the Gale-Shapley Deferred Acceptance Algorithm (see Abdulkadiroğlu and Sönmez (2003)), which is strategy-proof and Pareto efficient. Thus, truth-telling and strategic behavior are not a direct concern for the time being.[9]

In all, the school choice literature has relied almost exclusively on observable family and school attributes to model parents' decision-making process. I argue that accounting for heterogeneity and incomplete information, as some previously mentioned work, is vital when modeling parents' behavior

[7] Efficiency in terms of the allocation being Pareto-optimal.

[8] In terms of school seats allocation.

[9] It is worth noting that although an admission mechanism may be strategy-proof, agents making the decision must be aware of it for them to be truth-telling. nevertheless, analyzing possible strategic behaviors of parents goes beyond the scope of this paper.

because parents' choice sets can ultimately drive their choice. My contribution to the literature is studying the effect that accounting for parents' limited access to information has on preference estimation. I assume parents' choice sets are unobservable and estimate a model that allows for this using partial identification. These new insights could help shape future policies by highlighting the crucial role of information in parents' school choices.

3 Model and Partial Identification of its Parameters

I infer parents' preferences for schools without explicitly defining their choice sets using the random utility model developed by McFadden (1974) as my general framework. However, I modify the key assumption that states the econometrician observes agents' choice sets. In doing so, I do not find exact parameters (i.e., my model's parameters will not be point identified) but rather a partial identification region containing the set of parameter values compatible with the available data (hereafter the identified set).

Starting from the random utility model developed by McFadden(1974), consider that each parent i applies their child to school $c \in \mathcal{D}$, where $\mathcal{D}$ is the feasible set of schools.[10] There exists a function U_i drawn from $\mathcal{U}$ according to some probability distribution such that:

$$d \in^* C_i \Leftrightarrow U_i(d) \geqslant U_i(c) \qquad \forall c \in C_i, c \neq d \tag{1}$$

where ϵ^* denotes "is chosen from" and $C_i \subseteq \mathcal{D}$ denotes the agent's choice set.

Each student has a vector of observable attributes $\mathbf{x}_{ic} = (t_i, Z_{ic})$ where t_i denotes a socioeconomic indicator that classifies students as a priority, preferential or regular student. Z_{ic} is a student-school specific variable which represents the distance from student i's home to school c. Additionally, schools have an array of observable attributes: the schools' score on a national standardized test (SIMCE), the schools' tuition costs, and the schools' percentage of priority students, denoted by S_c, P_c and R_c respectively.

My baseline model assumes parents only value the quality of the school (for which I use SIMCE scores as a proxy) and its student body composition (the ratio of priority to regular students). Later I expand my baseline model to include tuition costs, a socioeconomic indicator, and parents' commuting time as explanatory variables.

The utility function of my baseline model is:

$$U_{ic} = \delta_1 \cdot S_c - \delta_2 \cdot R_c + \nu_{ic}$$

Additionally, I assume that each parent $i \in \mathcal{I}$ is further characterized by a real valued vector of unobservable attributes ν_i, which are idiosyncratic to the parent. Let $\mathcal{X}$ and $\mathcal{V}$ denote the supports of $\mathbf{x}_{ic}$ and ν_i respectively.

Following the distributional assumptions in standard discrete choice models [11], I impose the following restrictions on the distribution of U_i.

[10] This set consists of every school located in the students' city offering spaces during the 2020 admissions cycle.

[11] such as the conditional logit model (McFadden, 1974) and mixed logit model (McFadden and Train, 2000)

Assumption 1:

(I) *There exists a function* $W : \mathcal{X} \times \mathcal{V} \mapsto \mathbb{R}$, *known up to a finite dimensional parameter vector* $\boldsymbol{\delta} \in \Delta \subset \mathbb{R}^k$, *where* Δ *is a convex compact parameter space, and continuous in each of its arguments such that* $U_i(c) = W(\mathbf{x}_{ic}, \nu_i; \boldsymbol{\delta})$ *for all* $c \in \mathcal{D}, (\mathbf{x}_{ic}, \boldsymbol{\nu}_i)$ - a.s.

Most discrete choice models also make the two following assumptions:

(i) *A random sample of choice sets* C_i, *choices* d_i, *and attributes* $\mathbf{x}_i$ $\{(C_i, d_i, \mathbf{x}_i) : d_i \in^* C_i, i \in I \subset \mathcal{I}\}$, *is observed.*

(ii) $|C_i| \geqslant 2$ *for all* $i \in \mathcal{I}$, *where* $|\cdot|$ *denotes set cardinality.*

The key difference between my approach and the mainstream discrete choice literature lies in the econometrician's assumptions about choice set observability. I make the following assumption:

Assumption 2:

1. *A random sample of choices* d_i *and attributes* $\mathbf{x}_i, \{(d_i, \mathbf{x}_i) : i \in I \subset \mathcal{I}\}$, *is observed.*
2. $\Pr(|C_i| \geqslant \kappa) = 1$ *for all* $i \in \mathcal{I}$, *where* $\kappa \geqslant 2$ *is a known scalar.*

While Assumption 2.2(II) is comparable to (ii), assumption 2.2(I), omits the requirement that the agents' choice sets are observed, making it the key point of departure from McFadden.

I assume that $\Pr(\ell_i \geqslant \kappa) = 1$ for every student $i \in \mathcal{I}$, where $\ell_i = |C_i|$, $\kappa \geqslant 2$ and that ℓ_i conditional on $(\mathbf{x}_i, \nu_i)$ follows a discrete distribution.[12] Because the number of inequalities grow superlinearly with $|C|$, I can only work with relatively small choice sets for the time being. In particular each parents' choice sets consists of seven alternatives, i.e. $|C_i| = 7$ (See section 4 for details on the sample used for estimation).

Following Barseghyan et al., my final assumption is:

Assumption 3:

Agent i *draws a choice set of size* ℓ_i *such that:*

$$\Pr(\ell_i = q | \boldsymbol{\nu}_i) = \Pr(\ell_i = q) = \pi(q), \quad q = \kappa, \ldots, |\mathcal{D}|$$

where $\pi(q) \geqslant 0$ for $q \geqslant \kappa$ and $\sum_{q=\kappa}^{|\mathcal{D}|} \pi(q) = 1$.

It is worth mentioning that although $\pi(q)$ is a parameter in my model, [13] one could define a function in which families with different observable attributes observe choice sets of specific sizes with different probabilities.[14]

[12] This is, parents can observe different schools and have choice sets of different sizes.

[13] In section 5.3 I use comparative statics to show how the identified set changes with different values of $\pi(q)$.

[14] This is motivated by the idea that the parents of priority students have less access to information about schools (due to lack of technology, time, or skills, among many others) and will therefore choose, on average, from smaller choice sets compared to parents from higher socioeconomic classes.

The only restriction this assumption posits is that the distributional family of ℓ_i is a known parametric class and that ℓ_i is independent of ν_i. This independence assumption could be problematic because it implies that each parents' choice set size is independent of their unobservable attributes, which may not always be accurate. For example, parents who give more importance to their children's education will likely research available schools more and thus have larger choice sets.

Heroic as this assumption may be, it is somewhat analogous to many econometric models' common unconfoundedness and exogeneity assumption used for causal inference. Furthermore, conditional on ℓ_i, the model continues to allow for any dependence structure, without restriction, between parents' choice sets and their observable attributes and, conditional on observables, between parents' choice sets and their unobservable attributes. Moreover, agents may have choice sets that have different compositions even when they are the same size.

In all, I assume school choices and observable attributes, $\{(d_i, \mathbf{x}_i) : i \in I\}$, for a random sample of students $I \subset \mathcal{I}, |I| = n$, are observed, but that the parents' choice sets, $\{C_i : C_i \subseteq \mathcal{D}, i \in I\}$, are unobserved. Given $(\mathbf{x}_i, \nu_i)$ and choice set $C_i = G \subseteq \mathcal{D}$, if the model is correctly specified, the agent's observed choice d_i satisfies:

$$d_i^* \left(G; \mathbf{x}_i, \nu_i\right) = \arg\max_{c \in G} U \left(\mathbf{x}_{ic}, \nu_i\right)$$

for the data generating process of the model's parameters $\boldsymbol{\theta} = [\boldsymbol{\delta_1}, \boldsymbol{\delta_2}]$

Given κ, the set of optimal choices for all possible realizations $G \subseteq \mathcal{D}, |G| \geqslant \kappa$, is:

$$D_\kappa^* \left(\mathbf{x}_i, \nu_i\right) = \bigcup_{G \subseteq \mathcal{D} : |G| \geqslant \kappa} \left\{d_i^* \left(G; \mathbf{x}_i, \nu_i\right)\right\} = \bigcup_{G \sqsubseteq \mathcal{D} : |G| = \kappa} \left\{d_i^* \left(G; \mathbf{x}_i, \nu_i\right)\right\}$$

This model implies a set of multiple optimal choices $D_\kappa^* \left(\mathbf{x}_i, \nu_i\right)$ for each parent as a result of the multiple possible realizations of their choice set. This multiplicity is precisely what precludes point identification of the model's parameters in the absence of additional restrictions on the choice set formation process. I can only partially identify my model for this very reason.

Following Barseghyan et al., I use a result in Artstein (1983), that translates equation (1) into a finite number of conditional moment inequalities that fully characterize the sharp identification region $\boldsymbol{\theta}$ as the set of values of the parameter vector θ for which the inequalities hold. Thus, the sharp identification region Θ_I of the parameter vector $\boldsymbol{\theta} = [\boldsymbol{\delta_1}, \boldsymbol{\delta_2}]$ is given by:

$$\Theta_I = \left\{ \boldsymbol{\theta} \in \Theta : \Pr(d \in K \mid \mathbf{x}) \leqslant \sum_{q=\kappa}^{|\mathcal{D}|} \pi(q) P \left(D_q^*(\mathbf{x}, \nu; \boldsymbol{\delta}) \cap K \neq \varnothing ; \gamma \right), \forall K \subset \mathcal{D}, \mathbf{x} - \text{ a.s.} \right\} \tag{2}$$

For each $K \subset \mathcal{D}$, I estimate the left hand side of inequality (2) from data on students' applications by city from the Chilean centralized admission system. The right hand side is a model defined function of x_i known up to $\boldsymbol{\theta}$, and thus, a "theoretical probability" in the sense that data to compute such probabilities does not exist.[15] To compute $P\left(D_{\kappa}^{*}\left(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}\right) \cap K \neq \varnothing; \gamma\right)$ I exploit the logit closed-form choice probabilities. Note that 127 inequalities must be satisfied when $\mathcal{D} = |7|$.[16] The identified set contains all and only those values of the parameter for which these inequalities hold. The MATLAB code for the baseline model can be found in the Appendix.The codes for the model's extensions are analogous to that of the baseline model and are available upon request.[17]

4 Data Description

Because of the computational issues I discuss in section 5.1, I only use data from the city of Ovalle to estimate my model. Ovalle is a city in the Coquimbo Region with a population of more than 112.000.[18] The centralized admissions system platform had 4,689 establishments offering spaces for Pre-Kinder students in the 2019 application cycle, of which 42 were in Ovalle. I use data from the twenty two establishments located inside the city and leave out twenty schools located in rural areas. I do not use data from rural schools because they are smaller than non-rural schools and are located outside the city. On average, rural schools have classes that are less than half the size of non-rural schools.

During this admissions cycle, Ovalle had 655 students applying to pre-Kinder using the centralized admissions system and on average parents applied to three schools.[19] It is worth noting that the number of schools parents apply to does not necessarily coincide with the size of their choice sets. For example, parents may not rank and apply to all feasible schools if they believe their child has no chance of getting into a certain school[20] or if the child will be accepted by one of the first few schools on their list

I construct my database using several administrative data sources from the Ministry of Education of Chile (MINEDUC) which can be found at their website [21]. First, I use records containing student-level information of basic demographic information such as home address and parent's income. Using these demographic characteristics, I control for observed heterogeneity non-parametrically when estimating parents' school preferences.

My second source of data is individual-level eligibility for the Subvención Escolar Preferencial (SEP) targeted voucher system and the student's socio-economic level (priority, preferential or non-vulnerable student).

My third source of administrative data is the average test scores for different subjects from the 2nd, 4th, 6th, and 8th-grade SIMCE test.[22] In section 5, I use this variable as a proxy for school quality which is a school attribute that directly affects parents' utility function.

[15] I refer to these probabilities as the model implied probabilities.

[16] The number of possible subsets for a set of seven elements is $2^7 - 1$. The empty subset is redundant because the inequality for $K = \{\emptyset\}$ will always hold; therefore, there is one subset that is subtracted from the total.

[17] dmselman@uc.cl

[18] According to the 2017 census of the National Statistics Institute.

[19] compared to the two and a half country average.

[20] Schools offer limited spots each year and some schools are more popular than others.

[21] https://datosabiertos.mineduc.cl/

[22] These include math, language, social and natural science, history, and geography.

Finally, I use parents' application submissions made on the centralized admissions system platform (applications made in 2019 for 2020 admission). I only use the first preference declared in each student's application because exploiting the richness of the information in how parents rank schools within their applications goes beyond the scope of this paper.[23]

Although the data contains applications to Pre Kinder (PK), Kindergarten, 1st grade, 7th grade, and 9th grade, I only consider students enrolling in PK to exclude strategic decisions or students changing schools for endogenous reasons.[24] Finally, I only analyze the "regular" process applications, leaving out the "complementary" process ones. The "complementary" process is the second round of applications for families who want to change their previous one or did not apply in the first one. Parents' applications from these cycles are not comparable because the information available during each process is entirely different, leading to a possible strategic component behind parents' applications in the complementary round.

5 Main Results

5.1 Computational Challenges

The number of inequalities my algorithm must check to find the identified set increases non-linearly with every additional school the feasible set has. Since larger cities often have more schools, I cannot use data from big cities, and I can only estimate my model for a specific subset of them for the time being.

For example, if I were to expand the size of the choice set from $|\mathcal{D}| = 7$ to $|\mathcal{D}| = 8$, instead of having to check 128 inequalities, I would have to check 256. Further, 512 inequalities have to be checked if $|\mathcal{D}| = 9$, and so on. In all, my method quickly loses computational tractability as parents' feasible choice set grows.[25]

Another dimension in which my method loses computational tractability is the density of the parameters' grid. Since the algorithm must check every possible combination of parameters to find the partial identification region, marginally increasing the number of points in my grid disproportionately increases the number of times the algorithm must loop. Suppose I choose a grid for δ_1 and δ_2 of 10 points evenly spaced between (-1,1). This grid composition means that there are $10 * 10 = 100$ possible combinations of parameters that the algorithm must check to find the partial identification region. If I increase the density of both grids to 20, there would be 400 combinations of δ_1 and δ_2 that the algorithm must check.

Along the same lines, if I were to increase the number of parameters in my model, possible parameter combinations would increase even faster. When the model has three parameters, 1000 combinations must be checked for each inequality when grids have 10 points. For grids of 20 points, 8000 combinations must be checked and so on. This superlinear increase complicates marginal extensions of my model.

[23] It would be interesting to see if something like a Multinomial Logit Model fits in this application of partial identification.

[24] Parents may transfer their children to other schools for geographic, socio-economic or other personal reasons.

[25] It is worth noting that my algorithm works so that as soon as one inequality is not satisfied for a given set of parameters, the loop breaks and moves one to another combination of parameters. However, as κ decreases, more inequalities are satisfied because the model becomes more flexible, which means that although the loop eventually breaks, it still has to check a substantial amount of inequalities before it does.

Table 1 shows how computation time grows as the model becomes more complex. I estimated all models using three different grid sizes. The first three rows shows how long the algorithm takes to estimate the model when the feasible choice set has seven, eight and nine schools, using grids of 50 points for each model parameter. The following three rows show how long it takes when the grid has 100 points, and the last three, when each grid has 200 points. Columns denote the number of parameters being estimated in the model. As the size of the feasible set grows, so does computation time, however, this increase becomes more relevant as the number of parameters being estimated in the model grow. When grid density increases, the differences in time between estimating a model with an additional parameter or increasing the choice set size become significant.

Table 1: Time it takes to estimate different models

$\|\mathcal{D}\|$	δ_1	δ_1, δ_2	$\delta_1, \delta_2, \delta_3$	$\delta_1, \delta_2, \delta_3, \delta_4$
7	0.013 sec	1.11 sec	29.31 sec	9 hours
8	0.04 sec	1.13 sec	37.02 sec	10 hours
9	0.07 sec	1.4 sec	56.1 sec	13 hours
7	0.02 sec	3.26 sec	5.4 min	7 days
8	0.05 sec	3.9 sec	6.7 min	*
9	0.1 sec	4.6 sec	8.3 min	*
7	0.04 sec	12.6 sec	37.1 min	*
8	0.07 sec	13.39 sec	48.5 min	*
9	0.15 sec	18.44 sec	1.4 hours	*

*Estimation takes more than one week.

With this in mind, to be included in the sample, a city must meet two criteria; have at least 400 students applying to one or more schools, and its seven most popular schools must cover more than 70% of the total applications in that city. These seven schools compose the feasible set $\mathcal{D}$. Alto Hospicio, Buin and Ovalle are the only counties that meet both requirements. Thus far, I have only estimated my model using the data available from the city of Ovalle. Therefore, for context, all results showed hereafter are from this city's data set.

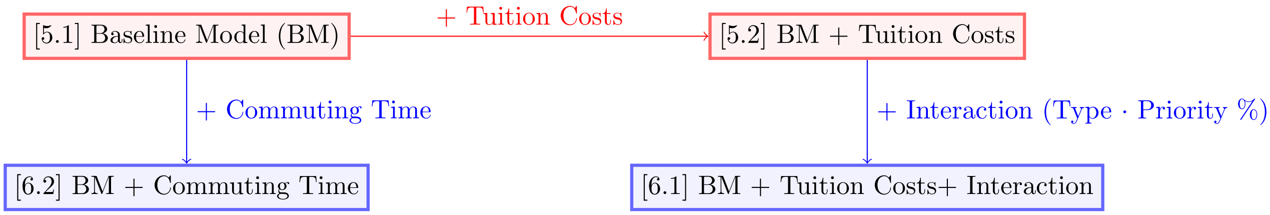
Finding cities that meet these criteria is challenging because counties that have a relatively small number of schools that cover 70% of applications often have less than 400 students, but bigger counties (with more than 400 students) have too many schools to find any seven of them that cover 70% of all applications. The ideal city is one with a lot of students but with big schools. On average, a city's nine most popular schools cover more than 70% of its applications.

If I modify the requirements and made 350 the minimum number of students per city, five counties would have seven schools covering more than 70% of the city‘s applications. Further, if the minimum number of students per city were 300, nine counties would meet the criteria. If I were to maintain the requirement of counties having a least 400 students, but I lowered the percentage of applications the seven most popular schools must cover from 70% to 65%, then six counties would be eligible. If I lowered that percentage to 60%, then sixteen counties would meet the criteria.

My results suggest that assumptions on agents' choice sets play a crucial role in my model's estimation. Under the assumption that agents choose from the full choice set ($\kappa = 7$),[26] there is no set of parameters for which any of the models below can rationalize the data.[27] However, as I relaxed the assumption that agents only draw full choice sets and introduced the possibility that agents' minimum choice set size is five (full minus two), my model can rationalize the data.[28] Nevertheless, the probabilities with which agents draw full, full minus two, or full minus one choice sets affect the existence and size of non-empty identified sets.

To better understand how the estimation behind my model works, I show how the parameters that can rationalize the data respond to changes in the amount and type of variables I add to the model. As illustrated in the diagram below, I start with a baseline model and extend it in four main ways.

The rest of this paper is organized as follows; in section 5.1, I use my baseline model to show how the probabilities of drawing choice sets of a specific sizes ($\pi(q)$) affect the partial identification region. In section 5.2, I continue by adding a third school-specific variable to illustrate how adding a (relevant) variable expands the identification region because the model gains explanatory power. In section 6.1 I extend the model to account for observed student heterogeneity by adding an interaction between the ratio of priority students and the type of student. Finally, in section 6.2 I extend the model by adding a student-school specific variable to my baseline model to account for parents commuting time to school.


[5.1] Baseline Model (BM)
+ Tuition Costs
[5.2] BM + Tuition Costs
+ Commuting Time
+ Interaction (Type · Priority %)
[6.2] BM + Commuting Time
[6.1] BM + Tuition Costs+ Interaction

5.2 Estimation Restrictions and Flexibility

Before presenting the results from my baseline model, I explain how I can make my model more flexible with κ or restrict it using student-specific or student-school-specific variables. Recall that every combination of δ_1 and δ_2 that satisfies inequality (3) forms the partial identification region. Its left-hand side is the empirical probability that a parent chooses a specific school from a subset K of schools, while the right-hand side is the model implied probability. As κ decreases, the right-hand side increases for a given set of parameters and schools, while the left-hand side remains constant.

Whenever agents do not observe the complete choice set, their probability of choosing a given school increases because the school removed from the choice set is its strongest competitor.[29] In other words,

[26] The standard assumption in most discrete choice models.

[27] To rule out the possibility of this being due to programming issues, I simulated data consistent with a set of model parameters to make sure my algorithm was correctly coded.

[28] When agents draw choice sets with $\kappa = 6$, my baseline model could not rationalize the data for any combination of probabilities, so I relaxed κ even more. My baseline model only begins to rationalize the data when the minimum choice set size is $|C| = 5$ (full minus two).

[29] In terms of their utility for a given set of model parameters. Thus, a school's strongest competitor may change for different sets of model parameters. Because every school has different observable attributes (such as tuition, test scores, and priority ratios), the "best" and "worst" competitors for a given school can change depending on which attributes have relatively more weight in the utility function.

as κ decreases, the probability of a parent choosing a given school increases because it competes against "worst" schools. For example, if an agent chooses from a set of only one school ($\kappa = 1$), inequality (2) is always satisfied because any set of parameters can rationalize a school being chosen if it is the only available one.

Furthermore, as the probability with which parents observe smaller choice sets increases ($\pi(q) < \pi(q-1)$), the right-hand side of inequality (2) also increases. The mechanism behind this is as follows; the probability of a given school being chosen grows as choice sets get smaller, thus, if the probability of drawing the relatively smaller choice sets is bigger, the weighted probability of that school being chosen is bigger. Recall that parents are more likely to choose a given school when the choice set is relatively smaller because it faces weaker competitors as it shrinks.

Therefore, the partial identification region grows when κ is smaller, or the cases in which a given school is chosen with higher probability have relatively more weight. The region grows because it is more likely that a combination of parameters will satisfy the inequality for every K as the right-hand side grows and the left-hand side does not change.

In the first extension of my baseline model I add tuition costs and in the second one I account for observed heterogeneity parametrically by adding an interaction between the schools' priority to non-priority student ratio and the type of student applying to that school.

For the third extension, I include a variable that accounts for parents' commuting time to each school. I do so by clustering students into ten groups based on where they live. Then, I calculate the distance between the center of each cluster and every school in the feasible set. Figure 1 shows Ovalle divided into the ten clusters I used to estimate the model in section 6.2. I use these clusters because I cannot estimate the model at an individual level[30] because I would lack statistical power, so I use clusters to represent different "types" of students based on their geographic location.

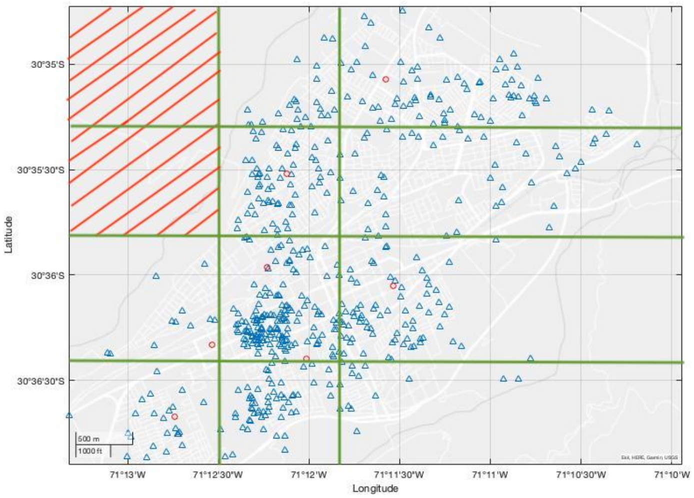


Figure 1: Map of Ovalle divided into the ten clusters used for the estimation. The first two quadrants were left out because no students nor schools were inside of either of them. Triangle and circle symbolize students and schools, respectively. The delimitation used to cluster Ovalle is a function of the distance between schools.

[30] Where each student would be a cluster and I would calculate the distance between their homes to every school, as most discrete choice models do.

Because these variables are student-school-specific, I must consider a trade-off before adding them to my baseline model. On the one hand, the model has more explanatory power because of the additional variable (assuming it is relevant to the model). On the other hand, accounting for said heterogeneity increases the number of moment inequalities that need to be satisfied for each set of parameters. In particular, the number of moment inequalities that need to be satisfied doubles when I add the interaction variable because every condition that needed to be satisfied before needs to be satisfied for the group of priority students and the non-priority students now.

To illustrate this trade-off, consider the extension in section 6.2, that consists on adding a proxy for commuting time to the baseline model. When this variable is included, the same set of parameters must allow the simultaneous rationalization of the choices made by every group.[31] Therefore, although including a variable that adds explanatory power seems like an obvious thing to do in some models, one must determine if the marginal information it provides out-weights the restrictions it imposes.

The increase of the amount of moment inequalities is not the only problem these student-school-specific variables posit. If some clusters are sufficiently small (in terms of students), a couple of students applying to one school makes the empirical probability of that school being chosen relatively high. If some clusters have schools that are chosen with disproportionately high probabilities, it is harder for a model to rationalize data from different clusters using the same set of parameters. The model's inability to rationalize data from small clusters when I add these student-school-specific variables shrinks the identification region, as I show in section 6.

5.3 Baseline Model

Recall my baseline model from section 3:

$$U_{ic} = \delta_1 \cdot S_c - \delta_2 \cdot R_c + \nu_{ic}$$

Figure 2 shows that the identification region is empty for $\kappa = 7$, which illustrates how my baseline model cannot rationalize the data under the assumption that parents observe every school in the feasible set.[32] Nevertheless, once I relax the assumption that parents' choice sets coincide with their feasible set (i.e., $\kappa < |\mathcal{D}|$), my model can rationalize the data, thus, a non-empty identification region exists.

[31] Computationally speaking, this means that a set of parameters must hold for all 1270 inequalities in order to form part of the identification region.

[32] This model cannot rationalize the data under the assumption that parents observe the full minus one choice set or the full minus two choice set either. As I show below, the model needs more flexibility in terms of κ because no combination of probabilities $\pi(q)$ can explain the data when $\kappa = 7$, $\kappa = 6$ or $\kappa = 5$.

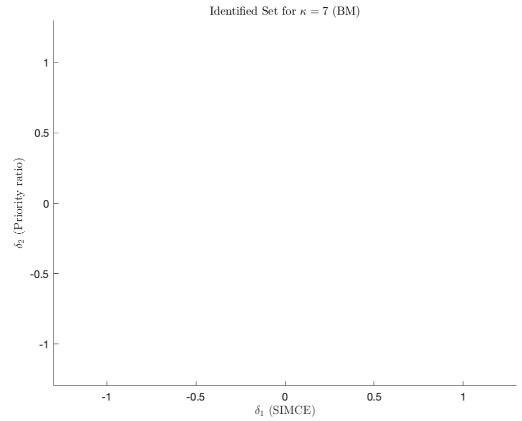


Figure 2: $\pi_1 = 1$

Figure 3 shows the identified set when $\kappa = 4$, where agents draw choice sets of a specific size with probability $\pi_1 = 0,1$, $\pi_2 = 0,1$, $\pi_3 = 0,4$, $\pi_4 = 0,4$. These probabilities mean that parents draw choice sets with seven schools (full choice set) 10% of the time. Along the same lines, parents draw choice sets of five schools (full minus two) 40% of the time, and so on. The positive slope of the identified set suggests that as the priority ratio matters more to parents, so does the schools' test scores, and vice-versa.

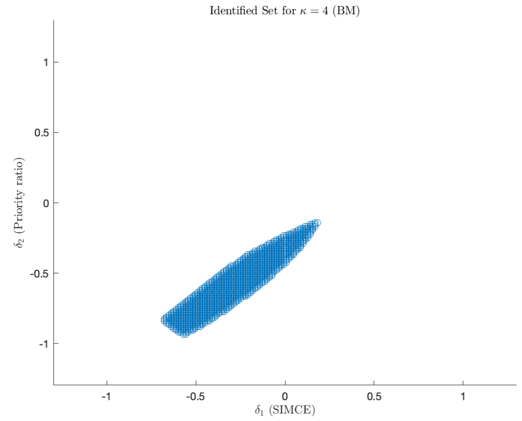


Figure 3: $\pi_1 = 0.1$; $\pi_2 = 0.1$; $\pi_3 = 0.4$; $\pi_4 = 0.4$

Figures 4, 5, 6 and 7 illustrate how the partial identification region changes when modifying the distribution of the probabilities (figures 4 and 5) and the support of the probabilities (figures 6 and 7). Figures 4 and 5 show that as it becomes less likely that parents observe the full choice set, the partial identification region grows but maintains its shape for the most part, so its interpretation is the same as that of figure 3. Note that these new partial identification regions grow, but there is no shift of the original identified set; all parameter values that could rationalize the data before, still can.

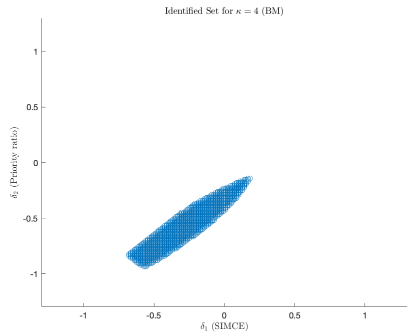


Figure 4: $\pi_1 = 0.1$; $\pi_2 = 0.1$; $\pi_3 = 0.4$; $\pi_4 = 0.4$

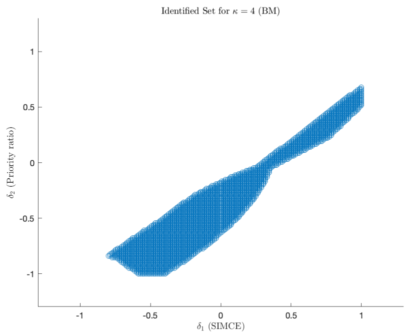


Figure 5: $\pi_1 = 0.025$; $\pi_2 = 0.025$; $\pi_3 = 0.45$; $\pi_4 = 0.5$

Figures 6 and 7 show how as the minimum choice set size decreases (now agents can draw smaller choice sets), the partial identification region also grows. Even though the (positive) relationship between parents' preferences on schools' priority ratio and test scores remains, the shape of the partial identification region does change. The new parameter combinations that become part of the partial identification region introduce the possibility of a trade-off in which parents may highly value schools' tests scores but do not care about its priority ratio, and vice-versa.

This new trade-off may seem paradoxical because, according to this figure, parents can have strong preferences for school test scores and strong preferences for a high priority ratio or a low one. However, in this case, the probabilities are such that the most likely scenario is that parents are only choosing schools from choice sets with three of the seven alternatives, which in the absence of additional restrictions gives so much flexibility to the mode that it can rationalize almost any parameter combination.

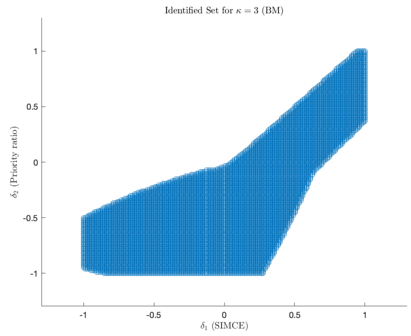


Figure 6: $\pi_1 = 0.05$; $\pi_2 = 0.1$; $\pi_3 = 0.2$; $\pi_4 = 0.25$; $\pi_5 = 0.4$

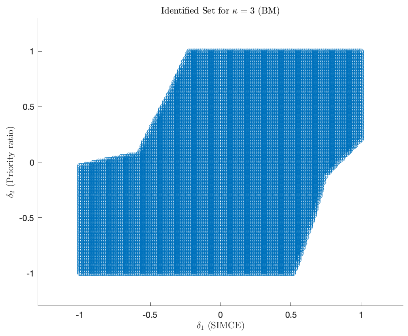


Figure 7: $\pi_1 = 0.025$; $\pi_2 = 0.025$; $\pi_3 = 0.05$; $\pi_4 = 0.1$; $\pi_5 = 0.8$

5.4 Baseline Model with tuition costs

For my first extension, I add tuition costs, a school-specific variable, to the baseline model. The new utility function is defined by:

$$U_{ic} = \delta_1 \cdot S_c - \delta_2 \cdot R_c - \delta_3 \cdot P_c + \nu_{ic}$$

Figure 8 shows the new model's identified set when $\kappa = 4$ and is analogous to figure 3. However now the figure is three-dimensional because this model partially identifies three parameters.

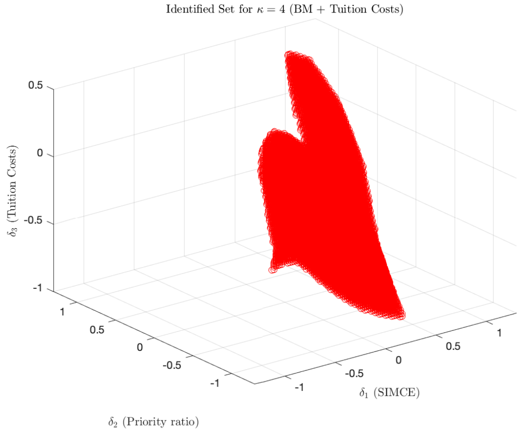


Figure 8: $\pi_1 = 0.025$; $\pi_2 = 0.025$; $\pi_3 = 0.45$; $\pi_4 = 0.5$

Figure 10 shows the partial identification region of this new model in two of its three dimensions to compare it to the baseline model. The positive correlation between preferences on a schools' priority ratio and test scores from the baseline model (Figure 9) remains when I add tuition costs to my baseline model (Figure 10). However, the identified region for δ_1 and δ_2 grows, which means this new variable adds relevant information to the model because now the model can rationalize the data for a broader set of parameters.

This extension has the same implications for the partial identification region as changing the probability support (see figure 6) but through a different mechanism. For this extension, I add a variable that can explain choices that could not be rationalized by a model that only considered priority ratio and test scores as relevant attributes for parents. After changing the minimum choice set size, the model can rationalize choices it could not before. For example, it can rationalize cases in which parents choose a school that is strictly dominated, by allowing for the possibility that parents were not aware that this better school existed or that it was a feasible option.

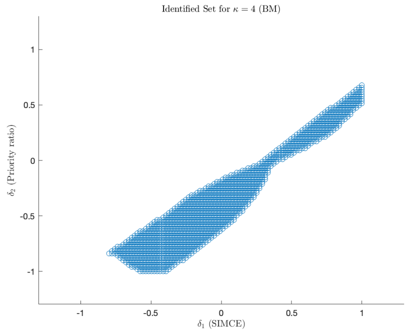


Figure 9: $\pi_1 = 0.025$; $\pi_2 = 0.025$; $\pi_3 = 0.45$; $\pi_4 = 0.5$

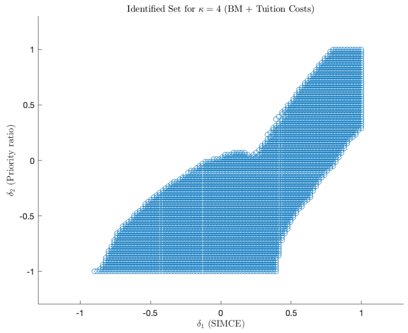


Figure 10: $\pi_1 = 0.025$; $\pi_2 = 0.025$; $\pi_3 = 0.45$; $\pi_4 = 0.5$

6 Extended Model Results

6.1 Baseline Model with tuition costs and student specific interaction

For my second extension, I add a student-school-specific variable that accounts for student heterogeneity. Using an interaction between a schools' priority ratio and the type of student applying to that school. The new utility function is:

$$U_{ic} = \delta_1 \cdot S_c - \delta_2 \cdot R_c - \delta_3 \cdot P_c + \delta_4 \cdot R_c \cdot t_i + \nu_{ic}$$

To account for heterogeneity in preferences of student body composition by type of student, I use a dummy variable t_i that equals one if the student applying is part of the "priority students group" or zero if the applicant is a regular student. Parameter δ_4 captures the additional effect that a schools' demographic composition has on a priority student choosing a school over regular students.

Comparing figures 11 and 12 we see that the partial identification region also grows with this extension, and the positive relationship between δ_1 and δ_2 remains.[33] The fact that the partial identification region grew with this extension illustrates how the explanatory power gained by this additional variable more than compensates its additional restrictions. Nevertheless, the partial identification region's shift suggests that parents value a schools' test scores more when the type of student is accounted for.

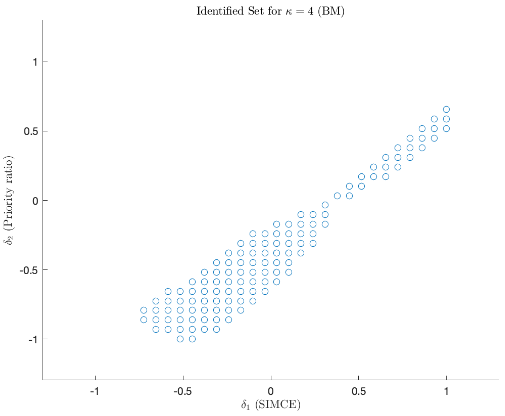


Figure 11: $\pi_1 = 0.025$; $\pi_2 = 0.025$; $\pi_3 = 0.45$; $\pi_4 = 0.5$

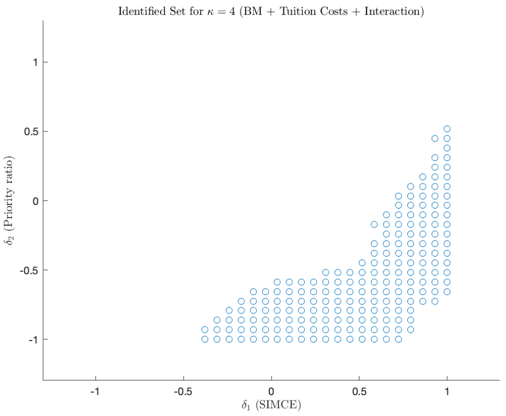


Figure 12: $\pi_1 = 0.025$; $\pi_2 = 0.025$; $\pi_3 = 0.45$; $\pi_4 = 0.5$

Note that this shift does not contradict that when extending a model, all points that were part of the partial identification region before must continue to be. In this case, I am adding an interaction between the student's type and the school's student body composition ratio student-specific variable, which introduces new restrictions to the model, as opposed to my first extension, where I added a new variable that was not specific to students nor schools.

[33] When running the algorithm, I used grids that were less dense I added more parameters with each extension of the baseline model because the time it took to run grew rapidly (see section 5.1). I run the baseline model again for each extension using the same grid density to make the figures comparable, which is why the baseline model identified set may appear to change throughout the paper.

6.2 Baseline Model with commuting time

My last extension includes a variable that accounts for parents' commuting time to and from school. This variable is often used in discrete choice models of school choice because evidence suggests it is an important driver behind the parents' decision (see section 2). The utility function, in this case, is given by:

$$U_{ic} = \delta_1 \cdot S_c - \delta_2 \cdot R_c - \delta_3 \cdot P_c - \delta_4 \cdot Z_{ic} + \nu_{ic}$$

Commuting distance, denoted by Z_{ic}, is a student-school-specific variable.[34] Note that commuting distance is different for each student across clusters and across schools. I group students in clusters because I use school-level data for my estimation. Logit models can account for each student's commuting distance and do not need to cluster them because their estimation is done using student-level data.

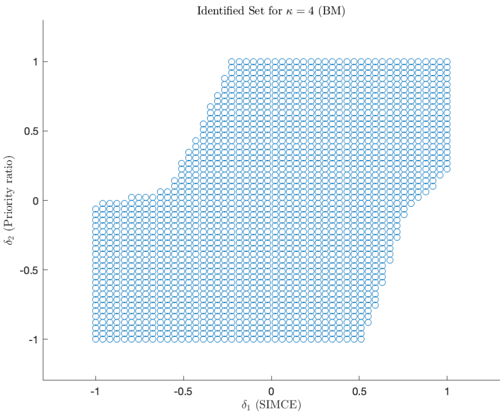


Figure 13: $\pi_1 = 0.025$; $\pi_2 = 0.025$; $\pi_3 = 0.05$; $\pi_4 = 0.1$; $\pi_2 = 0.8$

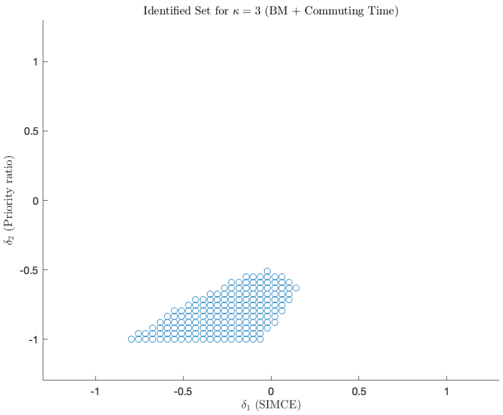


Figure 14: $\pi_1 = 0.025$; $\pi_2 = 0.025$; $\pi_3 = 0.05$; $\pi_4 = 0.1$; $\pi_2 = 0.8$

Comparing figures 13 and 14 it is clear that the explanatory power added by this new variable does not compensate for its additional restrictions. The identified set significantly shrinks when I account for commuting time (as opposed to the extension in section 6.1) because the amount of inequalities that need to be satisfied for each set of parameters significantly increases. In extension 6.1, the amount of inequalities doubles when accounting for student type, whereas it multiplies by ten when accounting for commuting time.[35] As I explain in section 5.1) , the mechanism behind this is that now the same set of parameters must rationalize ten times more data than before, which inevitably shrinks the set of parameters that can do so. Therefore, the restrictions that result from adding an interaction between student type and priority ratio are not enough to compensate the explanatory power gained. However, accounting for commuting distance, one can see that this variable imposes such strong restrictions on the estimation that the explanatory power it brings to the model cannot compensate them.

[34] Parents' commuting distance to school is the shortest distance between the center of the cluster they belong to and the school.

[35] The amount it multiplies by depends on the number of clusters. In this case it multiplies by ten because there are ten clusters

7 Concluding Remarks

The current literature on school choice has, for the most part, circumvented the challenges of preference estimation with unobserved choice sets by making strong assumptions about their formation process and composition. Using a frontier robust method of discrete choice analysis proposed by Barseghyan et al., I show how assumptions on parents' choice sets have important implications when estimating parents' preference parameters. I analyze these implications by studying the effects of marginally changing the minimum size of the choice sets parents drew from (and the probabilities they did so with) and extending my baseline model in two main ways.

With my first extension, I show how adding relevant variables to the model increase its explanatory power, and therefore, its partial identification region. My second extension illustrates the trade-off associated with adding student-school-specific interactions to the model. Moreover, when adding student-school-specific variables in my last extension, I find the model can no longer rationalize the data using the probability distribution from the other model's extensions. Further, none of my models can rationalize the data under the assumption that agents observe a full, full minus, or full minus two choice sets, no matter how much flexibility the probability distribution of choice set size gives to the model, when $|\mathcal{D}| = 7$.

These findings suggest that mainstream assumptions on the observability of choice sets in standard school choice models may be too strong. I find that small changes in my model's assumptions can significantly alter its results. With this I argue why discrete choice models might not always be robust to the general on agents' choice sets used in the literature. Standard Logit Models always find parameters to fit the data, therefore recognizing that a model is incorrectly specified is less obvious than when the estimation result is an empty set of parameters.

Although authors can (and often do) robustness checks by estimating the same model using different ways to define each parents' choice set and analyzing how these estimations change, the idea behind my model is conceptually different. Using different choice sets still assumes that parents are aware of all feasible schools, so it does not account for agents' imperfect information.

Assuming perfect information is problematic because it can affect the model's ability to rationalize the data; Parents may seem to behave irrationally only because the model does not take the possibility that parents did not know about all feasible schools into consideration. Some parents may choose schools that are strictly dominated by others because they did not know they could apply to the better one. My estimation method allows for this possibility, so what other models may categorize as an irrational choice can be completely rational in mine.

To account for parents' imperfect information on school supply, I estimate my model using partial identification, which allows me to work under assumptions that are much weaker than those used in most point identifying methods. Nevertheless, what makes my model flexible, translates into its estimation being computationally challenging. The algorithm can quickly solve the optimization problem if the model is constrained enough because few optimal choices exist for each agent. However, the amount of model implied optimal choices grows as the minimum number of schools parents observe grows. This means the algorithm must find and compute a larger array of solutions, which in turn increases estimation time.

Further, as discussed in section 5.1, I run into other limitations when accounting for students' observable heterogeneity (see models from sections 6.1 and 6.2). When using clusters for accounting for said heterogeneity, the algorithm must find a set of parameters that hold for each group of agents, which becomes more and more difficult as the number of clusters increase. Therefore, even though I can make my model flexible enough to rationalize data for different agents using the same parameters, I run into a wall where the model's flexibility no longer allows the algorithm to find an identification region when I use too many clusters.

Therefore, although my model helps argue against strong assumptions on agents' choice sets, it does not provide a plausible alternative way of estimating preferences under weaker assumptions. First, the model cannot be used in contexts where choice sets are too big. Be that as it may, one could argue that in some cases working with big choices is not necessary, nevertheless, not being able to account for observed heterogeneity is a big problem in most cases. Thus far, I have argued that valuable information is lost when the econometrician makes strong assumptions about agents' choice sets. However, I recognize that not making these assumptions and using alternative estimation methods leads to loss of information on preference heterogeneity, which is an essential aspect of consumer behavior.

In all, my contribution to the literature is methodological. Rather than focusing on the more studied aspects of school choice, such as quantifying the trade-off between different school attributes or specifying the determinants of school choice, I focus on the method behind all of it. Although I am aware of my model's many limitations due to its simplistic specification, applying the methodology proposed by Barseghyan et al. to the Chilean School Admission system brings attention to a critical aspect often overlooked in school choice models; choice sets are unobservable by the econometrician because parents have imperfect information about schools. To illustrate this I show how assumptions made by the econometrician have non-trivial implications when estimating different models. All things considered, I believe that we could learn a lot from a thorough examination and reconsideration of the literature's current mainstream practices.

References

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2009, December). Strategy-proofness versus efficiency in matching with indifferences: Redesigning the nyc high school match. *American Economic Review 99*(5), 1954–78.

Abdulkadiroğlu, A. and T. Sönmez (2003, June). School choice: A mechanism design approach. *American Economic Review 93*(3), 729–747.

Barseghyan, Levon, C. M. M. F. T. J. C. (2021). Heterogeneous choice sets and preferences. *Econometrica 89(5), 2015-2048*.

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

Billings, S. B., D. J. Deming, and J. E. Rockoff (2012, October). School segregation, educational attainment and crime: Evidence from the end of busing in charlotte-mecklenburg. Working Paper 18487, National Bureau of Economic Research.

Bosetti, L. (2004). Determinants of school choice: Understanding how parents choose elementary schools in alberta. *Journal of education policy 19*(4), 387–405.

Chumacero, R. A., D. Gómez, and R. D. Paredes (2011, October). I would walk 500 miles (if it paid): Vouchers and school choice in Chile. *Economics of Education Review 30*(5), 1103–1114.

Crawford, G. S., R. Griffith, and A. Iaria (2021). A survey of preference estimation with unobserved choice set heterogeneity. *Journal of Econometrics 222*(1, Part A), 4–43. Annals Issue: Structural Econometrics Honoring Daniel McFadden.

Gallego, F. A. and A. Hernando (2010). School choice in chile: Looking at the demand side. *Pontificia Universidad Catolica de Chile Documento de Trabajo* (356).

Goeree, M. S. (2008). Limited information and advertising in the u.s. personal computer industry. *Econometrica 76*(5), 1017–1074.

Hastings, J., T. J. Kane, and D. O. Staiger (2009). Heterogeneous preferences and the efficacy of public school choice. *NBER Working Paper 2145*, 1–46.

Hastings, J., R. Van Weelden, and J. Weinstein (2007, 04). Preferences, information, and parental choice behavior in public school choice.

Hastings, J. S., T. J. Kane, and D. O. Staiger (2005, November). Parental preferences and school competition: Evidence from a public school choice program. Working Paper 11805, National Bureau of Economic Research.

Holme, J. J. (2002). Buying homes, buying schools: School choice and the social construction of school quality. *Harvard Educational Review 72*(2), 177–206.

Matějka, F. and A. McKay (2015, January). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review 105*(1), 272–98.

McFadden, D. (1973). *Conditional Logit Analysis of Qualitative Choice Behavior*. BART impact studies final report series: Traveler behavior studies. Institute of Urban and Regional Development, University of California.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics 3*(4), 303–328.

McFadden, D. (2001, June). Economic choices. *American Economic Review 91*(3), 351–378.

McFadden, D. and K. Train (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics 15*(5), 447–470.

Neilson, C., C. Allende, F. Gallego, et al. (2019). Approximating the equilibrium effects of informed school choice. Technical report.

Appendix: Matlab Codes

Main Code:

```
clc
clear variables
close all

% Data
temp_prob_schools =  readtable('LHS_Ovalle.csv','ReadVariableNames',true);
prob_school= table2array(temp_prob_schools(:,2)); % Data vector of empirical
    probabilities
data_1 =  readtable('RHS_Ovalle_Simce_Priori.csv','ReadVariableNames',true);
priori=table2array(data_1(:,5)); % Data Vector of each school's tuition cost
simce=table2array(data_1(:,4)); % Data Vector of each school's SIMCE scores

% Parameters
C = {'1' '2' '3' '4' '5' '6' '7'}; % Feasible Set
nGroups = 2^numel(C) - 1; % Number of possible subsets
K_matrix= dec2bin(1:nGroups) == '1'; % Matriz of all possible values of K
full_choice_set=ones(1,length(C)) ;

% Variables
delta1=linspace(-1,1,50)'; % Possible parameter values for delta1 (SIMCE)
delta2=linspace(-1,1,50)'; %% Possible parameter values for delta2 (Priority
    ratio)

% Probabilities of drawing a choice set of size X
pi1=0.01;  % Prob of 7 schools (full choice set)
pi2=0.01;  % Prob of 6 schools (full-minus-one choice set)
pi3=0.08;  % Prob of 5 schools (full-minus-two choice set)
pi4=0.9;   % Prob of 4 schools
pi5=0;     % Prob of 3 schools
pi6=0;     % Prob of 2 schools
pi7=0;     % Prob of 1 schools

pi=[pi1 pi2 pi3 pi4]; % Vector of probabilities (Must contain all probabilities
     bigger than zero)

kappa=4; % Minimun choice set size

n_best=(length(C)-kappa); % Number of model implied optimal choices

rhs=zeros(nGroups,1); % Vector for the values from the RHS of equation (2)
lhs=zeros(nGroups,1); % Vector for the values from the LHS of equation (2)

% Matrixes used in the main loop
mat_full_minus_n=zeros(length(C),length(C),n_best+1);
aux_full_minus_k=zeros(length(C),n_best+1);
P=zeros(n_best+1,1);
```

```
condition=zeros(length(delta1),length(delta2));
values=zeros(7,7);

for d1=1:length(delta1) % For each possible value of delta1
    for d2=1:length(delta2) % For each possible value of delta2

            for j=1:length(C) % Find the best competitor of each school for a
                given set of parameters delta1 and delta2

                mat_full_minus_n(j,:,:)=f_baseline(j,delta1(d1),delta2(d2),
                    simce,priori,C,kappa)';
                [sets_kappa,ut] = f_baseline(j,delta1(d1),delta2(d2),simce,
                    priori,C,kappa);
                values(j,:,:)=ut;

            end

            for b=1:n_best+1 % For all the model implied optimal choices for a
                given choice set size
                for j=1:length(C)

                    % Probability that school j school is chosen

                    aux_full_minus_k(j,b)=p_rhs(j,mat_full_minus_n(j,:,b),
                        delta1(d1),delta2(d2),simce,priori);
                end
            end
            region=0;
            for i=1:nGroups % For all possible values of K
                for b=1:n_best+1

                    P(b,1)=K_matrix(i,:)*aux_full_minus_k(:,b);

                end

                rhs(i)= pi*P;
                lhs(i)=p_lhs(K_matrix(i,:),prob_school);

                if lhs(i)>rhs(i) && isreal(rhs(i))

                    break % The loop breaks if one inequality is not satisfied
                        for a given set

                elseif lhs(i)<=rhs(i)  && isreal(rhs(i)) % The inequality is
                    satisfied


                    region=region+1;
                end
            end
            condition(d1,d2)=region;
    end
end

% Find set of parameters that satisfy all 127 moment inequalities
idx = find(condition==127);
[i,j,k] = ind2sub(size(condition), idx);
B = [i j k condition(idx)];
set1=delta1(i);
set2=delta2(j);

% Plotting the identified set/ Partial identification region
figure(1)
scatter(set1,set2)
xlim([-1.3 1.3])
ylim([-1.3 1.3])
xlabel('$\delta_1$ (SIMCE)','interpreter', 'latex');
ylabel('$\delta_2$ (Priority ratio)','interpreter', 'latex');
title('Identified Set for $\kappa=7$ (BM)','interpreter', 'latex');
grid on;
```

Script for Function 1 (f_baseline.m)

```
% This function finds the "full minus k" choice sets for a given school and
% a given kappa when parents' utility function have 2 parameters (SIMCE score
% and priority ratio). For example, if we are working woth school "i" and
% kappa=5, this function generates a matrix that contains the full choice
% set, the full-minus-1 choice set (eliminates the best competitor of
% school "i"), and the full-minus-2 choice set (eliminates the best and
% second best competitor of school "i").

function [sets_kappa,ut] = f_baseline(rbd,delta1,delta2,simce,priori,C,kappa)
fullset=ones(length(C),1); % Vector for the full choice set
fullset(rbd)=NaN; % An intermediate choice set without school "i"
ut=((delta1*simce)+(delta2*priori)).*fullset; % Computes the objective
                                               % utility for each school in
                                               % the intermediate choice set



n_best=(length(C)-kappa); % Number of competitors that need to be removed
[~, positions] = maxk(ut,n_best); % Finds schools with the n_best utility
mat=ones(n_best,length(C)); % Matrix where each row is a choice set and
                            % each column is a school

% Now the n_best schools are removed from their respective choice sets

mat(1,positions(1))= 0;

for irow= 2:n_best
    mat(irow,:)= mat(irow-1,:);
    mat(irow,positions(irow))= 0;
end

fila=ones(1,7); % The number of columns is 7 because |C|=7
sets_kappa=[fila;mat]; % Matrix containing the 7 minus kappa choice sets
```

Script for Function 2 (p_rhs.m)

```
% This function computes the probability that school "n" is chosen from a
% given choice set given the parameters of a utility function that
% considers 2 school attributes (SIMCE and priority ratio).
% Note that "rbd" is used as a school id.

function [prob_rbd] = p_rhs(rbd,choice_set,delta1,delta2,simce,priori)
num=exp((delta1*simce)+(delta2*priori)); % Numerator of the logit formula
                                         % to compute the pobability that a
                                         % school is chosen.

den=choice_set*num; % Denominator of the logit formula used to compute
                    % the pobability that a school is chosen.

if choice_set(rbd)==0 % If school "i" is not part of the choice set,

    prob_rbd=0; % the probability that it will be chosen is zero.

else

    prob_rbd=num(rbd)/den; % Probability of choosing school "i"

end
end
```

Script for Function 3 (p_lhs.m)

```
% This function finds the empirical probability that a given school
% belongs to subset K.

function [prob_lhs] = p_lhs(subset,prob_school)
prob_lhs= subset * prob_school; % Multiplies the vector that represents
                                % subset K with the vector containing the
                                % empirical probability that a given school
                                % is chosen, for every school.
end
```