Identification and Estimation in Many-to-one Two-sided Matching without Transfers*

YingHua He† Shruti Sinha‡ Xiaoting Sun§

July 14, 2022

Abstract

In a setting of many-to-one two-sided matching with non-transferable utilities, e.g., college admissions, we study conditions under which preferences of both sides are identified with data on one single market. Regardless of whether the market is centralized or decentralized, assuming that the observed matching is stable, we show nonparametric identification of preferences of both sides under certain exclusion restrictions. To take our results to the data, we use Monte Carlo simulations to evaluate different estimators, including the ones that are directly constructed from the identification. We find that a parametric Bayesian approach with a Gibbs sampler works well in realistically sized problems. Finally, we illustrate our methodology in decentralized admissions to public and private schools in Chile and conduct a counterfactual analysis of an affirmative action policy.

Keywords: Many-to-one Two-sided Matching, Non-transferable Utility, Nonparametric Identification, College Admissions, School Choice.

*This paper subsumes Sun's job market paper "Identification and Estimation of Many-to-One Matching with an Application to the U.S. College Admissions" and He and Sinha's previously circulated paper "Identification and Estimation in Many-to-One Two-Sided Matching without Transfers." We thank Nikhil Agarwal, Cristina Gualdani, Thierry Magnac, Rosa Matzkin, Ismael Mourifie, Shuyang Sheng, and Xun Tang, for their useful comments. We would also like to thank the seminar and conference participants at 2018 California Econometrics Conference, 2018 Midwest Econometrics Group Conference, 2019 Canadian Econometric Study Group Meetings, 2019 Network Econometrics Junior's conference at Northwestern University, 2019 Seattle-Vancouver Econometrics Conference, Auburn University, Econometric Society World Congress 2020, EEA-ESEM 2019, ICEF Moscow, Jinan University, Johns Hopkins University, National University of Singapore, Peking University, Rice University, Simon Fraser University, Toulouse School of Economics, UCLA, UC Riverside, University of Florida, University of Toronto, UNSW Sydney, and Western University. Thanks also to Alípio Ferreira Cantisani for excellent research assistance. Sinha acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010.

†Email: yinghua.he@rice.edu, Rice University, Houston, Texas, USA.

‡Email: shruti.sinha@tse-fr.eu, Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France.

§Email: xiaoting_sun@sfu.ca, Simon Fraser University, Vancouver, Canada.

1 Introduction

In a many-to-one two-sided matching market, agents are categorized into two sides; everyone on one side has preferences over those on the other side; an agent on only one of the two sides can have multiple match partners from the other side. Many real-life markets fit this description, for example, the medical resident match (Roth, 1984; Agarwal, 2015) in the US, school admissions in Chile (Gazmuri, 2017) and Hungary (Aue et al., 2020), college admissions in the US, and graduate program admissions in France (He and Magnac, 2020). Such markets often exclude personalized transfers, even though limited monetary exchanges may exist. Hence, the literature defines it as matching without transfers or matching with non-transferable utility.

While the literature has extensively studied this type of matching theoretically (see, e.g., Roth and Sotomayor, 1992; Azevedo and Leshno, 2016), its econometrics is less explored. Our paper aims to make a contribution by answering the following questions: Are the preferences of both sides identified from data on who matches with whom? If so, how can the preferences be estimated?

To fix ideas, we proceed in the language of college admissions. We derive a set of sufficient conditions under which both student and college preferences are nonparametrically identified. Our results are obtained from a single market in which there are a continuum of students and a fixed number of colleges. We use that to approximate a single large market. Further, we provide an estimation procedure that is practical even in settings with many agents, allowing for rich observed and unobserved heterogeneity. Understanding agent preferences is often crucial for policymaking, and one may analyze a wide range of counterfactual policies with estimated preferences. Potentially, our results open a new avenue of research on such matching markets.

The main challenge in identifying student preferences is that each student's *actual* choice set is unobservable to the researcher. For student i to be able to enroll at college c, college c needs to accept i. The same difficulty exists in the identification of college preferences. Moreover, each student's and each college's choice sets are endogenously determined in equilibrium without market-clearing prices.

In our continuum setting, we assume that an observed matching is stable. That is, no college prefers to reject any of its currently matched students to vacate a seat, and no student prefers to leave her current match to become unmatched or matched with a college that is willing to accept her and, if necessary, reject one of its currently matched students. Stability is often imposed in the study of various matching markets (see, for a survey, Chiappori and Salanié, 2016) and is satisfied in equilibrium in our setting in certain game-theoretical models (Artemov et al., 2020; Fack et al., 2019).

Importantly, there is generically a unique stable matching that is characterized by the colleges' admission cutoffs (Azevedo and Leshno, 2016). When college preferences over individual students are represented by utility functions, a college's cutoff is the lowest utility level among its matched students. Cutoffs further define a student's actual choice set in equilibrium, called *feasible set*. A college is in a student's feasible set if the college's utility of being matched with her is higher than its cutoff. Stability implies that a student is matched with her most-preferred feasible college, similar to a discrete choice problem, except that feasible sets are unobservable and heterogeneous.

A simple equation, called the i-c match probability, is the key to understanding our identification result. Specifically, the conditional probability of student i being matched with college c is the sum of conditional probabilities of i choosing c from a given feasible set L weighted by the conditional probability of facing L:

$$\begin{aligned}&\mathbb{P}(\text{student } i \text{ is matched with college } c \mid x_i)\\ =&\sum_{\text{all possible feasible sets, } L} \underbrace{\mathbb{P}(L \text{ is } i\text{'s feasible set} \mid x_i)}_{\equiv A \text{ (college preferences)}} \cdot \underbrace{\mathbb{P}(c \text{ is } i\text{'s most-preferred college in } L \mid L, x_i)}_{\equiv B \text{ (student preferences)}},\end{aligned}$$

where x_i consists of all observed characteristics of student i (including pair-specific characteristics like distance to colleges). The equation provides a structured decomposition of the preferences of the two sides: piece A only depends on the preferences of *all* colleges, while piece B only depends on i's preferences over all colleges.

We then detail a set of exclusion restrictions, among other regularity conditions, such that the excluded variables act as "demand shifters" and "feasible-set shifters" (or supply shifters). Sufficient variation in these excluded variables identifies the preferences of colleges and students using the i-c match probability described above.

Here are some intuitions. For a college d, an (i, d)-specific demand shifter traces out how i's preferences for d affects the i-c match probability. Similarly, an (i, d)-specific feasible-set shifter traces out how d's preference for i affects the i-c match probability. A non-excluded variable affects the i-c match probability through preferences on both sides for all colleges. By taking derivatives of the i-c match probability with respect to (w.r.t.) all the variables, excluded and non-excluded, we derive systems of linear equations that link the effects of variations in demand and supply. Hence, the identification problem reduces to setting up systems of linear equations and ensuring the existence of a unique solution. Essentially, we extend Matzkin (2019)'s results on a nonseparable discrete choice model to two-sided matching.

The second objective of our paper is to provide practical methods that can be used to analyze real-life markets. We achieve this by deriving theoretical guidelines and showcasing a practical estimation method.

When taken to the data, the requirement of a large number of excluded variables may be difficult to meet. To address this, we theoretically characterize the tradeoff between exclusion restrictions and the degree of identifiable preference heterogeneity (Proposition 3.10). The researcher can use this result as a guideline for empirical studies when having insufficient excluded variables.

We also need a practical estimation method to take these identification results to the data. In fact, our identification arguments are constructive, leading to nonparametric and semiparametric estimators. Monte Carlo simulations suggest that estimating the matrices of partial derivatives in the linear systems using the average derivative estimators of Powell et al. (1989) performs well in finite samples only when the curse of dimensionality is not severe. In a reasonably sized problem, we resort to a parametric Bayesian approach with a Gibbs sampler (Rossi et al., 2012), resembling applications such as Logan et al. (2008) for one-to-one two-sided matching and Abdulkadiroğlu et al. (2017) for a one-sided problem. We demonstrate its good performance in Monte Carlo simulations with high dimensionality.

As an empirical application, we consider the decentralized admissions to secondary schools (grades 9–12) in Chile. To the best of our knowledge, this is one of the first attempts to estimate the preferences of both sides in a *decentralized* market of many-to-one two-sided matching without transfers. There is no clearinghouse, and students do not submit rank-order lists of schools. Although public schools do not select students and thus have no preferences, private schools, either subsidized or not subsidized by the government, can reject students according to their own preferences.

We focus on a market in which 9,304 students are matched with 125 schools and an outside option. The data include the matching outcome and a set of student characteristics and school attributes. By allowing highly flexible preference heterogeneity in the Bayesian approach, we estimate students' preferences over all schools and private schools' preferences over all students. For each school type (public, private subsidized, private non-subsidized), students have a distinct utility function. Similarly, private subsidized and non-subsidized schools have two different utility functions.

To illustrate the usefulness of the preference estimates, we consider a counterfactual policy in which students from low-income families (the bottom 40%) are prioritized for admissions to all schools. Segregation in terms of ability and income decreases, albeit slightly. On average, the policy benefits low-income students and hurts others. These changes are driven by low-income students moving from public schools to private subsidized schools, while crowding out other students to public or private non-subsidized schools. Notably, low-income students still do not choose non-subsidized schools, which may be driven by their high cost and the students' tastes. In sum, simply giving low-income students access to schools may not significantly change matching outcomes due to student preferences.

Related Literature. This paper is related to the literature on the identification of matching models; Table 1 provides an incomplete summary.

Table 1: Identification Results of Matching Models

	Transferable Utility (TU)	Non-Transferable Utility (NTU)
One-to-one	The match surplus is identified (see, e.g., Choo and Siow, 2006; Fox, 2010; Chiappori et al., 2017; Galichon and Salanie, 2020).	The match surplus is identified (see, e.g., Dagsvik, 2000; Menzel, 2015).
Many-to-one	The utility function and the distribution of the unobservables of both sides are identified in a homogeneous setting (see, e.g., Diamond and Agarwal, 2017).	The utility function and the distribution of the unobservables of both sides are identified in a homogeneous setting (see, e.g., Diamond and Agarwal, 2017). *Our paper*: The utility functions of both sides with heterogeneity and the distribution of the unobservables are identified.
Many-to-many	The match surplus and/or the distribution of the unobservables are identified (see, e.g., Fox, 2010; Fox et al., 2018).	The match surplus is identified (see, e.g., Menzel, 2017).

This literature is split into several strands depending on the preference structures of the agents – transferable utility (TU), non-transferable utilities (NTU), or imperfectly transferable utility (ITU) models;[1] and the maximum number of links an agent is permitted to form across sides – one-to-one, many-to-one, and many-to-many (see Chiappori and Salanié, 2016, for a survey).

There is also a close relationship between the one-to-one TU matching model (Choo and Siow, 2006; Fox, 2010, 2018; Graham, 2011; Sinha, 2015; Chiappori et al., 2017; Diamond and Agarwal, 2017; Galichon and Salanie, 2020; Gualdani and Sinha, 2020) and the many-to-one NTU matching model considered here. Market-clearing college cutoffs in our setting play the role of market-clearing shadow prices, although the endogenous cutoffs do not determine how the surplus is split among the agents.

Most of the work on identification within the NTU framework focuses on one-to-one markets (Dagsvik, 2000; Menzel, 2015; Uetake and Watanabe, 2020). Allowing for infinitely many agents on both sides of a many-to-one matching market (either NTU or TU), Agarwal (2015) and Diamond and Agarwal (2017) show identification under a homogeneity restriction on the preferences.[2] Our setting and theirs are non-nested, and we allow for richer preference heterogeneity. The additional identification power in our setting is a result of holding fixed the number of colleges, and thus the college cutoffs in equilibrium can be treated as fixed parameters. Agarwal and Somaini (2020a) provide a recent survey on empirical models of NTU matching.

Many-to-one NTU matching has been empirically studied in the context of secondary school admissions in Hungary (Aue et al., 2020) and graduate program admissions in France (He and Magnac, 2020). Their data include information on the preferences of both sides that is reported to a centralized mechanism. Therefore, they can independently identify and estimate the preferences of each side, essentially reducing the two-sided matching to two separate one-sided problems.

Centralized many-to-one NTU matching in the context of school choice has been studied extensively, both theoretically since Abdulkadiroğlu and Sönmez (2003) and empirically (e.g., Abdulkadiroğlu et al., 2017; Agarwal and Somaini, 2018; Calsamiglia et al., 2020; Fack et al., 2019; He, 2017; Kapor et al., 2020). In this literature, school preferences are (assumed to be) known because schools rank students according to certain pre-specified rules. The problem then reduces to identifying and estimating student

[1] See Galichon et al. (2019) for an example of ITU models.

[2] When studying the medical resident match, Agarwal (2015) discusses some intuitions of using exclusion restrictions to identify heterogeneous preferences on each side of the market.

preferences. See Agarwal and Somaini (2020b) for a survey.

Feasible sets in our setting resemble endogenous consideration sets that arise in one-sided decision problems. In this sense, our paper relates to the growing strand of literature that studies the econometrics of decision problems under consideration set formation (e.g., Abaluck and Adams, 2018; Barseghyan et al., 2019, 2021; Cattaneo et al., 2020). Our contribution here is that we provide a structural two-sided setting where the consideration probabilities in a student's decision problem are entirely determined by the college (supply side) preferences.

The remaining paper is organized as follows: Section 2 describes the model and data generating process; Section 3 discusses the identification of the preferences on both sides of the matching market; Section 4 illustrates an empirical analysis of the match between students and secondary schools in Chile; and Section 5 concludes.

2 Model

For the sake of exposition, our model is set up as a college admissions problem. Consider a single market with a continuum of students and finitely many colleges. The set of all students is $\mathbf{I}$, with a probability measure Q defined over it,[3] and the set of all colleges is $\mathbf{C} = \{1, 2, \ldots, C\}$. College $c \in \mathbf{C}$ has a capacity $q_c \in (0, 1)$.

Each student i is characterized by $x_i \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$, which is a vector of student-specific or pair-specific covariates observable to the researcher, and $\epsilon_i = (\epsilon_{i1} \ldots, \epsilon_{iC})$ and $\eta_i = (\eta_{1i} \ldots, \eta_{Ci})$, which are unobservable to the researcher. Further, for any $c \in \mathbf{C}$, ϵ_{ic} and η_{ci} are scalar random variables that denote taste shocks that students have for each college and vice-versa, respectively. Let (ϵ_i, η_i) be independent and identically distributed (i.i.d.) draws from a joint distribution F. We make no restrictions on this joint distribution, allowing for an arbitrary correlation structure within (ϵ_i, η_i).

For any $i \in \mathbf{I}$, the utility of being matched with college $c \in \mathbf{C}$ is $u_{ic} = U^c(x_i, \epsilon_{ic})$, where $U^c : \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ is a college-specific, nonparametric function mapping the observable and unobservable variables to a single index. Students can remain unmatched or, equivalently, be matched with an outside option denoted by "0." The

[3] The probability space is $(\mathbf{I}, \mathcal{B}(\mathbf{I}), Q)$ with $\mathcal{B}(\mathbf{I})$ being the Borel set of $\mathbf{I}$, $Q : \mathcal{B}(\mathbf{I}) \to [0, 1]$, and $Q(\mathbf{I}) = 1$.

utility of the outside option is normalized to 0, $u_{i0} = 0\ \forall i \in \mathbf{I}$.

Similarly, for $c \in \mathbf{C}$, the utility of being matched with student $i \in \mathbf{I}$ is $v_{ci} = V^c(x_i, \eta_{ci})$, where $V^c : \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ is a college-specific, nonparametric function. We assume that colleges have responsive preferences.[4] This implies that the total utility of a college from being matched with a subset of students (up to its capacity) is increasing in its utility from each student; for example, the total utility is the sum of the utility from each of its matched students. College c has an acceptability threshold, T_c, and finds student i unacceptable if $v_{ci} < T_c$.

With the data from one such continuum market on $\{x_i\}_i$, $\{q_c\}_c$, and who matches with whom, we aim to identify student and college preferences by identifying $\{U^c, V^c, T_c\}_c$ and F, although, as we shall see, $\{T_c\}_c$ are not always point identified. Note that x_i does not include college-specific variables that are constant across students, as such variables will be absorbed by the college-specific utility functions, U^c and V^c.

Remark 2.1. *We use the continuum market to approximate a data generating process in a large finite market as follows: student i's $(x_i, \epsilon_i, \eta_i)$ is an i.i.d. draw from their joint distribution, college c's capacity is a q_c-fraction of the total number of students, and $\{U^c, V^c, T_c\}_{c\in\mathbf{C}}$ determines both sides' preferences. This approximation is close in terms of equilibrium outcomes when we use the equilibrium concept in Section 2.1.*[5]

2.1 Matching and Stable Matching

We define a matching function or, simply, a matching, $\mu : \mathbf{I} \to \mathbf{C} \cup \{0\}$, such that (i) $\mu(i) = c \iff i \in \mu^{-1}(c)$, and (ii) $\forall\, c \in \mathbf{C}$, $\mu^{-1}(c) \subseteq \mathbf{I}$, where $0 \leq Q(\mu^{-1}(c)) \leq q_c$.

The following concepts are important for our analysis: individual rationality, blocking pairs, and stability. For notational reasons, we define them in the case

[4]For $\varepsilon > 0$, let $N_\varepsilon(i)$ and $N_\varepsilon(i')$ be a neighborhood of students around v_{ci} and $v_{ci'}$, respectively, such that $Q(N_\varepsilon(i)) = Q(N_\varepsilon(i'))$. Responsive preferences imply that, for any $\mathbf{I}^c \subset \mathbf{I}$ with $Q(\mathbf{I}^c) \leq q_c - Q(N_\varepsilon(i))$, $N_\varepsilon(i) \subset \mathbf{I} \setminus \mathbf{I}^c$, and $N_\varepsilon(i') \subset \mathbf{I} \setminus \mathbf{I}^c$, college c prefers $\mathbf{I}^c \cup N_\varepsilon(i)$ to $\mathbf{I}^c \cup N_\varepsilon(i')$ if and only if $v_{ci} > v_{ci'}$. See Roth and Sotomayor (1992) for a definition in a case with discrete students.

[5]In a usual study on identification, one takes the number of observations to infinity while keeping constant the "game." Our setting contains one single matching game whose nature changes with the market size. Fortunately, Proposition 3 of Azevedo and Leshno (2016), Proposition 4 of Fack et al. (2019), and Corollary 2 of Artemov et al. (2020) imply that, under certain conditions, the equilibrium outcome in the continuum approximates well an equilibrium outcome in a large finite market. Such an approximation is also used in the network literature (e.g., Menzel, 2022).

with discrete students, corresponding to our empirical application. The proper definitions for a model with a continuum of students can be found in Azevedo and Leshno (2016), with measure-zero sets of students appropriately dealt with.

A matching μ is *individually rational* if $u_{i\mu(i)} \geq u_{i0}$ and $v_{\mu(i)i} \geq T_{\mu(i)}$ for all $i \in \mathbf{I}$. A student-college pair $(i, c) \in \mathbf{I} \times \mathbf{C}$ *blocks* a matching μ if (i) student i strictly prefers college c to her current match $\mu(i)$, $u_{ic} > u_{i\mu(i)}$; and (ii) either college c has excess capacity, $Q(\mu^{-1}(c)) < q_c$, or college c prefers i to one of its matched students, $\exists\, i' \in \mu^{-1}(c)$, s.t., $v_{ci} > v_{ci'}$. Finally, a matching is *stable* if it is individually rational and not blocked by any pair $(i, c) \in \mathbf{I} \times \mathbf{C}$.

We assume that the matching in the data is stable.[6] A stable matching exists and is generically unique (Azevedo and Leshno, 2016).[7] Moreover, a stable matching is characterized by college "cutoffs." College c's cutoff is determined by its least-preferred matched student when its capacity constraint is binding; otherwise, it coincides with the acceptability threshold. Let δ_c be college c's cutoff. Then,

$$\forall\, c \in \mathbf{C},\ \delta_c = \inf_{j \in \mu^{-1}(c)} v_{cj} \text{ if } Q(\mu^{-1}(c)) = q_c;\ \delta_c = T_c \text{ if } Q(\mu^{-1}(c)) < q_c. \tag{1}$$

By definition, $\delta_c \geq T_c$. Under the assumption of responsive college preferences, to determine if student i can be accepted by c, we just need to compare v_{ci} and δ_c. With non-responsive preferences, how c ranks i and j would depend on who else c accepts, and δ_c alone would not be sufficient to determine if i could have been accepted by c.

In a stable matching of the continuum market, $\{U^c, V^c, T_c\}_{c \in \mathbf{C}}$, and F imply a unique vector of cutoffs, $\{\delta_c\}_c$. Therefore, $\{\delta_c\}_c$ is merely a shorthand notation for the expression in equation (1) rather than additional parameters.[8]

[6]Stability can be achieved in certain equilibrium if students apply to all acceptable colleges and if a stable mechanism, e.g., the deferred acceptance (Gale and Shapley, 1962), is used to find the matching. Theoretically, provided that students know what criteria colleges use to rank them, stability can still be satisfied in equilibrium, even if students choose not to apply to all acceptable colleges due to application costs (Fack et al., 2019) or if students make certain application mistakes (Artemov et al., 2020). Importantly, achieving stability does not require the market to be centralized, as shown in laboratory experiments (see, e.g., Pais et al., 2020), and steps in mechanisms such as the deferred acceptance can be implemented in a decentralized fashion (Grenet et al., 2022).

[7]We need the regularity condition that the set $\{i \in \mathbf{I} : \mu(i) \text{ is strictly less preferred than } c\}$ is open $\forall\, c \in \mathbf{C}$. This condition implies that a stable matching always allows an extra measure zero set of students into a college when this can be done without compromising stability.

[8]As mentioned in Remark 2.1 and footnote 5, the continuum approximates a large finite market. The literature cited therein shows that equilibrium cutoffs, hence matching outcomes, in the large

2.2 Feasible Sets in a Stable Matching

College c is defined *feasible* to student i if and only if $v_{ci} \geq \delta_c$. In effect, a student can "choose" to match with any of her feasible colleges, but not any infeasible college. We call the set of all feasible colleges of a student her *feasible set*. Let $\mathcal{L}$ be the collection of the 2^C possible feasible sets, $\mathcal{L} \equiv \{L : 0 \in L,\ L \setminus \{0\} \subseteq \mathbf{C}\}$. By construction, the outside option always belongs to every feasible set. A matching is stable if and only if every student is matched with her most-preferred feasible college. A classic issue in two-sided matching is that students' feasible sets are determined endogenously, unobserved by the researcher, and heterogeneous across students.

Let $\lambda_{L,i}$ be the probability that $L \in \mathcal{L}$ is student i's feasible set conditional on x_i. Given μ, $\lambda_{L,i}$ is a function of x_i only, $\lambda_{L,i} \equiv \lambda_L(x_i;\mu)$. Going forward, we suppress the dependence of λ_L on μ, as μ is observed and fixed in the data. Therefore,

$$\lambda_{L,i} = \lambda_L(x_i) = \mathbb{P}(\text{feasible set is } L|x_i) = \mathbb{P}(v_{ci} \geq \delta_c\ \forall\, c \in L;\ v_{di} < \delta_d\ \forall\, d \notin L|x_i). \quad (2)$$

If each student's feasible set was observed, $\lambda_{L,i}$ could be identified from the data, and recovering student and college preferences would follow from standard arguments in the discrete choice literature. However, we do not observe the feasible sets, and thus these conditional probability functions, $\lambda_L(x_i)$, are unknown. One could treat these as 2^C parameters and recover them directly from the data. This would restore the discrete choice arguments but quickly become intractable given its large dimension. Instead, we take an approach that does not rely on recovering $\lambda_{L,i}$.

3 Nonparametric Identification

We now turn to the nonparametric identification of student and college preferences given a stable matching, μ, and covariates, x_i, observed in *one* market.[9]

We begin with more details on the utility functions. Assume that for each college c, x_i has two continuous variables, one for each side of the market, that are *excluded* from the other colleges. Formally, $x_i = (y_i, w_i, z_i)$, where $y_i = (y_{i1} \ldots, y_{iC}) \in \mathcal{Y} \subseteq \mathbb{R}^C$

market can be close to $\{\delta_c\}_c$ with agents being practically "cutoff-takers." In other words, each agent's realized preferences have a negligible effect on cutoffs in large markets.

[9]See Matzkin (2007) for formal discussions and a definition of nonparametric identification.

and $w_i = (w_{i1} \dots, w_{iC}) \in \mathcal{W} \subseteq \mathbb{R}^C$ are vectors of excluded variables. For each $c \in \mathbf{C}$, the scalar random variable $y_{ic} \in \mathcal{Y}_c \subseteq \mathbb{R}$ enters U^c, while being excluded from all other student utility functions U^d for all $d \in \mathbf{C} \setminus \{c\}$ and all college utility functions V^d for all $d \in \mathbf{C}$. Similarly, $w_{ic} \in \mathcal{W}_c \subseteq \mathbb{R}$ enters V^c, while being excluded from all other college utility functions V^d for all $d \in \mathbf{C} \setminus \{c\}$ and all student utility functions U^d for all $d \in \mathbf{C}$. Hence, y_{ic} is a demand shifter and w_{ic} is a feasible-set, or supply, shifter. The row vector $z_i \in \mathcal{Z} \subseteq \mathbb{R}^{d_z}$ with $d_z \equiv d_x - 2C$ consists of student- and pair-specific random variables that affect the utilities of all agents on both sides.

Further, the utility functions are additively separable in excluded variables, i.e.,

$$u_{ic} = u^c(z_i) + r^c(y_{ic}) + \epsilon_{ic} \text{ and } v_{ci} = v^c(z_i) + w_{ic} + \eta_{ci}, \forall c \in \mathbf{C}. \tag{3}$$

Thus, for each $c \in \mathbf{C}$, the function U^c is separable into a nonparametric function $u^c : \mathcal{Z} \to \mathbb{R}$ and an index $r^c(y_{ic}) + \epsilon_{ic}$, where $r^c : \mathcal{Y}_c \to \mathbb{R}$ is a nonparametric function. The function V^c is separable into a nonparametric function $v^c : \mathcal{Z} \to \mathbb{R}$ and an additive index $w_{ic} + \eta_{ci}$. We impose scale normalization on each side. For students, there exists a known value $\overline{y}_c$ in the interior of $\mathcal{Y}_c$ such that $\frac{\partial r^c(\overline{y}_c)}{\partial y_{ic}} = 1$.[10] This holds trivially if $r^c(y_{ic}) = y_{ic}$. For colleges, we assume w_{ic} enters v_{ci} linearly with a coefficient normalized to one. As detailed below, this linearity assumption allows us to vary w_i to construct a sufficient number of equations without increasing the number of unknowns.[11] The assumptions on y_i and w_i can be switched, that is, having nonlinearity in w_i and linearity in y_i.

Remark 3.1. *The location normalization in the functions is worth highlighting because the joint distribution of (ϵ_i, η_i), F, is fully nonparametric. In student preferences, we already impose the normalization, $u_{i0} = 0$, in Section 2, but we need another location normalization for each c on either $u^c(z_i) + r^c(y_{ic})$ or ϵ_{ic} to separately identify the two. Similarly, for each college c's preferences, we need to location-normalize two*

[10]This normalization is stronger than necessary and is imposed for simplicity of exposition. For a relaxation, see Remark 3.2 and Appendix B. If y_{ic} is known to have a negative effect on student preferences, the partial is normalized to -1. The same applies to w_i.

[11]We can allow for nonlinearity of w_{ic} for a subset of colleges. That is, for $c \in \overline{\mathbf{C}} \subset \mathbf{C}$, $v_{ci} = v^c(z_i) + s^c(w_{ic}) + \eta_{ci}$, where s^c: $\mathcal{W}_c \to \mathbb{R}$ is a nonparametric function such that for some known value $\overline{w}_c$, $\frac{\partial s^c(\overline{w}_c)}{\partial w_{ic}} = 1$; and for $c \in \mathbf{C} \setminus \overline{\mathbf{C}}$, w_{ic} enters linearly with a coefficient normalized to one. In this case, our identification relies on varying w_{ic} for $c \in \mathbf{C} \setminus \overline{\mathbf{C}}$ instead of all $c \in \mathbf{C}$.

of the three model primitives $(v^c(z_i), \eta_{ci}, T_c)$ to pin down the third.

In Section 3.1, we are interested in recovering the derivatives of (u^c, r^c, v^c), so the above location normalization is not needed. However, it is necessary in Section 3.2 where we identify T_c, F, and hence δ_c. As we shall clarify, we location-normalize $u^c(z_i) + r^c(y_{ic})$, $v^c(z_i)$, and η_{ic}.

Remark 3.2. *Appendix B considers a more general nonseparable model. Specifically, there is full nonseparability for all but one (i.e., $2C-1$) utility functions, while for one student utility function, there is nonseparability between the observable z_i and an index $y_{ic} + \epsilon_{ic}$. That is, without loss of generality,*

$$u_{i1} = u^1 (z_i, y_{i1} + \epsilon_{i1}), \; u_{ic} = u^c (z_i, y_{ic}, \epsilon_{ic}) \; \forall c \in \mathbf{C} \setminus \{1\},$$
$$\text{and } v_{ci} = v^c (z_i, w_{ic}, \eta_{ci}) \; \forall c \in \mathbf{C}. \tag{4}$$

The additive index $y_{i1} + \epsilon_{i1}$ can be relaxed to some known function such as $y_{i1} \cdot \epsilon_{i1}$ (Matzkin, 2019). Appendix B presents a set of sufficient conditions under which our identification strategy applies to $\{u^c\}_c$, but identifying $\{v^c\}_c$ requires additional separability. This helps clarify the role of the additive separability in equation (3).

Recall that a matching is stable if and only if every student is matched with the most-preferred college in her feasible set. Thus, stability implies, for $c \in \mathbf{C} \cup \{0\}$,

$$\mathbb{P}(\mu(i) = c|x_i) = \sum_{L \in \mathcal{L}} \lambda_{L,i} \mathbb{P}\left(c = \arg\max_{d \in L} u_{id} | L, x_i\right). \tag{5}$$

The left-hand side is the conditional match probability, or the fraction of students with x_i matched with c, which is known in the data. Let $\sigma_{c,i} \equiv \sigma_c(x_i) = \mathbb{P}(\mu(i) = c|x_i)$. The right-hand side links $\sigma_{c,i}$ to the model through $\lambda_{L,i}$ and $\mathbb{P}(c = \arg\max_{d \in L} u_{id} | L, x_i)$. Below, we discuss these two types of terms one by one.

Recall that $\lambda_{L,i}$ is the probability that L is the feasible set of a student with x_i. It only depends on college preferences from which $\{y_{ic}\}_{c \in \mathbf{C}}$ is excluded. Thus, equation (2) implies $\lambda_{L,i} = \mathbb{P}(\iota_{ic} \geq \delta_c \; \forall \, c \in L; \; \iota_{id} < \delta_d \; \forall \, d \notin L | w_i, z_i) \equiv \lambda_L(\iota_{i1}, \ldots, \iota_{iC})$, where $\iota_{ic} \equiv v^c(z_i) + w_{ic}$ for every $c \in \mathbf{C}$.

In equation (5), $\mathbb{P}(c = \arg\max_{d \in L} u_{id} | L, x_i)$ is the probability that utility maximizing students with observables x_i "choose" c from feasible set L. Conditional on

L, this probability only depends on student preferences. For every $c \in \mathbf{C}$, we define $\tau_{ic} \equiv u^c(z_i) + r^c(y_{ic})$ and, further, $\mathbb{P}(c = \arg\max_{d \in L} u_{id} | L, x_i) \equiv g_{c,L}(\tau_{ic}; \tau_{id}, d \neq c)$, where the first argument of $g_{c,L}$ is always τ_{ic}. If $d \notin L$, $g_{c,L}$ does not vary with τ_{id}.

With these new notations, equation (5) can be rewritten as

$$\sigma_{c,i} = \sum_{L \in \mathcal{L}} \lambda_L(\iota_{i1}, \dots, \iota_{iC}) \cdot g_{c,L}(\tau_{ic}; \tau_{id}, d \neq c). \tag{6}$$

Below, we first study the conditions under which the functions $\{u^c, r^c, v^c\}_c$ are nonparametrically identified; we then identify the joint distribution of (ϵ_i, η_i), F. Later in Section 3.3, we present results that impose fewer requirements on the data.

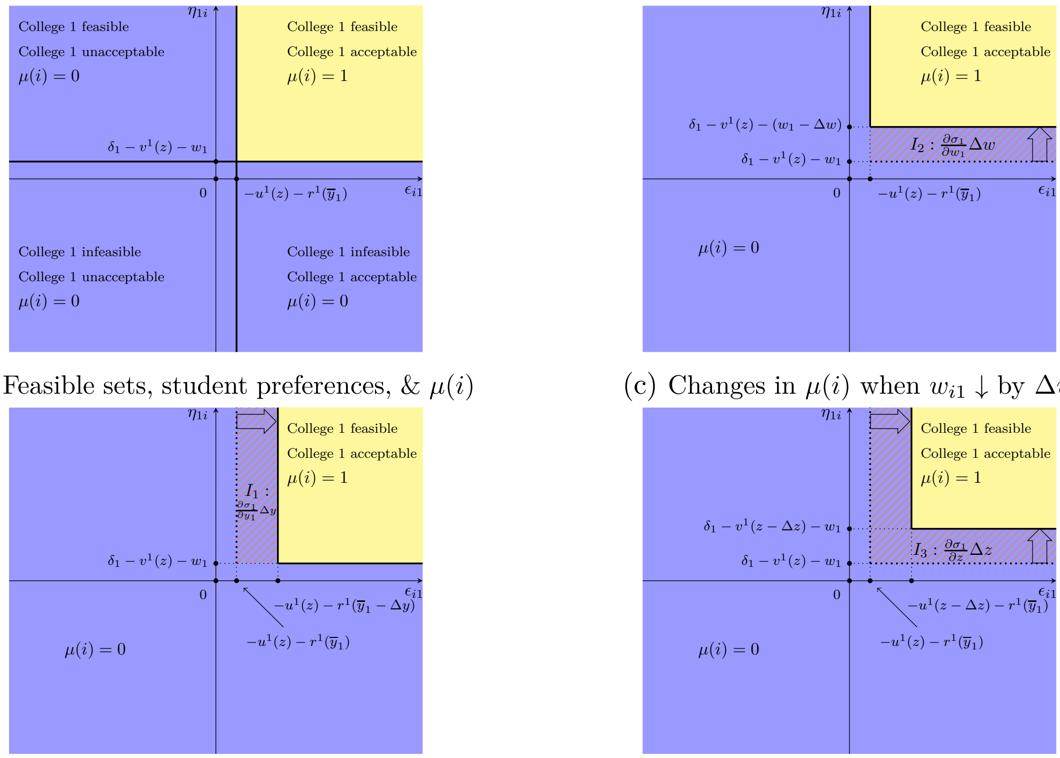
3.1 Identifying the Derivatives of the Utility Functions

We now nonparametrically identify the derivatives of the functions $\{u^c, r^c, v^c\}_c$ w.r.t. the observables. With these derivatives identified, the values of the functions are identified up to a constant, provided that z_i and y_i have full support. The idea is to use the variation in the excluded variables to trace out how each argument in equation (6) affects the conditional match probabilities. Specifically, the excluded variables in student preferences (y_i) only shift demand, while the excluded variables in college preferences (w_i) shift supply, or feasible sets; the effect of other variables (z_i) that affect both demand and supply can be written as a combination of the effects of y_i and w_i. This leads to a system of linear equations involving the derivatives of the conditional match probabilities whose solution is our parameters of interest.

3.1.1 A Simple Example with One College

We describe the intuition for identification in a one-college example, $\mathbf{C} = \{1\}$. Student utility functions are $u_{i1} = u^1(z_i) + r^1(y_{i1}) + \epsilon_{i1}$ for college 1 and $u_{i0} = 0$ for the outside option. Here, z_i is a scalar and $\frac{\partial r^1(\overline{y}_1)}{\partial y_1} = 1$ for a known value, $\overline{y}_1$. College 1's utility function is $v_{1i} = v^1(z_i) + w_{i1} + \eta_{1i}$. College 1's (unobserved) cutoff is δ_1.

To identify $\frac{\partial u^1}{\partial z_i}$ and $\frac{\partial v^1}{\partial z_i}$, we fix $y_{i1} = \overline{y}_1$ and consider any value (z, w_1) in the interior of $\mathcal{Z} \times \mathcal{W}_1$. Figure 1(a) shows that the space of $(\epsilon_{i1}, \eta_{1i})$ is partitioned into four parts based on the feasibility of college 1 and student i's preferences (i.e., the acceptability of college 1 to i). Moreover, $\mu(i) = 1$ if and only if college 1 is acceptable, or $\epsilon_{i1} > -u^1(z) - r^1(\overline{y}_1)$, and if college 1 is feasible to i, or $\eta_{1i} > \delta_1 - v^1(z) - w_1$.



(a) Feasible sets, student preferences, & $\mu(i)$

(c) Changes in $\mu(i)$ when $w_{i1} \downarrow$ by Δw

(b) Changes in $\mu(i)$ when $y_{i1} \downarrow$ by Δy

(d) Changes in $\mu(i)$ when $z_i \downarrow$ by Δz

Figure 1: Partitioning the Space of Unobservables in the One-College Case

Notes: Panel (a) describes the partition of the $(\epsilon_{i1}, \eta_{1i})$ space given $(z_i, y_{i1}, w_{i1}) = (z, \overline{y}_1, w_1)$ by student i's feasible set and preferences. The other panels show the changes in $\mu(i)$ when y_i decreases by Δy and affects only student preferences (panel b), when w_{i1} decreases by Δw and affects only feasible set (panel c), and when z_i decreases by Δz (panel d).

Figure 1(b)–(c) depict how the marginal effect of z_i on match probability is linked to the marginal effects of the excluded variables, y_{i1} and w_{i1}. Panel (b) describes the marginal effect of y_{i1}. Specifically, decreasing y_{i1} from $\overline{y}_1$ to $\overline{y}_1 - \Delta y$ makes college 1 less attractive to student i, and the region in which $\mu(i) = 1$ shrinks along the horizontal ϵ_{i1}-axis. The area I_1 depicts the set of students whose match differs when y_{i1} decreases. The induced change in the match probability, $\sigma_1 = \mathbb{P}(\mu(i) = 1 | z, \overline{y}_1, w_1)$, is the mass that the density of $(\epsilon_{i1}, \eta_{1i})$ puts on I_1, or, for $(z_i, y_{i1}, w_{i1}) = (z, \overline{y}_1, w_1)$,

$$\frac{\partial \sigma_1}{\partial y_{i1}} = \frac{\partial r^1}{\partial y_{i1}} \cdot \sum_{L \in \mathcal{L}} \lambda_L \cdot g'_{1,L} = \sum_{L \in \mathcal{L}} \lambda_L \cdot g'_{1,L}, \tag{7}$$

where $g'_{1,L}$ denotes the first derivative of $g_{1,L}$. By scale normalization, $\frac{\partial r^1(\overline{y}_1)}{\partial y_1} = 1$.

Panel (c) shows a similar graph in which decreasing w_{i1} from w_1 to $w_1 - \Delta w$ makes college 1 less likely to be feasible to student i. Hence, the region $\mu(i) = 1$ shrinks along the vertical η_{1i}-axis. The change in the match probability induced by the decrease in w_{i1} is the mass that the density of $(\epsilon_{i1}, \eta_{1i})$ puts on the area I_2, or,

$$\frac{\partial \sigma_1}{\partial w_{i1}} = \sum_{L \in \mathcal{L}} \lambda'_L \cdot g_{1,L}, \tag{8}$$

for $(z_i, y_{i1}, w_{i1}) = (z, \overline{y}_1, w_1)$, where λ'_L denotes the first derivative of λ_L.

In panel (d), the decrease in z_i makes college 1 less attractive and less likely to be feasible for student i, because z_i enters both student and college preferences. Besides, how z_i changes the region of $\mu(i) = 1$ depends on the shape of the functions $u^1(\cdot)$ and $v^1(\cdot)$. The change in the match probability induced by the change in z_i corresponds to the area I_3, or, for $(z_i, y_{i1}, w_{i1}) = (z, \overline{y}_1, w_1)$,

$$\frac{\partial \sigma_1}{\partial z_i} = \frac{\partial u^1}{\partial z_i} \cdot \sum_{L \in \mathcal{L}} \lambda_L \cdot g'_{1,L} + \frac{\partial v^1}{\partial z_i} \cdot \sum_{L \in \mathcal{L}} \lambda'_L \cdot g_{1,L}. \tag{9}$$

Our identification result relies on the changes caused by z_i, y_{i1}, and w_{i1}. Plugging equations (7) and (8) into equation (9), we have, for $(z_i, y_{i1}, w_{i1}) = (z, \overline{y}_1, w_1)$,

$$\frac{\partial \sigma_1}{\partial z_i} = \frac{\partial u^1}{\partial z_i} \cdot \frac{\partial \sigma_1}{\partial y_{i1}} + \frac{\partial v^1}{\partial z_i} \cdot \frac{\partial \sigma_1}{\partial w_{i1}}. \tag{10}$$

This equation reflects the chain rule: the effect of z_i on the match probability, $\frac{\partial \sigma_1}{\partial z_i}$, is realized through its effects on utilities u_{i1} and v_{1i}, captured by $\frac{\partial u^1}{\partial z_i}$ and $\frac{\partial v^1}{\partial z_i}$, and the effects of the utilities on the match probability, captured by $\frac{\partial \sigma_1}{\partial y_{i1}}$ and $\frac{\partial \sigma_1}{\partial w_{i1}}$. In equation (10), the derivatives of the match probability can be calculated from the data, and the two unknowns, $\frac{\partial u^1}{\partial z_i}$ and $\frac{\partial v^1}{\partial z_i}$ are the parameters of interest.

Importantly, when w_{i1} varies, the conditional match probability changes, but the two unknowns remain constant, which is a consequence of w_{i1} entering v_{1i} linearly. If, for any z, $(\epsilon_{i1}, \eta_{1i})$ has enough variation such that two distinct values of w_{i1} produce two linearly independent equations that have a unique solution, we identify the unknowns. This requirement is formalized as Condition 3.5 later, which imposes a mild restriction on the distribution of $(\epsilon_{i1}, \eta_{1i})$ as shown in Example 3.6 below.

We then identify $\frac{\partial r^1}{\partial y_{i1}}$, which is not one when $y_{i1} \neq \overline{y}_1$. For any value of (z_i, y_{i1}, w_{i1}), plugging equations (7) and (8) into equation (9), we rearrange and obtain

$$\left(\frac{\partial \sigma_1}{\partial z_i} - \frac{\partial v^1}{\partial z_i} \cdot \frac{\partial \sigma_1}{\partial w_{i1}}\right) \cdot \frac{\partial r^1}{\partial y_{i1}} = \frac{\partial u^1}{\partial z_i} \cdot \frac{\partial \sigma_1}{\partial y_{i1}}. \tag{11}$$

Because all the terms except for $\frac{\partial r^1}{\partial y_{i1}}$ are either identified or known, as long as $\frac{\partial \sigma_1}{\partial z_i} - \frac{\partial v^1}{\partial z_i} \cdot \frac{\partial \sigma_1}{\partial w_{i1}} \neq 0$ for some value of (z_i, w_{i1}), $\frac{\partial r^1}{\partial y_{i1}}$ is identified.

Below, we extend this example to the case with multiple colleges. We derive equation (10) for each college, in which the marginal effect of z_i on the probability of being matched with each college is the sum of its marginal effects on (u^c, v^c) for all c. By varying $\{w_{ic}\}_c$, we form a system of equations in $\{\frac{\partial u^c}{\partial z_i}, \frac{\partial v^c}{\partial z_i}\}_c$. The identification of $\frac{\partial r^c}{\partial y_{ic}}$ is the same as above and relies on a generalized version of equation (11) because we can identify $\frac{\partial r^c}{\partial y_{ic}}$ for each c separately by holding $\frac{\partial r^d}{\partial y_{id}} = 1$ for $d \neq c$.

3.1.2 Formal Identification Results

We now formalize the nonparametric identification of $\frac{\partial u^c}{\partial z_i}$, $\frac{\partial r^c}{\partial y_{ic}}$, and $\frac{\partial v^c}{\partial z_i}$, which extends Matzkin (2019) who exploits excluded variables to identify nonparametric nonseparable discrete choice models.

Assumption 3.3. *For each $c \in \mathbf{C}$, (i) $x_{ic}^k, \forall k = 1, \ldots, d_x$, are continuous random variables; (ii) the functions, u^c, r^c, and v^c, are continuously differentiable; and (iii) F is continuously differentiable.*

Part (i) of Assumption 3.3 requires that all covariates are continuous (but not necessarily full-support), which is relaxed in Section 3.3.

Assumption 3.4. *(ϵ_i, η_i) is distributed independently of x_i.*

This exogeneity assumption is made for simplicity. One way to relax this assumption is to adopt a control function approach (Heckman and Robb, 1985; Blundell and Powell, 2004; Imbens and Newey, 2009). See Appendix C for a discussion.

Additionally, we need a condition on the derivatives of match probabilities w.r.t. the excluded variables. Let $\boldsymbol{\sigma} \equiv (\sigma_1, \ldots, \sigma_C)'$ be a $C \times 1$ vector and define a $C \times 2C$ matrix $\Pi^*(z_i, y_i, w_i) \equiv (\frac{\partial \boldsymbol{\sigma}(z_i,y_i,w_i)}{\partial y_i'}, \frac{\partial \boldsymbol{\sigma}(z_i,y_i,w_i)}{\partial w_i'})$. The c^{th} row of Π^* is the derivatives of σ_c w.r.t. the vector of all excluded variables (y_i, w_i). We then consider a pair of distinct values of w_i, $\widehat{w}$ and $\widetilde{w}$, and define a $2C \times 2C$ matrix evaluated at $(z_i, y_i) = (z, \overline{y})$,

$$\Pi(z, \overline{y}, \widehat{w}, \widetilde{w}) \equiv \begin{pmatrix} \Pi^*(z, \overline{y}, \widehat{w}) \\ \Pi^*(z, \overline{y}, \widetilde{w}) \end{pmatrix}.$$

We impose the following testable condition on $\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})$.[12]

Condition 3.5. *For any z in the interior of $\mathcal{Z}$, there exist two values of w_i, $\widehat{w}$ and $\widetilde{w}$, in w_i's support conditional on $(z_i, y_i) = (z, \overline{y})$ such that $\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})$ has rank $2C$.*

Note that, for any value of z_i, Condition 3.5 only needs *two* values of w_i at which the matrix $\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})$ is full-rank. In other words, Condition 3.5 can hold even when there are infinitely many values of w_i at which $\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})$ is not full-rank. Intuitively, Condition 3.5 requires that the unobservables have reasonably sufficient variation such that w_i affects the supply (i.e., the feasible sets) and thus the matching probabilities in a "non-linear" way. The condition is violated if, for example, for some value of $(z, \overline{y})$, the student always chooses the outside option regardless of her feasible set. In this case, her conditional choice probabilities would not vary with w_i. The following one-college example shows that Condition 3.5 is plausible.[13]

Example 3.6. *Consider a one-college example: $\mathbf{C} = \{1\}$, and $\mathcal{L} = \{\{0\}, \{0, 1\}\}$. Equation (6) for $c = 1$ can be written as $\sigma_{1,i} = \lambda_{\{0,1\}}(\iota_{i1}) \cdot g_{1,\{0,1\}}(\tau_{i1})$ because $g_{1,\{0\}}(\tau_{i1}) = 0$. Recall that $\iota_{i1} = v^1(z_i) + w_{i1}$ and $\tau_{i1} = u^1(z_i) + r^1(y_{i1})$. We fix $y_{i1} = \overline{y}_1$ and have $r_1'(\overline{y}_1) = 1$. Condition 3.5 requires that, for any z in the interior of*

[12] For a formal statistical test of Condition 3.5 for a given value of z, one may use the method proposed by Chen and Fang (2019). Testing H_0: rank $(\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})) \leq 2C - 1$ against H_1: rank $(\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})) > 2C - 1$ is a special case of setup (1) in Chen and Fang (2019, p.1788).

[13] Our analysis of a nonparametric two-college model and common logit and probit models with two or more colleges also suggests that Condition 3.5 is not restrictive and that the *failure* of Condition 3.5 actually imposes strict restrictions on the supply, or the conditional probabilities of different feasible sets. For details, see online supplementary material available at `https://drive.google.com/file/d/1fJo4q4b3qZJ_PGycv1PF_Ag9FxS3TE4u/view?usp=sharing`.

$\mathcal{Z}$, there are two values of w_{i1}, $\widehat{w}_1$ and $\widetilde{w}_1$, such that the following matrix is full-rank:

$$\Pi(z, \overline{y}_1, \widehat{w}_1, \widetilde{w}_1) = \begin{pmatrix} \lambda_{\{0,1\}}(\widehat{\iota}_1) \cdot g'_{1,\{0,1\}}(\tau_1) & \lambda'_{\{0,1\}}(\widehat{\iota}_1) \cdot g_{1,\{0,1\}}(\tau_1) \\ \lambda_{\{0,1\}}(\widetilde{\iota}_1) \cdot g'_{1,\{0,1\}}(\tau_1) & \lambda'_{\{0,1\}}(\widetilde{\iota}_1) \cdot g_{1,\{0,1\}}(\tau_1) \end{pmatrix},$$

where $\widehat{\iota}_1 \equiv v^1(z) + \widehat{w}_1$, $\widetilde{\iota}_1 \equiv v^1(z) + \widetilde{w}_1$, and $\tau_1 \equiv u^1(z) + r^1(\overline{y}_1)$. A necessary condition for Condition 3.5 is $g'_{1,\{0,1\}}(\tau_1) \neq 0$, which is satisfied if ϵ_{i1} has a strictly increasing cumulative distribution function. Given that $g'_{1,\{0,1\}}(\tau_1) \neq 0$ and $g_{1,\{0,1\}}(\tau_1) \neq 0$, Condition 3.5 is satisfied if $\frac{\lambda'_{\{0,1\}}(\widehat{\iota}_1)}{\lambda_{\{0,1\}}(\widehat{\iota}_1)} \neq \frac{\lambda'_{\{0,1\}}(\widetilde{\iota}_1)}{\lambda_{\{0,1\}}(\widetilde{\iota}_1)}$, or $\frac{\partial \log \lambda_{\{0,1\}}(\widehat{\iota}_1)}{\partial \iota_{i1}} \neq \frac{\partial \log \lambda_{\{0,1\}}(\widetilde{\iota}_1)}{\partial \iota_{i1}}$. The violation of Condition 3.5 stringently restricts $\lambda_{\{0,1\}}(\iota_{i1})$, or the probability of college 1 being feasible to i. Specifically, for fixed z, Condition 3.5 is violated if the supply elasticity w.r.t. w_{i1} is linear in w_{i1}, or $\frac{\partial \log \lambda_{\{0,1\}}(\iota_{i1})}{\partial \iota_{i1}}$ is a constant for all w_{i1}. This means that $\lambda_{\{0,1\}}(\iota_{i1}) = \exp(a + b\iota_{i1})$ with constants a and b, which is unlikely, if not impossible, to be consistent with the definition that $\lambda_{\{0,1\}}(\iota_{i1}) = \mathbb{P}(\eta_{1i} \geq \delta_1 - \iota_{i1})$.

Proposition 3.7. *Under Assumptions 3.3-3.4 and Condition 3.5, for all $c \in \mathbf{C}$ and $k \in \{1, \ldots, d_z\}$, $\frac{\partial u^c(z)}{\partial z_i^k}$, $\frac{\partial v^c(z)}{\partial z_i^k}$, and $\frac{\partial r^c(y_c)}{\partial y_{ic}}$ are identified for all (z, y) in the interior of $\mathcal{Z} \times \mathcal{Y}$.*

Our identification arguments proceed in two steps. First, to identify $\frac{\partial u^c(z)}{\partial z_i^k}$ and $\frac{\partial v^c(z)}{\partial z_i^k}$, we exploit the variation in the excluded variables (y_i, w_i) and derive a system of linear equations that generalizes equation (10), or Figure 1, for the one-college example. Specifically, fixing $y_i = \overline{y}$, for any value z and w in the interior of $\mathcal{Z}$ and $\mathcal{W}$, for each college $d \in \mathbf{C}$, and for any component of z_i, z_i^k, we have

$$\frac{\partial \sigma_d}{\partial z_i^k} = \sum_{c \in \mathbf{C}} \frac{\partial \sigma_d}{\partial y_{ic}} \frac{\partial u^c(z)}{\partial z_i^k} + \sum_{c \in \mathbf{C}} \frac{\partial \sigma_d}{\partial w_{ic}} \frac{\partial v^c(z)}{\partial z_i^k}, \tag{12}$$

which gives C linear equations of $2C$ unknowns. Note that equation (12) is for one value of w_i. By evaluating equation (12) at $d = 1, \ldots, C$ and two different values of w_i, $\widehat{w}$ and $\widetilde{w}$, and stacking them together, we have

$$\begin{pmatrix} \frac{\partial \boldsymbol{\sigma}(z, \overline{y}, \widehat{w})}{\partial z_i^k} \\ \frac{\partial \boldsymbol{\sigma}(z, \overline{y}, \widetilde{w})}{\partial z_i^k} \end{pmatrix} = \Pi(z, \overline{y}, \widehat{w}, \widetilde{w}) \cdot \begin{pmatrix} \frac{\partial \mathbf{u}(z)}{\partial z_i^k} \\ \frac{\partial \mathbf{v}(z)}{\partial z_i^k} \end{pmatrix}, \tag{13}$$

where $\boldsymbol{u} \equiv \left(u^1, \ldots, u^C\right)'$ and $\boldsymbol{v} \equiv \left(v^1, \ldots, v^C\right)'$ are the vectors of u^c and v^c for all $c \in \mathbf{C}$. Both $\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})$ and the left-hand side are known from the data. Hence, the invertibility of $\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})$ in Condition 3.5 guarantees the existence of a unique solution to this system, leading to the identification of $\frac{\partial u^c(z)}{\partial z_i^k}$ and $\frac{\partial v^c(z)}{\partial z_i^k}$. Importantly, this only requires a pair of distinct values of w_i to satisfy Condition 3.5.

Second, we identify $\frac{\partial r^c(y_c)}{\partial y_{ic}}$, starting with $c = 1$. Fixing $y_{ic} = \overline{y}_c$ for all $c \in \mathbf{C}\backslash\{1\}$ and for any $y_1 \in \mathcal{Y}_1$, we can generalize equation (11) for the one-college example as

$$\left(\frac{\partial \sigma_d}{\partial z_i^k} - \sum_{c \in \mathbf{C}} \frac{\partial \sigma_d}{\partial w_{ic}} \frac{\partial v^c(z)}{\partial z_i^k} - \sum_{c \in \mathbf{C}\backslash\{1\}} \frac{\partial \sigma_d}{\partial y_{ic}} \frac{\partial u^c(z)}{\partial z_i^k}\right) \frac{\partial r^1(y_1)}{\partial y_{i1}} = \frac{\partial \sigma_d}{\partial y_{i1}} \frac{\partial u^1(z)}{\partial z_i^k}. \tag{14}$$

Because all other terms in equation (14) are either known or already identified, we identify $\frac{\partial r^1(y_1)}{\partial y_{i1}}$ as long as the coefficient of $\frac{\partial r^1(y_1)}{\partial y_{i1}}$ is nonzero for some (z, w). Similarly, $\frac{\partial r^c(y_c)}{\partial y_{ic}}$ can be identified for all $c \in \mathbf{C} \setminus \{1\}$.

In certain empirical applications, identifying the above derivatives is sufficient, in which case a large-support assumption on excluded variables is not needed. When the functions (u^c, v^c, r^c) must be identified, a full-support condition on their arguments, (z_i, y_i), is often imposed. This is the case for our next result of identifying the joint distribution F and cutoffs δ_c, because we need each of (u^c, v^c, r^c) identified up to a constant. Additionally, a full-support assumption on w_i is also required. In other words, we will assume all observables, (z_i, y_i, w_i), have full support.

3.2 Identifying the Cutoffs and Joint Distribution of Unobservables

We now formalize the assumptions that are needed for the identification of the cutoffs, $\{\delta_c\}_c$, and the joint distribution of unobservables, F.

Assumption 3.8. *For all $c \in \mathbf{C}$, (i) the functions $u^c + r^c$ and v^c are identified;*[14] *(ii) y_{ic} and w_{ic} possess an everywhere positive Lebesgue density conditional on z_i; (iii) the range of the function r^c is the whole real line, and $\mathcal{W} = \mathbb{R}^C$; and (iv) the ρ_c-quantile of the marginal distribution of η_{ci} is 0, i.e., $\text{Quantile}_{\eta_{ci}}(\rho_c) \equiv \inf\{\eta_c : F_{\eta_c}(\eta_c) \geq \rho_c\} = 0$,*

[14]This requires Proposition 3.7, a full support assumption on (z_i, y_i), and location normalization on $u^c + r^c$ and v^c for each c. See Remark 3.1 for a discussion on location normalization.

for an arbitrary $\rho_c \in (0, 1)$.

Part (iv) is a location normalization on college preferences, as mentioned in Remark 3.1, which can be done college-by-college. Alternatively, one may replace part (iv) by normalizing cutoffs $\{\delta_c\}_c$ to zero.

Before presenting our formal results, we give some intuitions. To identify cutoff δ_c, under the full-support assumption on student preferences (parts ii and iii of Assumption 3.8), we consider a mass of students to whom all colleges except for c are unacceptable. The probability of these students matching with c is $1-F_{\eta_c}(\delta_c - v^c(z_i) - w_{ic})$. Given the location normalization of F_{η_c} (part iv of Assumption 3.8), finding the maximum value of $v^c(z_i) + w_{ic}$ that sets this probability to $1 - \rho_c$ identifies δ_c.

To identify F, we use the conditional probability of being *unmatched*. Let us illustrate the intuition in the same one-college example as in Section 3.1.1. Given $(z, y_1, w_1) \in \mathcal{Z} \times \mathcal{Y}_1 \times \mathcal{W}_1$, Figure 2 shows the partition of the space of $(\epsilon_{i1}, \eta_{1i})$ by matching outcome, $\mu(i) = 0$ or 1. The conditional probability of $\mu(i) = 0$, highlighted in purple in the figure, can be decomposed into three parts, R_1, R_2, and R_3. Moreover,

$$\begin{aligned}
&\mathbb{P}(\mu(i) = 0 \mid z, y_1, w_1) \\
=&\mathbb{P}(\epsilon_{i1} < -u^1(z) - r^1(y_1) \text{ or } \eta_{1i} < \delta_1 - v^1(z) - w_1 \mid z, y_1, w_1) \\
=&\mathbb{P}(R_1 \cup R_2 \mid z, y_1, w_1) + \mathbb{P}(R_3 \cup R_2 \mid z, y_1, w_1) - \mathbb{P}(R_2 \mid z, y_1, w_1), \qquad (15)
\end{aligned}$$

where $\mathbb{P}(R_2 \mid z, y_1, w_1)$ is the joint CDF of $(\epsilon_{i1}, \eta_{1i})$, $F(-u^1(z) - r^1(y_1), \delta_1 - v^1(z) - w_1)$, our parameter of interest.

Further, $\mathbb{P}(R_1 \cup R_2 \mid z, y_1, w_1) = F_{\epsilon_1}(-u^1(z) - r^1(y_1))$ is the marginal CDF of ϵ_{i1}. It can be identified by considering the subset of students whose realization of w_{i1} is high enough so that the college will be feasible no matter what value η_{1i} takes. In other words, we can identify $\mathbb{P}(R_1 \cup R_2 \mid z, y_1, w_1)$ by "shutting down" the effects of college preferences. Similarly, $\mathbb{P}(R_3 \cup R_2 \mid z, y_1, w_1)$ can be identified by focusing on the subset of students whose realization of $r^1(y_{i1})$ is large enough so that those students find college 1 acceptable no matter what value ϵ_{i1} takes. As $\mathbb{P}(\mu(i) = 0 \mid z, y_1, w_1)$ is known from the data, once $\mathbb{P}(R_1 \cup R_2 \mid z, y_1, w_1)$ and $\mathbb{P}(R_3 \cup R_2 \mid z, y_1, w_1)$ are identified, equation (15) implies that $\mathbb{P}(R_2 \mid z, y_1, w_1)$ and thus F are identified.

We are now ready to present our formal results.

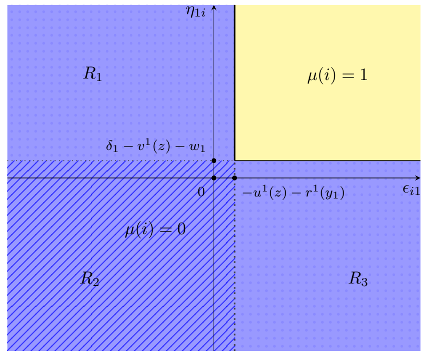


Figure 2: Partitioning the Space of Unobservables in the One-college Case

Notes: This figure shows the partition of the space of the unobservables $(\epsilon_{i1}, \eta_{1i})$ by the matching outcome ($\mu(i) = 0$ or 1) in a one-college setting given $(z_i, y_{i1}, w_{i1}) = (z, y_1, w_1)$.

Proposition 3.9. *Under Assumptions 3.4 and 3.8, (i) cutoffs $\{\delta_c\}_c$ are identified; (ii) the joint distribution of (ϵ_i, η_i), F, is identified; and (iii) acceptability threshold T_c for any college c with vacancies is identified, but T_c for c with a binding capacity constraint is only partially identified, $T_c \leq \delta_c$.*

To show part (ii) in a many-college setting, with the full, large support (parts ii and iii of Assumptions 3.8), we apply the same argument as in the one-college example to each college c sequentially, using extreme values of $r^c(y_{ic})$ or w_{ic} to "shut down" the effects of students' preference for college c or c's preference for students.

Part (iii) is a consequence of part (i) and the definition of cutoffs (equation 1). If c reaches its capacity, T_c can be any value below δ_c and result in the same stable matching. We can identify T_c by imposing additional assumptions, e.g., a full-capacity college having the same acceptability threshold as some college with vacancies.

3.3 Practical Issues

When our identification results are taken to the data, there can be practical issues. For example, vector z_i may include discrete variables such as gender, and the researcher may not have sufficient excluded variables. Below, we address these issues.

Discrete Random Variables. Our results can be extended to the case where z_i contains discrete variables but have at least one continuous variable.

Suppose that $z_i = (z_i^1, z_i^2)$, where z_i^1 is a vector of discrete variables and z_i^2 is a vector of continuous variables. Let the support of z_i^1 be a finite set of points $\{z^{1,1}, z^{1,2}, \ldots, z^{1,J}\}$. For $z_i^1 = z^{1,j}$, we define functions $(u^{c,j} + r^c, v^{c,j})$. Conditional on $z_i^1 = z^{1,1}$, we apply the results in Sections 3.1 and 3.2 to identify $\{u^{c,1} + r^c, v^{c,1}, \delta_c\}_c$ and F, requiring Assumptions 3.3(ii) and (iii), 3.4, 3.8(ii)-(iv), Condition 3.5, location normalization on $u^{c,1} + r^c$ and $v^{c,1}$ for each c, and (y_i, z_i^2) having full support. For $j \neq 1$, conditional on $z_i^1 = z^{1,j}$, using the conditional match probability of a mass of students for whom c is the only acceptable college, we can use F to identify $v^{c,j}$ for each c; similarly, using the conditional match probability of a mass of students whose feasible set is $\{0, c\}$, we can identify the function $u^{c,j} + r^c$.

Insufficient Excluded Variables. Allowing for college-level heterogeneity implies that we need to recover $3C$ preference parameters, $\{u^c, r^c, v^c\}_c$. As seen in Section 3.1, we need $2C$ excluded variables for identification of their derivatives.

The lack of sufficient excluded variables leads to a loss of identification, but not all is lost. We show that there is a trade-off between the identifiable degree of heterogeneity and the number of excluded variables. Consider $\mathbf{C}_1, \mathbf{C}_2 \subset \mathbf{C}$ with cardinality κ_1 and κ_2, respectively. For all $c \in \mathbf{C}_1$, there is a single excluded variable in student preferences, $y_{ic} = y_{i*}$; for all $c \in \mathbf{C}_2$, there is a single excluded variable in college preferences, $w_{ic} = w_{i*}$. We have the following identification result.

Proposition 3.10. *Suppose the preference heterogeneity is reduced: $u^c = u^*$ and $r^c = r^*$ for all $c \in \mathbf{C}_1$, and $v^c = v^*$ for all $c \in \mathbf{C}_2$. For any $c \in \mathbf{C}$, the derivatives of (u^c, r^c, v^c) are identified if: (i) Assumptions 3.3 and 3.4 hold; and (ii) the following rank conditions hold: (a) If $C \leq \kappa_1 + \kappa_2 - 2$, for any z in the interior of $\mathcal{Z}$, there exists a value of w_i, $\widehat{w}$, in w_i's support conditional on $(z_i, y_i) = (z, \overline{y})$ such that $\Pi^*(z, \overline{y}, \widehat{w})$ has column rank at least $2C - \kappa_1 - \kappa_2 + 2$; or (b) if $C > \kappa_1 + \kappa_2 - 2$, for any z in the interior of $\mathcal{Z}$, there exist two values of w_i, $\widehat{w}$ and $\widetilde{w}$, in w_i's support conditional on $(z_i, y_i) = (z, \overline{y})$ such that $\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})$ has column rank at least $2C - \kappa_1 - \kappa_2 + 2$.*

In other words, we need only $C - \kappa_1 + 1$ excluded variables on the student size and $C - \kappa_2 + 1$ on the college side to identify the less heterogeneous model.[15] Importantly, this still allows for coefficient heterogeneity across colleges in $\mathbf{C}_1$ and $\mathbf{C}_2$. For example,

[15] Note that the rank condition in Proposition 3.10 is weaker than Condition 3.5.

it can be incorporated parametrically by interacting college-specific observables with z_i. For $c \in \mathbf{C}_1$, let $u^c(z_i) = \beta p_c z_i$, where p_c is a college-specific observable. This amounts to z_i having a college-specific parameter $\beta_c = \beta p_c$.

Note that, even though we are moving towards homogeneity here, the model still remains non-nested with respect to the vertically differentiated models of Agarwal (2015) and Diamond and Agarwal (2017). In fact, our model allows for greater heterogeneity at the very least through pair-specific unobservables, ϵ_{ic} and η_{ci}.

Since our identification approach is constructive, it directly implies a nonparameteric estimator. In the Monte Carlo simulations in Appendix D, we apply this estimator in a semiparametric setting to avoid the well-known curse of dimensionality; yet the curse remains. Instead, we find that a parametric model based on a Bayesian approach works well. Below, we apply it to a real-life setting.

4 Secondary School Admissions in Chile

Guided by our identification results, we study the admissions to public and private secondary schools (grades 9-12) in Chile in 2007. The market is organized similarly to college admissions in the US. It is decentralized, both sides have preferences, and students do not submit rank-order lists of schools. We use the parametric Bayesian approach for preference estimation and then conduct counterfactual analysis.

4.1 Institutional Background and Data

Since 2003, secondary education has been compulsory for all Chileans up to 21 years of age. In principle, a public school must accept any student who is willing to enroll; a private school can be subsidized by the government or non-subsidized, but in either case, it can select students based on its preferences.

We focus on a relatively independent market, *Market Valparaiso*, that includes five municipalities (Valparaiso, Viña del Mar, Concon, Quilpue, and Villa Alemaña) as defined in Gazmuri (2017). Our data includes the municipality of a student's residence and the geographical coordinates of each school, which identifies everyone in the market. A student is defined to be in Market Valparaiso if in 2008 she resided in a municipality within the boundary of the market. A secondary school is in Market

Valparaiso if it is located in and admits students from the market.[16] In total, there are $9,304$ students and 125 schools. This reasonably large size makes it plausible that the continuum market in Section 3 is a good approximation.

We use the SIMCE dataset, provided by *La Agencia de Calidad de la Educaciõn* (Agency for the Quality of Education), on all 10th graders in 2008 to identify who started secondary school in 2007. The SIMCE is Chile's standardized testing program and tracks students' math and language performance. The data includes students' parental income, parental schooling, and other characteristics from a parental questionnaire sent home with students. Appendix E has a detailed description.

Table 2 summarizes the student characteristics. Among the students, 42% attend a public school, 44% attend a private subsidized school, 13% are in a private non-subsidized school, and less than 1% choose an outside option (i.e., attending a school outside Market Valparaiso). Students at each type of schools are very different on every dimension except gender as shown in Table 2.

Table 2: Summary Statistics of Student Characteristics

	All students (N=9,304)		Students enrolled in a secondary school of type: Public (N=3,951)		Private subsidized (N=4,083)		Private non-subsidized (N=1,211)		Outside Option (N=59)	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Female	0.51	0.50	0.54	0.50	0.48	0.50	0.52	0.50	0.46	0.50
Language score	0.49	0.29	0.36	0.26	0.55	0.27	0.75	0.23	0.52	0.29
Math score	0.50	0.29	0.34	0.24	0.56	0.26	0.78	0.20	0.47	0.27
Composite score	0.49	0.29	0.34	0.24	0.56	0.26	0.78	0.20	0.49	0.28
Mother's education (years)	13.97	3.20	12.43	2.79	14.33	2.81	17.78	1.86	14.05	2.83
Parental income (CLP)	430,551	494,025	194,710	147,046	357,899	283,635	1,447,069	541,758	387,288	389,583
Distance to enrolled school (km)	2.81	2.67	2.30	2.18	2.99	2.68	3.61	3.23	-	-

Notes: This table describes student characteristics in Market Valparaiso. Scores are measured in percentile rank (from 0 to 1). CLP stands for Chilean peso. Parental income is measured in 2008 when 1 USD was about 522 CLP.

We describe the 125 schools in Table 3. Most school attributes are calculated from student characteristics of *the 10th graders in a school in 2006*, and thus are pre-determined in the 2007 admissions that we study. A school's teacher quality is the average number of years the teachers have had in their teaching career. We also calculate the average tuition fees that are charged by each school in 2008. As tuition fees are largely fixed within each school and not completely flexible at the student level, we can still consider the problem as matching without transfers. Finally,

[16] There are 38 secondary schools located in Market Valparaiso that do not have any 9th graders from Market Valparaiso in 2007 and another 15 that have fewer than three 9th graders from Market Valparaiso on average in 2005, 2007, and 2009. We consider these 53 schools as outside options.

Table 3: Summary Statistics of School Attributes

	Public schools ($C = 25$)		All private schools: subsidized ($C = 67$)		All private schools: non-subsidized ($C = 33$)		Full capacity private schools: subsidized ($C = 29$)		Full capacity private schools: non-subsidized ($C = 6$)	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Average language score	0.33	0.12	0.54	0.15	0.71	0.14	0.57	0.16	0.66	0.21
Average math score	0.30	0.14	0.55	0.17	0.73	0.15	0.57	0.18	0.66	0.22
Average composite score	0.30	0.14	0.55	0.17	0.74	0.15	0.58	0.18	0.67	0.23
Average mother's edu. (years)	12.13	0.89	14.70	1.33	17.41	0.84	14.99	1.28	16.92	1.05
Fraction of female students	0.51	0.29	0.48	0.21	0.49	0.22	0.51	0.23	0.48	0.07
Median parental income (CLP)	154,000	20,000	331,343	147,411	1,278,788	481,863	346,552	163,079	950,000	440,454
Teacher experience (years)	17.51	6.21	13.54	7.66	18.07	8.90	13.55	8.59	13.85	9.80
Tuition (CLP)	3,283	1,409	17,444	10,576	59,894	8,658	19,078	9,894	57,648	12,107
Capacity	-	-	73.72	69.59	48.76	28.93	56.07	33.95	35.67	33.07
Valparaiso student enrollment[a]	158.04	134.60	60.94	58.55	36.70	26.82	50.83	31.90	34.17	32.36

Notes: This table describes the attributes of the schools in Market Valparaiso. Median parental income and tuition are measured in 2008 when 1 USD was about 522 CLP. [a] This excludes students who are not from Market Valparaiso.

we construct each private school's capacity from enrollment data in multiple years (Appendix E). As public schools cannot select students, their capacity is irrelevant.

4.2 Empirical Model

We allow student preferences to be school-type-specific. For student i, the utility of attending school c of type $t \in$ {public, private non-subsidized, private subsidized} is

$$u_{ict} = \alpha_{ft} \times female_i + \alpha_{mt} \times male_i + X'_{ic}\beta_t + \sigma_t \epsilon_{ic}, \tag{16}$$

where α_{ft} is a school-type fixed effect for female students; $female_i$ is a dummy variable for female; α_{mt} and $male_i$ are similarly defined; ϵ_{ic} is i.i.d. standard normal; σ_t (> 0) allows type-specific variances; and X_{ic} are student-school-specific variables, including (i) the distance between i's residence and school c; (ii) 6 school attributes, and (iii) 3 interactions between school attributes and student characteristics.[17]

Each student has an outside option, $u_{i0} = \epsilon_{i0}$, with ϵ_{i0} being standard normal. We impose the usual scale normalization through $\sigma_t^2 = 1$ for public school,[18] while

[17] Specifically, the 6 school attributes are tuition (in logarithm), average teacher experience, median parental income among students (in logarithm), fraction of female students, average composite score, and average mother's education. The last four variables are measured among the 2006 10th graders who are already in a secondary school in 2007. The 3 interactions include tuition interacted with student's parental income, school average composite score interacted with student composite score, and school average mother's education interacted with student mother's education.

[18] In addition, the variance of ϵ_{i0} is normalized to be one because there is insufficient variation to

the location normalization is imposed by setting the deterministic part of u_{i0} to zero.

School preferences are also type-specific, but public schools do not have a utility function because they cannot select students. For private school c of type t (subsidized or non-subsidized), its acceptability threshold is $T_c = 0$, and its utility function is

$$v_{cit} = \theta_t + Z'_{ci}\gamma_t + \eta_{ci}, \tag{17}$$

where θ_t is a type-specific intercept; η_{ci} is i.i.d. standard normal; and the vector Z_{ci} includes (i) 5 student characteristics and (ii) 3 interactions between student characteristics and school attributes.[19]

In school preferences, the variance of η_{ci} being one is the scale normalization and $T_c = 0$ is the location normalization. Allowing the type-specific intercept θ_t and assuming $T_c = 0$ imply that schools of the same type have the same acceptability threshold. Because each type has some schools with vacancies, we can separately identify δ_c and T_c. Otherwise, we would lose its identification (Proposition 3.9).

Guidance from Section 3. The above specification uses distance as an (i, c)*-pair-specific* excluded variable in student preferences. In school preferences, math and language scores are *student-specific* excluded variables, leading us to invoke Proposition 3.10 and limit heterogeneity. Our scale and location normalization is guided by Proposition 3.9 and Remark 3.1, while certain normalization is imposed via F.

4.3 Estimation and Results

We use a Bayesian approach with a Gibbs sampler for the estimation, which is first illustrated in Monte Carlo simulations (Appendix D.3). We provide the details on the updating of the Markov Chain in Appendices D.3 and F.

The estimation results are summarized in Table 4. A caveat is in order when we interpret the results. Because we do not deal with endogeneity issues that may arise

estimate the variance of u_{i0}. Only 59 students (out of 9,304) choose an outside option (see Table 2).

[19]The 5 student characteristics are female (a dummy variable), math score, language score, mother's education, and parental income (in logarithm). The 3 interactions include student gender interacted with the fraction of female students at the school, student math score interacted with school average math score, and student language score interacted with school average language score.

Table 4: Estimation Results: Student and School Preferences

	Public schools		Private Schools: Subsidized		Private Schools: Non-subsidized	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
Panel A. Student Preferences						
Female	0.458	(0.834)	10.500***	(0.611)	9.848*	(5.631)
Male	0.442	(0.834)	10.387***	(0.610)	9.089	(5.628)
Distance	-0.175***	(0.004)	-0.145***	(0.008)	-0.624***	(0.070)
log(tuition)	0.577***	(0.119)	-0.385***	(0.102)	-15.645***	(1.567)
log(tuition) × log(income)	-0.019*	(0.010)	0.044***	(0.008)	0.881***	(0.084)
log(median income)	-0.307***	(0.072)	-0.720***	(0.067)	2.454***	(0.544)
Teacher experience	0.003*	(0.002)	0.005***	(0.001)	0.020	(0.015)
Fraction of female students	-0.010	(0.034)	-0.320***	(0.054)	-1.428*	(0.773)
Average composite score	-1.948***	(0.124)	-2.354***	(0.132)	-16.104***	(2.241)
Average composite score × Composite score	6.282***	(0.232)	6.209***	(0.237)	18.167***	(1.616)
Average mother's education	0.038*	(0.023)	-0.283***	(0.024)	-1.572***	(0.298)
Average mother's education × Mother's education	0.007***	(0.001)	0.011***	(0.001)	0.056***	(0.007)
Standard deviation of the utility shock	Normalized to 1		1.093***	(0.063)	8.646***	(0.859)
Panel B. School Preferences						
Constant			3.686***	(0.852)	20.283***	(1.863)
Female			-3.545***	(0.193)	-1.438***	(0.274)
Female × Fraction of female			6.892***	(0.333)	1.968***	(0.443)
Math score			-3.251***	(0.541)	-7.053***	(0.920)
Average math score × Math score			7.955***	(0.938)	6.217***	(0.985)
Language score			-0.030	(0.460)	-4.593***	(0.859)
Average language score × Language score			1.053	(0.822)	5.087***	(1.135)
Mother's education			-0.010	(0.016)	-0.136***	(0.047)
log(income)			-0.225***	(0.072)	-1.030***	(0.146)

Notes: This table presents the posterior mean and standard deviation of each coefficient in student and school utility functions (equations 16 and 17). The Bayesian approach goes through a Markov Chain 1.75 million times, and the last 0.75 million iterations are used to calculate these statistics. *** p<0.01, ** p<0.05, * p<0.1

due to the correlation between preference shocks and school attributes or student characteristics, the estimates may not have a causal interpretation.

Panel A shows the estimates of student preferences. Most coefficients are of an expected sign. Interestingly, the coefficient on tuition is positive in the utility function for public schools, and parental income pushes it towards negative. This may reflect the fact that tuition at public schools is generally low (see Table 3) and may be correlated with unobserved school quality. There are also a few coefficients with an unexpected sign in school preferences. Panel B shows that non-subsidized schools negatively value a student's parental income and mother's education. This may be because, on average, students at those schools often have a high income (above 1.4 million CLP; see Table 2) and a highly educated mother (above 17 years).

Recall that each school's acceptability threshold is normalized to zero, so we can use the estimated school preferences to calculate if a student is acceptable to a private school. On average, a subsidized school finds 79% of the students acceptable, while a non-subsidized school finds 84% acceptable. The higher acceptability rate at non-subsidized schools does not imply that students are more often matched with them because their high tuition lowers their desirability to many students, especially those with a low parental income (see Table 4).

Before conducting counterfactual analysis, we evaluate model fit in Appendix F. It shows that our model fits the data reasonably well when we compare the observed matching with the one predicted based on our model.

4.4 Counterfactual: Prioritizing Low-income Students

We consider a counterfactual policy in which students from low-income families are prioritized for admissions to all schools. A student is of low income if her parental income is among the lowest 40%.[20] Each private school's preferences over students are made lexicographical: low-income students are above others, and within each group of students, a school ranks them as in the current regime; all low-income students are acceptable, while others' acceptability is the same as in the current regime.

To simulate the counterfactual outcome, we choose 1,500 draws of the coefficients in the Markov Chain in the Bayesian estimation.[21] We run the Gale-Shapley deferred acceptance with each draw and obtain 1,500 sets of counterfactual stable matchings. We then report the average of these counterfactual matchings; recall that we only have one matching outcome under the current regime—the observed one.

Table 5 presents the results. There are several noticeable patterns when we move from the current regime to the counterfactual. First, low-income students are in schools that have higher-ability and higher-income students in the same cohort, while the opposite is true for non-low-income students. Second, some low-income students leave public schools for private subsidized schools, while crowding out some other students to public schools. Lastly, the policy benefits low-income students and hurts

[20]This resembles a policy adopted in 2008 in Chile as documented by Gazmuri (2017). It benefits 44% of elementary school students in 2012 in terms of admission priorities and a tuition waiver. Appendix Table F.3 shows summary statistics of the students by income status.

[21]Specifically, there are 15 blocks of 100 draws. The blocks are equally spaced in the 0.75 million iterations in the Markov chain that are used to calculate the posterior means and standard deviations.

Table 5: Sorting and Student Welfare in the Current and Counterfactual Regimes

	Low-income students		Non-low-income students	
	Current	Counterfactual	Current	Counterfactual
	(1)	(2)	(3)	(4)
Average composite score (same cohort) at matched school	0.374	0.385	0.585	0.576
Average parental income (same cohort) at matched school	216,389	223,003	592,848	587,884
Fraction enrolled at each school type:				
Public	0.679	0.602	0.233	0.260
Private subsidized	0.315	0.392	0.533	0.504
Private non-subsidized	0.002	0.002	0.227	0.228
Outside option	0.004	0.003	0.008	0.008
Welfare effects of moving from current to counterfactual:				
Average utility change (reduction in distance, km)	0.332		−0.225	
Winners (fraction)	0.120		0.000	
Losers (fraction)	0.000		0.072	
Indifferent (fraction)	0.881		0.927	

Notes: The outcome in the current regime is the one observed in the data. To simulate the counterfactual outcome, we choose 1,500 draws of the coefficients in the Markov Chain in the Bayesian estimation and obtain 1,500 sets of stable matching outcomes. The statistics for the counterfactual regime are averages across the 1,500 outcomes. The average utility change is measured in terms of willingness to travel to a public school in kilometers.

others. On average, low-income students' welfare gain is equivalent to decreasing travel distance (to a public school) by 0.332 km. This gain is concentrated among 12% of the low-income students, while others are not affected. Correspondingly, 7.2% of the non-low-income students are worse off, and none is better off.

These results indicate that low-income students dislike private non-subsidized schools. We explore why it is the case. With the 1,500 draws of student preferences, we examine low-income students' favorite school of each type and measures how tuition and student characteristics contribute to student preferences. We find that low-income students on average value their favorite public school at 2.72 and private subsidized school at 2.56, while their favorite non-subsidized school is only valued at −21.76, in general unacceptable to them. We calculate the contribution of different variables to these differences by shutting down their effect in the utility functions. The results show that high tuition at non-subsidized schools is the main contributor. It is worth noting that tuition could be correlated with unobserved school quality. Hence, low-income students may dislike a non-subsidized school because of its high costs and/or their tastes. Additionally, mother's education, both a student's own and a school's average, is also an important factor. Student ability, measured by their composite score, and distance to each school do not appear to be as important.

In sum, giving low-income students access to schools fails to significantly change matching outcomes due to their own preferences. Low-income students are deterred from private non-subsidized schools by their high tuition. These findings are in line with the preference heterogeneity documented in public school choice (see, e.g., Abdulkadiroğlu et al., 2017; Kapor et al., 2020), although tuition plays no role there.

5 Concluding Remarks

We study nonparametric identification of agent preferences in many-to-one two-sided matching without transfers. We derive a set of sufficient conditions for identification and provide guidance for empirical studies. For example, our results clarify the data requirement for the identification of various levels of preference heterogeneity. To take our results to the data of a reasonably sized market, we propose a Bayesian approach with a Gibbs sampler whose performance is illustrated in Monte Carlo simulations. Our model encompasses many real-life matching markets, such as college admissions and school choice in many countries, in which our identification results and empirical method can be applied. Hence, this paper opens a new avenue for empirical research.

We illustrate our method in the context of secondary school admissions in Chile. The market is sizable with 9,304 students and 125 schools. Our estimates fit the data well. As an example of the usefulness of the estimates, we consider a counterfactual policy in which students from low-income families are prioritized for admissions to all schools. Although the policy benefits low-income students, its effects are small. Simply giving low-income students access to schools does not significantly change their outcomes due to their own preferences. Such insights are difficult to obtain without estimating the preferences of both sides. In this sense, our method can help provide an ex-ante evaluation of a range of alternative policies.

Our paper builds on an important assumption: an agent, such as a college, that accepts multiple match partners has responsive preferences and values each match partner independently. Relaxing this assumption may require new theoretical tools. In particular, a stable matching may not exist if college preferences are not responsive (Roth and Sotomayor, 1992).[22] We leave this extension to future research.

[22] Alternatively, substitutable preferences (Kelso and Crawford, 1982) can restore the existence of a stable matching. However, college utility functions will be defined on subsets of students, and cutoffs, which depend on how each college values each individual student, will not be well defined.

References

Abaluck, J. and A. Adams (2018): "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," *Working Paper*.

Abdulkadiroğlu, A., N. Agarwal, and P. A. Pathak (2017): "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match," *American Economic Review*, 107, 3635–3689.

Abdulkadiroğlu, A. and T. Sönmez (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747.

Agarwal, N. (2015): "An Empirical Model of the Medical Match," *American Economic Review*, 105, 1939–1978.

Agarwal, N. and P. Somaini (2018): "Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism," *Econometrica*, 86, 391–444.

——— (2020a): "Empirical Models of Non-Transferable Utility Matching," .

——— (2020b): "Revealed preference analysis of school choice models," *Annual Review of Economics*, 12, 471–501.

Artemov, G., Y.-K. Che, and Y. He (2020): "Strategic 'Mistakes': Implications for Market Design Research," Manuscript.

Aue, R., T. Klein, and J. Ortega (2020): "What happens when separate and unequal school districts merge?" *ZEW Discussion Paper*.

Azevedo, E. M. and J. D. Leshno (2016): "A supply and demand framework for two-sided matching markets," *Journal of Political Economy*, 124, 1235–1268.

Barseghyan, L., M. Coughlin, F. Molinari, and J. C. Teitelbaum (2019): "Heterogeneous Choice Sets and Preferences," *Working paper*.

Barseghyan, L., F. Molinari, and M. Thirkettle (2021): "Discrete Choice under Risk with Limited Consideration," .

Blundell, R. W. and J. L. Powell (2004): "Endogeneity in semiparametric binary response models," *The Review of Economic Studies*, 71, 655–679.

Calsamiglia, C., C. Fu, and M. Güell (2020): "Structural estimation of a model of school choices: The boston mechanism versus its alternatives," *Journal of Political Economy*, 128, 642–680.

Cattaneo, M. D., X. Ma, Y. Masatlioglu, and E. Suleymanov (2020): "A random attention model," *Journal of Political Economy*, 128, 2796–2836.

Chen, Q. and Z. Fang (2019): "Improved inference on the rank of a matrix," *Quantitative Economics*, 10, 1787–1824.

Chesher, A. (2003): "Identification in nonseparable models," *Econometrica*, 71, 1405–1441.

Chiappori, P.-A. and B. Salanié (2016): "The Econometrics of Matching Models," *Journal of Economic Literature*, 54, 832–861.

Chiappori, P.-A., B. Salanié, and Y. Weiss (2017): "Partner choice, investment in children, and the marital college premium," *American Economic Review*, 107, 2109–67.

Choo, E. and A. Siow (2006): "Who Marries Whom and Why," *Journal of Political Economy*, 114, 175–201.

Dagsvik, J. K. (2000): "Aggregation in matching markets," *International Economic Review*, 41, 27–58.

Diamond, W. and N. Agarwal (2017): "Latent indices in assortative matching models," *Quantitative Economics*, 8, 685–728.

Fack, G., J. Grenet, and Y. He (2019): "Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions," *American Economic Review*, 109, 1486–1529.

Fox, J. T. (2010): "Identification in matching games," *Quantitative Economics*, 1, 203–254.

——— (2018): "Estimating matching games with transfers," *Quantitative Economics*, 9, 1–38.

Fox, J. T., C. Yang, and D. H. Hsu (2018): "Unobserved Heterogeneity in Matching Games," *Journal of Political Economy*.

Gale, D. and L. Shapley (1962): "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 69, 9–15.

Galichon, A., S. D. Kominers, and S. Weber (2019): "Costly concessions: An empirical framework for matching with imperfectly transferable utility," *Journal of Political Economy*, 127, 2875–2925.

Galichon, A. and B. Salanie (2020): “Cupid’s Invisible Hand: Social Surplus and Identification in Matching Models,” *Working Paper*.

Gazmuri, A. (2017): “School Segregation in the Presence of Student Sorting and Cream-Skimming,” *Working paper, Toulouse School of Economics*.

Gelman, A. and D. B. Rubin (1992): “Inference from iterative simulation using multiple sequences,” *Statistical science*, 7, 457–472.

Graham, B. S. (2011): “Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers,” North-Holland, vol. 1 of *Handbook of Social Economics*, 965–1052.

Grenet, J., Y. He, and D. Kübler (2022): “Preference Discovery in University Admissions: The Case for Dynamic Multioffer Mechanisms,” *Journal of Political Economy*, 130, 1427–1476.

Gualdani, C. and S. Sinha (2020): “Partial Identification in Matching Models for the Marriage Market,” *Working paper, Toulouse School of Economics, https://arxiv.org/abs/1902.05610*.

He, Y. (2017): “Gaming the Boston School Choice Mechanism in Beijing,” *Working Paper, Rice University and Toulouse School of Economics*.

He, Y. and T. Magnac (2020): “Application costs and congestion in matching markets,” *Working Paper, Rice University and Toulouse School of Economics*.

Heckman, J. and R. Robb (1985): “Alternative methods for evaluating the impact of interventions: An overview,” *Journal of econometrics*, 30, 239–267.

Imbens, G. W. and W. K. Newey (2009): “Identification and estimation of triangular simultaneous equations models without additivity,” *Econometrica*, 77, 1481–1512.

Kapor, A. J., C. A. Neilson, and S. D. Zimmerman (2020): “Heterogeneous beliefs and school choice mechanisms,” *American Economic Review*, 110, 1274–1315.

Kelso, A. S. and V. P. Crawford (1982): “Job matching, coalition formation, and gross substitutes,” *Econometrica: Journal of the Econometric Society*, 1483–1504.

Logan, J. A., P. D. Hoff, and M. A. Newton (2008): "Two-Sided Estimation of Mate Preferences for Similarities in Age, Education, and Religion," *Journal of the American Statistical Association*, 103, 559–569.

Matzkin, R. L. (2007): "Chapter 73 Nonparametric identification," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. 6, 5307–5368.

——— (2019): "Constructive identification in some nonseparable discrete choice models," *Journal of Econometrics*, 211, 83 – 103.

Menzel, K. (2015): "Large Matching Markets as Two-Sided Demand Systems," *Econometrica*, 83, 897–941.

——— (2017): "Strategic network formation with many agents," *Working Paper, New York University.*

——— (2022): "Strategic network formation with many agents," .

Pais, J., Á. Pintér, and R. F. Veszteg (2020): "Decentralized matching markets with(out) frictions: a laboratory experiment," *Experimental Economics*, 23, 212–239.

Petrin, A. and K. Train (2010): "A control function approach to endogeneity in consumer choice models," *Journal of marketing research*, 47, 3–13.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

Rossi, P. E., G. M. Allenby, and R. McCulloch (2012): *Bayesian statistics and marketing*, John Wiley & Sons.

Roth, A. E. (1984): "The evolution of the labor market for medical interns and residents: a case study in game theory," *Journal of political Economy*, 92, 991–1016.

Roth, A. E. and M. Sotomayor (1992): "Two-sided matching," *Handbook of game theory with economic applications*, 1, 485–541.

Sinha, S. (2015): "Identification and Estimation in One-to-one Matching Models with Nonparametric Unobservables," *Toulouse School of Economics.*

Uetake, K. and Y. Watanabe (2020): "Entry by merger: Estimates from a two-sided matching model with externalities," *Available at SSRN 2188581.*

Online Appendix to

Identification and Estimation in Many-to-one Two-sided Matching without Transfers

YingHua He Shruti Sinha Xiaoting Sun

A Proofs and Discussions from Section 3

Proof of Proposition 3.7. Equation (6) can be rewritten as

$$\begin{aligned}\sigma_{c,i} &= \Lambda_c(\tau_{i1},\dots,\tau_{iC};\iota_{i1},\dots,\iota_{iC})\\ &= \Lambda_c(u^1(z_i)+r^1(y_{i1}),\dots,u^C(z_i)+r^C(y_{iC});v^1(z_i)+w_{i1},\dots,v^C(z_i)+w_{iC}),\end{aligned}$$

where Λ_c denotes some function. Under Assumption 3.3, Λ_c, u^c, r^c, and v^c are continuously differentiable and the observables are all continuously distributed.

To simplify notations, for $k=1,\dots,d_z$, let $u^c_{z_i^k}=\frac{\partial u^c}{\partial z_i^k}$, and similar notations are defined for v^c, r^c, and (y_{ic},w_{ic}). For colleges $c,\ d\in\mathbf{C}$, taking derivatives of $\sigma_{d,i}$ with respect to y_{ic}, w_{ic}, and z_i^k, respectively, one obtains

$$\frac{\partial\sigma_{d,i}}{\partial y_{ic}}=\frac{\partial\Lambda_d(x_i)}{\partial\tau_{ic}}r^c_{y_{ic}} \text{ and } \frac{\partial\sigma_{d,i}}{\partial w_{ic}}=\frac{\partial\Lambda_d(x_i)}{\partial\iota_{ic}}, \tag{A.1}$$

$$\frac{\partial\sigma_{d,i}}{\partial z_i^k}=\sum_{c\in\mathbf{C}}\frac{\partial\Lambda_d(x_i)}{\partial\tau_{ic}}u^c_{z_i^k}+\sum_{c\in\mathbf{C}}\frac{\partial\Lambda_d(x_i)}{\partial\iota_{ci}}v^c_{z_i^k}. \tag{A.2}$$

First, we show the identification of the derivatives of the functions u^c and v^c. We fix $y_i=\overline{y}$ and consider z in the interior of $\mathcal{Z}$. Recall that $r^c_{y_{ic}}=1$ when $y_{ic}=\overline{y}_c$ due to the scale normalization. Substituting equation (A.1) into equation (A.2), we get

$$\frac{\partial\sigma_{d,i}}{\partial z_i^k}=\sum_{c\in\mathbf{C}}\frac{\partial\sigma_{d,i}}{\partial y_{ic}}u^c_{z_i^k}+\sum_{c\in\mathbf{C}}\frac{\partial\sigma_{d,i}}{\partial w_{ic}}v^c_{z_i^k}. \tag{A.3}$$

Suppose that two different values of the C-dimensional vector of excluded regressors w_i, $\widehat{w}$ and $\widetilde{w}$, satisfy Condition 3.5. We define $\widehat{x}=(z,\overline{y},\widehat{w})$ and $\widetilde{x}=(z,\overline{y},\widetilde{w})$. Further, let $\widehat{\sigma}_{c,i}=\mathbb{P}(\mu(i)=c|\widehat{x})$ and $\widetilde{\sigma}_{c,i}=\mathbb{P}(\mu(i)=c|\widetilde{x})$. By evaluating equation (A.3) at

$d = 1, \ldots, C$ and $x_i = \widehat{x}$, $\widetilde{x}$, and stacking them together, we have

$$\underbrace{\begin{bmatrix} \frac{\partial \widehat{\sigma}_{1,i}}{\partial z_i^k} \\ \vdots \\ \frac{\partial \widehat{\sigma}_{C,i}}{\partial z_i^k} \\ \frac{\partial \widetilde{\sigma}_{1,i}}{\partial z_i^k} \\ \vdots \\ \frac{\partial \widetilde{\sigma}_{C,i}}{\partial z_i^k} \end{bmatrix}}_{\Sigma_{z_i^k}(z,\overline{y},\widehat{w},\widetilde{w})} = \underbrace{\begin{bmatrix} \frac{\partial \widehat{\sigma}_{1,i}}{\partial y_{i1}} & \cdots & \frac{\partial \widehat{\sigma}_{1,i}}{\partial y_{iC}} & \frac{\partial \widehat{\sigma}_{1,i}}{\partial w_{i1}} & \cdots & \frac{\partial \widehat{\sigma}_{1,i}}{\partial w_{iC}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \widehat{\sigma}_{C,i}}{\partial y_{i1}} & \cdots & \frac{\partial \widehat{\sigma}_{C,i}}{\partial y_{iC}} & \frac{\partial \widehat{\sigma}_{C,i}}{\partial w_{i1}} & \cdots & \frac{\partial \widehat{\sigma}_{C,i}}{\partial w_{iC}} \\ \frac{\partial \widetilde{\sigma}_{1,i}}{\partial y_{i1}} & \cdots & \frac{\partial \widetilde{\sigma}_{1,i}}{\partial y_{iC}} & \frac{\partial \widetilde{\sigma}_{1,i}}{\partial w_{i1}} & \cdots & \frac{\partial \widetilde{\sigma}_{1,i}}{\partial w_{iC}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \widetilde{\sigma}_{C,i}}{\partial y_{i1}} & \cdots & \frac{\partial \widetilde{\sigma}_{C,i}}{\partial y_{iC}} & \frac{\partial \widetilde{\sigma}_{C,i}}{\partial w_{i1}} & \cdots & \frac{\partial \widetilde{\sigma}_{C,i}}{\partial w_{iC}} \end{bmatrix}}_{\Pi(z,\overline{y},\widehat{w},\widetilde{w})} \times \begin{bmatrix} u_{z_i^k}^1 \\ \vdots \\ u_{z_i^k}^C \\ v_{z_i^k}^1 \\ \vdots \\ v_{z_i^k}^C \end{bmatrix}. \tag{A.4}$$

The first C rows of the matrix $\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})$ consist of the derivatives of the conditional match probabilities w.r.t. the excluded regressors, evaluated at $x_i = \widehat{x}$. The c^{th} row corresponds to the derivatives of $\sigma_{c,i}$ w.r.t. to the vector of excluded regressors (y_i, w_i). The second C rows of $\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})$ are constructed in a similar way with the derivatives being evaluated at $x_i = \widetilde{x}$.

Under Condition 3.5, there always exist $\widehat{w}$ and $\widetilde{w}$ such that $\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})$ in equation (A.4) is invertible. We then identify $u_{z_i^k}^c$ and $v_{z_i^k}^c$ for all c by solving a system of linear equations. Formally, let $\Pi_{z_i^k}^c$ be the matrix formed by replacing the c^{th} column of matrix $\Pi(z, \overline{y}, \widehat{w}, \widetilde{w})$ by the vector $\Sigma_{z_i^k}(z, \overline{y}, \widehat{w}, \widetilde{w})$ (defined in equation A.4). By the Cramer's rule, for any $c \in \mathbf{C}$,

$$u_{z_i^k}^c = \frac{|\Pi_{z_i^k}^c|}{|\Pi|}, \quad v_{z_i^k}^c = \frac{|\Pi_{z_i^k}^{c+C}|}{|\Pi|}.$$

Second, we identify the derivative of the function r^c for all c. We start with r^1 and fix $y_{ic} = \overline{y}_c$ for $c \in \mathbf{C} \backslash \{1\}$; under the scale normalization, $r_{y_{ic}}^c = 1$ for $c \in \mathbf{C} \backslash \{1\}$. For any $y_1 \in \mathcal{Y}_1$, Substituting equation (A.1) into equation (A.2), we obtain

$$\left(\frac{\partial \sigma_{d,i}}{\partial z_i^k} - \sum_{c \in \mathbf{C}} \frac{\partial \sigma_{d,i}}{\partial w_{ic}} v_{z_i^k}^c - \sum_{c \in \mathbf{C} \backslash \{1\}} \frac{\partial \sigma_{d,i}}{\partial y_{ic}} u_{z_i^k}^c \right) r_{y_{i1}}^1 = \frac{\partial \sigma_{d,i}}{\partial y_{i1}} u_{z_i^k}^1,$$

where all the terms except $r_{y_{i1}}^1$ are known or already identified. Therefore, as long as $\frac{\partial \sigma_{d,i}}{\partial z_i^k} - \sum_{c \in \mathbf{C}} \frac{\partial \sigma_{d,i}}{\partial w_{ic}} v_{z_i^k}^c - \sum_{c \in \mathbf{C} \backslash \{1\}} \frac{\partial \sigma_{d,i}}{\partial y_{ic}} u_{z_i^k}^c \neq 0$ for certain (z, w), $r_{y_{i1}}^1$ is identified. The derivative of the function r^c for $c \in \mathbf{C} \setminus \{1\}$ can be identified in the same manner.

So far, we have only considered z_i that shows up in the utility functions for all colleges and for both sides. For any element of z_i that is excluded from certain utility functions, the identification is a special case of the above proof by noting that some derivatives of the utility functions are zero. □

Proof of Proposition 3.9. We start with the identification of the cutoffs $\{\delta_c\}_c$, i.e., part (i). Because of the large support assumption on r^c (parts ii and iii of Assumption 3.8), for each $c \in \mathbf{C}$, there exists $\mathcal{J}_c \subseteq \mathcal{X}$ such that for any $x_i \in \mathcal{J}_c$, c is the only acceptable college with probability one, and that $Q(i \in \mathbf{I} : x_i \in \mathcal{J}_c) > 0$. Then,

$$\mathbb{P}(\mu(i) = 0 | x_i \in \mathcal{J}_c) = \mathbb{P}(v^c(z_i) + w_{ic} + \eta_{ci} < \delta_c | x_i \in \mathcal{J}_c) = F_{\eta_{ci}}(\delta_c - \iota_{ic}),$$

where the last equality is due to $\iota_{ic} = v^c(z_i) + w_{ic}$ and the independence between η_{ci} and x_i (Assumption 3.4). There exists a unique ι_c^* such that $\delta_c - \iota_c^* = \text{Quantile}_{\eta_{ci}}(\rho_c) = \inf\{(\delta_c - \iota_{ic}) : F_{\eta_{ci}}(\delta_c - \iota_{ic}) \geq \rho_c\}$. By Assumption 3.8(iv), $\delta_c - \iota_c^* = 0$, which identifies δ_c. To show part (iii), we use the definition of cutoffs, equation (1).

To prove part (ii), the identification of the distribution F, note that the conditional probability of being unmatched can be written as

$$\begin{aligned} \mathbb{P}(\mu(i) = 0 \mid x) &= \sum_{L \in \mathcal{L}} \lambda_{L,i} \mathbb{P}\left(0 = \arg\max_{d \in L} u_{id} \mid L, x\right) \\ &= \sum_{L \in \mathcal{L}} \lambda_L(\iota_1, \ldots, \iota_C) \cdot g_{0,L}(\tau_1, \ldots, \tau_C) \\ &\equiv \Lambda_0(\tau_1, \ldots, \tau_C, \iota_1, \ldots, \iota_C). \end{aligned} \tag{A.5}$$

Recall that for each $c \in \mathbf{C}$, $\tau_c = u^c(z) + r^c(y_c)$ and $\iota_c = v^c(z) + w_c$. Given that the functions $\{u^c + r^c, v^c\}_c$ are identified (part i of Assumption 3.8), for any given value (z, y, w), the arguments in equation (A.5), $\{\tau_c, \iota_c\}_c$, are known. Because $\mathbb{P}(\mu(i) = 0 \mid x)$ is observed from the data, the function Λ_0 is identified.

For each $c \in \mathbf{C}$, define $A_c \equiv \{u_{ic} < 0\} = \{u^c(z) + r^c(y_c) + \epsilon_{ic} < 0\} = \{\epsilon_{ic} < -\tau_c\}$, and $B_c \equiv \{v_{ci} < \delta_c\} = \{v^c(z) + w_c + \eta_{ci} < \delta_c\} = \{\eta_{ci} < \delta_c - \iota_c\}$. The parameter of interest can be written as

$$F(-\tau_1, \ldots, -\tau_C, \delta_1 - \iota_1, \ldots, \delta_C - \iota_C)$$

$$
\begin{aligned}
=&\mathbb{P}(\epsilon_{i1} \leq -\tau_1, \ldots, \epsilon_{iC} \leq -\tau_C, \eta_{1i} \leq \delta_1 - \iota_1, \ldots, \eta_{Ci} \leq \delta_C - \iota_C) \\
=&\mathbb{P}\left\{\cap_{c=1}^{C} (A_c \cap B_c)\right\} \\
=&\mathbb{P}\left\{\cap_{c=1}^{C} (A_c \cap B_c) \mid x\right\},
\end{aligned}
$$

where the last equality follows from the independence between (ϵ_i, η_i) and x_i (Assumption 3.4). Further, the conditional probability of being unmatched is

$$
\begin{aligned}
&\Lambda_0(\tau_1, \ldots, \tau_C, \iota_1, \ldots, \iota_C) \\
=&\mathbb{P}\left\{\cap_{c=1}^{C} (A_c \cup B_c) \mid x\right\} \\
=&\mathbb{P}\left\{\left[\cap_{c=1}^{C-1} (A_c \cup B_c)\right] \cap (A_C \cup B_C) \mid x\right\} \\
=&\mathbb{P}\left\{\left[\cap_{c=1}^{C-1} (A_c \cup B_c) \cap A_C\right] \cup \left[\cap_{c=1}^{C-1} (A_c \cup B_c) \cap B_C\right] \mid x\right\} \\
=&\mathbb{P}\left\{\cap_{c=1}^{C-1} (A_c \cup B_c) \cap A_C \mid x\right\} + \mathbb{P}\left\{\cap_{c=1}^{C-1} (A_c \cup B_c) \cap B_C \mid x\right\} \\
&- \mathbb{P}\left\{\cap_{c=1}^{C-1} (A_c \cup B_c) \cap A_C \cap B_C \mid x\right\} \\
=&\Lambda_0(\tau_1, \ldots, \tau_C, \iota_1, \ldots, \infty) + \Lambda_0(\tau_1, \ldots, \infty, \iota_1, \ldots, \iota_C) \\
&- \mathbb{P}\left\{\cap_{c=1}^{C-1} (A_c \cup B_c) \cap A_C \cap B_C \mid x\right\}. \qquad \text{(A.6)}
\end{aligned}
$$

Let $H_C(\tau_1, \ldots, \tau_C, \iota_1, \ldots, \iota_C) \equiv \mathbb{P}\left\{\cap_{c=1}^{C-1} (A_c \cup B_c) \cap A_C \cap B_C \mid x\right\}$. It is identified from equation (A.6) because Λ_0 is identified. Moreover, we have

$$
\begin{aligned}
&H_C(\tau_1, \ldots, \tau_C, \iota_1, \ldots, \iota_C) \\
=&\mathbb{P}\left\{\cap_{c=1}^{C-2} (A_c \cup B_c) \cap A_{C-1} \cap A_C \cap B_C \mid x\right\} + \mathbb{P}\left\{\cap_{c=1}^{C-2} (A_c \cup B_c) \cap B_{C-1} \cap A_C \cap B_C \mid x\right\} \\
&- \mathbb{P}\left\{\cap_{c=1}^{C-2} (A_c \cup B_c) \cap A_{C-1} \cap B_{C-1} \cap A_C \cap B_C \mid x\right\} \\
=&H_C(\tau_1, \ldots, \tau_C, \iota_1, \ldots \iota_{(C-2)}, \infty, \iota_C) + H_C(\tau_1, \ldots, \tau_{(C-2)}, \infty, \tau_C, \iota_1, \ldots, \iota_C) \\
&- \mathbb{P}\left\{\cap_{c=1}^{C-2} (A_c \cup B_c) \cap A_{C-1} \cap B_{C-1} \cap A_C \cap B_C \mid x\right\}, \qquad \text{(A.7)}
\end{aligned}
$$

which identifies $\mathbb{P}\left\{\cap_{c=1}^{C-2} (A_c \cup B_c) \cap A_{C-1} \cap B_{C-1} \cap A_C \cap B_C \mid x\right\}$. Note that in the last two lines in equation (A.7), the effects of ι_{C-1} and τ_{C-1} are "shut down" in the first and second terms, respectively. We further define

$$
H_{C-1}(\tau_1, \ldots, \tau_C, \iota_1, \ldots, \iota_C) \equiv \mathbb{P}\left\{\cap_{c=1}^{C-2} (A_c \cup B_c) \cap A_{C-1} \cap B_{C-1} \cap A_C \cap B_C \mid x\right\}.
$$

Repeat the above argument and define a sequence of functions recursively until $H_2(\tau_1,\ldots,\tau_C,\iota_1,\ldots,\iota_C) = H_2(\tau_1,\ldots,\tau_C,\infty,\ldots,\iota_C) + H_2(\infty,\ldots,\tau_C,\iota_1,\ldots,\iota_C)$ $-H_1(\tau_1,\ldots,\tau_C,\iota_1,\ldots,\iota_C)$, where on the RHS, the effects of ι_1 and τ_1 are "shut down" in the first and second terms, respectively. Every function in the sequence is identified.

It then follows that $F(-\tau_1,\ldots,-\tau_C,-\iota_1,\ldots,-\iota_C) = \mathbb{P}\left\{\cap_{c=1}^{C}(A_c \cap B_c) \mid x\right\} = H_1(\tau_1,\ldots,\tau_C,\iota_1,\ldots,\iota_C)$ is identified. □

Proof of Proposition 3.10. We use the same argument as in the proof of Proposition 3.7, except that the matrix in equation (A.4) reduces in dimension due to the additional homogeneity restrictions. Specifically, with some abuse of notations, suppose that Π^* has eliminated the duplicated elements due to the homogeneity restriction, that vector $\boldsymbol{u}$ consists of u^* and $u^c\ \forall c \in \mathbf{C}\setminus\mathbf{C}_1$, and that $\boldsymbol{v}$ consist of v^* and $v^c\ \forall c \in \mathbf{C}\setminus\mathbf{C}_2$. The vectors y_i and w_i are defined similarly. Fix $y_i = \overline{y}$ and consider (z,w) in the interior of $\mathcal{Z}\times\mathcal{W}$. We can rewrite equation (A.3) as

$$\frac{\partial\sigma_{d,i}}{\partial z_i^k} = \frac{\partial\sigma_{d,i}}{\partial y_{i*}}u^*_{z_i^k} + \sum_{c\in\mathbf{C}\setminus\mathbf{C}_1}\frac{\partial\sigma_{d,i}}{\partial y_{ic}}u^c_{z_i^k} + \frac{\partial\sigma_{d,i}}{\partial w_{i*}}v^*_{z_i^k} + \sum_{c\in\mathbf{C}\setminus\mathbf{C}_2}\frac{\partial\sigma_{d,i}}{\partial w_{ic}}v^c_{z_i^k}. \tag{A.8}$$

Stacking equation (A.8) for all $d\in\mathbf{C}$, we obtain

$$\begin{pmatrix}\frac{\partial\sigma_{1,i}}{\partial z_i^k}\\ \vdots \\ \frac{\partial\sigma_{C,i}}{\partial z_i^k}\end{pmatrix} = \Pi^*(z,\bar{y},w)\cdot\begin{pmatrix}\frac{\partial\boldsymbol{u}(z)}{\partial z_i^k}\\ \frac{\partial\mathbf{v}(z)}{\partial z_i^k}\end{pmatrix}. \tag{A.9}$$

When the number of parameters $2C-\kappa_1-\kappa_2+2$ is at most C, equation (A.9) implies identification; otherwise, we can identify the parameters by considering a pair of distinct values of w_i, as in the proof of Proposition 3.7. □

B Identification of a Nonseparable Model

We now discuss the nonparametric identification of the nonseparable model in equation (4) based on the arguments in Matzkin (2019). Recall that the utility functions are $u_{i1} = u^1(z_i, y_{i1}+\epsilon_{i1})$, $u_{ic} = u^c(z_i, y_{ic}, \epsilon_{ic})\ \forall c \in \mathbf{C}\setminus\{1\}$, and $v_{ci} = v^c(z_i, w_{ic}, \eta_{ci})\ \forall c \in$

C. For notational simplicity, we also use $u^1(z_i, y_{i1}, \epsilon_{i1})$ to denote $u^1(z_i, y_{i1} + \epsilon_{i1})$. The utility of the outside option u_{i0} is assumed to be a continuous random variable. [B.1]

Assumption B.1. *For each $c \in \mathbf{C}$, (i) $x_{ic}^k, \forall k = 1, \ldots, d_x$, are continuous random variables; (ii) the functions, u^c and v^c, are continuously differentiable; (iii) F is continuously differentiable; (iv) u^c and v^c are strictly increasing in their last argument; and (v) for $c \in \mathbf{C}\backslash\{1\}$, when $u^c(z_i, y_{ic}, \epsilon_{ic}) = u_{i0}$, $\frac{\partial u^c(z_i, y_{ic}, \epsilon_{ic})}{\partial y_{ic}} \neq 0$, and for $c \in \mathbf{C}$, when $v^c(z_i, w_{ic}, \eta_{ci}) = \delta_c$, $\frac{\partial v^c(z_i, w_{ic}, \eta_{ci})}{\partial w_{ic}} \neq 0$.*

Assumption B.2. (ϵ_i, η_i) *is independent of x_i.*

Assumption B.3. *(i) The utility of the outside option is $u_{i0} = h(y_{i0})$, where $y_{i0} \in \mathcal{Y}_0 \subseteq \mathbb{R}^{d_{y0}}$ is a vector of observed covariates and h is a known function; (ii) The support of u_{i0}, $\mathcal{U}_0 \subseteq \mathbb{R}$, is a superset of the range of the function u^c, $\forall c \in \mathbf{C}$.*

Parts (i)–(iii) of Assumption B.1 and Assumption B.2 impose smoothness and exogeneity similar to Assumptions 3.3 and 3.4 in Section 3. Part (iv) of Assumption B.1 guarantees that there is a one-to-one relationship between the value of each utility function and its unobservable. Part (v) of Assumption B.1 guarantees that u_{ic} and v_{ci} are not constant w.r.t. y_{ic} and w_{ic}, respectively, such that a change in y_{ic} or w_{ic} generates a change in the conditional probability of being unmatched. Assumption B.3 guarantees that u_{i0} is observed by the researcher and has a large support.

By the monotonicity assumption (part iv of Assumption B.1), for each $c \in \mathbf{C}$, the inverse of u^c and v^c w.r.t. their last argument exists. Let $\tilde{u}^c$ and $\tilde{v}^c$ denote the inverse of u^c and v^c w.r.t. their last argument, respectively. That is, for any $a \in \mathbb{R}$,

$$u^1\left(z_i, \tilde{u}^1(z_i, a)\right) = a, \text{ and } v^1\left(z_i, w_{i1}, \tilde{v}^1(z_i, w_{i1}, a)\right) = a,$$
$$u^c(z_i, y_{ic}, \tilde{u}^c(z_i, y_{ic}, a)) = a, \text{ and } v^c(z_i, w_{ic}, \tilde{v}^c(z_i, w_{ic}, a)) = a, \text{ for any } c \in \mathbf{C}\backslash\{1\}$$

Then, we have

$$\begin{aligned}\lambda_{L,i} &= \mathbb{P}(v_{ci} \geq \delta_c \; \forall c \in L;\, v_{di} < \delta_d \; \forall d \notin L \mid x_i; \mu) \\ &= \mathbb{P}(v^c(z_i, w_{ic}, \eta_{ci}) \geq \delta_c \; \forall c \in L;\, v^d(z_i, w_{id}, \eta_{di}) < \delta_d \; \forall d \notin L \mid w_i, z_i; \mu)\end{aligned}$$

[B.1] In separable models, $u_{i0} = 0$ is a location normalization because the conditional match probability only depends on the difference in the utility shocks. However, in this nonseparable model, it would impose an additional restriction.

$$= \mathbb{P}(\eta_{ci} \geq \tilde{v}^c(z_i, w_{ic}, \delta_c) \ \forall c \in L; \ \eta_{di} < \tilde{v}^d(z_i, w_{id}, \delta_d) \ \forall d \notin L \mid w_i, z_i; \mu)$$
$$= \lambda_L(\iota_{i1}, \ldots, \iota_{iC}),$$

where $\iota_{ic} = \tilde{v}^c(z_i, w_{ic}, \delta_c)$ for $c \in \mathbf{C}$. Since $\tilde{v}^c$ is a c-specific nonparametric function, the following analysis does not rely on the identification of δ_c. Similarly,

$$\begin{aligned}
\mathbb{P}(0 = \arg\max_{d \in L} u_{id} \mid L, x_i, u_{i0}) &= \mathbb{P}(u_{i0} > u_{id} \text{ for all } d \in L \mid L, x_i, u_{i0}) \\
&= \mathbb{P}(u_{i0} > u^d(z_i, y_{id}, \epsilon_{id}) \text{ for all } d \in L \mid L, x_i, u_{i0}) \\
&= \mathbb{P}(\epsilon_{id} < \tilde{u}^d(z_i, y_{id}, u_{i0}) \text{ for all } d \in L \mid L, x_i, u_{i0}) \\
&= g_{0,L}(\tau_{i1}, \ldots, \tau_{iC}),
\end{aligned}$$

where $\tau_{i1} = \tilde{u}^1(z_i, u_{i0}) - y_{i1}$ and for $c \in \mathbf{C} \backslash \{1\}$, $\tau_{ic} = \tilde{u}^c(z_i, y_{ic}, u_{i0})$. Note that if $c \notin L$, $g_{0,L}$ does not change with the argument τ_{ic}.

Further, following equation (6) for $c = 0$, we have

$$\begin{aligned}
\sigma_0(x_i, u_{i0}) &= \sum_{L \in \mathcal{L}} \lambda_L(\iota_{i1}, \ldots, \iota_{iC}) \cdot g_{0,L}(\tau_{i1}, \ldots, \tau_{iC}) \\
&\equiv \Lambda_0(\tau_{i1}, \ldots, \tau_{iC}, \iota_{i1}, \ldots, \iota_{iC}),
\end{aligned} \tag{B.10}$$

where Λ_0 is a nonparametric function.

To identify the derivatives of $\{u^c, v^c\}_c$, we extend the argument in Matzkin (2019). Our identification depends on conditions on the derivatives of the probability of being *unmatched* w.r.t. the excluded variables. Let $y_{i,-1} = (y_{i2}, \ldots, y_{iC}) \in \mathcal{Y}_{-1} \subseteq \mathbb{R}^{2C-1}$ denote the vector of y_i excluding y_{i1}. For a given point (z, w, y_{-1}, u_0) in the interior of $\mathcal{Z} \times \mathcal{W} \times \mathcal{Y}_{-1} \times \mathcal{U}_0$, consider $2C$ different values, $y_1^1, \ldots, y_1^{2C}$, in the interior of the support of y_{i1} conditional on (z, w, y_{-1}, u_0). We define a $C \times C$ matrix

$$\Pi_1\left(y_1^1, \ldots, y_1^C; z, w, y_{-1}, u_0\right) \equiv \begin{pmatrix} \frac{\partial \sigma_0\left(z, w, y_{-1}, u_0, y_1^1\right)}{\partial y_i'} \\ \vdots \\ \frac{\partial \sigma_0\left(z, w, y_{-1}, u_0, y_1^C\right)}{\partial y_i'} \end{pmatrix},$$

where for $m = 1, \ldots, C$, the mth row of the matrix Π_1 consists of the derivatives of conditional probability of being unmatched w.r.t. the C excluded variables y_i, evaluated at $(z, w, y_{-1}, u_0, y_1^m)$. Further, we define a $2C \times 2C$ matrix

$$\Pi_2\left(y_1^1, \ldots, y_1^{2C}; z, w, y_{-1}, u_0\right) \equiv \begin{pmatrix} \frac{\partial \sigma_0\left(z,w,y_{-1},u_0,y_1^1\right)}{\partial y_i'} & \frac{\partial \sigma_0\left(z,w,y_{-1},u_0,y_1^1\right)}{\partial w_i'} \\ \vdots & \vdots \\ \frac{\partial \sigma_0\left(z,w,y_{-1},u_0,y_1^{2C}\right)}{\partial y_i'} & \frac{\partial \sigma_0\left(z,w,y_{-1},u_0,y_1^{2C}\right)}{\partial w_i'} \end{pmatrix},$$

where for $m = 1, \ldots, 2C$, the mth row of the matrix Π_2 consists of the derivatives of conditional probability of being unmatched w.r.t. the $2C$ excluded variables (y_i, w_i), evaluated at $(z, w, y_{-1}, u_0, y_1^m)$.

Condition B.4. *For a given point (z, w, y_{-1}, u_0) in the interior of $\mathcal{Z} \times \mathcal{W} \times \mathcal{Y}_{-1} \times \mathcal{U}_0$, there exists C different values, $y_1^1, \ldots, y_1^C$, in the interior of the support of y_{i1} conditional on (z, w, y_{-1}, u_0) such that $\Pi_1\left(y_1^1, \ldots, y_1^C; z, w, y_{-1}, u_0\right)$ has rank C.*

Condition B.5. *For a given point (z, w, y_{-1}, u_0) in the interior of $\mathcal{Z} \times \mathcal{W} \times \mathcal{Y}_{-1} \times \mathcal{U}_0$, there exists $2C$ different values, $y_1^1, \ldots, y_1^{2C}$, in the interior of the support of y_{i1} conditional on (z, w, y_{-1}, u_0) such that $\Pi_2\left(y_1^1, \ldots, y_1^{2C}; z, w, y_{-1}, u_0\right)$ has rank $2C$.*

Note that we can choose C different values of y_{i1} to satisfy Condition B.4 and then independently choose another $2C$ values of y_{i1} to satisfy Condition B.5.

Let ϵ_c^ρ denote the ρ-quantile of ϵ_{ic}, i.e., $\epsilon_c^\rho = \text{Quantile}_{\epsilon_{ic}}(\rho) = \inf\{\epsilon_c : F_{\epsilon_{ic}}(\epsilon_c) \geq \rho\}$ for $\rho \in (0, 1)$, where $F_{\epsilon_{ic}}$ denote the marginal CDF of ϵ_{ic}.

Proposition B.6. *Suppose that Assumptions B.1-B.3 and Conditions B.4 and B.5 are satisfied. We have (i) for each $c \in \mathbf{C}\backslash\{1\}$, for any point (z, y_c) in the interior of $\mathcal{Z} \times \mathcal{Y}_c$, for any $\rho \in (0, 1)$, and for any coordinate $k = 1, \ldots, d_z$, $\frac{\partial u^c\left(z,y_c,\epsilon_c^\rho\right)}{\partial z_i^k}$ and $\frac{\partial u^c\left(z,y_c,\epsilon_c^\rho\right)}{\partial y_{ic}}$ are identified; for $c = 1$, $\frac{\partial u^1\left(z,y_1+\epsilon_1^\rho\right)}{\partial z_i^k}$ and $\frac{\partial u^1\left(z,y_1+\epsilon_1^\rho\right)}{\partial y_{i1}} = \frac{\partial u^1\left(z,y_1+\epsilon_1^\rho\right)}{\partial \epsilon_{i1}}$ are identified; and (ii) for each $c \in \mathbf{C}$, for any point (z, w_c) in the interior of $\mathcal{Z} \times \mathcal{W}_c$, for any coordinate $k = 1, \ldots, d_z$, $\frac{\partial v^c(z,w_c,\eta_{ci})}{\partial z_i^k} / \frac{\partial v^c(z,w_c,\eta_{ci})}{\partial w_{ic}}$ is identified, where η_{ci} is such that $v^c(z, w_c, \eta_{ci}) = \delta_c$.*

We group the proofs at the end of this section. Using the variation in u_{i0}, we identify the derivatives of student utility functions at all quantiles of the unobservable ϵ_i. For the college utility functions, without additional assumptions, we only identify the ratio of the derivatives at certain values of the unobservable (i.e., η_c such that

$v^c(z, w_c, \eta_c) = \delta_c)$. This is because, on the college side, the probability of being unmatched is determined by comparing v_{ci} with δ_c, while δ_c is unobserved and fixed. This lack of variation restricts the identification of the derivatives of v_{ci}.

With a more restrictive functional form of v_{ci}, the following corollary identifies these derivatives. For that, we let η_c^ρ be the ρ-quantile of η_{ci}, i.e., $\eta_c^\rho = \text{Quantile}_{\eta_{ci}}(\rho) = \inf\{\eta_c : F_{\eta_{ci}}(\eta_c) \geq \rho\}$ for $\rho \in (0,1)$, where $F_{\eta_{ci}}$ is the marginal CDF of η_{ci}.

Corollary B.7. *Suppose that $v_{ci} = v^c(z_i, \eta_{ci}) + w_{ic}$, that w_{ic} has a large support, and that Assumptions B.1, B.2, and B.3(i) and Conditions B.4 and B.5 are satisfied. For any point z in the interior of $\mathcal{Z}$, for all $c \in \mathbf{C}$, any $\rho \in (0,1)$, and $k = 1, \ldots, d_z$, $\frac{\partial v^c(z, \eta_c^\rho)}{\partial z_i^k}$ is identified.*

For this corollary, we do not need Assumption B.3(ii), which is required only for identifying the derivatives of u^c for all possible values of ϵ_{ic}.

Proof of Proposition B.6. To simplify notations, for $k = 1, \ldots, d_z$, let $u^c_{z_i^k} = \frac{\partial u^c}{\partial z_i^k}$ and similar notations are defined for v^c, $\tilde{u}^c$, $\tilde{v}^c$, and the other variables, and let $\sigma_0^m = \sigma_0(z, w, y_{-1}, y_0, y_1^m)$ for $m = 1, \ldots, 2C$. Let t^m be the value of $(\tau_{i1}, \ldots, \tau_{iC}, \iota_{i1}, \ldots, \iota_{iC})$ evaluated at $(z, w, y_{-1}, y_0, y_1^m)$. Under Assumption B.1(i)-(iii), in equation (B.10), Λ_0, u^c, and v^c are continuously differentiable and the observables are all continuously distributed. Taking derivatives of equation (B.10) on both sides w.r.t. y_{ic} and w_{ic}, and evaluating them at $(z, w, y_{-1}, y_0, y_1^m)$, we have, for $c = 1$,

$$\frac{\partial \sigma_0^m}{\partial y_{i1}} = -\frac{\partial \Lambda(t^m)}{\partial \tau_{i1}} \text{ and } \frac{\partial \sigma_0^m}{\partial w_{i1}} = \frac{\partial \Lambda(t^m)}{\partial \iota_{i1}} \tilde{v}^1_{w_{i1}}, \tag{B.11}$$

and, for $c \neq 1$,

$$\frac{\partial \sigma_0^m}{\partial y_{ic}} = \frac{\partial \Lambda(t^m)}{\partial \tau_{ic}} \tilde{u}^c_{y_{ic}} \text{ and } \frac{\partial \sigma_0^m}{\partial w_{ic}} = \frac{\partial \Lambda(t^m)}{\partial \iota_{ic}} \tilde{v}^c_{w_{ic}}. \tag{B.12}$$

Further, taking derivatives of equation (B.10) on both sides w.r.t. u_{i0} and z_i^k, and evaluating them at $(z, w, y_{-1}, y_0, y_1^m)$, we have

$$\frac{\partial \sigma_0^m}{\partial u_{i0}} = \sum_{c=1}^{C} \frac{\partial \Lambda(t^m)}{\partial \tau_{ic}} \tilde{u}^c_{u_{i0}}, \tag{B.13}$$

$$\frac{\partial \sigma_0^m}{\partial z_i^k} = \sum_{c=1}^{C} \frac{\partial \Lambda(t^m)}{\partial \tau_{ic}} \tilde{u}^c_{z_i^k} + \sum_{c=1}^{C} \frac{\partial \Lambda(t^m)}{\partial \iota_{ic}} \tilde{v}^c_{z_i^k}. \tag{B.14}$$

Substituting equations (B.11) and (B.12) into equations (B.13) and (B.14), we have

$$\frac{\partial \sigma_0^m}{\partial u_{i0}} = -\frac{\partial \sigma_0^m}{\partial y_{i1}} \tilde{u}^1_{u_{i0}} + \sum_{c=2}^{C} \frac{\partial \sigma_0^m}{\partial y_{ic}} (\tilde{u}^c_{y_{ic}})^{-1} \tilde{u}^c_{u_{i0}}, \tag{B.15}$$

$$\frac{\partial \sigma_0^m}{\partial z_i^k} = -\frac{\partial \sigma_0^m}{\partial y_{i1}} \tilde{u}^1_{z_i^k} + \sum_{c=2}^{C} \frac{\partial \sigma_0^m}{\partial y_{ic}} (\tilde{u}^c_{y_{ic}})^{-1} \tilde{u}^c_{z_i^k} + \sum_{c=1}^{C} \frac{\partial \sigma_0^m}{\partial w_{ic}} (\tilde{v}^c_{w_{ic}})^{-1} \tilde{v}^c_{z_i^k}. \tag{B.16}$$

To get the relationship between the derivatives of u^c and $\tilde{u}^c$, for $\forall c \in \mathbf{C}\backslash\{1\}$, taking derivatives on both side of the equations, $u^c(z_i, y_{ic}, \tilde{u}^c(z_i, y_{ic}, u_{i0})) = u_{i0}$, w.r.t. y_{ic}, u_{i0}, and z_i^k, one gets, $u^c_{y_{ic}} + u^c_{\epsilon_{ic}} \tilde{u}^c_{y_{ic}} = 0$, $u^c_{\epsilon_{ic}} \tilde{u}^c_{u_{i0}} = 1$, and $u^c_{z_i^k} + u^c_{\epsilon_{ic}} \tilde{u}^c_{z_i^k} = 0$; it then follows that $\tilde{u}^c_{y_{ic}} = -\frac{u^c_{y_{ic}}}{u^c_{\epsilon_{ic}}}$, $\tilde{u}^c_{u_{i0}} = \frac{1}{u^c_{\epsilon_{ic}}}$, and that $\tilde{u}^c_{z_i^k} = -\frac{u^c_{z_i^k}}{u^c_{\epsilon_{ic}}}$ at the value of ϵ_{ic}, ϵ_c, such that $u^c(z, y_c, \epsilon_c) = u_0$. Similarly, for $c = 1$, $\tilde{u}^1_{u_{i0}} = \frac{1}{u^1_{\epsilon_{i1}+y_{i1}}}$ and $\tilde{u}^1_{z_i^k} = -\frac{u^1_{z_i^k}}{u^1_{\epsilon_{i1}+y_{i1}}}$ at the value of $\epsilon_{i1} + y_{i1}$ such that $u^1(z, \epsilon_1 + y_1) = u_0$. Importantly, y_1 does not need to satisfy Conditions B.4 and B.5 because for any y_1, one can find an ϵ_1 so that the above equation holds.

Similarly, taking derivatives of the equation, $v^c(z_i, w_{ic}, \tilde{v}^c(z_i, w_{ic}, \delta_c)) = \delta_c$, w.r.t. w_{ic} and z_i^k and making rearrangements, we obtain, for $c \in \mathbf{C}$, $\tilde{v}^c_{w_{ic}} = -\frac{v^c_{w_{ic}}}{v^c_{\eta_{ci}}}$ and $\tilde{v}^c_{z_i^k} = -\frac{v^c_{z_i^k}}{v^c_{\eta_{ci}}}$ at the value of η_{ci} such that $v^c(z, w_c, \eta_c) = \delta_c$.

Plugging the above relationships among the utility functions and their inverse into equations (B.15) and (B.16), we obtain

$$\frac{\partial \sigma_0^m}{\partial u_{i0}} = -\frac{\partial \sigma_0^m}{\partial y_{i1}} \frac{1}{u^1_{\epsilon_{i1}+y_{i1}}} - \sum_{c=2}^{C} \frac{\partial \sigma_0^m}{\partial y_{ic}} \frac{1}{u^c_{y_{ic}}}, \tag{B.17}$$

$$\frac{\partial \sigma_0^m}{\partial z_i^k} = \frac{\partial \sigma_0^m}{\partial y_{i1}} \frac{u^1_{z_i^k}}{u^1_{\epsilon_{i1}+y_{i1}}} + \sum_{c=2}^{C} \frac{\partial \sigma_0^m}{\partial y_{ic}} \frac{u^c_{z_i^k}}{u^c_{y_{ic}}} + \sum_{c=1}^{C} \frac{\partial \sigma_0^m}{\partial w_{ic}} \frac{v^c_{z_i^k}}{v^c_{w_{ic}}}. \tag{B.18}$$

Next, stacking equation (B.17) for $m = 1, \ldots, C$, we have

$$\begin{pmatrix} \frac{\partial \sigma_0^1}{\partial u_{i0}} \\ \vdots \\ \frac{\partial \sigma_0^C}{\partial u_{i0}} \end{pmatrix} = -\begin{pmatrix} \frac{\partial \sigma_0\left(z, w, y_{-1}, y_0, y_1^1\right)}{\partial y_i'} \\ \vdots \\ \frac{\partial \sigma_0\left(z, w, y_{-1}, y_0, y_1^C\right)}{\partial y_i'} \end{pmatrix} \begin{pmatrix} \frac{1}{u^1_{\epsilon_{i1}+y_{i1}}} \\ \frac{1}{u^2_{y_{i2}}} \\ \vdots \\ \frac{1}{u^C_{y_{iC}}} \end{pmatrix},$$

where the vector $(\frac{1}{u^1_{\epsilon_{i1}+y_{i1}}}, \frac{1}{u^2_{y_{i2}}}, \ldots, \frac{1}{u^C_{y_{iC}}})'$ is finite due to part (v) of Assumption B.1. Note that the derivatives of σ_0 in the above system can be observed from the population data. Then, by Condition B.4, $\frac{1}{u^1_{\epsilon_{i1}+y_{i1}}}$ and $\frac{1}{u^c_{y_{ic}}}$ for each $c \in \mathbf{C}\backslash\{1\}$ are identified.

Similarly, stacking equation (B.18) for $m = 1, \ldots, 2C$, we obtain

$$\begin{pmatrix} \frac{\partial \sigma_0^1}{\partial z_i^k} \\ \vdots \\ \frac{\partial \sigma_0^{2C}}{\partial z_i^k} \end{pmatrix} = \begin{pmatrix} \frac{\partial \sigma_0\left(z, w, y_{-1}, y_0, y_1^1\right)}{\partial y_i'} & \frac{\partial \sigma_0\left(z, w, y_{-1}, y_0, y_1^1\right)}{\partial w_i'} \\ \vdots & \vdots \\ \frac{\partial \sigma_0\left(z, w, y_{-1}, y_0, y_1^{2C}\right)}{\partial y_i'} & \frac{\partial \sigma_0\left(z, w, y_{-1}, y_0, y_1^{2C}\right)}{\partial w_i'} \end{pmatrix} \begin{pmatrix} u^1_{z_i^k}/u^1_{\epsilon_{i1}+y_{i1}} \\ u^2_{z_i^k}/u^2_{y_{i2}} \\ \vdots \\ u^C_{z_i^k}/u^C_{y_{iC}} \\ v^1_{z_i^k}/v^1_{w_{i1}} \\ \vdots \\ v^C_{z_i^k}/v^C_{w_{iC}} \end{pmatrix}.$$

Then, by Condition B.5, for all $c \in \mathbf{C}$, $\frac{v^c_{z_i^k}}{v^c_{w_{ic}}}$ is identified at the value of η_{ci} such that $v^c(z, w_c, \eta_c) = \delta_c$. Also, $\frac{u^1_{z_i^k}}{u^1_{\epsilon_{i1}+y_{i1}}}$, and for each $c \in \mathbf{C}\backslash\{1\}$, $\frac{u^c_{z_i^k}}{u^c_{y_{ic}}}$ are identified. Combining this with the first identification result, we identify $u^c_{z_i^k}$ for all c, at the value of ϵ_{ic} such that $u^c(z, y_c, \epsilon_c) = u_0$ for $c \in \mathbf{C}\backslash\{1\}$, and at the value of $\epsilon_{i1} + y_{i1}$ such that $u^1(z, \epsilon_1 + y_1) = u_0$ for $c = 1$.

Further, for each c and for any $\rho \in (0, 1)$, define the conditional ρ-quantile of u_{i0} given (z_i, y_{ic}) as $\text{Quantile}_{u_{i0}|(z_i,y_{ic})}(\rho) = \inf\{u_0 : F_{u_{i0}|(z_i,y_{ic})}(u_0) \geq \rho\}$. Because of part (iv) of Assumption B.1, for any (z, y_c), for ϵ_{ic} such that $u^c(z, y_c, \epsilon_{ic}) = u_{i0}$, the equivariance property of quantiles (e.g., Chesher, 2003) implies that

$$\text{Quantile}_{u_{i0}|(z,y_c)}(\rho) = u^c(z, y_c, \epsilon_c^\rho),$$

where the LHS is known from the joint distribution of (u_{i0}, z_i, y_{ic}). Therefore, the above identification result indicates that for all c, we can identify $u^c_{z_i^k}$ for any given (z, y) and ϵ_c^ρ. □

Proof of Corollary B.7. Proposition B.6 implies that $\frac{\partial v^c(z_i,\eta_{ci})}{\partial z_i^k}$ is identified, where η_{ci} is such that $v^c(z_i, \eta_{ci}) + w_{ic} = \delta_c$. For any z and $\rho \in (0, 1)$, the equivariance property of quantiles (e.g., Chesher, 2003) implies that $\text{Quantile}_{-w_{ic}|z}(\rho) = v^c(z, \eta_c^\rho)$, where the

LHS is known from the joint distribution of (w_{ic}, z_i). Hence, $\frac{\partial v^c(z, \eta_c^\rho)}{\partial z_i^k}$ is identified. $\square$

C A Control Function Approach

This appendix discusses a control function approach that relaxes Assumption 3.4 in the identification of the derivatives of $\{u^c, r^c, v^c\}_c$.

For simplicity, we consider the case where there is one endogenous variable. That is, $z_i = (z_{1i}, z'_{2i})'$, where z_{1i} is a scalar endogenous random variable and z_{2i} is a vector of exogenous random variables. Suppose that z_{1i} can be written as a nonparametric function of exogenous variables z_{2i}, a vector of exogenous variables t_i that is not contained in z_{2i}, and a scalar unobserved random variable ξ_i:

$$z_{1i} = h(t_i, z_{2i}, \xi_i). \tag{C.19}$$

Assume that the unobservables ξ_i and (ϵ_i, η_i) are independent of all the exogenous variables (t_i, z_{2i}, y_i, w_i) but are not independent of each other. The endogeneity of z_{1i} arises due to the correlation between ξ_i and (ϵ_i, η_i).

The following approach exploits a control variable e_i such that conditional on e_i, z_{1i} and (ϵ_i, η_i) are independent. In a nonadditive setting described in equation (C.19), suppose that the CDF of ξ_i is strictly increasing and continuous, and that h is strictly monotone in its last argument. Then the control variable $e_i = F_{z_{1i}|(t_i, z_{2i})}(z_i, t_i) = F_{\xi_i}(\xi_i)$, where $F_{z_{1i}|(t_i, z_{2i})}(z_i, t_i)$ is the conditional CDF of z_{1i} given (t_i, z_{2i}) and $F_{\xi_i}(\xi_i)$ is the CDF of ξ_i (Imbens and Newey, 2009). In an additive setting where $z_{1i} = h(t_i, z_{2i}) + \xi_i$ and $\mathbb{E}(\xi_i | t_i, z_{2i}) = 0$, the control variable $e_i = \xi_i$.[C.2]

Suppose that each element in (ϵ_i, η_i) can be decomposed into a function of e_i and a residual that is independent of e_i. Specifically, for each $c \in \boldsymbol{C}$, we obtain

$$\epsilon_{ic} = \varphi^c(e_i) + \tilde{\epsilon}_{ic} \text{ and } \eta_{ci} = \phi^c(e_i) + \tilde{\eta}_{ci}. \tag{C.20}$$

Note that $\tilde{\epsilon}_{ic}$ and $\tilde{\eta}_{ci}$ are independent of (t_i, z_{2i}, y_i, w_i) because ξ_i (and thus e_i) and (ϵ_i, η_i) are both independent of (t_i, z_{2i}, y_i, w_i). Besides, $\tilde{\epsilon}_{ic}$ and $\tilde{\eta}_{ci}$ are independent of

[C.2] For examples of parametric specifications in consumer choice models and in matching models, see Petrin and Train (2010) and Agarwal (2015).

z_{1i} because z_{1i} is a function of (t_i, z_{2i}) and ξ_i.

Plugging equation (C.20) into the utility functions in equation (3), we have

$$u_{ic} = u^c(z_i) + r^c(y_{ic}) + \varphi^c(e_i) + \tilde{\epsilon}_{ic} \text{ and } v_{ci} = v^c(z_i) + w_{ic} + \phi^c(e_i) + \tilde{\eta}_{ci}, \forall c \in \mathbf{C}.$$

We can treat e_i as observed because it can be identified from the joint distribution of (z_i, t_i). A similar argument as that in Proposition 3.7 then can be used to identify the derivatives of the functions $\{u^c, v^c, r^c, \varphi^c, \phi^c\}_c$.

D Monte Carlo Simulations

In a series of Monte Carlo simulations, this appendix shows (i) that a semiparametric approach based on the results in Section 3 suffers from the curse of dimensionality, and (ii) that a parametric model based on a Bayesian approach works well.

D.1 Setup

There are 3000 students competing for admissions to 3 colleges. The capacities of the colleges are $\{750, 700, 750\}$. Every student has access to an outside option of value ϵ_{i0} (i.i.d. $N(0, 1)$). Student i's utility when being admitted to college c is given by,

$$u_{ic} \quad = \beta_c^d \times d_{ic} + \beta_c^s \times s_i + \beta_c^z \times z_i + \epsilon_{ic}, \tag{D.21}$$

where d_{ic} is student-college-specific and follows i.i.d. (across colleges and across students) $N(0, 36)$, s_i is one of the characteristics of student i (i.i.d. $N(5, 36)$), z_i is another characteristic of i (i.i.d. $N(0, 36)$), and ϵ_{ic} is i.i.d. standard normal. $\beta_c^d = -1$ for all c is a normalization, while $\beta_c^s = \beta_c^z = 1$ for $c = 1, 2, 3$.

College c values each student as follows:

$$v_{ci} \quad = \gamma_c^w \times w_{ic} + \gamma_c^m \times m_i + \gamma_c^z \times z_i + \eta_{ci}, \tag{D.22}$$

where w_{ic} is a student-college-specific characteristic (i.i.d. $N(0, 36)$), m_i is another characteristic of student i (i.i.d. $N(0, 36)$), and η_{ic} is i.i.d. standard normal. z_i appears in both student and college preferences. $\gamma_c^w = 1$ for all c is a normalization, while

$\gamma_c^m = \gamma_c^z = 1$ for $c = 1, 2, 3$. For simplicity, we assume that $T_c = -\infty$ or, equivalently, every college finds every student acceptable.

There are in total 150 MC samples (markets). The capacity constraint is always binding. Note that we obtain a set of estimates from each sample/market.

D.2 Estimation: Average Derivatives

To operationalize our nonparametric results, we impose three additional assumptions. First, the true functional form is known except for the distribution of (ϵ_i, η_i), which gives us a semiparametric setting. Second, in student preferences, the parameters to be estimated are $\beta_c^s = 1$ for $c = 1, 2, 3$ and β^z such that $\beta_c^z = \beta^z = 1$ (i.e., we have prior knowledge that β_c^z is constant across colleges). Third, in college preferences, the parameters to be estimated are $\gamma_c^m = 1$ for $c = 1, 2, 3$ and γ^z such that $\gamma_c^z = \gamma^z = 1$ (i.e., we have prior knowledge that γ_c^z is constant across colleges).

Let $x_i = (d_i, w_i, s_i, z_i, m_i)$, with $d_i = (d_{1i}, d_{2i}, d_{i3})$ and $w_i = (w_{1i}, w_{2i}, w_{i3})$. We rewrite equation (12) in the semiparametric setting for s_i and m_i, respectively, integrate over the entire support of x_i to obtain unconditional expectations $\mathbb{E}$:

$$\begin{pmatrix} \mathbb{E}\left(\frac{\partial \sigma_1(x_i)}{\partial s_i}\right) \\ \mathbb{E}\left(\frac{\partial \sigma_2(x_i)}{\partial s_i}\right) \\ \mathbb{E}\left(\frac{\partial \sigma_3(x_i)}{\partial s_i}\right) \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left(-\frac{\partial \sigma_1(x_i)}{\partial d_{i1}}\right) & \mathbb{E}\left(-\frac{\partial \sigma_1(x_i)}{\partial d_{i2}}\right) & \mathbb{E}\left(-\frac{\partial \sigma_1(x_i)}{\partial d_{i3}}\right) \\ \mathbb{E}\left(-\frac{\partial \sigma_2(x_i)}{\partial d_{i1}}\right) & \mathbb{E}\left(-\frac{\partial \sigma_2(x_i)}{\partial d_{i2}}\right) & \mathbb{E}\left(-\frac{\partial \sigma_2(x_i)}{\partial d_{i3}}\right) \\ \mathbb{E}\left(-\frac{\partial \sigma_3(x_i)}{\partial d_{i1}}\right) & \mathbb{E}\left(-\frac{\partial \sigma_3(x_i)}{\partial d_{i2}}\right) & \mathbb{E}\left(-\frac{\partial \sigma_3(x_i)}{\partial d_{i3}}\right) \end{pmatrix} \cdot \begin{pmatrix} \beta_1^s \\ \beta_2^s \\ \beta_3^s \end{pmatrix}. \tag{D.23}$$

$$\begin{pmatrix} \mathbb{E}\left(\frac{\partial \sigma_1(x_i)}{\partial m_i}\right) \\ \mathbb{E}\left(\frac{\partial \sigma_2(x_i)}{\partial m_i}\right) \\ \mathbb{E}\left(\frac{\partial \sigma_3(x_i)}{\partial m_i}\right) \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left(\frac{\partial \sigma_1(x_i)}{\partial w_{i1}}\right) & \mathbb{E}\left(\frac{\partial \sigma_1(x_i)}{\partial w_{i2}}\right) & \mathbb{E}\left(\frac{\partial \sigma_1(x_i)}{\partial w_{i3}}\right) \\ \mathbb{E}\left(\frac{\partial \sigma_2(x_i)}{\partial w_{i1}}\right) & \mathbb{E}\left(\frac{\partial \sigma_2(x_i)}{\partial w_{i2}}\right) & \mathbb{E}\left(\frac{\partial \sigma_2(x_i)}{\partial w_{i3}}\right) \\ \mathbb{E}\left(\frac{\partial \sigma_3(x_i)}{\partial w_{i1}}\right) & \mathbb{E}\left(\frac{\partial \sigma_3(x_i)}{\partial w_{i2}}\right) & \mathbb{E}\left(\frac{\partial \sigma_3(x_i)}{\partial w_{i3}}\right) \end{pmatrix} \cdot \begin{pmatrix} \gamma_1^m \\ \gamma_2^m \\ \gamma_3^m \end{pmatrix}. \tag{D.24}$$

The derivatives with respect to z_i lead to:

$$\begin{pmatrix} \mathbb{E}\left(\frac{\partial \sigma_1(x_i)}{\partial z_i}\right) \\ \mathbb{E}\left(\frac{\partial \sigma_2(x_i)}{\partial z_i}\right) \\ \mathbb{E}\left(\frac{\partial \sigma_3(x_i)}{\partial z_i}\right) \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left(\sum_{c=1}^3 \frac{\partial \sigma_1(x_i)}{\partial w_{ic}}\right) & -\mathbb{E}\left(\sum_{c=1}^3 \frac{\partial \sigma_1(x_i)}{\partial d_{ic}}\right) \\ \mathbb{E}\left(\sum_{c=1}^3 \frac{\partial \sigma_2(x_i)}{\partial w_{ic}}\right) & -\mathbb{E}\left(\sum_{c=1}^3 \frac{\partial \sigma_2(x_i)}{\partial d_{ic}}\right) \\ \mathbb{E}\left(\sum_{c=1}^3 \frac{\partial \sigma_3(x_i)}{\partial w_{ic}}\right) & -\mathbb{E}\left(\sum_{c=1}^3 \frac{\partial \sigma_3(x_i)}{\partial d_{ic}}\right) \end{pmatrix} \cdot \begin{pmatrix} \gamma^z \\ \beta^z \end{pmatrix}. \tag{D.25}$$

We now have 3 equations in 2 unknowns specified by equation (D.25). Using any two of the equations leads to an estimator. Moreover, we can formulate an estimator based on the generalized method of moments (GMM) that uses all three equations.

In sum, our estimation of β's and γ's relies on equation systems (D.23)–(D.25).

Results. The estimation results from the 150 MC samples are in the left part of Table D.1 (columns 1–3). We observe that the estimated coefficients are not close to their true values. The performance does not improve significantly when we double the sample size. Our explanation for this poor performance in the estimation is the curse of dimensionality. When calculating partial derivatives in equation systems (D.23) and (D.24), we deal with 4-dimensional objects (i.e., $(s_i, d_{1i}, d_{2i}, d_{i3})$ or $(m_i, w_{i1}, w_{i2}, w_{i3})$); in equation system (D.25), it is 7-dimensional (i.e., $(z_i, d_{1i}, d_{2i}, d_{i3}, w_{i1}, w_{i2}, w_{i3})$), which may explain that the estimators for β^z and γ^z perform the worst. This explanation is confirmed when we reduce the dimensionality in the model.

Reduced dimensionality. In student preferences (equation D.21), we further impose that the parameters to be estimated are $\beta_c^s = 1$ for $c = 1, 2, 3$ and $\beta_1^z = 1$, while we assume, and know, that $\beta_2^z = \beta_3^z = 0$ (i.e., z_i does not enter i's utility for college 2 or 3). In college preferences (equation D.22), the parameters to be estimated are $\gamma_c^m = 1$ for $c = 1, 2, 3$ and $\gamma_3^z = 1$, while we assume, and know, that $\gamma_1^z = \gamma_2^z = 0$ (i.e., colleges 1 and 2 do not use z_i to evaluate students). Based on these new parameter values, *we re-generate another 150 MC samples* for estimation.

We now have a simplified version of equation (D.25) with a reduced dimension:

$$\begin{pmatrix} \mathbb{E}\left(\dfrac{\partial \sigma_1(x_i)}{\partial z_i}\right) \\ \mathbb{E}\left(\dfrac{\partial \sigma_2(x_i)}{\partial z_i}\right) \\ \mathbb{E}\left(\dfrac{\partial \sigma_3(x_i)}{\partial z_i}\right) \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left(\dfrac{\partial \sigma_1(x_i)}{\partial w_{i3}}\right) & -\mathbb{E}\left(\dfrac{\partial \sigma_1(x_i)}{\partial d_{i1}}\right) \\ \mathbb{E}\left(\dfrac{\partial \sigma_2(x_i)}{\partial w_{i3}}\right) & -\mathbb{E}\left(\dfrac{\partial \sigma_2(x_i)}{\partial d_{i1}}\right) \\ \mathbb{E}\left(\dfrac{\partial \sigma_3(x_i)}{\partial w_{i3}}\right) & -\mathbb{E}\left(\dfrac{\partial \sigma_3(x_i)}{\partial d_{i1}}\right) \end{pmatrix} \cdot \begin{pmatrix} \gamma_3^z \\ \beta_1^z \end{pmatrix}. \tag{D.26}$$

The estimation results are presented in the right half of Table D.1 (columns 4–6). We observe that all estimates are centered around their corresponding true value.

Table D.1: Semiparametric Estimation: The General and Reduced Models

	General: higher dimensionality $\beta_c^z = \beta^z$, $\gamma_c^z = \gamma^z$				Reduced: lower dimensionality $\beta_2^z = \beta_3^z = \gamma_1^z = \gamma_2^z = 0$		
	Median	Mean	Std. Dev.		Median	Mean	Std. Dev.
	(1)	(2)	(3)		(4)	(5)	(6)
	A. Coefficients on s in student preferences (true value = 1)						
β_1^s	0.98	1.11	0.50	β_1^s	0.98	1.21	0.98
β_2^s	1.00	1.11	0.52	β_2^s	0.91	1.16	1.06
β_3^s	0.99	1.12	0.50	β_3^s	1.01	1.18	0.82
	B. Coefficients on m in college preferences (true value = 1)						
γ_1^m	1.04	1.64	3.21	γ_1^m	1.00	1.02	0.27
γ_2^m	0.94	1.30	3.71	γ_2^m	1.02	1.05	0.30
γ_3^m	1.12	1.47	3.41	γ_3^m	0.98	1.09	0.44
	C. Coefficients on z in student and college preferences (true value = 1)						
	GMM with all conditions in equation (D.25)						
β^z	0.12	0.41	2.54	β_1^z	0.97	0.99	0.16
γ^z	0.16	0.08	3.13	γ_3^z	0.97	1.00	0.21
	Using conditions 1 & 2 in equation (D.25)						
β^z	0.05	1.11	10.37	β_1^z	0.97	1.01	0.26
γ^z	0.17	-0.59	5.71	γ_3^z	0.97	1.11	1.30
	Using conditions 1 & 3 in equation (D.25)						
β^z	0.30	0.08	25.20	β_1^z	0.97	1.00	0.15
γ^z	0.19	8.37	92.48	γ_3^z	0.98	1.00	0.20
	Using conditions 2 & 3 in equation (D.25)						
β^z	-0.06	0.84	7.34	β_1^z	0.99	1.03	0.35
γ^z	0.08	0.30	6.35	γ_3^z	0.95	1.03	0.30

Notes: This table presents estimates for the coefficients in student or college utility functions (equations D.21 and D.22). The statistics are calculated using 150 MC samples. In the general model, we assume that $\beta_c^z = \beta^z$ (i.e., we have prior knowledge that β_c^z is constant across colleges) and $\gamma_c^z = \gamma^z$. The estimation is based on equation systems (D.23), (D.24), and (D.25). In the reduced model, we assume that we know $\beta_2^z = \beta_3^z = 0$ (i.e., z_i does not enter i's utility for college 2 or 3) and $\gamma_1^z = \gamma_2^z = 0$ (i.e., colleges 1 and 2 do not use z_i to evaluate students). The estimation is based on equation systems (D.23), (D.24), and (D.26).

D.3 A Parametric Approach: Bayesian Estimation

The practical difficulties of the semiparametric method motivate us to consider a parametric approach. We again focus on the utility functions as in equations (D.21) and (D.22) and use the 150 MC samples generated in Section D.1. In other words, z_i enters each college's preferences and each student's preferences over all colleges.

We assume that we know the functional form and the distributions of ϵ_{ic} and η_{ci}; however, we do not know, and thus will estimate, the standard deviation of ϵ_{i3} (the shock in students' utility for college 3), denoted by σ_ϵ. The other parameters to be estimated are β_c^d, β_c^s and β_c^z for all c in student preferences and γ_c^w, γ_c^m and γ_c^z for all c in college preferences. Collectively, we denote them by $(\beta, \gamma, \sigma_\epsilon)$.

Bayesian Estimation Procedure. We use a Gibbs sampler to implement the Bayesian estimation. The priors for β, γ, σ_ϵ^2 are:

$$\beta \sim N(0, \Sigma_\beta),\ \gamma \sim N(0, \Sigma_\gamma),\ \text{and } \sigma_\epsilon^2 \sim IW(\overline{\sigma}_\epsilon^2, \nu_\epsilon).$$

where IW is the inverse Wishart distribution. Following Chapter 5 of Rossi et al. (2012), we set diffuse priors as follows: The prior variances of β and γ (Σ_β and Σ_γ) are 100 times the identity matrix, and $(\overline{\sigma}_\epsilon^2, \nu_\epsilon) = (1, 2)$.

In each iteration, the Gibbs sampler goes through the following steps (for notational simplicity, we omit the index for iterations):

1. Conditional on student preferences, u_{ic}, from the previous iteration, we update college preferences, v_{ci}, by invoking the restrictions implied by the stability of the observed matching. For each college c, let $\mathcal{I}_c$ be the set of students with $u_{i\mu(i)} > u_{ic}$ (i.e., students who like their own match more than c) and $\mathcal{I}^c$ be the set of students with $u_{i\mu(i)} < u_{ic}$. The updating of college c's utilities and cutoff has four parts.

(a) c's preferences over those who are matched with it: Given v_{ci} from the previous iteration, we find $\underline{v}_c = \max_{i \in \mathcal{I}^c} v_{ci}$. For each i such that $\mu(i) = c$, v_{ci} is drawn from the standard normal distribution truncated below by $\underline{v}_c$.
(b) c's cutoff: It is the lowest utility among those who are matched with c.
(c) c's preferences over those in $\mathcal{I}^c$: c's utility for any student $i \in \mathcal{I}^c$ is drawn from the standard normal truncated above by c's cutoff.
(d) c's preferences over those in $\mathcal{I}_c$: c's utility for any student $i \in \mathcal{I}_c$ is drawn from the standard normal (without any truncation).

2. Conditional on the updated college preferences v_{ci} in this iteration, we update student preferences, u_{ic}, again by invoking the restrictions implied by stability of the observed match. Note that v_{ci} determines all colleges' cutoffs and their feasibility to each student. The updating of student preferences has three parts:[D.3]

(a) i's preferences over infeasible colleges: For an infeasible college c (i.e., v_{ci} is below c's cutoff), student i's utility is drawn from a normal distribution with variance 1 if $c \neq 3$ or σ_ϵ^2 if $c = 3$.

[D.3] In the estimation, a student's outside option is an always feasible college. The student's preference for her outside option is also updated according to the following steps.

(b) i's utility for her matched college: Given u_{ic} from the previous iteration, we find the highest utility among all feasible colleges other than $\mu(i)$, denoted by $\underline{u}_i$. i's utility for $\mu(i)$ is drawn from a normal distribution truncated below by $\underline{u}_i$ with variance 1 if $c \neq 3$ or σ_ϵ^2 if $c = 3$.

(c) i's preferences over her unmatched feasible colleges: i's utility for a feasible college c ($\neq \mu(i)$) is drawn from a normal distribution truncated above by $u_{i\mu(i)}$ with variance 1 if $c \neq 3$ or σ_ϵ^2 if $c = 3$.

3. Following the standard procedure as detailed in Chapter 5 of Rossi et al. (2012), we then update the distribution of β, γ, and σ_ϵ^2 conditional on the updated v_{ci} and u_{ic} as well as the data.

For each MC sample, we iterate through the Markov Chain 1.5 million times, and discard the first 0.55 million draws as "burn in" to ensure mixing. We compute the Potential Scale Reduction Factor (PSRF) following Gelman and Rubin (1992). For all the 19 parameters across the 150 MC samples, 92.4% of the PSRFs are below 1.1, while less than 1% of them are above 1.3.

Results. This parametric approach leads to the results in Table D.2. We observe that the estimator works well.

Table D.2: Results from Bayesian Estimation

Statistics on the 150 posterior means of each coefficient from the 150 MC samples							
	Median	Mean	Std. Dev.		Median	Mean	Std. Dev.
True value = 1				Coefficients on d (true value = -1)			
β_1^s	1.03	1.04	0.08	β_1^d	-1.03	-1.04	0.08
β_2^s	1.03	1.04	0.08	β_2^d	-1.04	-1.04	0.08
β_3^s	1.03	1.04	0.08	β_3^d	-1.03	-1.04	0.08
γ_1^m	1.05	1.10	0.23	Coefficients on w (true value = 1)			
γ_2^m	1.04	1.06	0.15	γ_1^w	1.04	1.10	0.22
γ_3^m	1.07	1.08	0.15	γ_2^w	1.04	1.06	0.15
β_1^z	1.03	1.04	0.08	γ_3^w	1.06	1.08	0.14
β_2^z	1.03	1.04	0.08				
β_3^z	1.02	1.04	0.08	Std. dev. of student utility shock (ϵ_{i3})			
γ_1^z	1.06	1.10	0.23	σ_ϵ	1.04	1.03	0.20
γ_2^z	1.04	1.06	0.14				
γ_3^m	1.06	1.08	0.15				

Notes: This table presents statistics on the posterior means of the coefficients in student and college utility functions (equations D.21 and D.22). There are 150 posterior means from the 150 Monte Carlo samples. For each sample, the Bayesian approach with a Gibbs sampler goes through the Markov Chain 1.5 million times, and we take the first 0.55 million iterations as "burn in." The last 0.95 million iterations are used to calculate the posterior means in a sample.

E Data Construction

For student and school characteristics, the main dataset we have used is the SIMCE test result dataset which is accompanied by parent and teacher questionnaires. To extract tuition data and location of students and schools, we have used publicly available data on the Ministry's website, `http://datos.mineduc.cl/dashboards/19731/bases-de-datos-directorio-de-establecimiento` (last accessed on March 28, 2021).

Here we briefly outline the construction of some key variables:

1. `Distance.` The data does not include the home address of each student. Instead, the distance is calculated as follows. We obtain the latitude and longitude of each school and those of each student's comuna. The former is contained in the data, whereas the latter is obtained from an online tool (`http://www.gpsvisualizer.com/geocoder/`). Using a Matlab package (*distance*) to calculate geodesic distances, we obtain the distances between each comuna and each school, measured in kilometers.
2. `Tuition.` Datasets with average monthly tuition (per student) are publicly available for most public and private subsidized schools in the years 2004-12. Interval data is available for most schools in 2013. To impute the missing tuition values in 2008, we first regressed tuitions in year `t` on tuitions in year `t+1`, and then predicted the missing values of year `t` using this fitted regression. We started with `t=2012`, and iteratively proceeded until `t=2008`.
3. `Teacher Quality.` This is measured by the average teacher experience at the school level. A teacher's tenure includes the years spent in other schools.
4. `Average percentile scores.` We first studentize the test scores of students in 2008 and compute their individual percentile rank in the whole market. This is used as a student characteristic. We take an average over the percentile ranks for each school in 2006 and use this as a school characteristic in 2008.
5. `Average parental education.` The average mother's education in 2006 is considered a school-level characteristic in 2008.
6. `Median parental Income.` Parental income is reported in 13 intervals. For each school, we first compute the proportion of households in each of the 13 intervals; then,

we find the median income interval based on the 13 proportions and use the midpoint of the median income interval as the median parental income.

7. `School enrollments and capacity.` We compute enrollments for each school for grade 10 in the years 2006, 2008, and 2010. We also compute enrollments for each school for grade 11 in 2010.[E.4] We take the maximum of these enrollments across each school and set it as the capacity unless it is less than 20 (in which case the capacity is set to 20). We use this variable to determine which schools have a binding capacity constraint for grade 10 in the year 2008.

Finally, missing values are imputed. For students, missing values for variable X are imputed by matching the observations to a group of similar observations (similar in dimensions other than X), respectively. The missing values are then assigned the median values of X for that matched group. For schools, missing values are replaced by analogous aggregated variables at the school level in 2008.

F Additional Details on Data Analysis

Estimation The same as our Monte Carlo simulations, we use a Bayesian approach with a Gibbs sampler to estimate student and school preferences in the Chilean data. In addition to the procedure of updating the Markov Chain as described in Section D.3 for the Monte Carlo, this appendix describes some unique features in this empirical exercise. In particular, we emphasize that (i) some schools are girls or boys only and thus are never feasible to the other gender in the updating of the Markov Chain, (ii) a student can be unacceptable to a school, and (iii) there are some students who are not from Market Valparaiso but attending a school in Market Valparaiso and contributing to the determination of school cutoffs.

There are 527 students who are not from Market Valparaiso but attend a private school in Market Valparaiso. Among them, 161 students attend a private school with binding capacity constraint. When updating the Markov Chain, these 161 students are included in the calculation of school cutoffs, but their preferences are not the focus of our paper. Therefore, to simplify the procedure, we assume that they only find their matched school acceptable (i.e., better than their outside option).

[E.4] We use grade 11 in 2010 as a proxy for grade 10 in 2009.

We iterate through two distinct chains from dispersed initial values 1.75 million times, and take the first 1 million as "burn in." The posterior means and standard deviations of the last 0.75 million iterations are similar between the chains. We check convergence by calculating the Potential Scale Reduction Factor (PSRF) as proposed by Gelman and Rubin (1992). The PSRFs are below 1.1 for all the parameters.

Model Fit Our model fits the data reasonably well when we compare the observed matching with the one predicted based on our model.

We use the average of 1,000 simulations of the matching market to calculate the model prediction. In each simulation, we take the posterior means in Table 4 and the observables of each student and each school, randomly draw the utility shocks in equations (16) and (17) according to the estimated distributions, and calculate each student and each school's preferences. A stable matching is found by the Gale-Shapley deferred acceptance in each simulation and is compared to the observed matching.

As a benchmark, we calculate a random prediction that is similarly constructed for 1,000 simulations, except that each agent's utility for a school/student is a draw from the standard normal. Its fit is then evaluated against the observed matching.

We present two sets of model fit measures. The first is how often among the 1,000 simulations an observed outcome is correctly predicted. For their matched school, the random prediction is correct for merely 1.24% of the students. In contrast, our model correctly predicts for 5.88% of the students, 4.74 times the rate from the random prediction.[F.5] Moreover, the model correctly predicts the type of their matched school for 61.73% of the students, 1.63 times the rate from the random prediction (37.82%).

The second set of model fit measures focuses on the average characteristics of each school's matched students and the attributes of each student's matched school. For a given student characteristic (evaluated as an average at each school), we calculate the root-mean-square errors (RMSEs, hereafter) across the 1,000 simulations with the "error" being the difference between each school's predicted average and its observed average.[F.6] Hence, a high RMSE indicates a poor fit. Compared with the random

[F.5] This seemingly low number is understandable: the matching market resembles a discrete choice with 125 options, so correctly predicting a student's choice is challenging.

[F.6] Specifically, for student characteristic x, $RMSE_x = \sqrt{\frac{1}{M \cdot C}\sum_{m=1}^{M}\sum_{c=1}^{C}\left(\bar{x}_{c,m}^{pred} - \bar{x}_c^{obs}\right)^2}$, where $\bar{x}_{c,m}^{pred}$ is the average characteristic among the students matched with school c in the m-th simulated prediction, the model prediction leads to RMSEs that are 55–71% lower except for the characteristic, female. In the data, a student's gender does not play an important role in the utility functions (see Table 4), while being weakly correlated with the student's composite score and uncorrelated with other characteristics. This might explain the poor fit of the model for this characteristic.

Similarly, for a given school attribute, the RMSEs from the model are 32–45% lower than those from the random prediction except for two attributes, teacher experiences and the fraction of female students. The poor fit on those two dimensions may be due to their relative irrelevance in student and school preferences.[F.7]

Low-income versus Non-low-income Students Our counterfactual policy prioritize students from low-income families for admissions to all schools. A student is of low income if the student's parental income is among the lowest 40%. Table F.3 shows some summary statistics of the students by their income status.

Table F.3: Summary Statistics of Student Characteristics by Income Status

	Low Income (N=3,997)		Non-low Income (N=5,307)	
	mean	s.d.	mean	s.d.
Mother's education (years)	12.29	2.74	15.23	2.92
Female	0.52	0.50	0.51	0.50
Language score	0.37	0.26	0.58	0.28
Math score	0.37	0.25	0.59	0.28
Composite score	0.36	0.25	0.59	0.28
Parental income (CLP)	133,613	37,021	654,193	557,166
Distance to the enrolled school (km)	2.69	2.42	2.90	2.85

Notes: This table describes the student characteristics by income status. A student is of low income if the student's parental income is among the bottom 40%. Parental income is measured in 2008 when 1 USD was about 522 CLP.

market and $\bar{x}_c^{obs}$ is the average characteristic among those who are matched with c in the data.

[F.7]These two attributes do not significantly contribute to the utility functions (see Table 4) and are only weakly correlated with other school attributes. Specifically, a school's fraction of females is uncorrelated with all the school attributes, and a school's teacher experience is weakly correlated with average student score but uncorrelated with all other school attributes.