Online Appendix to

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

YingHua He Shruti Sinha Xiaoting Sun

A Identification of a Nonseparable Model

We now discuss the nonparametric identification of a more general nonseparable utility specification based on the arguments in Matzkin (2019). As we shall see, compared with those for the separable model, our identification results for this model are based on an additional assumption (Assumption A.3) and two different rank conditions (Conditions A.4 and A.5). Hence, the sufficient conditions below do not nest those in the main text.

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,

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

The additive index $y_{i1}+\epsilon_{i1}$ can be relaxed to some known function such as $y_{i1}\cdot\epsilon_{i1}$ (Matzkin, 2019). Below we discuss a set of sufficient conditions under which our identification strategy applies to $\{u^c\}_c$. Moreover, we show $\{v^c\}_c$ is identified under additional separability. This helps clarify the role of the additive separability in equation (1). For notational simplicity, we also use $u^1(z_i,y_{i1},\epsilon_{i1})$ to denote $u^1\left(z_i,y_{i1}+\epsilon_{i1}\right)$. The utility of the outside option u_{i0} is assumed to be a continuous random variable.[A.1]

Assumption A.1. *(i) z_i, y_i, and w_i are continuously distributed; (ii) for each $c\in\mathbf{C}$, the functions, u^c and v^c, are continuously differentiable; (iii) F is continuously*

[A.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.

differentiable; (iv) for each $c \in \mathbf{C}$, 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 $v^c(z_i, w_{ci}, \eta_{ci}) = \delta_c$, $\frac{\partial v^c(z_i,w_{ci},\eta_{ci})}{\partial w_{ci}} \neq 0$.

Assumption A.2. *(ϵ_i, η_i) is independent of (z_i, y_i, w_i).*

Assumption A.3. *(i) The utility of the outside option is $u_{i0} = h(y_{i0})$, where $y_{i0} \in \mathcal{Y}_0 \subseteq \mathbb{R}^{d_{y_0}}$ 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 A.1 and Assumption A.2 impose smoothness and exogeneity similar to Assumptions 3.1 and 3.2 in Section 3. Part (iv) of Assumption A.1 guarantees that there is a one-to-one relationship between the value of each utility function and its unobservable. Part (v) of Assumption A.1 guarantees that u_{ic} and v_{ci} are not constant w.r.t. y_{ic} and w_{ci}, respectively, such that a change in y_{ic} or w_{ci} generates a change in the conditional probability of being unmatched. Assumption A.3 guarantees that u_{i0} is observed by the researcher and has a large support.

By the monotonicity assumption (part iv of Assumption A.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 $a \in \mathbb{R}$,

$$u^1\left(z_i, \tilde{u}^1(z_i, a)\right) = a, \text{ and } v^1\left(z_i, w_{1i}, \tilde{v}^1(z_i, w_{1i}, a)\right) = a,$$

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

Then, we have

$$\begin{aligned}
\lambda_L(\iota_{1i}, \ldots, \iota_{Ci}) &= \mathbb{P}(v_{ci} \geq \delta_c \; \forall\, c \in L;\; v_{di} < \delta_d \; \forall\, d \notin L | z_i, w_i; \mu) \\
&= \mathbb{P}(v^c(z_i, w_{ci}, \eta_{ci}) \geq \delta_c \; \forall\, c \in L;\; v^d(z_i, w_{di}, \eta_{di}) < \delta_d \; \forall\, d \notin L | z_i, w_i; \mu) \\
&= \mathbb{P}(\eta_{ci} \geq \tilde{v}^c(z_i, w_{ci}, \delta_c) \;\; \forall\, c \in L;\; \eta_{di} < \tilde{v}^d(z_i, w_{di}, \delta_d) \;\; \forall\, d \notin L | z_i, w_i; \mu),
\end{aligned}$$

where $\iota_{ci} = \tilde{v}^c(z_i, w_{ci}, \delta_c)$ 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,

$$\begin{aligned}
\mathbb{P}(0 = \arg\max_{c \in L} u_{ic} | L, z_i, y_i, u_{i0}) &= \mathbb{P}(u_{i0} > u_{ic} \text{ for all } c \in L | L, z_i, y_i, u_{i0}) \\
&= \mathbb{P}(u_{i0} > u^c(z_i, y_{ic}, \epsilon_{ic}) \text{ for all } c \in L | L, z_i, y_i, u_{i0}) \\
&= \mathbb{P}(\epsilon_{i1} < \tilde{u}^1(z_i, u_{i0}) - y_{i1} \text{ if } 1 \in L; \\
&\qquad \epsilon_{ic} < \tilde{u}^c(z_i, y_{ic}, u_{i0}) \text{ for all } c \in L \text{ and } c \neq 1 | L, z_i, y_i, u_{i0}) \\
&= g_{0,L}(\tau_{i1}, \ldots, \tau_{iC}),
\end{aligned}$$

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 (5) for $c = 0$, we have

$$\begin{aligned}\sigma_0(z_i, y_i, w_i, u_{i0}) &= \sum_{L\in\mathcal{L}} \lambda_L(\iota_{1i}, \ldots, \iota_{Ci}) \cdot g_{0,L}(\tau_{i1}, \ldots, \tau_{iC}) \\ &\equiv \Lambda_0(\tau_{i1}, \ldots, \tau_{iC}, \iota_{1i}, \ldots, \iota_{Ci}), \end{aligned} \tag{A.2}$$

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 value (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 $(z_i, w_i, y_{i,-1}, u_{i0}) = (z, w, y_{-1}, u_0)$. We define a $C \times C$ matrix

$$\Pi_1\left(y_1^1, \ldots, y_1^C; z, w, y_{-1}, u_0\right) \equiv \begin{pmatrix} \frac{\partial\sigma_0\left(z,w,y_{-1},u_0,y_1^1\right)}{\partial y_{i1}} & \cdots & \frac{\partial\sigma_0\left(z,w,y_{-1},u_0,y_1^1\right)}{\partial y_{iC}} \\ \vdots & \ddots & \vdots \\ \frac{\partial\sigma_0\left(z,w,y_{-1},u_0,y_1^C\right)}{\partial y_{i1}} & \cdots & \frac{\partial\sigma_0\left(z,w,y_{-1},u_0,y_1^C\right)}{\partial y_{iC}} \end{pmatrix},$$

where for $m = 1, \ldots, C$, the m^{th} 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

$$\Pi_2\left(y_1^1, \ldots, y_1^{2C}; z, w, y_{-1}, u_0\right) \equiv \begin{pmatrix} \frac{\partial\sigma_0(z,w,y_{-1},u_0,y_1^1)}{\partial y_{i1}} & \cdots & \frac{\partial\sigma_0(z,w,y_{-1},u_0,y_1^1)}{\partial y_{iC}} & \frac{\partial\sigma_0(z,w,y_{-1},u_0,y_1^1)}{\partial w_{1i}} & \cdots & \frac{\partial\sigma_0(z,w,y_{-1},u_0,y_1^1)}{\partial w_{Ci}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\sigma_0(z,w,y_{-1},u_0,y_1^{2C})}{\partial y_{i1}} & \cdots & \frac{\partial\sigma_0(z,w,y_{-1},u_0,y_1^{2C})}{\partial y_{iC}} & \frac{\partial\sigma_0(z,w,y_{-1},u_0,y_1^{2C})}{\partial w_{1i}} & \cdots & \frac{\partial\sigma_0(z,w,y_{-1},u_0,y_1^{2C})}{\partial w_{Ci}} \end{pmatrix},$$

where for $m = 1, \ldots, 2C$, the m^{th} 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 and w_i, evaluated at $(z, w, y_{-1}, u_0, y_1^m)$.

Condition A.4. *For a given value (z, w, y_{-1}, u_0) in the interior of $\mathcal{Z} \times \mathcal{W} \times \mathcal{Y}_{-1} \times \mathcal{U}_0$, there exist 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 A.5. *For a given value (z, w, y_{-1}, u_0) in the interior of $\mathcal{Z} \times \mathcal{W} \times \mathcal{Y}_{-1} \times \mathcal{U}_0$, there exist $2C$ 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 A.4 and then independently choose another $2C$ values of y_{i1} to satisfy Condition A.5.

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

Proposition A.6. *Suppose that Assumptions A.1-A.3 and Conditions A.4 and A.5 are satisfied. We have (i) for each $c \in \mathbf{C}\backslash\{1\}$, for any value (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 value (z, w_c) in the interior of $\mathcal{Z} \times \mathcal{W}_c$, for any coordinate $k = 1, \ldots, d_z$, $\frac{\partial v^c(z, w_c, \eta_{ci})}{\partial z_i^k} / \frac{\partial v^c(z, w_c, \eta_{ci})}{\partial w_{ci}}$ is identified, where η_c is such that $v^c\left(z, w_c, \eta_c\right) = \delta_c$.*

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 $v^c\left(z, w_c, \eta_c\right) = \delta_c$). 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., $\eta_c^\rho = \text{Quantile}_{\eta_{ci}}(\rho) \equiv \inf\{\eta_c : F_{\eta_{ci}}(\eta_c) \geq \rho\}$ for $\rho \in (0,1)$, where $F_{\eta_{ci}}$ is the marginal CDF of η_{ci}.

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

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

Proof of Proposition A.6. To simplify notations, for $k = 1, \ldots, d_z$, let $u_{z_i^k}^c = \frac{\partial u^c}{\partial z_i^k}$ and similar notations are defined for v^c, $\tilde{u}^c$, $\tilde{v}^c$, and the other variables, and let $\sigma_0^m = \sigma_0\left(z, w, y_{-1}, u_0, y_1^m\right)$ for $m = 1, \ldots, 2C$. Let t^m be the value of $(\tau_{i1}, \ldots, \tau_{iC}, \iota_{1i}, \ldots, \iota_{Ci})$

evaluated at $(z, w, y_{-1}, u_0, y_1^m)$. Under Assumption A.1(i)-(iii), in equation (A.2), Λ_0, u^c, and v^c are continuously differentiable and the observables are all continuously distributed. Taking derivatives of equation (A.2) on both sides w.r.t. y_{ic} and w_{ci}, and evaluating them at $(z, w, y_{-1}, u_0, y_1^m)$, we have, for $c = 1$,

$$\frac{\partial \sigma_0^m}{\partial y_{i1}} = -\frac{\partial \Lambda_0(t^m)}{\partial \tau_{i1}} \text{ and } \frac{\partial \sigma_0^m}{\partial w_{1i}} = \frac{\partial \Lambda_0(t^m)}{\partial \iota_{1i}} \tilde{v}^1_{w_{1i}}, \tag{A.3}$$

and, for $c \neq 1$,

$$\frac{\partial \sigma_0^m}{\partial y_{ic}} = \frac{\partial \Lambda_0(t^m)}{\partial \tau_{ic}} \tilde{u}^c_{y_{ic}} \text{ and } \frac{\partial \sigma_0^m}{\partial w_{ci}} = \frac{\partial \Lambda_0(t^m)}{\partial \iota_{ci}} \tilde{v}^c_{w_{ci}}. \tag{A.4}$$

Further, taking derivatives of equation (A.2) on both sides w.r.t. u_{i0} and z_i^k, and evaluating them at $(z, w, y_{-1}, u_0, y_1^m)$, we have

$$\frac{\partial \sigma_0^m}{\partial u_{i0}} = \sum_{c=1}^{C} \frac{\partial \Lambda_0(t^m)}{\partial \tau_{ic}} \tilde{u}^c_{u_{i0}}, \tag{A.5}$$

$$\frac{\partial \sigma_0^m}{\partial z_i^k} = \sum_{c=1}^{C} \frac{\partial \Lambda_0(t^m)}{\partial \tau_{ic}} \tilde{u}^c_{z_i^k} + \sum_{c=1}^{C} \frac{\partial \Lambda_0(t^m)}{\partial \iota_{ci}} \tilde{v}^c_{z_i^k}. \tag{A.6}$$

Substituting equations (A.3) and (A.4) into equations (A.5) and (A.6), we have

$$\frac{\partial \sigma_0^m}{\partial u_{i0}} = -\frac{\partial \sigma_0^m}{\partial y_{i1}} \tilde{u}^1_{u_{i0}} + \sum_{c=2}^{C} \frac{\partial \sigma_0^m}{\partial y_{ic}} (\tilde{u}^c_{y_{ic}})^{-1} \tilde{u}^c_{u_{i0}}, \tag{A.7}$$

$$\frac{\partial \sigma_0^m}{\partial z_i^k} = -\frac{\partial \sigma_0^m}{\partial y_{i1}} \tilde{u}^1_{z_i^k} + \sum_{c=2}^{C} \frac{\partial \sigma_0^m}{\partial y_{ic}} (\tilde{u}^c_{y_{ic}})^{-1} \tilde{u}^c_{z_i^k} + \sum_{c=1}^{C} \frac{\partial \sigma_0^m}{\partial w_{ci}} (\tilde{v}^c_{w_{ci}})^{-1} \tilde{v}^c_{z_i^k}. \tag{A.8}$$

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 equation, $u^c(z_i, y_{ic}, \tilde{u}^c(z_i, y_{ic}, u_{i0})) = u_{i0}$, 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 A.4 and A.5 because for any y_1, one can find an ϵ_1 so that the above equation holds.

Similarly, taking derivatives of the equation, $v^c(z_i, w_{ci}, \tilde{v}^c(z_i, w_{ci}, \delta_c)) = \delta_c$, w.r.t. w_{ci} and z_i^k and making rearrangements, we obtain, for $c \in \mathbf{C}$, $\tilde{v}^c_{w_{ci}} = -\frac{v^c_{w_{ci}}}{v^c_{\eta_{ci}}}$ 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 (A.7) and (A.8), we obtain

$$\frac{\partial \sigma_0^m}{\partial u_{i0}} = -\frac{\partial \sigma_0^m}{\partial y_{i1}} \frac{1}{u^1_{\epsilon_{i1}+y_{i1}}} - \sum_{c=2}^{C} \frac{\partial \sigma_0^m}{\partial y_{ic}} \frac{1}{u^c_{y_{ic}}}, \tag{A.9}$$

$$\frac{\partial \sigma_0^m}{\partial z_i^k} = \frac{\partial \sigma_0^m}{\partial y_{i1}} \frac{u^1_{z_i^k}}{u^1_{\epsilon_{i1}+y_{i1}}} + \sum_{c=2}^{C} \frac{\partial \sigma_0^m}{\partial y_{ic}} \frac{u^c_{z_i^k}}{u^c_{y_{ic}}} + \sum_{c=1}^{C} \frac{\partial \sigma_0^m}{\partial w_{ci}} \frac{v^c_{z_i^k}}{v^c_{w_{ci}}}. \tag{A.10}$$

Next, stacking equation (A.9) for $m = 1, \ldots, C$, we have

$$\begin{pmatrix} \frac{\partial \sigma_0^1}{\partial u_{i0}} \\ \vdots \\ \frac{\partial \sigma_0^C}{\partial u_{i0}} \end{pmatrix} = -\begin{pmatrix} \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^1)}{\partial y_{i1}} & \cdots & \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^1)}{\partial y_{iC}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^C)}{\partial y_{i1}} & \cdots & \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^C)}{\partial y_{iC}} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{u^1_{\epsilon_{i1}+y_{i1}}} \\ \frac{1}{u^2_{y_{i2}}} \\ \vdots \\ \frac{1}{u^C_{y_{iC}}} \end{pmatrix},$$

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 A.1. Note that the derivatives of σ_0 in the above system can be observed from the population data. Then, by Condition A.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 (A.10) for $m = 1, \ldots, 2C$, we obtain

$$\begin{pmatrix} \frac{\partial \sigma_0^1}{\partial z_i^k} \\ \vdots \\ \frac{\partial \sigma_0^{2C}}{\partial z_i^k} \end{pmatrix} = \begin{pmatrix} \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^1)}{\partial y_{i1}} & \cdots & \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^1)}{\partial y_{iC}} & \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^1)}{\partial w_{1i}} & \cdots & \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^1)}{\partial w_{Ci}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^{2C})}{\partial y_{i1}} & \cdots & \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^{2C})}{\partial y_{iC}} & \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^{2C})}{\partial w_{1i}} & \cdots & \frac{\partial \sigma_0(z,w,y_{-1},u_0,y_1^{2C})}{\partial w_{Ci}} \end{pmatrix} \cdot \begin{pmatrix} u^1_{z_i^k}/u^1_{\epsilon_{i1}+y_{i1}} \\ u^2_{z_i^k}/u^2_{y_{i2}} \\ \vdots \\ u^C_{z_i^k}/u^C_{y_{iC}} \\ v^1_{z_i^k}/v^1_{w_{1i}} \\ \vdots \\ v^C_{z_i^k}/v^C_{w_{Ci}} \end{pmatrix}.$$

Then, by Condition A.5, for all $c \in \mathbf{C}$, $\frac{v^c_{z_i^k}}{v^c_{w_{ci}}}$ 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 A.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 $u^c_{z^k_i}$ for any given (z, y_c) and $\rho \in (0,1)$. $\square$

Proof of Corollary A.7. Proposition A.6 implies that $\frac{\partial v^c(z_i,\eta_{ci})}{\partial z^k_i}$ is identified, where η_{ci} is such that $v^c(z_i, \eta_{ci}) + w_{ci} = \delta_c$. For any z and $\rho \in (0,1)$, the equivariance property of quantiles (e.g., Chesher, 2003) implies that $\text{Quantile}_{-w_{ci}|z}(\rho) = v^c(z, \eta^\rho_c)$, where the LHS is known from the joint distribution of (w_{ci}, z_i). Hence, $\frac{\partial v^c\left(z,\eta^\rho_c\right)}{\partial z^k_i}$ is identified. $\square$

B A Control Function Approach

This appendix discusses a control function approach (Heckman and Robb, 1985; Blundell and Powell, 2004; Imbens and Newey, 2009) that relaxes Assumption 3.2 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:

$$z_{1i} = h(t_i, z_{2i}, \xi_i). \tag{B.11}$$

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 (B.11), 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 $e_i = \xi_i$.[B.2]

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

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

$$\epsilon_{ic} = \varphi^c(e_i) + \tilde{\epsilon}_{ic} \text{ and } \eta_{ci} = \phi^c(e_i) + \tilde{\eta}_{ci}. \tag{B.12}$$

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

Plugging equation (B.12) into the utility functions in equation (1), we have

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

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 3.4 then can be used to identify the derivatives of the functions $\{u^c,\ v^c,\ r^c,\ \varphi^c,\ \phi^c\}_c$.

C Evaluating Condition 3.3

C.1 A Nonparametric One-college Example

The following example shows that in a one-college case, Condition 3.3 holds for *all* but the exponential distribution on η_{1i}.

Example C.1. *Consider a one-college example:* $\mathbf{C} = \{1\}$, *and* $\mathcal{L} = \{\{0\}, \{0,1\}\}$. *Equation* (5) *for* $c = 1$ *can be written as* $\sigma_1(z_i, y_i, w_i) = \lambda_{\{0,1\}}(\iota_{1i}) \cdot g_{1,\{0,1\}}(\tau_{i1})$ *because* $g_{1,\{0\}}(\tau_{i1}) = 0$. *Recall that* $\iota_{1i} = v^1(z_i) + w_{1i}$ *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 3.3 requires that, for any* z *in the interior of* $\mathcal{Z}$, *there are two values of* w_{1i}, $\widehat{w}_1$ *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 3.3 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 3.3 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* $\frac{\partial \log \lambda_{\{0,1\}}(\widehat{\iota}_1)}{\partial \iota_{1i}} \neq \frac{\partial \log \lambda_{\{0,1\}}(\widetilde{\iota}_1)}{\partial \iota_{1i}}$. *The violation of Condition 3.3 stringently restricts* $\lambda_{\{0,1\}}(\iota_{1i})$, *or the probability of college 1 being feasible to* i. *Specifically, for fixed* z, *Condition 3.3 is violated if the supply* *elasticity w.r.t.* w_{1i} *is linear in* w_{1i}*, or* $\frac{\partial \log \lambda_{\{0,1\}}(\iota_{1i})}{\partial \iota_{1i}}$ *is a constant for all* w_{1i}*. This means that* $\lambda_{\{0,1\}}(\iota_{1i}) = \exp(a + b\iota_{1i})$ *with constants* a *and* b*, which only occurs when* η_{1i} *has an exponential distribution.*

C.2 Parametric Analysis of Probit and Logit Models

C.2.1 Two Colleges

We now parameterize a two-college model with $\mathbf{C} = \{1, 2\}$. Student i's utility when attending college c for $c = 1, 2$ is specified as

$$u_{ic} = u^c(z_i) + r^c(y_{ic}) + \epsilon_{ic} = z_i + y_{ic} + \epsilon_{ic}.$$

And $u_{i0} = \epsilon_{i0}$. Because $r^c(y_{ic}) = y_{ic}$, we can choose any value to be $\overline{y}_c$ at which $\frac{\partial r^c(\overline{y}_c)}{\partial y_{ic}} = 1$ as required by the scale normalization.

College c values student i at

$$v_{ci} = v^c(z_i) + w_{ci} + \eta_{ci} = z_i + w_{ci} + \eta_{ci}.$$

We will consider ϵ_{i0}, ϵ_{ic}, and η_{ci} being i.i.d. $N(0, 1)$ or type I extreme values.

Let δ_1 and δ_2 be the cutoffs of the two colleges given the stable matching in the continuum economy. Given the parametric assumptions, for a wide range of (δ_1, δ_2) in $\mathbb{R}^2$, there exist a vector of college capacities and joint distributions of (z_i, y_i, w_i) such that (δ_1, δ_2) are the cutoffs given the stable matching.

We start with a probit model in which ϵ_{i0}, ϵ_{ic}, and η_{ci} are i.i.d. $N(0, 1)$. We use Mathematica to derive an expression for $\Pi(z, y, \widehat{w}, \widetilde{w})$ and calculate its determinant.

To show that we can choose $(\widehat{w}, \widetilde{w})$ to make $|\Pi(z, y, \widehat{w}, \widetilde{w})|$ non-zero for given values of (z_i, y_i), we consider a more adversarial case by fixing the values of $(\widehat{w}_2, \widetilde{w}_1, \widetilde{w}_2)$ while letting $\widehat{w}_1$ change freely. In this example, we let $\delta_1 = 1$ and $\delta_2 = 0.75$.

Panel (a) in Figure C.1 shows how $|\Pi(z, y, \widehat{w}, \widetilde{w})|$ changes with $\widehat{w}_1$ for 4 different vectors of $(z, y, \widehat{w}_2, \widetilde{w}_1, \widetilde{w}_2)$. In each of the four cases, for a wide range of $\widehat{w}_1$, $|\Pi(z, y, \widehat{w}, \widetilde{w})| \neq 0$. We have also experimented with more values of $(z, y, \widehat{w}_2, \widetilde{w}_1, \widetilde{w}_2)$ as well as different values of (δ_1, δ_2) and found similar evidence for Condition 3.3.

We then repeat the same analysis in a logit model. That is, ϵ_{i0}, ϵ_{ic}, and η_{ci} are i.i.d. type I extreme values. Again, $\delta_1 = 1$ and $\delta_2 = 0.75$. Panel (b) in Figure C.1 shows how $|\Pi(z, y, \widehat{w}, \widetilde{w})|$ changes with $\widehat{w}_1$ for 4 different vectors of $(z, y, \widehat{w}_2, \widetilde{w}_1, \widetilde{w}_2)$. For all cases, there is again a wide range of $\widehat{w}_1$ such that Condition 3.3 is satisfied.

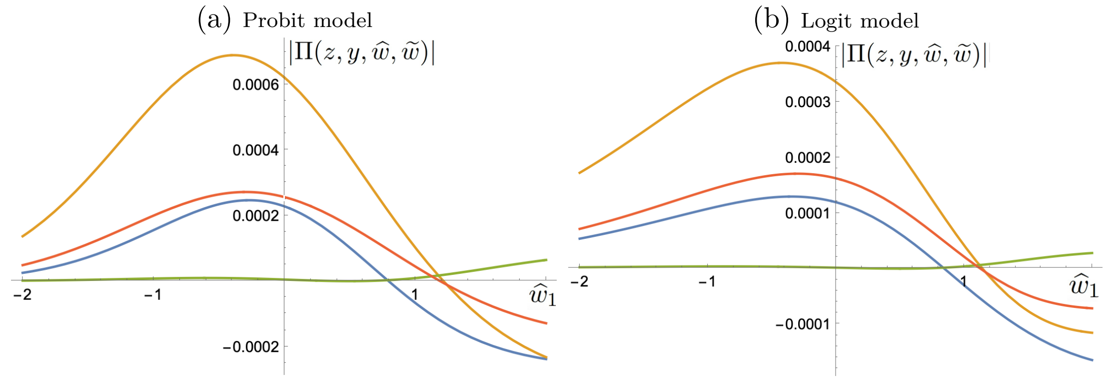

Figure C.1: Probit and Logit Models with Two Colleges

Notes: This figure shows how $|\Pi(z, y, \widehat{w}, \widetilde{w})|$ changes with $\widehat{w}_1$ for 4 different vectors of $(z, y_1, y_2, \widehat{w}_2, \widetilde{w}_1, \widetilde{w}_2)$ in a probit model (Panel a) and in a logit model (Panel b). The cutoffs are fixed at $\delta_1 = 1$ and $\delta_2 = 0.75$. For both panels, the four vectors of $(z, y_1, y_2, \widehat{w}_2, \widetilde{w}_1, \widetilde{w}_2)$ (from the top line to the bottom line at $\widehat{w}_1 = -2$) are: (i) $(1, -0.5, 0.5, -0.5, 1, 0.5)$; (i) $(1, -1, 1, -0.5, 1, 0.5)$; (iii) $(0.5, 0.5, -0.5, -0.5, 0.5, 1)$; (iv) $(0.5, 1, -1, 1, 0.5, 1)$.

C.2.2 Logit Models with Three or Four Colleges

We further expand the example to three or four colleges. Due to computational issues, it becomes infeasible to consider probit models. We therefore focus on logit models. Figure C.2 shows how $|\Pi(z, y, \widehat{w}, \widetilde{w})|$ changes with $\widehat{w}_1$ in the 2-, 3-, and 4-college examples for 4 different values of other variables. From Panel (a) to (c), there is no evidence that Condition 3.3 becomes more difficult to satisfy as there are more colleges.

Remark 1. *In Figure C.2, the absolute value of $|\Pi(z, y, \widehat{w}, \widetilde{w})|$ decreases (exponentially) with the number of colleges, but it is not a sign of possible violations of the full-rank condition. In fact, such a pattern is implied by the definition of $\Pi(z, y, \widehat{w}, \widetilde{w})$ because each element in the matrix is a partial derivative of a match probability and thus tends to be a small value. By the Leibniz formula for determinants, we have*

$$|\Pi(z, y, \widehat{w}, \widetilde{w})| = \sum_{\varrho \in S_{2C}} sgn(\varrho) \prod_{j=1}^{2C} \pi_{\varrho(j), j},$$

where sgn is the sign function of permutations in the permutation group S_{2C}, which returns $+1$ and -1 for even and odd permutations, respectively; $\pi_{\varrho(j), j}$ is the element of Π in the $\varrho(j)$-th row and j-th column. Based on the discussion above, $\prod_{j=1}^{2C} \pi_{\varrho(j), j}$ tends to be small and decrease when C increases, so does $|\Pi(z, y, \widehat{w}, \widetilde{w})|$.

As a piece of evidence that is consistent with this observation, when we express

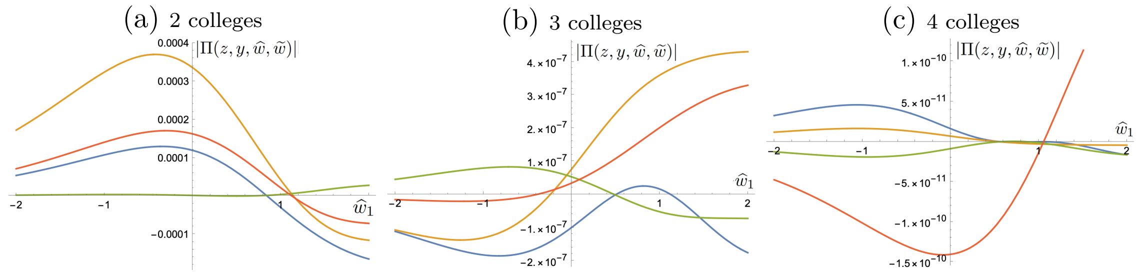

Figure C.2: Logit Models with 2–4 Colleges

Notes: This figure shows how $|\Pi(z, y, \widehat{w}, \widetilde{w})|$ changes with $\widehat{w}_1$ given 4 different vectors of other variables in each logit model with a different number of colleges. Panel (a) is the same as Panel (b) in Figure C.1. In Panel (b), there are three colleges; the cutoffs are $\delta_1 = 1$, $\delta_2 = 0.75$, and $\delta_3 = 0.5$; and the four vectors of $(z, y, \widehat{w}_2, \widehat{w}_3, \widetilde{w}_1, \widetilde{w}_2, \widetilde{w}_3)$ are: $(1, -1, 0.5, 1, 0.5, 0.5, 0.5, 1, 0.5)$, $(-1, 1, 0.5, -1, 0.5, 1, 0.5, 1, 0.5)$, $(-0.5, 0.5, 0.5, -0.5, 0.5, 0.5, 1, 0.5, 1)$, and $(0.5, -0.5, 0.5, -0.5, 0.5, 0.5, 1, 0.5, 0.5)$. In Panel (c), there are four colleges; the cutoffs are $\delta_1 = 1$, $\delta_2 = 0.75$, $\delta_3 = 0.5$, and $\delta_4 = 0.6$; and the four vectors of $(z, y, \widehat{w}_2, \widehat{w}_3, \widehat{w}_4, \widetilde{w}_1, \widetilde{w}_2, \widetilde{w}_3, \widetilde{w}_4)$ are: $(0.5, -0.5, 0.5, 0.5, -0.5, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 0.5)$, $(-0.5, 0.5, 0.5, -0.5, -0.5, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 1)$, $(1, -1, 0.5, 0.5, 1, 0.5, 1, 0.5, 0.5, 1, 1, 0.5)$, and$(-1, 1, 0.5, -0.5, -1, 0.5, 1, 1, 0.5, 1, 0.5, 0.5)$.

match probabilities in percentage points, the determinants corresponding to the three in Figure C.2 are 100^{2C} *times of those in Figure C.2 and thus increase in* C *exponentially.*

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,

$$u_{ic} \; = \beta_c^y \times y_{ic} + \beta_c^s \times s_i + \beta_c^z \times z_i + \epsilon_{ic}, \tag{D.13}$$

where y_{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. For

$c = 1, 2, 3$, $\beta_c^y = -1$ and $\beta_c^s = \beta_c^z = 1$.

College c values each student as follows:

$$v_{ci} \;\; = \gamma_c^w \times w_{ci} + \gamma_c^m \times m_i + \gamma_c^z \times z_i + \eta_{ci}, \tag{D.14}$$

where w_{ci} 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. For $c = 1, 2, 3$, $\gamma_c^w = \gamma_c^m = \gamma_c^z = 1$. 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 $x_i = (y_i, w_i, s_i, z_i, m_i)$, with $y_i = (y_{i1}, y_{i2}, y_{i3})$ and $w_i = (w_{1i}, w_{2i}, w_{3i})$. We rewrite equation (11) in the semiparametric setting for s_i and m_i, respectively, integrate over the entire support of x_i to obtain unconditional expectations $\mathbb{E}$:

$$\begin{pmatrix} \mathbb{E}\left(\dfrac{\partial\sigma_1(x_i)}{\partial s_i}\right) \\ \mathbb{E}\left(\dfrac{\partial\sigma_2(x_i)}{\partial s_i}\right) \\ \mathbb{E}\left(\dfrac{\partial\sigma_3(x_i)}{\partial s_i}\right) \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left(-\dfrac{\partial\sigma_1(x_i)}{\partial y_{i1}}\right) & \mathbb{E}\left(-\dfrac{\partial\sigma_1(x_i)}{\partial y_{i2}}\right) & \mathbb{E}\left(-\dfrac{\partial\sigma_1(x_i)}{\partial y_{i3}}\right) \\ \mathbb{E}\left(-\dfrac{\partial\sigma_2(x_i)}{\partial y_{i1}}\right) & \mathbb{E}\left(-\dfrac{\partial\sigma_2(x_i)}{\partial y_{i2}}\right) & \mathbb{E}\left(-\dfrac{\partial\sigma_2(x_i)}{\partial y_{i3}}\right) \\ \mathbb{E}\left(-\dfrac{\partial\sigma_3(x_i)}{\partial y_{i1}}\right) & \mathbb{E}\left(-\dfrac{\partial\sigma_3(x_i)}{\partial y_{i2}}\right) & \mathbb{E}\left(-\dfrac{\partial\sigma_3(x_i)}{\partial y_{i3}}\right) \end{pmatrix} \cdot \begin{pmatrix} \beta_1^s \\ \beta_2^s \\ \beta_3^s \end{pmatrix}. \tag{D.15}$$

$$\begin{pmatrix} \mathbb{E}\left(\dfrac{\partial\sigma_1(x_i)}{\partial m_i}\right) \\ \mathbb{E}\left(\dfrac{\partial\sigma_2(x_i)}{\partial m_i}\right) \\ \mathbb{E}\left(\dfrac{\partial\sigma_3(x_i)}{\partial m_i}\right) \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left(\dfrac{\partial\sigma_1(x_i)}{\partial w_{1i}}\right) & \mathbb{E}\left(\dfrac{\partial\sigma_1(x_i)}{\partial w_{2i}}\right) & \mathbb{E}\left(\dfrac{\partial\sigma_1(x_i)}{\partial w_{3i}}\right) \\ \mathbb{E}\left(\dfrac{\partial\sigma_2(x_i)}{\partial w_{1i}}\right) & \mathbb{E}\left(\dfrac{\partial\sigma_2(x_i)}{\partial w_{2i}}\right) & \mathbb{E}\left(\dfrac{\partial\sigma_2(x_i)}{\partial w_{3i}}\right) \\ \mathbb{E}\left(\dfrac{\partial\sigma_3(x_i)}{\partial w_{1i}}\right) & \mathbb{E}\left(\dfrac{\partial\sigma_3(x_i)}{\partial w_{2i}}\right) & \mathbb{E}\left(\dfrac{\partial\sigma_3(x_i)}{\partial w_{3i}}\right) \end{pmatrix} \cdot \begin{pmatrix} \gamma_1^m \\ \gamma_2^m \\ \gamma_3^m \end{pmatrix}. \tag{D.16}$$

The derivatives with respect to z_i lead to:

$$\begin{pmatrix} \mathbb{E}\left(\frac{\partial \sigma_1(x_i)}{\partial z_i}\right) \\ \mathbb{E}\left(\frac{\partial \sigma_2(x_i)}{\partial z_i}\right) \\ \mathbb{E}\left(\frac{\partial \sigma_3(x_i)}{\partial z_i}\right) \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left(\sum_{c=1}^{3} \frac{\partial \sigma_1(x_i)}{\partial w_{ci}}\right) & -\mathbb{E}\left(\sum_{c=1}^{3} \frac{\partial \sigma_1(x_i)}{\partial y_{ic}}\right) \\ \mathbb{E}\left(\sum_{c=1}^{3} \frac{\partial \sigma_2(x_i)}{\partial w_{ci}}\right) & -\mathbb{E}\left(\sum_{c=1}^{3} \frac{\partial \sigma_2(x_i)}{\partial y_{ic}}\right) \\ \mathbb{E}\left(\sum_{c=1}^{3} \frac{\partial \sigma_3(x_i)}{\partial w_{ci}}\right) & -\mathbb{E}\left(\sum_{c=1}^{3} \frac{\partial \sigma_3(x_i)}{\partial y_{ic}}\right) \end{pmatrix} \cdot \begin{pmatrix} \gamma^z \\ \beta^z \end{pmatrix}. \tag{D.17}$$

We now have 3 equations in 2 unknowns specified by equation (D.17). 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.15)–(D.17).

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.15) and (D.16), we deal with 4-dimensional objects (i.e., $(s_i, y_{i1}, y_{i2}, y_{i3})$ or $(m_i, w_{1i}, w_{2i}, w_{3i})$); in equation system (D.17), it is 7-dimensional (i.e., $(z_i, y_{i1}, y_{i2}, y_{i3}, w_{1i}, w_{2i}, w_{3i})$), 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.13), 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.14), 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.

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

	General: higher dimensionality $\beta^z_c = \beta^z$, $\gamma^z_c = \gamma^z$				Reduced: lower dimensionality $\beta^z_2 = \beta^z_3 = \gamma^z_1 = \gamma^z_2 = 0$		
	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^s_1	0.98	1.11	0.50	β^s_1	0.98	1.21	0.98
β^s_2	1.00	1.11	0.52	β^s_2	0.91	1.16	1.06
β^s_3	0.99	1.12	0.50	β^s_3	1.01	1.18	0.82
B. Coefficients on m in college preferences (true value = 1)							
γ^m_1	1.04	1.64	3.21	γ^m_1	1.00	1.02	0.27
γ^m_2	0.94	1.30	3.71	γ^m_2	1.02	1.05	0.30
γ^m_3	1.12	1.47	3.41	γ^m_3	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.17)							
β^z	0.12	0.41	2.54	β^z_1	0.97	0.99	0.16
γ^z	0.16	0.08	3.13	γ^z_3	0.97	1.00	0.21
Using conditions 1 & 2 in equation (D.17)							
β^z	0.05	1.11	10.37	β^z_1	0.97	1.01	0.26
γ^z	0.17	-0.59	5.71	γ^z_3	0.97	1.11	1.30
Using conditions 1 & 3 in equation (D.17)							
β^z	0.30	0.08	25.20	β^z_1	0.97	1.00	0.15
γ^z	0.19	8.37	92.48	γ^z_3	0.98	1.00	0.20
Using conditions 2 & 3 in equation (D.17)							
β^z	-0.06	0.84	7.34	β^z_1	0.99	1.03	0.35
γ^z	0.08	0.30	6.35	γ^z_3	0.95	1.03	0.30

Notes: This table presents estimates for the coefficients in student or college utility functions (equations D.13 and D.14). The statistics are calculated using 150 MC samples. In the general model, we assume that $\beta^z_c = \beta^z$ (i.e., we have prior knowledge that β^z_c is constant across colleges) and $\gamma^z_c = \gamma^z$. The estimation is based on equation systems (D.15), (D.16), and (D.17). In the reduced model, we assume that we know $\beta^z_2 = \beta^z_3 = 0$ (i.e., z_i does not enter i's utility for college 2 or 3) and $\gamma^z_1 = \gamma^z_2 = 0$ (i.e., colleges 1 and 2 do not use z_i to evaluate students). The estimation is based on equation systems (D.15), (D.16), and (D.18).

We now have a simplified version of equation (D.17) with a reduced dimension:

$$\begin{pmatrix} \mathbb{E}\left(\dfrac{\partial\sigma_1(x_i)}{\partial z_i}\right) \\ \mathbb{E}\left(\dfrac{\partial\sigma_2(x_i)}{\partial z_i}\right) \\ \mathbb{E}\left(\dfrac{\partial\sigma_3(x_i)}{\partial z_i}\right) \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left(\dfrac{\partial\sigma_1(x_i)}{\partial w_{3i}}\right) & -\mathbb{E}\left(\dfrac{\partial\sigma_1(x_i)}{\partial y_{i1}}\right) \\ \mathbb{E}\left(\dfrac{\partial\sigma_2(x_i)}{\partial w_{3i}}\right) & -\mathbb{E}\left(\dfrac{\partial\sigma_2(x_i)}{\partial y_{i1}}\right) \\ \mathbb{E}\left(\dfrac{\partial\sigma_3(x_i)}{\partial w_{3i}}\right) & -\mathbb{E}\left(\dfrac{\partial\sigma_3(x_i)}{\partial y_{i1}}\right) \end{pmatrix} \cdot \begin{pmatrix} \gamma^z_3 \\ \beta^z_1 \end{pmatrix}. \tag{D.18}$$

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.

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.13) and (D.14) 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 ζ_ϵ. The other parameters to be estimated are β_c^y, 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 $(\beta, \gamma, \zeta_\epsilon)$.

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

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

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 $(\overline{\zeta}_\epsilon^2, \nu_\epsilon) = (1, 2)$.

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 $N(\gamma_c^w w_{ci} + \gamma_c^m m_i + \gamma_c^z z_i, 1)$ 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 $N(\gamma_c^w w_{ci} + \gamma_c^m m_i + \gamma_c^z z_i, 1)$ 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 $N(\gamma_c^w w_{ci} + \gamma_c^m m_i + \gamma_c^z z_i, 1)$ (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 mean $\beta_c^y y_{ic} + \beta_c^s s_i + \beta_c^z z_i$ and variance 1 if $c \neq 3$ or ζ_ϵ^2 if $c = 3$.

(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 mean $\beta_c^y y_{ic} + \beta_c^s s_i + \beta_c^z z_i$ and variance 1 if $c \neq 3$ or ζ_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 mean $\beta_c^y y_{ic} + \beta_c^s s_i + \beta_c^z z_i$ and variance 1 if $c \neq 3$ or ζ_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 ζ_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, 92.04% of the PSRFs are below 1.1, while only 0.46% 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 as the posterior means are close to the true values. Moreover, we conclude that the posterior standard deviation is a reasonable measure of estimation precision. Comparing column (3), which represents the estimation precision, with column (4), which is the median of the posterior standard deviations, we

[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.

find that they are close to each other, although some of the values in column (4) tend to be smaller. Reassuringly, no value in column (3) is larger than the corresponding one in column (7), which is the 95th percentile among the 150 posterior standard deviations for each coefficient.

Table D.2: Results from Bayesian Estimation

	Posterior mean			Posterior Std. Dev.			
	Median	Mean	Std. Dev.	Median	Mean	5th Perc.	95th Perc.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
True value = 1							
β_1^s	1.02	1.02	0.10	0.07	0.07	0.06	0.09
β_2^s	1.03	1.02	0.11	0.07	0.07	0.06	0.09
β_3^s	1.03	1.03	0.10	0.07	0.07	0.06	0.09
γ_1^m	1.07	1.10	0.19	0.13	0.14	0.10	0.22
γ_2^m	1.05	1.06	0.35	0.13	0.15	0.10	0.22
γ_3^m	1.07	1.10	0.14	0.14	0.14	0.10	0.21
β_1^z	1.02	1.02	0.10	0.07	0.07	0.06	0.09
β_2^z	1.02	1.02	0.10	0.07	0.07	0.06	0.09
β_3^z	1.02	1.02	0.10	0.07	0.07	0.06	0.09
γ_1^z	1.07	1.10	0.19	0.14	0.14	0.10	0.22
γ_2^z	1.03	1.02	0.69	0.13	0.14	0.10	0.22
γ_3^z	1.07	1.10	0.15	0.14	0.14	0.10	0.22
Coefficients on y (true value $= -1$)							
β_1^y	-1.02	-1.03	0.09	0.07	0.07	0.06	0.09
β_2^y	-1.02	-1.03	0.10	0.07	0.07	0.06	0.08
β_3^y	-1.02	-1.02	0.10	0.07	0.07	0.06	0.09
Coefficients on w (true value = 1)							
γ_1^w	1.06	1.10	0.18	0.13	0.14	0.10	0.22
γ_2^w	1.05	1.03	0.68	0.13	0.16	0.10	0.22
γ_3^w	1.08	1.10	0.15	0.14	0.14	0.10	0.21
Std. dev. of student utility shock (ϵ_{i3})							
ζ_ϵ	1.05	1.04	0.19	0.17	0.17	0.15	0.19

Notes: This table presents statistics on the posterior means and standard deviations of the coefficients in student and college utility functions (equations D.13 and D.14). For each coefficient, there are 150 posterior means and 150 posterior standard deviations 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 and standard deviations 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, `https://datosabiertos.mineduc.cl` (last accessed on December 07, 2023).

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 number of years the teachers have had in their teaching career 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 as 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

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

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. As public schools cannot select students, their capacity is irrelevant.

Table E.3 and Table E.4 summarize the student characteristics and school attributes, respectively.

Table E.3: Summary Statistics of Student Characteristics

			Students enrolled in a secondary school of type							
	All students (N=9,314)		Public (N=3,911)		Private subsidized (N=4,048)		Private non-subsidized (N=1,211)		Outside Option (N=144)	
	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	0.49	0.50
Language score	0.50	0.29	0.36	0.26	0.55	0.27	0.75	0.23	0.46	0.27
Math score	0.49	0.29	0.34	0.24	0.56	0.26	0.78	0.20	0.39	0.25
Composite score	0.49	0.29	0.34	0.24	0.56	0.26	0.78	0.20	0.42	0.26
Mother's education (years)	13.97	3.19	12.43	2.78	14.33	2.81	17.78	1.86	13.54	2.82
Parental income (CLP)	430,336	493,814	194,861	147,491	358,906	284,207	1,447,069	541,758	283,333	282,595
Distance to enrolled school (km)	2.71	2.50	2.21	1.96	2.93	2.60	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.

Table E.4: Summary Statistics of School Attributes

			All private schools				Full capacity private schools			
	Public schools ($C = 20$)		subsidized ($C = 64$)		non-subsidized ($C = 33$)		subsidized ($C = 27$)		non-subsidized ($C = 6$)	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Average language score	0.32	0.13	0.54	0.15	0.70	0.14	0.58	0.16	0.65	0.20
Average math score	0.29	0.15	0.55	0.16	0.73	0.15	0.58	0.17	0.66	0.22
Average composite score	0.29	0.15	0.55	0.17	0.73	0.15	0.59	0.17	0.67	0.22
Average mother's edu. (years)	12.01	0.91	14.73	1.34	17.39	0.81	15.06	1.29	16.90	0.99
Fraction of female students	0.53	0.30	0.49	0.21	0.49	0.22	0.53	0.23	0.47	0.08
Median parental income (CLP)	155,000	22,361	335,156	149,783	1,284,848	475,573	353,704	166,944	950,000	440,454
Teacher experience (years)	17.89	6.41	13.68	7.81	18.07	8.90	13.79	8.84	13.85	9.80
Tuition (CLP)	4,034	1,750	17,899	10,548	57,780	8,117	19,858	9,711	55,673	11,351
Capacity	-	-	73.94	71.30	48.91	29.06	56.04	34.86	35.67	33.07
Valparaiso student enrollment[a]	195.55	124.80	63.25	58.97	36.70	26.82	53.59	31.29	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.

Note that for the school attributes in Table E.4, the following four variables are measured among the 2006 10th graders who are already in a secondary school in 2007: median parental income among students (in logarithm), fraction of female students, average composite score, and average mother's education.

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 375 students who are not from Market Valparaiso but attend a private school in Market Valparaiso. Among them, 75 students attend a private school with binding capacity constraint. When updating the Markov Chain, these 75 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).

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 but two parameters and below 1.2 for all 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 2 and the observables of each student and each school, randomly draw the utility shocks in equations (14) and (15) 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 1.36% of the students. In contrast, our model correctly predicts for 5.37% of the students, 3.95 times the rate from the random prediction.[F.5] Moreover, the model correctly predicts the type of their matched school for 56.48% of the students, 1.63 times the rate from the random prediction (34.60%).

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 prediction, the model prediction leads to RMSEs that are 45–72% 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 2), 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 33–45% 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]

[F.5]This seemingly low number is understandable: the matching market resembles a discrete choice with 117 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 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 2) and are only weakly correlated with other school attributes. Specifically, a school's fraction of females

Low-income versus Non-low-income Students Our counterfactual policy prioritizes 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.5 shows summary statistics of the students by their income status.

Table F.5: Summary Statistics of Student Characteristics by Income Status

	Low Income (N=4,002)		Non-low Income (N=5,312)	
	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	0.38	0.26	0.59	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)	133,633	37,002	653,869	557,009
Distance to the enrolled school (km)	2.59	2.24	2.80	2.67

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.

References

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

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

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

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

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.

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.

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

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

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