How do Parents Value Schools?
School Valuation in Strategic School Choice Mechanisms

Alessandra Guerrero
December 16, 2019

Abstract

This paper provides an econometric model with which to analyze the individual household valuation of schools, dependent on the admission probabilities of students (a function of household characteristics and household's ranking of the schools) and school characteristics. I observe an aggregated variable of school characteristics, an aggregated variable of household characteristics, household's ranking of the schools, as well as the eventual admission decision made by the schools based on the admission probabilities. I are interested in learning how the household's valuation of school characteristics changes when school characteristics change. I first estimate the admission probability using maximum likelihood estimation, and then estimate the latent school valuation parameter that I are interested in using multinomial logistic regression and maximum likelihood estimation. I find that the estimates for admission probabilities are close to the true values, but the estimate of the school valuation parameter (a function of probabilities and school characteristics) is downward biased.

I. Introduction

Existing work on school choice mechanisms seek to model the choice behaviors of parents in choosing from a selection of schools to which they send their children, utilizing an array of microeconometric methods with which to derive the optimal choice ranking [Beuermann et al., 2018], the effects of parent heterogeneity on school choice [Hastings et al., 2008], and the potential measured effects of changes in assignment mechanisms. School choice literature widely implements multinomial logistic regression and maximum likelihood estimation to derive the parameter estimates for certain household and school factors on school choice [Nguyen and Taylor, 2003]. I deviate from this traditional estimation of parameters of factors, to the estimation of the household valuation of schools in a strategic choice mechanism.

I seek to recover the valuation of different schools in a school choice mechanism, where parents (households) choose rankings from a set of schools. These households make strategic ranking decisions based on their admission probabilities and observed school characteristics.[1] The probability of acceptance is dictated by an algorithm utilized by the schools in the mechanism, which takes into account the household's ranking of the school and the household's distance to the school. I observe a set of participating schools in the mechanism, household distances to each school, household ranking of schools, student admission outcomes per household, and school-specific characteristics.

The households maximize the latent utility of their children attending a school given the admission probability at that school and school-specific characteristics, over the household-specific ranking of the schools. I derive the expected admissions probability at each school through the implementation of maximum likelihood estimation, given household characteristics. After recovering these probabilities, I implement an multinomial logistic regression and maximum likelihood estimation to recover the valuation of each school, given school-specific characteristics and the estimated admission probabilities.

[1] I follow the example of [Beuermann et al., 2018] in utilizing admission probabilities in the set-up of my model, although I utilize a different estimation method to include the probabilities and derive my estimation of valuation.

II. Formalization of Setup

I assume households have three schools to choose from in which to enroll their children. Households are indexed by $i \in \{1, 2, ...N\}$ and schools by $j \in J = \{A, B, C\}$. To be considered as applicants to a school the household must rank all schools, thus the available options to the parents are all possible rankings (permutations) of schools, each denoted by $r = (r_1, r_2, r_3)$, as a specified permutation A, B, C.

I assume that a portion of the utility of attending a certain school j is the same across households, and is identified up to a set of parameters. I assume the utility of attending a school is linear in parameters such that

$$U_j = U(Z_j) = \alpha + \beta Z_j \tag{1}$$

where Z_i is a vector of observable school characteristics. I denote the ranking of school j by household i as r_i^j ($r_i^j \in \{1, 2, 3\}$), where the three schools receive three separate ranks. The probability of acceptance of household i's child into school j, $P_{i,j}^r$, is given by $P_{i,j}^r = g(X_{i,j}, r_i^j)$ for some continuous g, where $X_{i,j}$ is a vector of observable household characteristics. I parameterize g as a linear function of $X_{i,j}$ when r_i^j is fixed:[2]

$$\begin{aligned} g(X_{i,j}, r_i^j) = \theta_{r_i^j} + \eta_{r_i^j} X_{i,j} &= \sum_{R=1}^{3} \mathbb{1}\{r_i^j = R\} \cdot (\theta_R + \eta_R X_{i,j}) &(2)\\ &= \sum_{R=1}^{3} \mathbb{1}\{r_i^j = R\} \cdot \theta_R + \left(\sum_{R=1}^{3} \mathbb{1}\{r_i^j = R\} \cdot \eta_R \right) X_{i,j} &(3) \end{aligned}$$

Each household maximizes utility over all possible rankings, given the characteristics of the schools, the probability of acceptance given each ranking, and unobserved utility specific to the household, i.e.,

$$\max_r \left(U_i | r, \{Z_j\}_{j \in J}, \{P_{i,j}^r\}_{j \in J}, \varepsilon_{i,r} \right) = \max_r U_{i,r} \quad \forall i \tag{4}$$

where

$$U_{i,r} = \sum_{j \in J} P_{i,j}^r U_j + \varepsilon_{i,r} \tag{5}$$

Then, the optimal ranking will be a function of household characteristics, school characteristics, and unobserved utility such that, for some function h,

$$r_i^* = h(\{X_{i,j}\}_{j \in J}, \{Z_j\}_{j \in J}, \varepsilon_{i,r}) \tag{6}$$

For simplicity, Z_j and $X_{i,j}$ would be variables of dimension one, representing an aggregate school statistic and the distance betIen household i and school j respectively. I also assume $\varepsilon_{i,r} \sim$ T1EV(0) as in a standard multinomial logistic model.[3]

[2] I assume a uniform distribution of $X_{i,j}$. The values of θ and η will make sure P_j^r falls within [0,1].

[3] I will follow the example of all multinomial logistic models in dictating that the errors are distributed along the Type One Extreme Value (i.e. Extreme Value or Gumbell) distribution with the location parameter set at 0 as is custom.

III. Theoretical Solution

We can calculate the probability of households choosing the observed rankings $\{r_i\}_{i=1}^{N}$, and recover the parameter of interest, β, by maximizing the log likelihood of the standard multinomial logit.

i. Recovering the Acceptance Probabilities

First I recover $\{\theta_R, \eta_R\}_{R=1}^{3}$ by ML estimation. Denote the observed outcome of student i by school j as $a_{i,j}$ ($a_{i,j} = 1$ if accepted, 0 if rejected). For each $i \in \{1, ..., N\}$ and $j \in \{A, B, C\}$, I have[4]

$$\mathbb{P}\left(a_{i,j}|r_i^j, X_{i,j}, \{\theta_R, \eta_R\}_{R=1}^{3}\right) = (\theta_{r_i^j} + \eta_{r_i^j}, X_{i,j})^{a_{i,j}} \cdot \Big(1 - (\theta_{r_i^j} + \eta_{r_i^j}, X_{i,j})\Big)^{1-a_{i,j}} \tag{7}$$

I assume that for each individual i, $X_{i,j}$'s are i.i.d. Then $a_{i,j}$'s are also independent across j, and I have the following individual likelihood function:

$$\begin{aligned}\ell\left(\{\theta_R, \eta_R\}_{R=1}^{3}|\{a_{i,j}, r_i^j, X_{i,j}\}_{j\in\{A,B,C\}}\right) &= \mathbb{P}\left(\{a_{i,j}\}_{j\in\{A,B,C\}}|\{r_i^j, X_{i,j}\}_{j\in\{A,B,C\}}, \{\theta_R, \eta_R\}_{R=1}^{3}\right)\\ &= \prod_{j\in\{A,B,C\}} (\theta_{r_i^j} + \eta_{r_i^j}, X_{i,j})^{a_{i,j}} \cdot \Big(1 - (\theta_{r_i^j} + \eta_{r_i^j}, X_{i,j})\Big)^{1-a_{i,j}}\end{aligned}$$

I also assume $X_{i,j}$ are i.i.d. across i, then I have the following sample likelihood function

$$\begin{aligned}L\left(\{\theta_R, \eta_R\}_{R=1}^{3}|\{a_{i,j}, r_i^j, X_{i,j}\}_{\substack{i\in\{1,...,N\}\\ j\in\{A,B,C\}}}\right) &= \prod_{i=1}^{N} \ell\left(\{\theta_R, \eta_R\}_{R=1}^{3}|\{a_{i,j}, r_i^j, X_{i,j}\}_{j\in\{A,B,C\}}\right)\\ &= \prod_{\substack{i\in\{1,...,N\}\\ j\in\{A,B,C\}}} (\theta_{r_i^j} + \eta_{r_i^j}, X_{i,j})^{a_{i,j}} \cdot \Big(1 - (\theta_{r_i^j} + \eta_{r_i^j}, X_{i,j})\Big)^{1-a_{i,j}}\end{aligned}$$

which gives us the log-likelihood function

$$\log L(\{\theta_R,\eta_R\}_{R=1}^{3}) = \sum_{\substack{i\in\{1,...,N\}\\ j\in\{A,B,C\}}} \Big(a_{i,j}\log(\theta_{r_i^j} + \eta_{r_i^j} X_{i,j}) + (1 - a_{i,j})\log(1 - (\theta_{r_i^j} + \eta_{r_i^j} X_{i,j}))\Big)$$

Theoretically, I can take the partial derivative of each parameter, and solve for the first order conditions. However, such derivation is out of the scope of the paper, and I utilized a numerical approach to find the maximum log likelihood, as demonstrated in the simulation section.

ii. Recovering the Valuation of Schools

Second, I want to recover the parental valuation of each school, denoted by the vectors of α and β. For any given rank $r = r_i$, I assume $\varepsilon_{i,r} \sim \mathrm{T1EV}(0)$ and are i.i.d. I define

$$\delta_{i,r} = \sum_{j\in J} P_{i,j}^{r} U_J = \alpha \sum_{j\in J} P_{i,j}^{r} + \beta \sum_{j\in J} P_{i,j}^{r} Z_j \tag{8}$$

[4] See Appendix ii.1 for additional notes on recovering the admission probabilities.

and further denote

$$Y_{1,i}^r = \sum_{j\in J} P_{i,j}^r \quad , \quad Y_{2,i}^r = \sum_{j\in J} P_{i,j}^r Z_j$$

such that $\delta_{i,r} = \alpha Y_{1,i}^r + \beta Y_{2,i}^r$ and $U_{i,r} = \delta_{i,r} + \varepsilon_{i,r}$.[5]

I denote all possible permutations of rankings R, and for each $r \in R$, I denote the set of all alternatives as $R\backslash\{r\}$. Households solve the maximization problem

$$\max_r \sum_{j\in J} P_{i,j}^r U_j + \varepsilon_{i,r}$$

and I have $U_{i,-r} = \max_{r'\in R\backslash\{r\}} U_{i,r'} \sim \text{T1EV}(\log \sum_{r'\in R\backslash\{r\}} \exp(\delta_{i,r'}))$. Given my preference rankings, I have

$$\begin{aligned}
\Pr[U_{i,r} \geq U_{i,-r}] &= \Pr[U_{i,-r} - U_{i,r} \leq 0] \\
&= \frac{\exp(\delta_{i,r})}{\sum_{r'\in R} \exp(\delta_{i,r'})} \\
&= \frac{\exp(\alpha \sum_{j\in J} P_{i,j}^r + \beta \sum_{j\in J} P_{i,j}^r Z_j)}{\sum_{r'\in R} \exp(\alpha \sum_{j\in J} P_{i,j}^{r'} + \beta \sum_{j\in J} P_{i,j}^{r'} Z_j)}
\end{aligned} \tag{9}$$

Thus, I have the following likelihood function

$$\begin{aligned}
L(\alpha, \beta | \left\{r, \{Y_{1,i}^r, Y_{2,i}^r\}_{r\in R}\right\}_{i=1}^N) &= \prod_{i=1}^N \left\{ \prod_{r\in R} \Pr[i \to r | \alpha, \beta]^{\mathbb{1}(i\to j)} \right\} \\
&= \prod_{\substack{i\in\{1,\ldots,N\} \\ r\in R}} \left[\frac{\exp(\alpha \sum_{j\in J} P_{i,j}^r + \beta \sum_{j\in J} P_{i,j}^r Z_j)}{\sum_{r'\in R} \exp(\alpha \sum_{j\in J} P_{i,j}^{r'} + \beta \sum_{j\in J} P_{i,j}^{r'} Z_j)} \right]^{\mathbb{1}(i\to j)}
\end{aligned}$$

Which gives us the log-likelihood function

$$\begin{aligned}
\log L(\alpha, \beta) &= \sum_{\substack{i\in\{1,\ldots,N\} \\ r\in R}} \mathbb{1}(i \to r) \left[(\alpha Y_{1,i}^r + \beta Y_{2,i}^r) - \log \sum_{r'\in R} \exp(\alpha Y_{1,i}^{r'} + \beta Y_{2,i}^{r'}) \right] \\
&= \sum_{\substack{i\in\{1,\ldots,N\} \\ r\in R}} \mathbb{1}(i \to r) \left[\alpha \sum_{j\in J} P_{i,j}^r + \beta \sum_{j\in J} P_{i,j}^r Z_j - \right. \\
&\qquad \left. \log \left(\sum_{r'\in R} \exp(\alpha \sum_{j\in J} P_{i,j}^{r'} + \beta \sum_{j\in J} P_{i,j}^{r'} Z_j) \right) \right]
\end{aligned} \tag{10}$$

To find $\hat{\alpha}_{\mathbf{MLE}}$ and $\hat{\beta}_{\mathbf{MLE}}$, I would need to find the partial derivatives of the log-likelihood with respect to both α and β, and solve for the first order conditions.[6] However, as in the previous part, actually solving the conditions theoretically would be beyond the scope of this paper, and we maximize the log likelihood function numerically in the simulation.

[5] Following the example of multinomial logistic model notation as done in lecture on October 31, 2019.

[6] See Appendix ii.2 for the full derivations of the MNL model and ML estimation. See Appendix iii for a discussion about bias and standard errors.

IV. Simulation

i. Initiation

I simulated 5,000 households, with $X_{i,j} \sim \text{Unif}(0, 100)$, representing the distance between each household and each school. I initialized $\theta = (0.9, 0.7, 0.5)$ and $\eta = (-0.009, -0.007, -0.005)$, which means

$$P^r_{i,j} = 0.009 \cdot (100 - X_{i,j}) \quad \text{if } r_i^j = 1, \tag{11}$$
$$P^r_{i,j} = 0.007 \cdot (100 - X_{i,j}) \quad \text{if } r_i^j = 2, \tag{12}$$
$$P^r_{i,j} = 0.005 \cdot (100 - X_{i,j}) \quad \text{if } r_i^j = 3. \tag{13}$$

I drew three values from $N(80, 100)$ as Z_j. α and β are set to 200 and 1 respectively, which gives us

$$U_j = 200 + Z_j \tag{14}$$

Afterwards, aggregated utility was calculated as specified in the model. I can observe the households' distance from the the schools ($X_{i,j}$), the ranking of schools by households (r_i), and the acceptance decisions made by each school according to the distance and ranking ($a_{i,j}$).

ii. Estimating θ and η

I created a function to calculate the log likelihood given θ and η, as detailed in Section IIIi. For each parameter in θ and η, I evaluated the log-likelihood function on a grid of different values. See results in Figure 1. The red lines represent the true values, and the blue ones represent the value which gives the largest log-likelihood on the grid. If there is only one line on the graph, the true value and the estimated overlap. Since $\theta_{r_i^j} + \eta_{r_i^j} X_{i,j}$ must be bounded between 0 and 1, when other parameters are fixed, I tend to get corner solutions for the remaining parameter.

I also maximized the log-likelihood function using the *optim* function, and obtained the following estimates:

$$\widehat{\theta} = (0.903, 0.696, 0.480), \quad \widehat{\eta} = (-0.0092, -0.0070, -0.0048) \tag{15}$$

which are close to the true parameter values.

iii. Estimating α and β

First, I recovered the probabilities by constructing

$$\widehat{P^r_{i,j}} = \min(\max(\widehat{\theta_{r_i^j}} + \widehat{\eta_{r_i^j}} X_{i,j}, 0), 1) \tag{16}$$

Then, for each possible r in all six possible permutations of schools, I constructed

$$\widehat{Y^r_{1,i}} = \sum_{j \in J} \widehat{P^r_{i,j}}, \quad \widehat{Y^r_{2,i}} = \sum_{j \in J} \widehat{P^r_{i,j}} Z_j \tag{17}$$

Using the constructed $\widehat{Y^r_{1,i}}$ and $\widehat{Y^r_{2,i}}$, I created a function to calculate the log-likelihood function for a given pair of α and β as detailed in Section IIIii. I evaluated the log-likelihood function on a grid of different values for α and β, holding the other parameter constant. See results in Figure 2. The red lines represent the true values, and the blue ones represent the value which gives the largest log likelihood on the grid. I can see that the estimate for α is close to the true value, while the estimate for β seems to have a negative bias. I also evaluated the log-likelihood function on a grid of combinations of α and β. See Figure 3 for results. The blue dot, which has coordinates (189, 0.7), represents the pair of α and β which gives the maximized-log likelihood.

I also maximized the log-likelihood using $optim$, and obtained the following estimates:

$$\widehat{\alpha} = 189.0, \quad \widehat{\beta} = 0.679$$

which are consistent with the estimates I recovered from evaluating the grid, but are both smaller than the true values. I have tried different initial values of α and β and observed similar patterns of downward bias. The standard error estimated from the Hessian is 0.0016. I note the properties of the maximum likelihood estimator $\widehat{\beta}$ in Appendix iii.

iv. Interpretation

In my model set-up, I have a set of the single valuation parameters α and β, where observable school characteristics Z_j are evaluated in the same way across households. I consider an aggregate variable of school characteristics Z_j and household characteristics variable X_i such that I only have a single characteristic of households and a single characteristic of schools. my school valuation estimator $\hat{\beta}_{\textbf{MLE}}$ tells us how the value of the schools to households changes depending on a change in the aggregate measure of observable school characteristics. Possible extensions of the model include having more school characteristics variables and estimating how each of them would affect household's valuation of the schools, or having unobserved heterogeneity across households' valuations, which I further discuss in the Appendix.

v. Monte Carlo Simulation

In order to learn more about the properties of $\widehat{\beta}$, I ran a Monte Carlo simulation, where the generation of data and the estimation process Ire repeated 500 times. See Figure 4 for the distribution of $\widehat{\beta}$. The blue line indicates the position of the mean ($mean(\widehat{\beta}) = 0.578$). The standard deviation is 0.464. Therefore I have an empirical estimate for the bias and standard error of $\widehat{\beta}$:

$$\widehat{Bias} = mean(\widehat{\beta}) - \beta = 0.578 - 1 = -0.421$$
$$\widehat{se} = \frac{sd(\widehat{\beta})}{\sqrt{N}} = \frac{0.464}{\sqrt{5000}} = 0.0066$$

References

[Allison and Christakis, 1994] Allison, P. D. and Christakis, N. A. (1994). Logit Models for Sets of Ranked Items. *Sociological Methodology*, 24:199–228.

[Beuermann et al., 2018] Beuermann, D., Jackson, C. K., Navarro-Sola, L., and Pardo, F. (2018). What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output. Working Paper 25342, National Bureau of Economic Research.

[Chapman and Staelin, 1982] Chapman, R. G. and Staelin, R. (1982). Exploiting Rank Ordered Choice Set Data within the Stochastic Utility Model. *Journal of Marketing Research*, 19(3):288–301.

[Fok et al., 2012] Fok, D., Paap, R., and Van Dijk, B. (2012). A Rank-Ordered Logit Model with Unobserved Heterogeneity in Ranking Capabilities. *Journal of Applied Econometrics*, 27(5):831–846.

[Hastings et al., 2008] Hastings, J. S., Kane, T. J., and Staiger, D. O. (2008). Heterogeneous Preferences and the Efficacy of Public School Choice.

[Hausman and Ruud, 1987] Hausman, J. A. and Ruud, P. A. (1987). Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics*, 34(1-2):83–104.

[Nguyen and Taylor, 2003] Nguyen, A. N. and Taylor, J. (2003). Post-High School Choices: New Evidence from a Multinomial Logit Model. *Journal of Population Economics*, 16(2):287–306.

Appendix

i. Figures

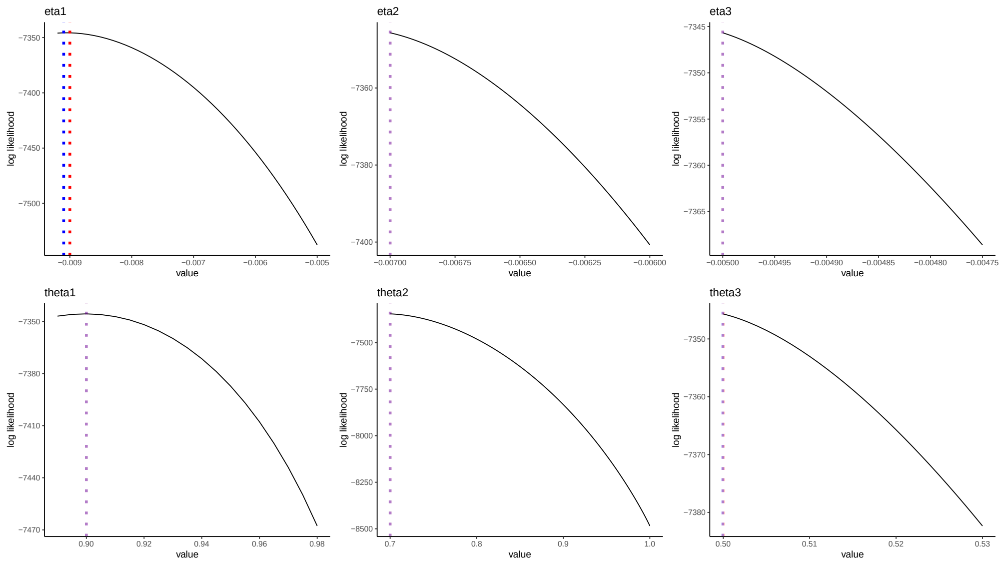

Figure 1: *Log-likelihood on a grid of each parameter in θ and η when holding other parameters at true value*

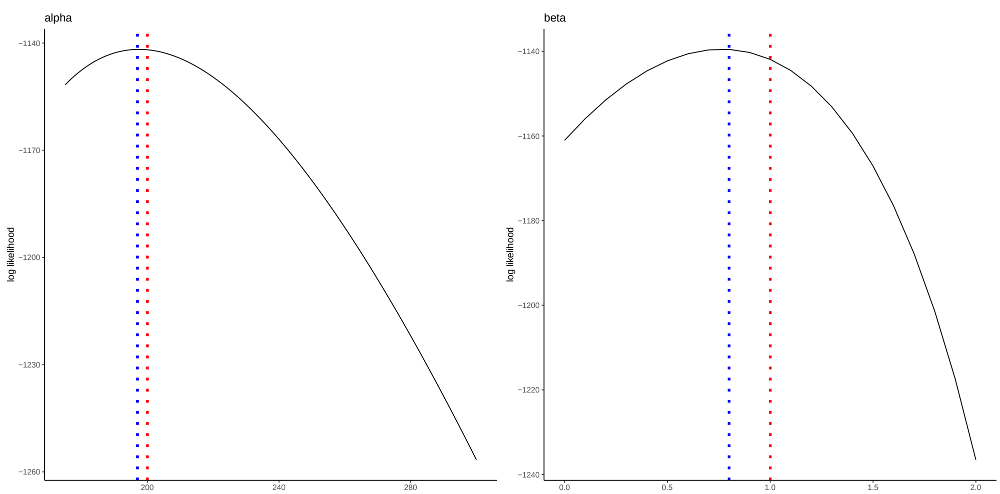

Figure 2: *Log-likelihood on a grid of α and β when holding other parameters at true value*

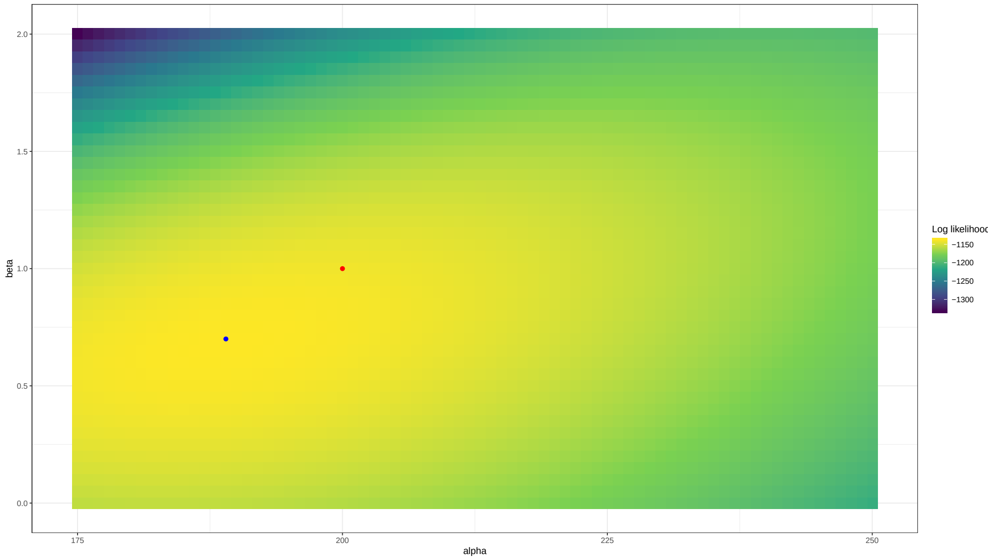

Figure 3: *Heat map of log-likelihood on a grid of α and β*

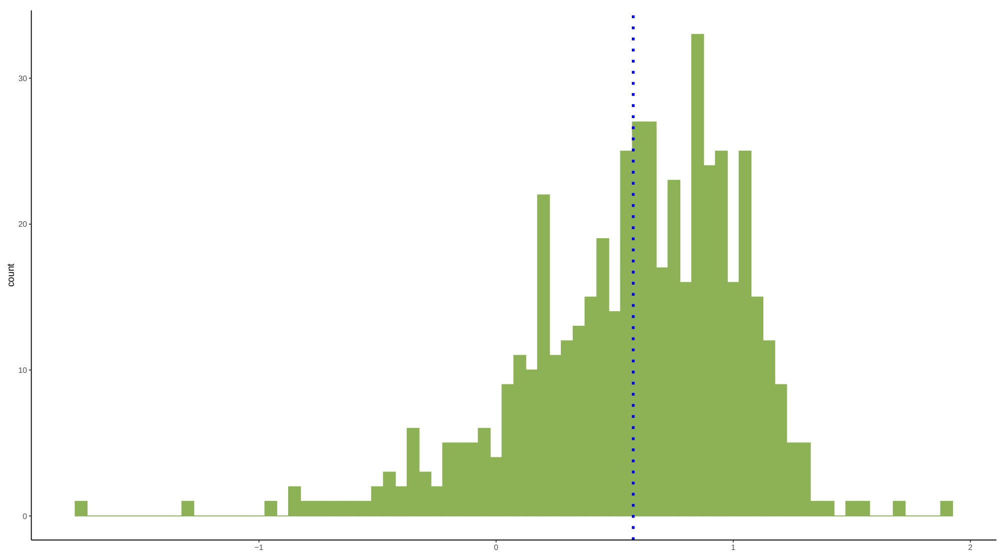

Figure 4: *Distribution of $\widehat{\beta}$ from Monte Carlo*

ii. Derivations

ii.1 Recovering the Acceptance Probabilities

By construction, for each household i and school j, I have

$$\begin{aligned}\mathbb{P}(a_{i,j}=1|r_i^j, X_{i,j}, \{\theta_R, \eta_R\}_{R=1}^3) &= p_{i,j}^r = \theta_{r_i^j} + \eta_{r_i^j} X_{i,j} \\ &= (\theta_{r_i^j} + \eta_{r_i^j} X_{i,j})^1 \cdot \Big(1-(\theta_{r_i^j} + \eta_{r_i^j} X_{i,j})\Big)^0 \\ \mathbb{P}(a_{i,j}=0|r_i^j, X_{i,j}, \{\theta_R, \eta_R\}_{R=1}^3) &= 1 - p_{i,j}^r = 1-(\theta_{r_i^j} + \eta_{r_i^j} X_{i,j}) \\ &= (\theta_{r_i^j} + \eta_{r_i^j} X_{i,j})^0 \cdot \Big(1-(\theta_{r_i^j} + \eta_{r_i^j} X_{i,j})\Big)^1\end{aligned}$$

Therefore, I can more compactly write

$$\mathbb{P}(a_{i,j}|r_i^j, X_{i,j}, \{\theta_R, \eta_R\}_{R=1}^3) = (\theta_{r_i^j} + \eta_{r_i^j} X_{i,j})^{a_{i,j}} \cdot \Big(1-(\theta_{r_i^j} + \eta_{r_i^j} X_{i,j})\Big)^{1-a_{i,j}}$$

Since the acceptance results are independent across schools for each household, I construct the individual likelihood function

$$\begin{aligned}\mathbb{P}(a_{i,A}, a_{i,B}, a_{i,C}|r_i^j, X_{i,j}, \{\theta_R, \eta_R\}_{R=1}^3) &= \prod_{j\in\{A,B,C\}} \mathbb{P}(a_{i,j}|r_i^j, X_{i,j}, \{\theta_R, \eta_R\}_{R=1}^3) \\ &= \prod_{j\in\{A,B,C\}} (\theta_{r_i^j} + \eta_{r_i^j} X_{i,j})^{a_{i,j}} \cdot \Big(1-(\theta_{r_i^j} + \eta_{r_i^j} X_{i,j})\Big)^{1-a_{i,j}}\end{aligned}$$

and the sample likelihood function follows by multiplying each individual's likelihood function.

ii.2 Recovering the Valuation of Schools

I will let

$$\delta_{i,r} = \sum_{j\in J} P_{i,j}^r U_J = \alpha \sum_{j\in J} P_{i,j}^r + \beta \sum_{j\in J} P_{i,j}^r Z_j$$

$$Y_{1,i}^r = \sum_{j\in J} P_{i,j}^r \quad , \quad Y_{2,i}^r = \sum_{j\in J} P_{i,j}^r Z_j$$

Thus, $\delta_{i,r} = \alpha Y_{1,i}^r + \beta Y_{2,i}^r$ and $U_{i,r} = \delta_{i,r} + \varepsilon_{i,r}$.

And I seek to solve the problem

$$\max_r \sum_{j\in J} P_{i,j}^r U_j + \varepsilon_{i,r}$$

where $U_{i,-r} = \max_{r' \in R \setminus \{r\}} U_{i,r'} \sim \text{T1EV}(\log \sum_{r' \in R \setminus \{r\}} \exp(\delta_{i,r'}))$.

$$\begin{aligned}
\Pr[U_{i,r} \geq U_{i,-r}] &= \Pr[U_{i,-r} - U_{i,r} \leq 0] \\
&= \frac{1}{1 + \exp(-(\delta_{i,r'} - \delta_{i,r}))} \\
&= \frac{\exp(\delta_{i,r})}{\exp(\delta_{i,r}) + \exp(\delta_{i,r'}))} \\
&= \frac{\exp(\delta_{i,r})}{\sum_{r' \in R} \exp(\delta_{i,r'})} \\
&= \frac{\exp(\delta_{i,r})}{\sum_{r' \in R} \exp(\delta_{i,r'})} \\
&= \frac{\exp(\alpha Y_{1,i}^{r} + \beta Y_{2,i}^{r})}{\sum_{r' \in R} \exp(\alpha Y_{1,i}^{r'} + \beta Y_{2,i}^{r'})} \\
&= \frac{\exp(\alpha Y_{1,i}^{r} + \beta \sum_{j \in J} P_{i,j}^{r} Z_j)}{\sum_{r' \in R} \exp(\alpha \sum_{j \in J} P_{i,j}^{r'} + \beta \sum_{j \in J} P_{i,j}^{r'} Z_j)}
\end{aligned}$$

$$\begin{aligned}
L(\alpha, \beta | \left\{r, \{Y_{1,i}^{r}, Y_{2,i}^{r}\}_{r \in R}\right\}_{i=1}^{N}) &= \prod_{i=1}^{N} \left\{ \prod_{r \in R} \Pr[i \to r | \alpha, \beta]^{\mathbb{1}(i \to r)} \right\} \\
&= \prod_{\substack{i \in \{1,\ldots,N\} \\ r \in R}} \left[\frac{\exp(\alpha \sum_{j \in J} P_{i,j}^{r} + \beta \sum_{j \in J} P_{i,j}^{r} Z_j)}{\sum_{r' \in R} \exp(\alpha \sum_{j \in J} P_{i,j}^{r'} + \beta \sum_{j \in J} P_{i,j}^{r'} Z_j)} \right]^{\mathbb{1}(i \to r)}
\end{aligned}$$

$$\begin{aligned}
\log L(\alpha, \beta) &= \sum_{\substack{i \in \{1,\ldots,N\} \\ r \in R}} \mathbb{1}(i \to r) \left[(\alpha Y_{1,i}^{r} + \beta Y_{2,i}^{r}) - \log \sum_{r' \in R} \exp(\alpha Y_{1,i}^{r'} + \beta Y_{2,i}^{r'}) \right] \\
&= \prod_{\substack{i \in \{1,\ldots,N\} \\ r \in R}} \mathbb{1}(i \to r) \log \left[\frac{\exp(\alpha \sum_{j \in J} P_{i,j}^{r} + \beta \sum_{j \in J} P_{i,j}^{r} Z_j)}{\sum_{r' \in R} \exp(\alpha \sum_{j \in J} P_{i,j}^{r'} + \beta \sum_{j \in J} P_{i,j}^{r'} Z_j)} \right] \\
&= \sum_{\substack{i \in \{1,\ldots,N\} \\ r \in R}} \mathbb{1}(i \to r) [\log(\exp(\delta_{i,r}) - \log \left[\sum_{r' \in R} \exp(\delta_{i,r'}) \right] \\
&= \sum_{\substack{i \in \{1,\ldots,N\} \\ r \in R}} \mathbb{1}(i \to r) \left[\delta_{i,r} - \log \left(\sum_{r' \in R} \exp(\delta_{i,r'}) \right) \right]
\end{aligned}$$

$$= \sum_{\substack{i \in \{1,\ldots,N\} \\ r \in R}} \mathbb{1}(i \to r) \left[\alpha Y_{1,i}^{r} + \beta Y_{2,i}^{r} - \log \left(\sum_{r' \in R} \exp(\alpha Y_{1,i}^{r'} + \beta Y_{2,i}^{r'}) \right) \right]$$

$$l(\alpha, \beta) = \sum_{\substack{i \in \{1,\ldots,N\} \\ r \in R}} \mathbb{1}(i \to r) \left[\alpha \sum_{j \in J} P_{i,j}^{r} + \beta \sum_{j \in J} P_{i,j}^{r} Z_j - \right.$$
$$\left. \log \left(\sum_{r' \in R} \exp(\alpha \sum_{j \in J} P_{i,j}^{r'} + \beta \sum_{j \in J} P_{i,j}^{r'} Z_j) \right) \right]$$

So I find the ML estimators for α and β where the log-likelihood is maximized

$$\hat{\alpha}_{\textbf{MLE}}, \hat{\beta}_{\textbf{MLE}} = \arg\max_{\alpha, \beta} \log L(\alpha, \beta)$$
$$= \arg\max_{\alpha, \beta} \sum_{\substack{i \in \{1,\ldots,N\} \\ r \in R}} \mathbb{1}(i \to r) \left[\alpha \sum_{j \in J} P_{i,j}^{r} + \beta \sum_{j \in J} P_{i,j}^{r} Z_j - \right.$$
$$\left. \log \left(\sum_{r' \in R} \exp(\alpha \sum_{j \in J} P_{i,j}^{r'} + \beta \sum_{j \in J} P_{i,j}^{r'} Z_j) \right) \right]$$

Where $s(\alpha) = \frac{\partial}{\partial \alpha} l(\alpha, \beta)$ and $s(\beta) = \frac{\partial}{\partial \beta} l(\alpha, \beta)$, I then set

$$\frac{\partial}{\partial \alpha} \sum_{\substack{i \in \{1,\ldots,N\} \\ r \in R}} \mathbb{1}(i \to r) \left[\alpha \sum_{j \in J} P_{i,j}^{r} + \beta \sum_{j \in J} P_{i,j}^{r} Z_j - \log \left(\sum_{r' \in R} \exp(\alpha \sum_{j \in J} P_{i,j}^{r'} + \beta \sum_{j \in J} P_{i,j}^{r'} Z_j) \right) \right] = 0$$
$$\frac{\partial}{\partial \beta} \sum_{\substack{i \in \{1,\ldots,N\} \\ r \in R}} \mathbb{1}(i \to r) \left[\alpha \sum_{j \in J} P_{i,j}^{r} + \beta \sum_{j \in J} P_{i,j}^{r} Z_j - \log \left(\sum_{r' \in R} \exp(\alpha \sum_{j \in J} P_{i,j}^{r'} + \beta \sum_{j \in J} P_{i,j}^{r'} Z_j) \right) \right] = 0$$

to find my First Order Conditions, giving us $\hat{\alpha}_{\textbf{MLE}}, \hat{\beta}_{\textbf{MLE}}$.

iii. Properties of the Estimators

There are two stages in my estimation, and I do not have enough information on whether and how having the first-stage estimates in the second-stage data would change the properties of my estimators. I discuss the following properties based on my knowledge of ML estimators in general, and for $\widehat{\beta}$ in particular, I performed a Monte Carlo simulation in the simulation section to learn about its properties.

iii.1 Bias

I know that the maximum likelihood estimators are not necessarily unbiased. Due to the complexity of the model, I cannot explicitly derive the theoretical bias. From the simulation results, it seems like I have a small bias (if any) for the parameters I estimated in the first stage, and a negative bias for $\widehat{\alpha}$ and $\widehat{\beta}$. I estimated the bias of $\widehat{\beta}$, the parameter of interest, using Monte Carlo simulation.

iii.2 Consistency

I know that maximum likelihood estimators are always (within distributional regularities) consistent. Therefore, $\widehat{\beta} \to \beta$ in probability as $n \to \infty$. Equivalently, I denote the probability limit of my consistent maximum likelihood estimator as

$$\operatorname*{plim}_{n\to\infty} \widehat{\beta} = \beta$$

iii.3 Asymptotic Normality

I know that the maximum likelihood estimator has an asymptotically normal distribution. Taking $\widehat{\beta}$ as an example, I have

$$\sqrt{N}(\hat{\beta}_{MLE} - \beta) \xrightarrow{d} N(0, I(\beta)^{-1})$$

Therefore, I have

$$\begin{aligned} \hat{\beta}_{MLE} &\xrightarrow{d} N(\beta, \frac{1}{N} I(\beta)^{-1}) \\ \Rightarrow se^2(\hat{\beta}) &= Var(\hat{\beta}) = \frac{1}{N} diag(I(\beta)^{-1}) \\ \Rightarrow \widehat{se}(\hat{\beta}) &= \sqrt{\widehat{se}^2(\hat{\beta})} = \sqrt{\frac{1}{N} diag(\widehat{I(\beta)}^{-1})} \end{aligned}$$

Since the Hessian of the log likelihood function is the negative of Information matrix, I would have an estimate for the Information matrix, which would allow us to have an estimated standard error.

iii.4 Efficiency and Equivariance

The two last properties of maximum likelihood estimators are efficiency and equivariance. The variance of any given unbiased estimator $\widehat{\beta}$ cannot be lower than the Cramer-Rao Lower Bound, so an estimator whose variance is equal to this lower bound is considered efficient. However, as I have shown, my maximum likelihood estimator $\widehat{\beta}$ is not unbiased. Therefore, I cannot conlcude that my estimator is efficient. The last property is that of equivariance, holds despite estimator bias and gives us that my maximum likelihood estimator is robust to reparameterization. Thus, for my estimator $\widehat{\beta}$ for the school valuation parameter β and any given reparameterization $h(\widehat{\beta})$, $h(\widehat{\beta})$ is also an estimator of $h(\beta)$.

iv. Issues with Current Model & Model Extensions

iv.1 Strict Utility Preference Rankings

In my model, for simplicity, I assumed an aggregate error term that follows T1EV distribution for each ranking. However, such error term can be hard to interpret because it is hard to think of a way that the error term is related to the ranking itself instead of to each individual school's quality. A more realistic model would allow each school to have their own associated utility shocks for each household, and enforce rational, strict utility ranking of schools by households, using the following set-up:

From [Chapman and Staelin, 1982] [Allison and Christakis, 1994], I know that even with strategic choices (where the rankings affect the probabilities of acceptance), so long as parents are rational, conditioning on the probability of admission, the top choice school must have a higher ex post utility than if any other school were to be ranked first. Similarly, in the remaining schools, the school ranked second would have a higher ex post utility than than if any other school were to be ranked second, and so on. Using a similar set-up as [Beuermann et al., 2018], conditional on admission probabilities, household i prefers their first ranked school over any other school such that

$$U_{i,j}^{r_{i,1}} > U_{i,k}^{r_{i,1}} | P_{i_k} \forall k \neq j \in J$$

Given the three school choices for parents, and assuming rational choices, the probability of household i submitting a given ranking is given by

$$\begin{aligned}\Pr[r_{i_1} = j_1, r_{i_2} = j_2, r_{i_3} = j_3] = \Pr[&(U_{i,j_1}^{r_{i,1}} | p_{i,j_1}^r \geq U_{i,k}^{r_{i,1}} | p_{i,k}^r \forall k \neq j_1 \in J) \\ &\cap (U_{i,j_2}^{r_{i,2}} | p_{i,j_2}^r \geq U_{i,k}^{r_{i,2}} | p_{i,k}^r \forall k \neq j_1, j_2 \in J) \\ &\cap (U_{i,j_3}^{r_{i,3}} | p_{i,j_3}^r \geq U_{i,k}^{r_{i,3}} | p_{i,k}^r \forall k \neq j_1, j_2, j_3 \in J)\end{aligned}$$

which defines the likelihood of observing the set of schools as a function of household utilities for the schools and some error. Theoretically, I can derive the (log-)likelihood function following the above, and recover the valuation for school characteristics using maximum likelihood estimation. However, given the distribution of my errors for each individual school, the addition of three Gumbell-distributed errors does not result in a logistic distribution. Using an "exploded" multinomial logistic regression utilizes the condition of the strict preference ranking. However, this is unfortunately out of the technical scope of this paper.

iv.2 Inclusion of Unobserved Heterogeneity

A possible and reasonable extension of the model involves the assumption of unobserved heterogeneity within the classes, or types, of households. I planned to derive the admission probabilities using the same method as is used in this present paper, and utilize a Latent Class (LC) multinomial logistic regression model to estimate how three different classes of households value schools. my model set-up was similar to the one in this paper, however included three classes $k \in K = \{1, 2, 3\}$ where the utility of attending a schools is given by

$$u_{j,k} = u(Z_j, k) = \alpha_k + \beta_k Z_j$$

And the probability of acceptance is given by

$$g\left(X_{k,j}, r_k^j\right) = \theta_{r_k^j} + \eta_{r_k^j} X_{k,j}$$

Thus, I would have estimated, using an latent class model and maximum likelihood estimation,

$$\{\hat{\alpha}_k, \hat{\beta}_k, \{\hat{P}_{j,k}\}_{j \in \{A,B,C\}}\}_{k=1}^3$$

This extension would be helpful in deriving how households that fit into three "latent classes" -care more about future peers, care more about school safety and inclusivity, or care more about school educational achievement- value schools upon changes in school characteristics and admission probabilities separately.

iv.3 Model Implementation on Real-World School Choice Data

Although I simulated my own data set in this paper, a selection of papers on school choice mechanism have implemented choice model methodology on a variety of data sets, the most popular among them being US data on the Boston Mechanism and New Haven Mechanism outcomes. I would be able to apply my model with any data that allows us to observe household characteristics, school characteristics, the household-specified preference rankings of schools, and admission outcomes for students. I would require a choice mechanism that allows for strategic actions on the behalf of the household to solve my model. A number of models have utilized data on strategic mechanisms from developing countries [Beuermann et al., 2018], which could as well be an interesting direction of the extension of this paper. In general, my methodology seems to be widely applicable to a range of school choice data.