Selection Model of School Choice

Michael Shashoua
Rice University
shashoua@rice.edu

October, 2018

In this paper, I create a model of school choice with endogenous selection. I show that a general model of selection is identified without any additional assumptions from the standard school choice framework. I perform a monte carlo to show that this model pins down the correct parameter values, while not incorporating selection may severely bias the estimates. I also present a potential counterfactual where an alternative is removed, and the model with selection is able to correctly predict the results. Incorporating selection is essential for school choice research, and this paper provides a framework to study it.

One of the most critical decisions many families face is determining the best school for their children. The elementary, middle, and high school experience can set students on very different life trajectories. Parents may invest a lot of resources researching schools, and an entire industry exists to help families navigate this process. The goal of a school district is to provide its families more options, improve existing options, and make the school matching process satisfactory for the families that participate.

The participation decision, however, is often overlooked by school districts. When only a subset of the population takes part in a school choice mechanism, the analysis and recommendations might be heavily biased. While school districts aim to improve schools and see enrollment rise, this bias might lead to worse results. Accounting for selection is very important yet it is often ignored in school choice research.

Matching, and in particular assigning students to schools, has a rich economic literature. Gale and Shapley's 1962 matching paper on marriage formalized the Deferred Acceptance (DA) algorithm. The DA algorithm has become one of the most popular methods to match students to schools. Pure DA is attractive because it is a strategy-proof mechanism in that students have no incentive to misreport their true preferences.

A main contribution of this work will be addressing endogenous entry in school choice. I will use the deferred acceptance algorithm for it's popular and simple truth-telling qualities, yet this work is generalizable beyond the simple DA framework.

The matching literature recognizes that students might not apply to schools because they take into account the probability of attending, and furthermore it finds instances where students do not apply everywhere they prefer. Under the same logic, a natural extension is that students do not enter the mechanism because they believe they have a zero percent chance of admission or the value from the outside option is better than all options within the mechanism. These non-participating students exist outside of the mechanism, yet they should still be accounted for when a school district makes decisions. Any counterfactual that affects a student's (perceived) probability of admission or (perceived) utility from any of the options within the mechanism might induce entry into the mechanism.

Entry in the school choice literature has largely been ignored. Entry will change the set of students actually matched which might have important welfare implications. There are examples of entry/exit in matching mechanisms unraveling markets. In the gastroenterology resident matching an unanticipated shock to the number of participants made the mechanism unravel.[1]

One of the main findings of this paper is that even under simple selection environments, the results can be heavily biased. This bias carries through to counterfactuals, which may predict large discrepancies from the true results of regime change.

The remainder of the paper is organized as follows. In Section I, I present an illustrative model of school choice with endogenous selection which I use in section II to discuss briefly the relative advantages and limitations of different approaches to the empirical study of school choice and selection models in general. In Section III, I discuss the identification and estimation of the model, and I apply those techniques with a monte carlo study and policy experiments in section IV. Section V contains future extensions to the model. Concluding remarks are included in Section VI. More technical details on estimating the model, generating the monte carlo study, and adapting the model to different primitives are contained in the appendix.

I. Model

This model of school choice includes both endogenous selection and correlation between the selection decision and the observed outcome.

[1] See McKinney, Niederle, Roth (2005).

This is a two stage problem. In the first stage students draw a cost shock and decide whether to enter the mechanism or to take their safety school. In the second stage the students who entered the mechanism draw school related utility shocks, where school utility is correlated with the cost draw, and they submit rank ordered lists.

In this model the student beliefs are exogenously determined. The next version of this model will be an equilibrium model with endogenous beliefs, but for the time being any beliefs can be entered into the model.

I. Second Stage

In the second stage we will only observe the students who decide to submit a rank order list. Each student carries their cost shock η_i from the first period, where η_i is drawn from $F_\eta(\cdot)$. Let S be the set of available schools.

The utility from enrolling in school $s \in S$ is denoted as follows:

$$U_{is} = u_{is} + \eta_i \beta_{\eta,s} + \varepsilon_{i,s}, \tag{1}$$

where u_{is} denotes the observable utility,

$$u_{is} = \beta_{0,s} + X_{is}\beta_x + H_i \beta_{H,s}. \tag{2}$$

The $\beta_{0,s}$ acts as a school specific intercept, so this will encompass things that only vary by school, such as percent male. The observable covariates X_{is} vary by student by school. This includes data such as distance from home to school. The final covariate, H_i, denotes student varying characteristics, such as gender or ethnicity. There is a school specific coefficient for each of the H covariates, so the model is general in that particular student attributes might give higher utility at different schools. The cost shock, η_i, is an example of something that only varies by student and it is correlated with each school differently. The cost shock is unobserved to the econometrician. Finally the student-school utility shock, $\varepsilon_{i,s}$, is drawn from $F_\varepsilon(\cdot)$.

Once a student views all his school specific utility shocks in the second stage, he is tasked with creating the list that he submits to the mechanism. The optimal list under the Gale-Shapely algorithm is truth-telling, which does not require the student to consider the admission probabilities.[2]

Assume there are only 4 schools. A student ranks schools as follows: $3 \succ_i 2 \succ_i 4 \succ_i 1$. The probability of observing that ranking is equivalent to: $Pr(U_3 > U_2 > U_4 > U_1)$.

This can be broken down into a product of conditional probabilities:

$$Pr(U_3 > U_2, U_4, U_1) \times Pr(U_2 > U_4, U_1 | U_3 > U_2, U_4, U_1) \times$$
$$...Pr(U_4 > U_1 | U_3 > U_2, U_4, U_1; U_2 > U_4, U_1).$$

II. First Stage

In the first stage, students must decide if they are going to apply to the mechanism. While the students do not need to take into account admission probabilities in the Gale-Shapely deferred acceptance algorithm, they will take admission probabilities into account in the decision to participate.

A student will enter the mechanism if the expected value from paying the cost and entering the mechanism is higher than the expected utility of the outside option. Let U_{i0} denote the utility of student i's safety school. Let $d_i = 1$ denote that student i submits an application. I can express an application as

$$\begin{aligned} d_i = 1 \text{ if } EV(X_i, \eta_i) - \eta_i &\geq E[U_{i0}] \\ \equiv EV(X_i, \eta_i) - \eta_i &\geq u_{i0} + \beta_{\eta,0}\eta_i + E[\varepsilon_{i,0}] \end{aligned} \tag{3}$$

where $EV(X_i, \eta_i)$ is the expected utility from entering the mechanism which takes into account admission probabilities in the mechanism. I discuss how to compute the $EV(\cdot)$ function in the next subsection. The expectation on both sides of the inequality is in reference to the unobserved school utility shocks.

If it were possible for the econometrician to view the η_i for each student i, then it would be standard procedure to uncover the parameter estimates. However, because the η_i is unobserved, it will require integrating it out, with complex integrals that require the use of simulation. The estimation details are in the section III.

[2]Under a Gale-Shapely Deferred Acceptance algorithm, truth-telling is a weakly dominate strategy, and if the probability of acceptance is strictly between 0 and 1, then truth-telling is strictly dominant. It is an simple extension to forgo the truth-telling assumption in place of a stability assumption.

III. Calculating the Expected Value Function

In this subsection I present a simple example that walks through calculation of the expected value function. Given 3 options with respective utility $\delta_s + \varepsilon_s$ for $s = \{1, 2, 3\}$, and ε_s is drawn from $F_\varepsilon(\cdot)$.

Under guaranteed admission, I can calculate the expected value of entry by going through the expected value of each option if it were chosen. Begin by looking over the range of ϵ such that option 1 is chosen when the student is guaranteed admission to all schools:

$$E_{\epsilon \in M_1}[\max_s(\delta_s + \epsilon_i)] = \int_{-\infty}^{\infty} (\delta_1 + \epsilon_1) \left[\int_{-\infty}^{\delta_1+\epsilon_1-\delta_2} \int_{-\infty}^{\delta_1+\epsilon_1-\delta_3} f\varepsilon(\epsilon_1, \epsilon_2, \epsilon_3) d\epsilon_3 d\epsilon_2 \right] d\epsilon_1 \tag{4}$$

The more interesting context, however, is that the student is not guaranteed admission. Taking into account non-trivial admission probabilities, the expected latent utility will be strictly less than the expected latent utility under guaranteed admission. If we need to take admission probabilities into account, then equation (4) becomes

$$\begin{aligned} E_{\epsilon \in M_1}[\max_i(\delta_s + \epsilon_s)] = & \int_{-\infty}^{\infty} \Bigg(\int_{-\infty}^{\delta_1+\epsilon_1-\delta_2} \int_{-\infty}^{\delta_2+\epsilon_2-\delta_3} \Big[P_1(\delta_1 + \epsilon_1) + (1 - P_1)P_2(\delta_2 + \epsilon_2) \ldots \\ & + (1 - P_1)(1 - P_2)P_3(\delta_3 + \epsilon_3) \Big] f\varepsilon(\epsilon_1, \epsilon_2, \epsilon_3) d\epsilon_3 d\epsilon_2 \ldots, \\ & + \int_{-\infty}^{\delta_1+\epsilon_1-\delta_3} \int_{-\infty}^{\delta_3+\epsilon_3-\delta_2} \Big[P_1(\delta_1 + \epsilon_1) + (1 - P_1)P_3(\delta_3 + \epsilon_3) \ldots \\ & + (1 - P_1)(1 - P_3)P_2(\delta_2 + \epsilon_2) \Big] f\varepsilon(\epsilon_1, \epsilon_2, \epsilon_3) d\epsilon_2 d\epsilon_3 \Bigg) d\epsilon_1 \end{aligned} \tag{5}$$

where P_s is the probability of being accepted by alternative s. When the probability is less than one, then we are also concerned with the 2nd and 3rd option at every point. The integral expands with every additional option, and the more possible options, the more complicated it becomes. There is no simple closed form, and the expected value must be simulated. See appendix section

VII for details.

II. Departure from Previous Literature

This work touches on two important strands of economics, school choice and selection models. Much of the school choice literature focuses on the implication from switching mechanisms, in particular from a Boston mechanism to a Gale-Shapely algorithm. Other research uses a constant mechanism, but studies the effects of changing specific aspects, such as priority scores or the number of schools a student can apply to.

This paper focuses on a problem that has not been adequately addressed in the literature: determining the effects of an application process when only a subset of students apply, while taking the entire student population into consideration. This essentially creates endogenous entry in school choice.

The school choice literature has largely ignored the selection aspect. The analysis tends to focus on districts that do not have many outside options available to their students, such as the school system in Barcelona, whose mechanism includes 96% of all Barcelona schools (Casalmiglia, Fu, Guell 2014), or the school sorting in Beijing which declared that "outside options were not relevant" (He 2014). In many papers that discuss American school districts such as New York, Cambridge, and Boston, the individuals who enter matching mechanisms and those who decide not to enter are different in both observable and unobservable ways. Few papers discuss entry into a school matching mechanism. Among them Ferreyra and Kosenok (2014), discuss students' decisions to apply to charter schools through charter school location decisions, and Mehta (2014) describes a model of charter school input decisions. The most relevant work to this project is Kapor, Neilson, Zimmerman (2016), yet they use the model to explain achievement and assume the students know the school utility shocks before entering the mechanism. I relax the full information framework and allow for correlation between the cost and the school utilities. Additionally Walters (2012) estimated a flexible demand-side model of charter school application and attendance decisions. However, that model does not use the decision to apply and the ranked list of schools as the outcomes, but rather student achievement, as a function of school choice, is the outcome of interest. Those models requires data on student achievement, assumptions on the the achievement production function, and the assumption that achievement scores are a major outcome that determines student selection. My model is more general, not requiring any special covariates like achievement scores yet it has the ability to encompass them.

Applying selection techniques to a school choice mechanism is not trivial. The school choice context not only includes multiple, correlated alternatives and censored observations, but it also contains a discrete outcome.[3] In school choice we have no such outcome that is so inextricably linked to utility as wage for labor selection. Only the decision to apply and the rank ordered list are outcomes. Standard selection techniques, such as a control function approach, are not applicable with this type of data.

If selective entry is not accounted for, then models are under the assumption that everyone who can participate in a market does participate. However, that is the wrong way to look at it, as individuals make choices based on their unobservables. This means that observed relationships, or matches, should be viewed through a lens of endogenous decision making and not an exogenous relationship. Ignoring selection would bias the coefficient estimates and the counterfactual predictions might be very far from reality.

Selective entry can have important effects in matching markets. By incorrectly assuming that everyone is participating in a market, any dynamics of markets growing or shrinking are ignored. Muriel and Roth (2003) and McKinney, Niederle, Roth (2005) examine the unravelling of the gastroenterology resident matching in 1997 due to an unanticipated shock to the number of applicants. A more general model which endogenizes selection into the specific gastroenterology market would better encompass market size and the decisions of those that self-select into the market.

There is also a growing body of work of endogenous entry and self-selection in the auction literature. The decision to enter and which auction to chose among a menu of auctions has important considerations in maximizing revenue. There is empirical evidence that individuals self-select into different types of auctions by their unobserved value, and bidding behavior can be rationalized through auctions with entry.[4]

Finally, recent work on non-parametric identification of school choice models attempts to relax the parametric assumptions that has dominated the literature.[5] In the monte carlo simulation in

[3]For examples of selection models which include multivariate choices see Yen (2005), and for censored data see Lin and Yen (2006).

[4]See Engelbrecht-Wiggans, 1993; Smith and Levin, 1996, 2002; Pevnitskaya, 2004, and Palfrey and Pevnitskaya, 2008 for this literature.

[5]Agarwal and Somaini 2014 is an example of an non-parametric approach.

section IV, I use the same parametric specification to estimate as the true data generating process, but ignoring selection will bias the estimates. The next section shows how accounting for selection does not require any additional normalizations beyond those of the standard multinomial logit model.

III. Identification and Estimation

I. Identification of School Choice with Endogenous Selection

The discussion for identification of this model follows the identification strategy of the Tobit II models from Amemiya (1985).

Tobit II models are of the form:

$$y_{2i} = \begin{cases} y^*_{2i} & \text{if } y^*_{1i} \geq 0 \\ 0 & \text{if } y^*_{1i} < 0 \end{cases}$$

where the outcome, y^*_{2i} is observed if $y^*_{1i} \geq 0$. This can be further generalized to

$$y_{2i} = \begin{cases} y^*_{2i} & \text{if } y^*_{1i} \geq 0 \text{ and } y^*_{2i} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Which is sometimes called the "double hurdle" (Cragg 1971) as it requires $y^*_{1i} \geq 0$ and conditional on that, $y^*_{2i} \geq 0$. Consider for example,

$$\begin{aligned} y_{1i} = & \quad z_i'\beta + \varepsilon_i \\ y_{2i} = & \quad x_i'\gamma + \nu_i. \end{aligned}$$

To identify this model, given any structure of correlation between the shocks, an exclusion restriction is generally made of the form that there exist some variable in z that is not in x, that is something that affects the propensity to be observed while not changing the underlying outcome observed.

In a parametric model which specifies the distributions of the unobservables, a control function approach can be used to identify the model. Under normality, the inverse mills ratio has nonlinearities which are key to identifying the model. The body of the inverse mills ratio is linear, but the tails introduce nonlinearities, so a condition for identification in the absence of an exclusion restriction is that at least one of the x's displays sufficient variation to induce tail behavior in the inverse mills ratio.

The model presented in this paper does not require any exclusion restriction z for identification. The key to identification in this framework is the nonlinearities in the expected value function.[6] Furthermore, the probabilities of acceptance, while providing more variation in the expected value function, are also not necessary.

To walk through the identification argument, I will begin with the most basic setting in which the utility from being matched to a certain school is $U_s = \beta_{\eta,s}\eta + \varepsilon_s$. I will apply the following parametric specification to the errors, $\eta \sim N(\mu_\eta, \sigma_\eta)$ and ε_s is distributed type 1 extreme value (T1EV). Interestingly, this model does not require any normalizations beyond the T1EV (location and scale normalizations) of the school specific shocks and the location of some of the coefficients. If there is only one school and the student is guaranteed admission, then the decision to enter the mechanism is:

$$\begin{aligned} d_i = 1 \text{ if} \quad & EV(\eta_i) - \eta_i \geq \beta_{\eta,0}\eta_i + E[\varepsilon_{i,0}] \\ & \equiv \beta_{\eta,0}\eta_i + \gamma - \eta_i \geq \beta_{\eta,0}\eta_i + \gamma \;, \\ & \equiv -\eta_i \geq 0 \end{aligned} \tag{6}$$

where γ is the expectation of a T1EV random variable, known as Euler's constant. It is clear that with only 1 school and guaranteed admission, we have no hope of pinning down the variance of η. Even with an exclusion restriction that only affects the probability of entry but not the utility from the match, we cannot pin down the scale.[7] If we add another school to the setting with admission probability 1, then the entry decision becomes:

$$\begin{aligned} d_i = 1 \text{ if} \quad & EV(\eta_i) - \eta_i \geq \beta_{\eta,0}\eta_i + E[\varepsilon_{i,0}] \\ & \equiv log(exp(\beta_{\eta,0}\eta_i) + exp(\beta_{\eta,1}\eta_i)) + \gamma - \eta_i \geq \beta_{\eta,0}\eta_i + \gamma \\ & \equiv log(exp(\beta_{\eta,0}\eta_i) + exp(\beta_{\eta,1}\eta_i)) - \eta_i \geq \beta_{\eta,0}\eta_i \end{aligned} \tag{7}$$

The only requirement for identification in this model is that the $\beta_{\eta,s}$ coefficients are not all equal. Under those assumptions, I can identify both the location and scale of η. In terms of the

[6] See section V for details on incorporating an exclusion restriction into the model.

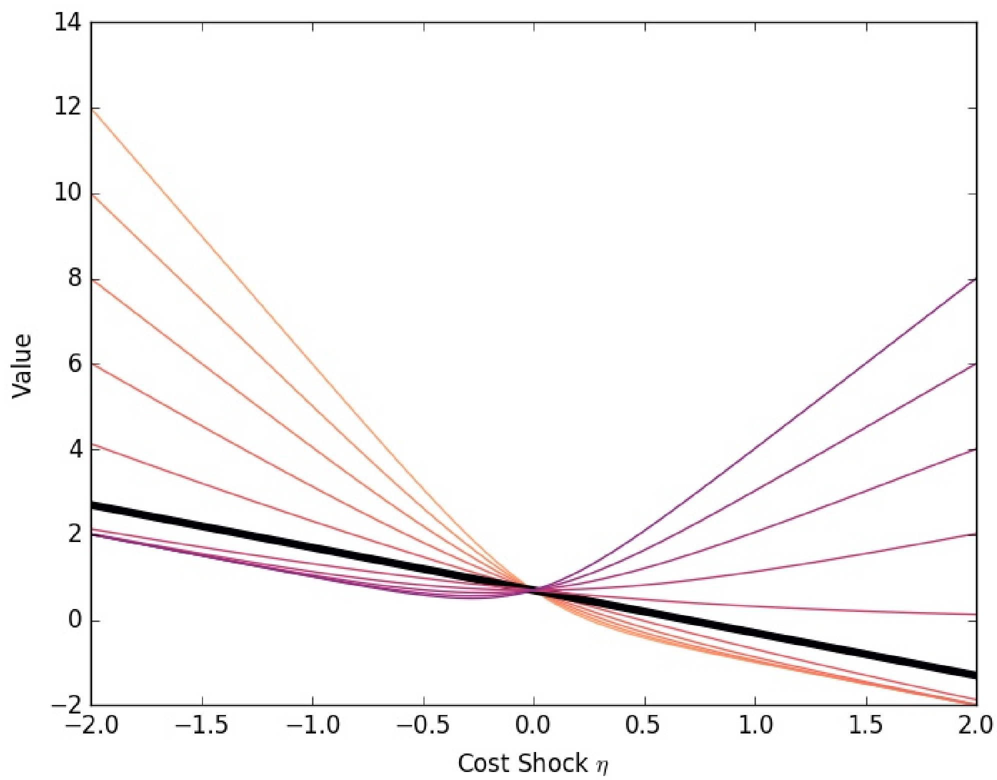
[7] If we had some exclusion restriction Z, then equation (6) would be $d_i = 1$ if $Z_i\beta_z - \eta_i \geq 0$, which is observational equivalent to $Z_i\frac{\beta_z}{c} - \frac{\eta_i}{c} \geq 0$.

β_η coefficients, the scale is identified but not the location. Because of this, one of the coefficients can be set to zero.[8] The only thing that matters for these coefficients is their value relative to each other. In Figure (1) I show how the nonlinearities enter by graphing the function

$$f(\eta) = log(exp(\beta_{\eta,0}\eta) + exp(\beta_{\eta,1}\eta)) - \eta - \beta_{\eta,0}\eta,$$

at different levels of $\beta_{\eta,1}$. I set $\beta_{\eta,0} = 0$ and I vary $\beta_{\eta,1}$ from -5 to 5, where the lighter lines are lower values of $\beta_{\eta,1}$. The cost shock η varies from -2 to 2. The thick black line is when $\beta_{\eta,0} = \beta_{\eta,1}$, which makes this a linear function in η.

Figure 1: *The Decision to Enter the Mechanism, Varying the β Values*



For identification, a requirement is that we observe some people not participating, either from the realization of the cost shock or their covariates, which in Figure (1) is denoted as a realized

[8]This also applies to the $\beta_{H,s}$ coefficients for each H variable that only varies by individual and not by alternative. See section VI in the appendix for details and the proof.

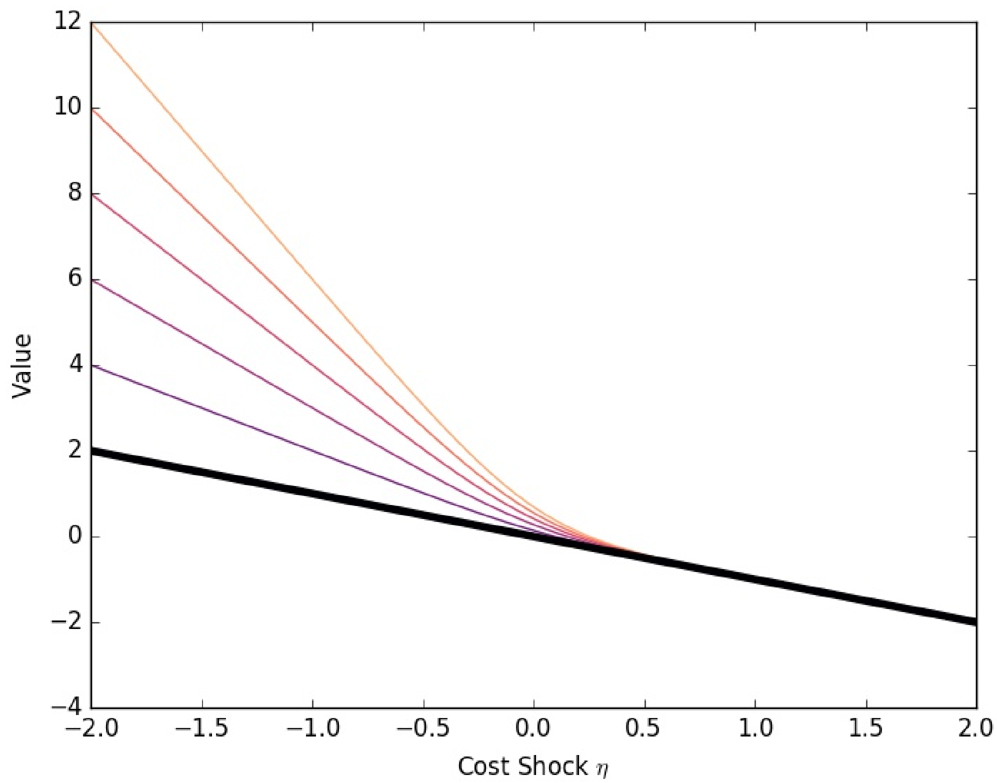
value below zero. This is essentially stating that if we have individuals with the same observable characteristics, some choose to enter the mechanism and some remain out of it. Otherwise selection decisions could entirely be attributed to selection on observables.

Different probabilities of admission also affect the value function. As an example, in Figure (2) I will plot the following function,

$$f(\eta) = EV(\eta) - \eta - \beta_{\eta,0}\eta - \gamma,$$

with the probability of acceptance to school 1 ranging from 1 to 0, with lighter colors representing the higher probability. Again, the thick black line is when the probability of admission to school 2 is zero. I hold $\beta_{\eta,0} = 0$ and $\beta_{\eta,1} = -5$.

Figure 2: *The Decision to Enter the Mechanism, Varying the Admission Probability*



As the probability of admission to the second option gets closer to zero, the value function becomes more and more linear. It should be clear that we can pin down the location and scale of η as long as 1) η has some values in the nonlinear region, 2) some people choose not to enter the mechanism, and 3) there are non-degenerate admission probabilities for at least one school beside the safety school.

As more schools are added the function becomes even more nonlinear. Through the nonlinearities and the varying covariates among the individual students, the η parameters can be identified.

If the distribution of η is identified, and indeed if η were observed in the data, the probability of observing a list becomes the standard multinomial logit. I now quickly recap the identification and interpretation of parameters from multinomial logit models.

In logit models, individuals only care about the differences in utility across alternatives. If something boosts the utility level of all alternatives the same amount, then it cannot be used to explain decision making behavior. On the other hand, if the covariates vary by individual by alternative, then their coefficients are fully identified. If covariates vary only by individual, and it affects the utility of each alternative differently, then the scale of those parameters is identified but not the location. It is without loss of generality to normalize one of alternative's coefficients to zero. Similarly, anything that varies by alternative but not by individual is indistinguishable from a school fixed effect. These covariates should be left out of estimation, as the parameter estimates are observationally equivalent to a fixed effect, so it is only necessary to include alternative specific intercepts and set one of them equal to zero. Finally, anything that does not vary with alternatives nor individuals will not be identified and should be left out of the model.

Interpretation of the parameters of a logit model requires taking into account the normalizations for specific covariates. For covariates that vary by individual by alternative, the sign of a parameter can be interpreted as the direction of influence of its covariate. However, the absolute magnitude of the parameters is interpretable only through an odds ratio. For intercepts and covariates that only vary by individual, there is a required normalization so the sign of the parameter is completely arbitrary and just the utility scale between the different alternatives is important. Again, the utility difference is not itself interpretable because it enters into an odds ratio.

A useful statistic is the marginal effect of an independent variable on choice probabilities, common in the IO literature. It is also interesting to predict choice probabilities for a certain individual with specific covariate values, or to compute market share under counterfactual regimes.

Commonplace in school choice literature is to put all the variables in reference to the parameter for distance to school. This makes an intuitive interpretation of the covariates in the form of “willingness to travel." For settings in which there is a monetary cost that varies by school and across individuals, a “willingness to pay" can also be used.

II. Calculating the Probability of Applying

The decision to apply is represented in equation (3). For a given X_i, it is possible to estimate the $EV(\cdot)$ function through simulation.[9] The probability of a student choosing to participate becomes:

$$Pr(d_i = 1|X_i) = \int_{-\infty}^{\infty} 1\{EV(X_i, \eta_i) - \eta_i \geq u_{i0} + \beta_{\eta,0}\eta_i + \gamma\} dF_\eta(\eta_i). \tag{8}$$

Even with η_i normally distributed, this has no closed form as it is not possible to pull η out of all the terms. To simulate this probability, I draw η_j J times from the $F_\eta(\cdot)$ distribution and average over them:

$$Pr(d_i = 1|X_i) = \frac{1}{J}\sum_{j=1}^{J} 1\{EV(X_i, \eta_j) - \eta_j \geq u_{i0} + \beta_{\eta,0}\eta_j + \gamma\}. \tag{9}$$

III. Probability of Observing a List ℓ_i

The T1EV school utility shocks provide a very convenient way to calculate application probabilities. Conditional on the η_i, the likelihood of a certain ranking has a closed form solution, which is generally referred to as an exploded logit. Returning to the example in section I, assume a student ranks schools as follows: $3 \succ_i 2 \succ_i 4 \succ_i 1$.

Taking advantage of the IIA property of T1EV allows us to take the product of the unconditional choice probabilities. The example above is expressed as:

$$\frac{exp\{u_{i3} + \beta_{\eta,3}\eta_i\}}{\sum_{s' \in \{1,2,3,4\}} exp\{u_{is'} + \beta_{\eta,s'}\eta_i\}} \times \frac{exp\{u_{i2} + \beta_{\eta,2}\eta_i\}}{\sum_{s' \in \{1,2,4\}} exp\{u_{is'} + \beta_{\eta,s'}\eta_i\}} \times \frac{exp\{u_{i4} + \beta_{\eta,4}\eta_i\}}{\sum_{s' \in \{1,4\}} exp\{u_{is'} + \beta_{\eta,s'}\eta_i\}}.$$

[9] See appendix section VII for the calculation details of the $EV(\cdot)$ function.

Ignoring selection, the probability of observing a list ℓ_i is,

$$Pr(\ell_i|\eta_i) = \prod_{s\in\ell_i} \frac{exp\{u_{is} + \beta_{\eta,s}\eta_i\}}{\sum\limits_{s':s\succeq_i s'} exp\{u_{is'} + \beta_{\eta,s'}\eta_i\}}, \tag{10}$$

where $s \in \ell_i$ denotes that school s is ranked in the list by student i, and $s' : s \succeq_i s'$ is the set of schools s' ranked at or below s by student i.

In equation (10), there was a simple closed form for the probability of a list ℓ_i conditional on observing η_i. However, the econometrician does not observe η_i and it must be integrated out. It is important to remember that not all values of η_i will lead to lists being observed, so we must condition only on the η_i values that yield a list. Suppressing the conditioning on covariates X_i,

$$Pr(\ell_i|d_i = 1) = \frac{Pr(\ell_i, d_i = 1)}{Pr(d_i = 1)},$$

where the numerator can go through different η_i values and calculate the exploded logit, multiplied by the indicator if that value of η yields an application. The denominator is the value from equation (8). Plugging this in gives

$$Pr(\ell_i|d_i = 1) = \frac{\int_{-\infty}^{\infty} 1\{EV(X_i, \eta_i) - \eta_i \geq u_{i0} + \beta_{\eta,0}\eta_i + \gamma\} \left(\prod_{s\in\ell_i} \frac{exp\{u_{is}+\beta_{\eta,s}\eta_i\}}{\sum\limits_{s':s\succeq_i s'} exp\{u_{is'}+\beta_{\eta,s'}\eta_i\}} \right) dF_\eta(\eta_i)}{\int_{-\infty}^{\infty} 1\{EV(X_i, \eta_i) - \eta_i \geq u_{i0} + \beta_{\eta,0}\eta_i + \gamma\} dF_\eta(\eta_i)}$$

Which we can simulate as before with a draw of η_j,

$$Pr(\ell_i|d_i = 1) = \frac{\sum_{j=1}^{J} 1\{EV(X_i, \eta_j) - \eta_j \geq u_{i0} + \beta_{\eta,0}\eta_j + \gamma\} \left(\prod_{s\in\ell_i} \frac{exp\{u_{is}+\beta_{\eta,s}\eta_j\}}{\sum\limits_{s':s\succeq_i s'} exp\{u_{is'}+\beta_{\eta,s'}\eta_j\}} \right)}{\sum_{j'=1}^{J} 1\{EV(X_i, \eta_{j'}) - \eta_{j'} \geq u_{i0} + \beta_{\eta,0}\eta_{j'} + \gamma\}}.$$

Let $d_i = 1$ denote the decision of individual i to apply and $d_i = 0$ denotes the decision to not apply. The parameters we want to estimate are $\Theta = \{\beta_x, \beta_\eta, \beta_s, \beta_H\}$, where the last three are vectors of length $|S| - 1$, and $\Sigma = \{\mu_\eta, \sigma_\eta\}$, which determine the distribution of η.[10] The likelihood

[10] Recall that one of the parameters can be set to zero for covariates that do not vary by both alternative and individual.

for i can be expressed as

$$\mathcal{L}_i(\Theta, \Sigma) = Pr(d_i = 0)^{1\{d_i=0\}} \big(Pr(d_i = 1)Pr(\ell_i | d_i = 1)\big)^{1\{d_i=1\}},$$

where the simulated estimates of $Pr(d_i = 0)$ and $Pr(\ell_i|d_i = 1)$ are used. The full simulated likelihood over all individuals is simply the product across everyone in the data:

$$L(\Theta, \Sigma) = \prod_{i=1}^{N} \mathcal{L}_i(\Theta, \Sigma).$$

IV. Monte Carlo and Policy Experiments

I. Monte Carlo

In this section, I use 3 schools and generate X variables that vary by individual and school. The probability of acceptance to the schools are $P_1 = .2$, $P_2 = 1$, and $P_3 = .9$, with further details on the data generating process provided in section I of the appendix.

With an underlying DGP, I simulate the decisions for 1000 students, both the decision to apply and, conditional on applying, a ranked order list. I then estimate the parameters and repeat this 100 times, reporting RMSE, mean absolute error, the parameter estimates and the standard deviation.

In addition, for each of the 100 simulations I will also estimate the parameters using a "naive" approach under the belief that selection is not important. Two naive approaches will be estimated: 1) a model with just a β_x coefficient, and 2) a model with both β_x and school specific intercepts.

The naive parameters will be estimated from multinomial logit and just based on the application lists provided. This ignores the students who did not enter, as the underlying logit is under the assumption that selection is unimportant. I report the mean and standard deviation of these naive parameters as well.

In Table 1 I present the results for this model. The model's estimates are very close to the true data generating process. Recall that $\beta_{\eta,1}$, the alternative specific cost shock coefficient, has been normalized to zero. The standard deviation is largest on the $\beta_{\eta,3}$ parameter. In Figure (3) I graph the density of the deviation from the truth of $\beta_{\eta,3}$ across different levels of simulations. Even after just 50 simulations, the values appear to cluster around the true parameter values.

Table 1: *Accounting for Selection*

	true	mean	std. dev	l.q.	med	h.q.	r.m.s.e	m.a.e.
β_X	3.00	3.012	0.426	2.745	2.963	3.223	0.424	0.318
$\beta_{\eta,2}$	-0.30	-0.283	0.188	-0.403	-0.312	-0.166	0.188	0.146
$\beta_{\eta,3}$	3.10	3.169	0.455	2.923	3.120	3.445	0.457	0.345
μ_η	0.50	0.504	0.111	0.424	0.513	0.586	0.110	0.089
σ_η	1.75	1.703	0.271	1.531	1.672	1.851	0.274	0.219

The table reports the mean, standard deviation, 25th percentile, median, 75th percentile, root mean square error, and mean absolute error of the parameter estimates over 100 simulations.

Under the first naive approach, the model estimated is $U_{i,s} = X_s\beta_x + \varepsilon_{i,s}$. The parameter estimate and standard deviation are in Table 2. It is clear that β_x is biased, as it is almost a third of the real value, and it has a very small standard deviation.

Table 2: *No Selection, no Intercepts*

	true	mean	std. dev	l.q.	med	h.q.
β_X	3.00	1.136	0.064	1.087	1.136	1.180

The table reports the mean, standard deviation, 25th percentile, median, and 75th percentile of the parameter estimates over 100 simulations.

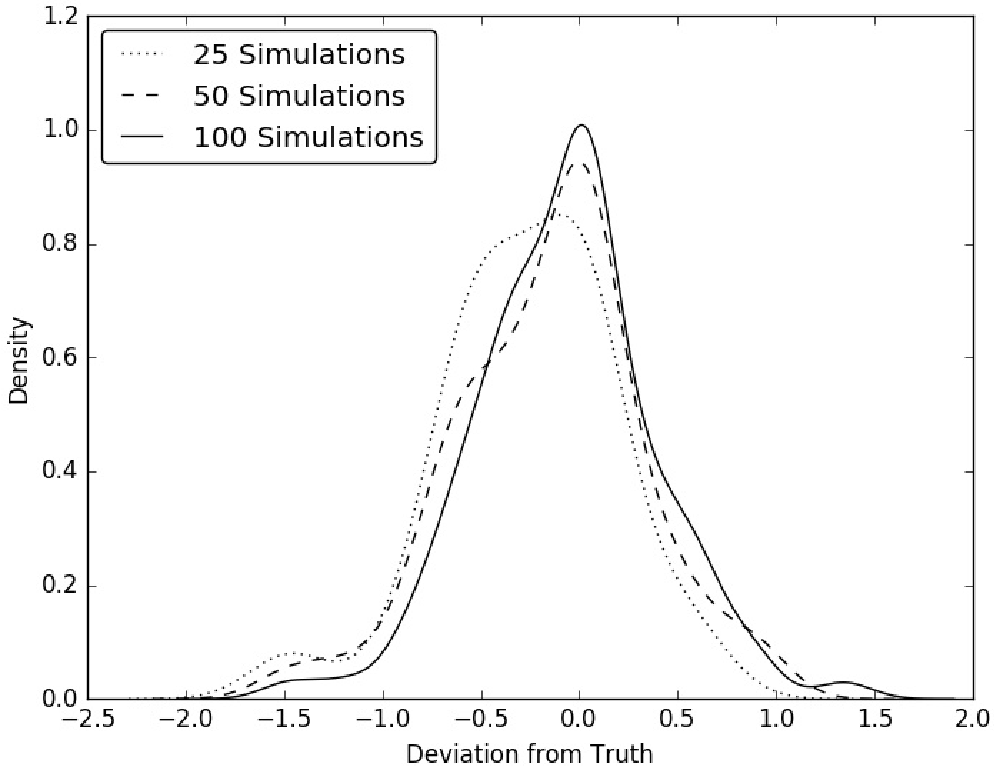
An initial thought would be to include alternative specific intercepts. Indeed, when the variance of the cost shock is small, it is nearly the same as a school specific intercept. When the cost shock variance is precisely zero, all individuals have the same cost shock, and the model appears the exact same as a school specific intercept. In Table 3 I include school specific intercepts, normalizing school 1's intercept to zero. The estimate of β_x remains severely biased, and the school specific intercepts are not statistically significant. With the cost shocks distributed $N(.5, 1.75)$, the majority of the shocks will be positive, which lend a higher utility to alternative 3 as $\beta_{\eta,3}$ is the largest β. Because of this, when I estimate the model with alternative specific intercepts, it has to explain more individuals choosing alternative 3 with a higher intercept as the X coefficients are similar across the options. These estimates are far from the true process, and including school specific intercepts do little to alleviate the bias.

II. Policy Analysis

Once the parameters are identified, it is possible to conduct the following counterfactuals:

- Adding a new school

Figure 3: *The Deviation from the Truth of $\beta_{\eta,3}$*



- Removing schools
- Changing the characteristics of existing schools
- Changing the school capacity/probability of admission for certain groups

These counterfactuals are not able to be performed without endogenous entry. It it generally unrealistic to assume that the population that applies to the mechanism is the same that applies under all counterfactual simulations unless there is forced participation.

It is also possible to look at changing the mechanism or how changes to the priority score would affect participation. Another counterfactual can detail the effect of targeted interventions to get more students to apply. Finally, a district can gauge interest in new programs before introducing them (a la McFadden's (1974) BART estimation).

The counterfactual I present for illustration is the removal of the third alternative and estimate

Table 3: *No Selection, with Intercepts*

	true	mean	std. dev	l.q.	med	h.q.
β_X	3.00	1.134	0.065	1.086	1.133	1.182
β_2		-0.151	0.107	-0.223	-0.138	-0.085
β_3		0.200	0.147	0.091	0.184	0.313

The table reports the mean, standard deviation, 25th percentile, median, and 75th percentile of the parameter estimates over 100 simulations.

the effects of both how many applicants leave the mechanism and what the new lists look like. The naive approach has nothing to say on the entry or exit of individuals–it will simply suggest that everyone that applied to the 3rd school will apply to the next highest alternative in its place.

I randomly select from one of the monte carlo simulations to estimate the parameters and compare the truth to my estimated model and also compare it to the naive estimates. Table 4 presents the estimated effects under the counterfactual regime, both under a naive approach and incorporating this model.

Table 4: *Policy Results of Removing 3rd Alternative, Estimated vs Actual*

	Baseline	**CF Naive**	**CF Model**	**True Results**
Applicants	829	829	606.99	604
List – % Submitted	[1 2 3] – 0.12	[1 2] – 0.58	[1 2] – 0.68	[1 2] – 0.69
	[1 3 2] – 0.2	[2 1] – 0.42	[2 1] – 0.32	[2 1] – 0.31
	[2 1 3] – 0.08			
	[2 3 1] – 0.08			
	[3 1 2] – 0.26			
	[3 2 1] – 0.26			

The “Baseline" column is what is observed in the data. The “CF Naive" are the estimates of removing the 3rd alternative under the naive approach. The “CF Model" column are the estimated results of the counterfactual applying the model in this paper, while the last column are the actual results that would occur. This generated dataset was randomly selected from one of the monte carlo simulations.

Column 2 of Table 4 presents the baseline data. When school 3 is available as an alternative, over 400 people chose school 3 as their top choice. Column 3 presents the naive estimates. In particular, no will exit the mechanism under the naive approach, and in the absence of school 3, students will now just select their next highest option.

Column 4 contains the model’s estimated impact. It uses the parameter estimates from the baseline data, where school 3 was available, and then uses those parameters as a DGP to average the outcome over many draws.

This model predicts that nearly 25% of the individuals will drop out, and for those that participate, school 1 will be applied to at twice the rate of school 2. This model takes into account students who have X values such that they participate, and others who participate just because of the cost shock. When school 3 was removed, the expected value declined and many individuals stayed out of the mechanism.

The true effect of removing the third alternative from the generated data is in column 5. The model comes extremely close to the actual results because it was able to closely pin down the parameter values. The naive approach, which is the standard in school choice literature, has very biased estimates and cannot predict the large entry or exit under regime changes.

This section presented a simple example that did not even make use of the fact that the naive parameter estimates are biased. Under counterfactual policies of adding a new school or changing some characteristics of an existing school, the naive approach would suffer from both the lack of entry and exit as well as biased preference parameters.

V. Model Extensions

In this section I discuss how the model can be adapted for use under many frameworks and how recent methodological techniques might be applied. I focus on the following topics: Outside options and learning, more general application behavior, a different timing of error draws, adding exclusion restrictions, incorporating richer beliefs, leaving the pure deferred acceptance framework, and the long term effects of regime change.

I. Allowing Outside Option

Following Kapor, Neilson, Zimmerman (2016), this model can account for students who decide not to take the outcome of the selection mechanism.

There are then two scenarios. First, assume that the student was assigned a school that was not his safety school. There is then an unexpected utility shock and the student can choose between the assigned school, the safety school, or the outside option. This can explain situations where the data contains students that apply through the mechanism yet still take the safety school, which was available to them throughout. This also makes sense logistically as students go through the application process for both public schools and private schools, so they might not learn their outside options until the end of the matching process.

Under the second scenario, the student is only assigned to his safety school. The options when the unexpected shocks are drawn are only between the safety school and the outside option.

It is standard to allow for multiple rounds of information that explain changing opinions as the process continues. Many school districts allow students to resubmit lists after initial matches have been made, and this can involve students wanting to get into schools they previously rejected.[11] This model can explain that behavior, which allows for richer analysis.

It is not necessary that the final shocks be unexpected, but if they are expected we would require a certain type of behavior. In particular, once a student has paid the cost and applies through the mechanism, he should apply to as many schools as possible. If there are three schools with each admission probability less than 1, then we expect him to apply to all 3. If he only applies to 1 or 2 schools, his expected utility would be strictly higher if he applied to all 3. That is because if he were matched to a non-safety school from the mechanism, then he gets to draw an additional error shock when considering the outside option. It is not possible to rationalize a student's decision to refrain from applying to as many schools as possible if he anticipates a final shock after the mechanism assigns preliminary matches.

II. A More General Behavior of Application

A student who ex-ante decides to participate may not end up submitting an application. If after paying the cost and viewing all the school utility shocks, he observes that the safety school yields the highest utility, then he will just take that option and not provide any rank ordered list. Similarly, if after observing the utilities he sees that only one school gives him a higher utility than the safety school, then he may only rank that school.

The probability of observing a list is equal to the probability of a cost shock such that a student participates times the probability of observing utility shocks such that the safety school is not the top pick. This can easily be accounted for in the likelihood function. See appendix section V for details.

[11] See Narita (2016) for a model with multiple rounds of application.

III. Realizing all Errors before Entry Decision

If the entire vector of costs and errors is drawn before the entry decision is made, then the new decision is to enter if:

$$V(X_i, \varepsilon) - \eta_i \geq u_{i0} + \beta_{\eta,0}\eta_i + \varepsilon_{i,0}.$$

Note that there is no longer an expected value term conditional on the cost draw, η_i, but rather the actual vector of utility shocks, ε. It is not 'expected' in that the shocks are unknown, but rather that the final placement is unknown. In particular, we can solve for $V(X_i, \varepsilon)$ multiplying the probabilities of admission times the utility values.

The major difference between this model and the one with unknown utility shocks is that no one would pay the entry cost unless they were sure that the safety school will not give the highest utility (conditional on the cost draw, as always).

If the individuals observe all the cost and utility shocks, then the probability that the econometrician observes an application can be written as

$$\int\int 1\{V(X_i, \varepsilon) - \eta_i \geq u_{i0} + \beta_{\eta,0}\eta_i + \varepsilon_{i,0}\} f(\eta_i, \varepsilon) d\eta_i d\varepsilon$$

Where the ε is a vector, so this involves an $|S| + 1$ dimension integral. The probability of observing a list ℓ_i is equivalent to $Pr(d_i = 1) \times Pr(\ell_i | d_i = 1)$ which can be expressed as

$$\int\int 1\{V(X_i, \varepsilon) - \eta_i \geq u_{i0} + \beta_{\eta,0}\eta_i + \varepsilon_{i,0}\} Pr(\ell_i | \eta_i, \varepsilon) f(\eta_i, \varepsilon) d\eta_i d\varepsilon$$

Where the $Pr(\ell_i | \eta_i, \varepsilon)$ is simply an accounting exercise, to see if that particular combination will yield the highest utility.

To simulate this, you draw the entire vector of errors, and at once determine both the entry decision and the list submitted (only if there is a decision to enter do you consider the list). Simulating and averaging the outcomes over many draws for each student will give us probability of entry and optimal list conditional on entry. It is just an exercise to count how many simulations fall within each list.

It is not clear that this model is identified without very strong parametric assumptions and normalizations. The decision to enter could be from high utility shocks or low cost shocks, and trying to pin down the mean of the cost shock would be difficult without an exclusion restriction. In the next section I show how to incorporate an exclusion restriction into the model.

IV. Exclusion Restrictions

Exclusion restrictions in a school choice model are covariates which affect an individual's decision to enter the mechanism, but do not affect the distribution of utility. There are two ways of accomplishing this. The first is to find some covariate, Z, that only affects the decision to enter. The decision to enter would become

$$EV(X_i, \eta_i) - \eta_i + Z_i\beta_z \geq u_{i0} + \beta_{\eta,0}\eta_i + \varepsilon_{i,0},$$

where Z does not enter into $EV(\cdot)$ nor into $u_{i,0}$. These are covariates that mitigate the cost of entry but do not affect the utility of the alternatives.

The second approach is to adjust the probability of admission in a way that does not affect the utility of each option. For example, in Agarwal and Somaini (2014) the authors use the fact that students who are eligible for free and reduced lunch face different admission probabilities and assume that being a free and reduced lunch student does not affect the distribution of utility for the alternatives.

In practice it is not easy to find suitable exclusion restrictions for school choice. This paper establishes identification without relying on exclusion restrictions, yet they can always be incorporated if there are adequate restrictions.

V. Richer Beliefs

Admission beliefs are a central component to school choice, and there is a recent and growing literature that uses subjective expectations data to understand decision making behavior. While the early literature almost exclusively assumed expectations are either myopic or rational, this approach is problematic as almost any behavior can be consistent with a certain set of beliefs (Manski 1993). This makes it extremely important to use accurate beliefs for performing both the estimation and counterfactuals.

There are many strategies for obtaining objective beliefs. They include simulating rounds of admission and lists, looking at past years admission rates, and schools themselves publishing expected admission rates for certain types of students. Papers can then determine if they want to assume correct and rational beliefs, naive beliefs in which students assume they are accepted to every school, or anything in-between. In addition they can add heterogeneity in that some individuals are sophisticated with correct beliefs, and others are naive.[12]

It is prohibitively expensive to solicit the subjective beliefs of every individual who takes part in the mechanism and those who do not. The work that has been done on incorporating expectations data helps clarify the individual's thought process under uncertainty.[13] The approach of these papers is to collect data on the expectations of the alternative chosen and the alternatives that were not chosen, in addition to beliefs under counterfactual regimes. In a more recent article Wiswall and Zafar (2014) solicit subjective beliefs both before and after an intervention in which they provide objective information. This experiment allows them to see the updating that students do, and, among other things, can be used to see how interventions might change beliefs.

This is an active and large area of research, so I leave studying the beliefs for future work. What is important in this model is that any beliefs can be entered into the model and the corresponding probabilities can be estimated. Another potential path is that if beliefs are collected on a subset of the students, those students can be used to estimate the structural parameters of the model. Applying those parameter estimates to the rest of the student population will allow estimation of some bounds on their subjective beliefs.

VI. Beyond Pure Deferred Acceptance

If there existed a binding constraint on the amount of schools you could apply to, then both the expected value and probabilities would change. The new expected value formula would be

$$EV(X,\eta) = \max_a \left(\sum_s \pi_{isa} \mathbb{E}(U_s) \right), \tag{11}$$

where a is the optimal list submitted, and $|a|$ is less than or equal to the maximum amount of

[12] Calsamiglia, Fu and Guell (2015) have naive and sophisticated types, Agarwal and Somaini (2015) assume naive as the baseline, He 2015 allows for heterogeneous levels of specification.

[13] See Manski, 2004, for a survey of this literature. In the context of schooling choices, Zafar (2011, 2013), Giustinelli (2010), Arcidiacono, Hotz, and Kang (2011), Kaufmann (2012), and Stinebrickner and Stinebrickner (2012, 2014) incorporate subjective expectations into models of choice behavior.

schools that can be applied to. Let π_{isa} denote the admission probability for student i to school s if they apply there, and 0 otherwise.

The submitted rank order lists will also reflect this probability of admission. Truth-telling is no longer a dominant strategy, even if the underlying mechanism is still Deferred Acceptance. In general, once the pure deferred acceptance setting is abandoned, then admission probabilities become important at every stage of the decision making process.

For other mechanisms beyond pure deferred acceptance, admission probabilities are taken into account. For example, in the setting of deferred acceptance where the length of the list is restricted to just one school, we are in the Boston mechanism. In the Boston mechanism with list size greater than one, we once again use equation (11). This time, π_{isa} will be the beliefs for acceptance to school s both depending if it was listed and in what order it was listed.

VII. Long Term Effects

Following recent work by Fack, Grenet, He (2015), I can evaluate both the short-run and long-run equilibrium effects. The short-run equilibrium effects concern immediate behavior changes when I change some aspect of the mechanism such as removing alternatives or altering the probability of admission. The long-run effects take into account both the changes in the mechanism, and the changing covariates of schools as the population that attends it changes.

For example, if a school decides to guarantee admission for low-socioeconomic students, I can estimate how the applications decisions this year would change. In addition, I can link those decisions to changing population demographics at the school and iterate forward to see how the student body and application decisions might change over time.

VI. Conclusion

A fundamental step in the empirical analysis of school choice models is accounting for selection bias. There are innate differences between the students that enter these school choice mechanisms and those that abstain. Without accounting for the selection aspect of school choice, the parameters may be biased and the analysis can be misleading.

This paper presented a simple, yet generalizable model of school choice with endogenous selection. The monte carlo study showed how not accounting for selection can severely bias the estimates and the estimated counterfactual results can be far from the truth. Almost the entire school choice literature can be viewed through this selection model, and it is necessary to consider school choice from a larger perspective, not just from those that provide applications.

There are many extensions this model can take. In section V I outlined some current branches of research and how to apply the selection model to specific scenarios. Probably the most important and interesting extension is applying correct beliefs to these models, which will also affect the results. Modeling both selection and beliefs requires additional data which is not part of the traditional school choice framework. This analysis is necessary to get a more complete view of the true decision making process of individuals in school choice.

VII. Literature

Agarwal, N., Somaini, P. (2014). Demand Analysis using Strategic Reports: An application to a school choice mechanism. Cambridge, MA. doi:10.3386/w20775

Amemiya, Takeshi. Advanced econometrics. Harvard university press, 1985.

Berry, Steven T., and Philip A. Haile. Nonparametric identification of multinomial choice demand models with heterogeneous consumers. No. w15276. National Bureau of Economic Research, 2009.

Casalmiglia, C., Fu, C., Guell, M. (2014). Structural Estimation of a Model of School Choices: the Boston Mechanism vs. Its Alternatives. Documentos de Trabajo (FEDEA), ISSN 1696-7496, No. 21, 2014, Pags. 1-63, (21), 1-63.

Engelbrecht-Wiggans, R. (1993). Optimal Auctions Revisited. Games and Economic Behavior, 5(2), 227-239. doi:10.1006/game.1993.1013

Fack, G., Grenet, J., He, Y. (2015). Beyond Truth-Telling: Preference Estimation with Centralized School Choice.

Ferreyra, M. M., Kosenok, G. (2015). Charter School Entry and School Choice: The Case of Washington, D.C.

Gale, D., Shapley, L. S. (1962). College Admissions and the Stability of Marriage. The American Mathematical Monthly, 69(1), 9. doi:10.2307/2312726

He, Y. (2015). Gaming the Boston School Choice Mechanism in Beijing. Working Paper.

Kapor, A., Neilson, C., Zimmerman, S. (2016). Heterogeneous Beliefs and School Choice.

Levin, D., Smith, J. L. (1996). Optimal Reservation Prices in Auctions. The Economic Journal, 106(438), 1271. doi:10.2307/2235520

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. The Review of Economic Studies, 60(3), 531. doi:10.2307/2298123

McFadden, D. (1974). The measurement of urban travel demand. Journal of Public Economics, 3(4), 303-328. doi:10.1016/0047-2727(74)90003-6

Mckinney, C. N., Niederle, M., Roth, A. E. (n.d.). The Collapse of a Medical Labor Clearinghouse (and Why Such Failures Are Rare).

Mehta, N. (n.d.). Competition in Public School Districts: Charter School Entry, Student Sorting, and School Input Determination. Retrieved from http://ir.lib.uwo.ca/economicscibc

Narita, Y. (2016). Match or Mismatch: Learning and Inertia in School Choice.

Niederle, M., Roth, A. E. (2003). Unraveling Reduces Mobility in a Labor Market: Gastroenterology with and without a Centralized Match. Journal of Political Economy, 111(6), 1342-1352. doi:10.1086/378530

Palfrey, T. R., Pevnitskaya, S. (2008). Endogenous entry and self-selection in private value auctions: An experimental study. Journal of Economic Behavior Organization, 66(3), 731-747. doi:10.1016/j.jebo.2006.06.005

Pevnitskaya, S. (n.d.). Endogenous Entry in First - Price Private Value Auctions : the Self - Selection Effect.

Train, K. (2002). Discrete Choice Methods with Simulation.

Walters, C. (2012). A Structural Model of Charter School Choice and Academic Achievement.

Wiswall, M., Zafar, B. (2015). Determinants of College Major Choice: Identification using an Information Experiment. The Review of Economic Studies, 82(2), 791-824. doi:10.1093/restud/rdu044

Yen, S. T. (2005). A Multivariate Sample-Selection Model: Estimating Cigarette and Alcohol Demands with Zero Observations. American Journal of Agricultural Economics, 87(2), 453-466. doi:10.1111/j.1467-8276.2005.00734.x

Yen, S. T., Lin, B.-H. (2006). A Sample Selection Approach to Censored Demand Systems. American Journal of Agricultural Economics, 88(3), 742-749. doi:10.1111/j.1467-8276.2006.00892.x

A. Appendices

I. Monte Carlo Details

The monte carlo in section IV contains the following primitives. There are three alternatives, with the probability of acceptance: $P_1 = .2$, $P_2 = 1$, and $P_3 = .9$, where school 2 is the safety school.

The utility is specified as:

$$U_{i,s} = X_{i,s}\beta_x + \eta_i\beta_{\eta,s} + \varepsilon_{i,s}$$

with $\beta_X = 3$, $\beta_{\eta,1} = 0$, $\beta_{\eta,2} = -.3$, and $\beta_{\eta,3} = 3.1$. The cost shock $\eta \sim N(.5, 1.75)$, and $\varepsilon_{i,s}$ is distributed T1EV. The covariates are distributed independently with $X_{i,1} \sim N(2,5)$, $X_{i,2} \sim N(2,5)$, and $X_{i,3} \sim N(1,5)$.

Each simulation contains 1000 students, and I view for every student i: $X_{i,s}$ for $s = \{1,2,3\}$, the decision to enter d_i, and if $d_i = 1$ I also observe the full ranked list of schools 1, 2, and 3.

I repeat the simulation with a new draw of X, η, and ε for each simulation and I report the results.

The counterfactual is performed by randomly selecting one of the 100 monte carlo simulations, and treating that as the true data. I estimate the parameters of interest, and treating the estimated parameters as the true DGP, I draw a series of ε and η and estimate the results of a regime change. The overall estimated impact is the average across all these draws.

II. Normality Assumption on School Utility Shocks

We can repeat the above analysis with a normality assumption instead of T1EV. In the normally distributed setting, the η_i cost shock does not directly enter into the utility, but it is correlated with the utility error. I leave this section in the most general terms, but in practice the variance-covariance matrix requires normalizations in order to identify this model.[14]

III. Second Stage

In the second stage we will only observe the students who decide to submit a rank order list. The cost shock η_i from the first period is no longer relevant after it has been paid.

[14]See Train 2002 for a discussion of identification pitfalls

The utilities from enrolling in each school $s \in S$ is denoted as follows:

$$U_{is} = u_{is} + \varepsilon_{i,s}, \tag{12}$$

where u_{is} denotes the observable utility,

$$u_{is} = \beta_{0,s} + X_{is}\beta_x + H_i\beta_{H,s}. \tag{13}$$

The school errors $\varepsilon_{i,s}$ are the utility shocks from attending school s. The cost shock and utility shocks are distributed multivariate random normal:

$$\begin{pmatrix} \eta \\ \varepsilon_1 \\ \vdots \\ \varepsilon_S \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\eta,\eta} & \gamma_1 & \cdots & \gamma_S \\ & \sigma_{11} & \cdots & \gamma_{1,S} \\ \vdots & & \vdots & \\ & & & \sigma_{SS} \end{pmatrix} \right]. \tag{14}$$

Once a student views all his school specific utility shocks in the second stage, he is tasked with creating the list that he submits to the mechanism. The optimal list under the Gale-Shapely mechanism is truth-telling, which does not require the student to consider the admission probabilities.[15]

Following normality, it is possible to decompose ε_s in terms of η:

$$\varepsilon_s = \frac{\gamma_s}{\sigma_{\eta\eta}}\eta + \nu_s$$

with $\nu_s \sim N(0, \sigma_{ss} + \frac{\gamma_s^2}{-\sigma_{\eta\eta}} - 2\gamma_s)$ and ν_s is independent of η.

We can rewrite the utility for school s as:

$$U_{is} = u_{is} + \frac{\gamma_s}{\sigma_{\eta\eta}}\eta_i + \nu_{is}$$

and we will note the S-length vector of unobserved utility shocks, ν, has mean 0 and variance matrix Ω_v.

[15] Under a Gale-Shapely Deferred Acceptance Algorithm, truth-telling is a weakly dominate strategy, and if the probability of acceptance is strictly between 0 and 1, then truth-telling is strictly dominant. It is an simple extension to forgo the truth-telling assumption in place of stability assumption.

Now, conditional on the η_i, I will show how to get the probability of a certain ranking. For illustration assume there are only 4 schools. A student ranks schools as follows: $\ell_i = 3 \succ_i 2 \succ_i 4 \succ_i 1$. The probability of observing that ranking is equivalent to: $Pr(U_3 > U_2 > U_4 > U_1)$. As we are only interested in the utility differences, this probability is equivalent to $Pr(U_2 - U_3 < 0, U_4 - U_2 < 0, U_1 - U_4 < 0)$. Let $\tilde{U}_{s,s'} = U_s - U_{s'}$.

If we denote the deterministic part of utility, $u_{is} + \frac{\gamma_s}{\sigma_{\eta\eta}}\eta_i$, as V_s, then we can represent the vector of school utility stacked 1 to 4, as $U_i = V_i + \nu_i$, where $\nu_i \sim N(0, \Omega_v)$. Define the 3×4 matrix

$$M = \begin{pmatrix} 0 & 1 & \text{-1} & 0 \\ 0 & \text{-1} & 0 & 1 \\ 1 & 0 & 0 & \text{-1} \end{pmatrix}.$$

With this matrix, the probability of the ranked alternatives becomes

$$\begin{aligned} Pr(\ell_i = 3,2,4,1|\eta) &= Pr(\tilde{U}_{2,3} < 0, \tilde{U}_{4,2} < 0, \tilde{U}_{1,4} < 0|\eta) \\ &= Pr(MU_i < 0|\eta) \\ &= Pr(MV_i + M\nu_i < 0|\eta) \\ &= Pr(M\nu_i < -MV_i|\eta), \end{aligned}$$

and $M\nu_i$ is distributed jointly normal with zero mean and covariance matrix $M\Omega_\nu M$. Note that the matrix M will change for every individual based on the preferences submitted. The $Pr(M\nu_i < -MV_i)$ can be simulated by GHK. To identify the model I will have to make some assumptions on the variance structure, because although the variance covariance matrix of S schools will contain $S(S+1)/2$ parameters, generally only $[(S-1)S/2] - 1$ parameters are identified on account of the normalizations of scale and the fact we are only interested in utility differences.[16]

IV. First Stage

While we are able to simulate $Pr(\ell_i|\eta_i)$, not all values of η_i will lead to entering the mechanism. A student will enter the mechanism if the expected value from paying the cost and entering the

[16]See Train section 5.6.3 for details on running the algorithm and discussion of identification.

mechanism is higher than the expected utility of the outside option. Let U_{i0} denote the utility of student i's safety school. Let $d_i = 1$ denote that student i submits an application. I can express an application as

$$\begin{aligned} d_i = 1 \text{ if } EV(X_i, \eta_i) - \eta_i &\geq E[U_{i0}|\eta_i] \\ \equiv EV(X_i, \eta_i) - \eta_i &\geq u_{i0} + E[\varepsilon_{i,0}|\eta_i] \\ \equiv EV(X_i, \eta_i) - \eta_i &\geq u_{i0} + \tfrac{\gamma_s}{\sigma_{\eta\eta}}\eta_i + E[\nu_{is}|\eta_i] \\ \equiv EV(X_i, \eta_i) - \eta_i &\geq u_{i0} + \tfrac{\gamma_s}{\sigma_{\eta\eta}}\eta_i, \end{aligned} \tag{15}$$

where $Pr(d_i = 1)$ can be simulated with many draws of η_i. The likelihood contribution of an individual who submits a rank order list ℓ_i is

$$\begin{aligned} Pr(\ell_i|d_i = 1) =& \frac{Pr(\ell_i, d_i = 1)}{Pr(d_i = 1)} \\ =& \frac{\int_{-\infty}^{\infty} 1\{EV(X_i, \eta_i) - \eta_i \geq u_{i0} + \frac{\gamma_s}{\sigma_{\eta\eta}}\eta_i\} Pr(\ell_i|\eta_i) dF(\eta_i)}{\int_{-\infty}^{\infty} 1\{EV(X_i, \eta_i) - \eta_i \geq u_{i0} + \frac{\gamma_s}{\sigma_{\eta\eta}}\eta_i\} dF(\eta_i)}, \end{aligned}$$

which can be simulated as in the standard model in section III. The utility function can be parametrized with parameters θ_u. The unknown parameters we want to solve for are thus (θ_u, Σ) where Σ denotes the parameters in the variance-covariance matrix. The likelihood for individual i can be expressed as:

$$\mathcal{L}_i(\Theta, \Sigma) = Pr(d_i = 0)^{1\{d_i=0\}} \big(Pr(d_i = 1) Pr(\ell_i|d_i = 1)\big)^{1\{d_i=1\}}.$$

V. List is Not Always Observed

If we only observe a rank ordered list given 1) an individual paid the cost and 2) utility of safety school is not the highest, then this would change the lists we observe in the data. In particular, we would never see a list with the safety school ranked highest.

Let $t_i = 1$ denote individual i's decision to submit a list. The likelihood for that individual will be the following:

$$\mathcal{L}_i(\Theta, \Sigma) = Pr(t_i = 0)^{1\{t_i=0\}} \big(Pr(t_i = 1) Pr(\ell_i|t_i = 1)\big)^{1\{t_i=1\}},$$

where

$$\begin{aligned} Pr(t_i = 0) = \quad & Pr(\eta \text{ is too high}) + Pr(\eta \text{ is low enough})... \\ & \times Pr(\text{Top Choice is safety school}|\eta \text{ is low enough}) \\ = \quad & \int_{-\infty}^{\infty} 1\{EV(X_i, \eta_i) - \eta_i \leq u_{i0} + \beta_{\eta,0}\eta_i + \gamma\} f(\eta) d\eta \ldots \\ & +(\int_{-\infty}^{\infty} 1\{EV(X_i, \eta_i) - \eta_i \geq u_{i0} + \beta_{\eta,0}\eta_i + \gamma\} f(\eta) d\eta) \ldots \\ & \times \frac{\int_{-\infty}^{\infty} 1\{EV(X_i,\eta_i) - \eta_i \geq u_{i0} + \beta_{\eta,0}\eta_i + \gamma\} \frac{\exp(\delta_0(\eta))}{\sum_k \exp(\delta_k(\eta))} f(\eta) d\eta}{\int_{-\infty}^{\infty} 1\{EV(X_i,\eta_i) - \eta_i \geq u_{i0} + \beta_{\eta,0}\eta_i + \gamma\} f(\eta) d\eta} \end{aligned} \tag{16}$$

and $Pr(\ell_i)$ takes the standard exploded logit form.

VI. Logit Coefficient Normalizations

If the covariates do not vary by both alternative **and** individual, the location of these parameters cannot be identified in a logit model. Consider the following covariate H which only varies by individual. I will compare the true parameters β_a for alternative $a = \{1, 2\}$, with another set of parameters $\tilde{\beta}_a = \beta_a + c$ for $a = \{1, 2\}$ and some constant $c \in \mathcal{R}$. The probability of choosing option 1 under the true parameters is

$$Pr(a_i = 1|\beta) = \frac{\exp\{\beta_1 H_i\}}{\exp\{\beta_1 H_i\} + \exp\{\beta_2 H_i\}},$$

while under the alternative parameters,

$$\begin{aligned} Pr(a_i = 1|\tilde{\beta}) \quad & = \frac{\exp\{\tilde{\beta}_1 H_i\}}{\exp\{\tilde{\beta}_1 H_i\} + \exp\{\tilde{\beta}_2 H_i\}} \\ & = \frac{\exp\{(\beta_1 + c) H_i\}}{\exp\{(\beta_1 + c) H_i\} + \exp\{(\beta_2 + c) H_i\}} \\ & = \frac{\exp\{\beta_1 H_i\} \exp\{c H_i\}}{\exp\{\beta_1 H_i\} \exp\{c H_i\} + \exp\{\beta_2 H_i\} \exp\{c H_i\}} \\ & = \frac{\exp\{\beta_1 H_i\}}{\exp\{\beta_1 H_i\} + \exp\{\beta_2 H_i\}} \qquad = Pr(a_i = 1|\beta) \end{aligned}$$

This same proof applies for school specific intercepts, which only vary by alternative and not by individual, by setting $H_i = 1$ for all i.

VII. Calculate Expected Value Fuction

To simulate the expected values, I need the utility vector δ and acceptance probabilities vector P. I take a draw of ε of the same length as δ from the $F_\varepsilon(\cdot)$ distribution and I order the P and $\delta + \varepsilon$

vectors according to $\delta + \varepsilon$.

Finally I need to create a vector that is the cumulative product of $(1 - P)$ with a 1 placed on top of the vector and the last element is deleted. Call this vector NP.

Assuming that with this specific draw of ϵ, the options are ranked 1, 2, 3, then

$$NP = [1, (1 - P_1), (1 - P_1)(1 - P_2)]'$$

the expected utility for this draw of error values is thus

$$(\delta + \epsilon) \circ P \times NP$$

where $\circ$ denotes element-wise multiplication. I then average this over many draws of ε to simulate the expected value.

For a numerical example, consider a draw of ε such that $\delta + \varepsilon = \{6, 7, 5\}$ and $P = \{.6, .3, 1\}$. I will re-order both vectors such that $\delta + \varepsilon = \{7, 6, 5\}$ and $P = \{.3, .6, 1\}$.

Next construct $NP = \{1, (1 - .3), (1 - .3)(1 - .6)\} \equiv \{1, .7, .28\}$. Then $(\delta + \epsilon) \circ P = \{7 \times .3, 6 \times .6, 5 \times 1\} \equiv \{2.1, 3.6, 5\}$.

Finally, to get $(\delta + \epsilon) \circ P \times NP$ I take the inner product of $\{2.1, 3.6, 5\}$ and NP, which is $2.1 \times 1 + 3.6 \times .7 + 5 \times .28 = 6.02$. The expected utility will then be the average over many draws of ε.