Estimating Welfare Effects in a Nonparametric Choice Model: The Case of School Vouchers*

Vishal Kamat
Toulouse School of Economics
University of Toulouse Capitole
vishal.kamat@tse-fr.eu

Samuel Norris
Harris School of Public Policy
University of Chicago
samnorris@uchicago.edu

June 16, 2021

Abstract

We develop new robust discrete choice tools to learn about the average willingness to pay and average cost of a school voucher in a program that randomly allocates vouchers. We consider a nonparametric, nonseparable choice model that places no restrictions on the functional form of utilities or the distribution of unobserved heterogeneity. We exploit the insight that the welfare parameters in this model can be expressed as functions of the demand for the different schools. However, while the random allocation of the voucher reveals the value of demand at two prices, the parameters generally depend on its values beyond these prices. We show how to sharply characterize what we can learn when demand is specified to be entirely nonparametric or to be parameterized in a flexible manner, both of which imply that the parameters are not necessarily point identified. We use our tools to analyze the welfare effects of voucher provision in the DC Opportunity Scholarship Program, a school voucher program in Washington, DC. We find that the provision of both the status-quo voucher and a wide range of counterfactual vouchers of different amounts have positive benefits net of costs. In comparison, traditional logit models produce estimates towards the lower end of our bounds, and hence may understate the benefits. We also find that the positive results can be explained by the popularity of low-tuition schools in the program; removing them from the program can result in a negative net benefit.

KEYWORDS: Discrete choice analysis, welfare analysis, demand analysis, nonparametrics, partial identification, school vouchers, Opportunity Scholarship Program.

JEL classification codes: C14, C25, D12, D61, I21.

*We thank Ivan Canay, Isis Durrmeyer, Thierry Magnac, Adam Rosen, Max Tabord-Meehan, Alex Torgovitsky, and participants at several seminars and conferences for useful comments and discussions. Vishal Kamat gratefully acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program). First version: arXiv:2002.00103 dated January 31, 2020.

1 Introduction

School vouchers are a topic of active policy debate across several countries. In their basic form, they are government-funded certificates of a certain amount that students can use to offset tuition at an eligible private school of their choice. By reducing the price of private schools and making them more affordable, advocates argue that vouchers foster school choice and make recipients better off (Friedman, 1962).

To empirically investigate this claim, a number of studies have estimated the effects of vouchers on various outcomes using data from programs that randomly allocate vouchers (e.g., Abdulkadiroğlu et al., 2018; Angrist et al., 2002; Dynarski et al., 2018; Howell et al., 2000; Krueger and Zhu, 2004; Mayer et al., 2002; Mills and Wolf, 2017; Muralidharan and Sundararaman, 2015; Wolf et al., 2010). However, as surveyed in Epple et al. (2017), the evidence from these studies is mixed. Some find positive effects, while others find null or even negative effects. Nonetheless, despite this mixed evidence on the effects on outcomes, the data in each of these studies indicate that a non-trivial proportion of recipients choose to use the voucher. By revealed preference arguments, this suggests that recipients in general value vouchers and, in turn, that vouchers may be welfare-enhancing. Yet, little empirical work has attempted to quantify these welfare benefits and analyze whether they can justify the costs of providing vouchers.

In this paper, we develop new discrete choice tools to quantify the welfare effects of school vouchers, and use them to study a voucher program in Washington, DC. We make two main contributions: a methodological and an empirical one.

Methodologically, we show how to robustly learn about the average welfare benefit to recipients and cost to the government of a voucher in programs that randomly allocate vouchers. We measure benefits by the willingness to pay for the voucher, a natural welfare money metric that relates to the compensating variation of the decrease in school prices induced by the voucher. We show how to characterize what we can learn about these quantities when the choice model remains entirely nonparametric as well as under flexible parameterizations of demand. As we discuss below, these tools and the theoretical arguments behind them are novel and of independent interest in discrete choice analysis beyond the voucher setup we consider in this paper.

Empirically, we use our tools to analyze the welfare effects of voucher provision in the DC Opportunity Scholarship Program (OSP). Our estimates reveal that both the status-quo voucher and a wide range of counterfactual vouchers of different amounts would have positive benefits net of costs, and that these positive results can be explained by the popularity of low-tuition schools in the program.

Our methodological analysis is motivated by a natural concern one may have with using traditional parametric discrete choice approaches (e.g., Berry et al., 1995; McFadden, 1974; Train, 2009) to learn about our quantities of interest. These approaches are based on fully-parameterized models chosen to guarantee point identification. The parameterizations include known functional forms on the utilities and distributions such as the logistic or normal on the unobserved heterogeneity. However, these parameterizations might be restrictive, and produce results driven by them. Our analysis therefore aims to alternatively learn about our quantities in an entirely nonparametric model, or when the parameterization is allowed to be flexibly chosen.

We begin our analysis by considering a nonparametric, nonseparable model of school choice that imposes no restrictions on the functional form of utilities or on the distribution of unobserved heterogeneity. The sole noteworthy assumption we make is that the voucher affects choices only through decreasing prices of eligible private schools. As we argue in Section 2.2, an implication of this assumption is that our analysis in general can be interpreted as a partial equilibrium one and, in turn, that our analysis can more concretely be viewed as evaluating the welfare effects of policies that marginally change voucher provision.

We then exploit the insight from Bhattacharya (2015, 2018) that in our model, the welfare parameters can be expressed as functions of the demand for the different schools. Given the random allocation of vouchers, the data on choices of recipients and non-recipients reveal the value of demand at two prices, namely the prices with and without the application of the status-quo voucher. The identification problem, however, is that the parameters of interest generally depend on demand values beyond these two prices.

The tools we develop sharply characterize what we can learn about these parameters under a given specification of demand. Our baseline specification is completely nonparametric and takes the demand for each school to be only decreasing with its own price and increasing with the prices of other schools. We also consider auxiliary specifications that additionally allow demand to be generally parameterized through a flexible functional form restriction on how it varies with prices. For both specifications, we develop tractable computational procedures that characterize what we can learn about our parameters.

An important novelty of our procedures is that they account for the complication that under both specifications there may not exist a single point identified demand but multiple demand functions consistent with the data. Indeed, this is generally the case unless one focuses attention solely on arguably restrictive parametric specifications of demand. Our procedures formally show how to exploit the geometry of our parameters and demand specifications so that we can computationally search through the space of admissible demand functions and generate the set of all parameter values consistent with them in the general case, while continuing to generate the unique parameter value in the restrictive case. In the case of the nonparametric specification, our procedure involves an additional technical novelty that entails showing how the nonparametric space of admissible demands can without loss of information be reduced to a parametric space. This importantly ensures that the computational problems are finite dimensional and tractable.

Empirically applying the developed tools to the OSP, our estimates reveal that the provision of the \$7,500 status-quo voucher as well as a wide range of counterfactual vouchers of different amounts have positive welfare benefits net of the costs the government faces to provide them. This conclusion is robust to several choices of flexible parametric demand specifications and continues to hold even under the nonparametric specification. For example, under our most flexible parametric specification, we find that the average benefit net of costs is bounded between \$645 and \$2,887, whereas, under the nonparametric specification, it is bounded between \$213 and \$5,088.

For comparison, we also estimate our parameters using traditional discrete choice tools based on various fully-parameterized logit specifications of our model. We find that the parameterizations in these specifications can play an important role in the conclusions one reaches. In particular, the various specifications all produce estimates of welfare benefits that are towards the lower bounds of our nonparametric specification and even below those of some of our parametric specifications. This suggests that the logit specifications potentially understate the response to price changes, a feature of logit documented in several other empirical settings (e.g., Compiani, 2019; Ho and Pakes, 2014; Tebaldi et al., 2019). As a result, we find that they may substantially understate the benefits of voucher provision in comparison to our estimates.

Our analysis also reveals which features of the program might be driving the positive welfare effects. A closer inspection of the data reveals a high number of popular, low-tuition schools in the program. These schools potentially induce a high welfare benefit for recipients relative to the net costs the government faces to fund a voucher when redeemed at them. Indeed, a key rationale for school vouchers cited by Friedman (1962) is that they may subsidize private schools that provide services individuals value more efficiently than government-funded schools. To measure the importance of these low-tuition schools in the OSP, we estimate how the welfare effects of the program would change if these schools were removed from the program. Our estimates reveal the presence of such schools plays an essential role in explaining our positive findings: absent schools with tuition of at most \$3,500, the program can result in a negative net benefit.

Our analysis contributes to several literatures. Methodologically, we contribute to the growing literature on identification in nonparametric discrete choice models. Similarly to our analysis, this literature is motivated by the concern that results based on the traditional approach of using fully-parameterized models may be driven by specific choices of parameterizations. One approach pursued in this literature is to argue point identification, which is often based on requiring large amounts of exogenous variation in the data (e.g., Berry and Haile, 2009, 2014; Briesch et al., 2010; Chiappori and Komunjer, 2009; Matzkin, 1993). In our empirical setting, however, the random allocation of the voucher induces only binary exogenous variation in prices. We therefore develop tools that allow us to more generally partially identify various welfare effects of price changes under binary variation in prices.

Several papers have similarly developed tools to nonparametrically evaluate alternative discrete choice questions—such as estimating the effect of different prices and choice sets on demand, characterizing the underlying utility functions, and testing the premise of utility maximization—in setups that permit partial identification (e.g., Chesher et al., 2013; Kamat, 2019; Kitamura and Stoye, 2018; Manski, 2007; Tebaldi et al., 2019). As in our analysis, a key theoretical challenge in these papers is how to develop a sharp and tractable characterization of what we can learn about the question of interest given the nonparametric nature of the model. The tools we develop and the theoretical arguments to justify them, however, are novel and distinct from those in these papers, which rely on specific features of their respective setups and do not straightforwardly apply to our setting.

In order to develop our results, we build on recent arguments from the literature on nonparametric welfare analysis. In particular, Bhattacharya (2015, 2018) show that various welfare effects of price changes can be nonparametrically expressed as known functions of each good's demand. If demand is point identified, then one can indeed directly apply these results. The basis of our identification analysis is to show how to identify various such welfare effects even when demand is not necessarily point identified. Recently, Bhattacharya (2019) also derives analytic nonparametric bounds for welfare effects in such cases with two goods where one of them is a numeraire good. These results however do not straightforwardly extend to the case with multiple goods and prices present in our setup. We show how the geometry of the known functions characterizing the welfare effects along with that of our nonparametric demand specification can be exploited to develop a computational procedure that delivers bounds. In addition, we also show how flexible parametric specifications can be incorporated into the analysis. In this direction, we exploit ideas from Mogstad et al. (2018), who show how parametric restrictions can be incorporated in the alternative setting of a treatment effect model.

From an empirical standpoint, we contribute to the literature on the evaluation of school voucher programs. As highlighted above, most papers in this area estimate the effects of vouchers on outcomes. However, these estimates do not address the welfare implications of these programs. We complement these papers by providing welfare estimates of a specific program, and developing general tools that can be used to analyze programs beyond the one we study.

A smaller group of papers uses choice models to study various voucher-related school choice questions of interest (e.g., Allende, 2019; Arcidiacono et al., 2016; Carneiro et al., 2019; Gazmuri, 2019; Neilson, 2013). As we highlight in Section 2.2, these papers analyze questions that generally go beyond the scope of our analysis, which can be interpreted as asking how much individuals value a voucher in a partial equilibrium setup. However, they do so using fully-parameterized, point identified models of choice. Our analysis can therefore be viewed as complementary as we analyze a narrower, yet relevant question, but do so using arguably more robust, credible tools. Indeed, our empirical results reveal that welfare estimates based on fully parameterized versions of our model can be potentially misleading in comparison to those our tools provide.

We organize our analysis as follows. Section 2 describes our model of school choice and demand specifications. Section 3 defines our parameters measuring the welfare effects of voucher provision. Section 4 presents our identification analysis. Section 5 presents our empirical results. Section 6 concludes and discusses directions for future research. Proofs and additional details are presented in the Supplementary Appendix.

2 Model and Demand Functions

2.1 Model of School Choice

Suppose the set of schools where individuals can enroll can be partitioned into government-funded schools, and private schools that do and do not participate in the voucher program. Let $\mathcal{J}_g$ denote the set of government-funded schools, $\mathcal{J}_n$ denote the set of private schools not participating in the voucher program, and $\mathcal{J}_v$ denote the set of private schools participating in the voucher program. The status-quo voucher program provides an amount of at most $\tau_{\text{sq}} \in \mathbf{R}_+$ to cover the price (the tuition) for any school in $\mathcal{J}_v$. For the jth school in $\mathcal{J}_v$, let $p_j^* \in \mathbf{R}_+$ denote its original price before applying the voucher and let $p_j(\tau)$ denote its price after applying a voucher of amount $\tau \in \mathbf{R}_+$, where these two prices are related by the relationship $p_j(\tau) = \max\{0, p_j^* - \tau\}$. Under this notation, the original price and that under the status-quo amount for the jth school in $\mathcal{J}_v$ are given by $p_j(0)$ and $p_j(\tau_{\text{sq}})$, respectively. For notational convenience, we use $\mathcal{J}_s = \mathcal{J}_g \cup \mathcal{J}_n \cup \mathcal{J}_v$ to denote the set of all schools. In addition, we take $\mathcal{J}_v = \{1, \ldots, J\}$, where the schools in this list are ordered in terms of their original prices, i.e. $p_1^* \leq \ldots \leq p_J^*$, and we take $p(\tau) = (p_1(\tau), \ldots, p_J(\tau))$ to denote the vector of prices for these schools under a voucher of amount τ.

Each individual i in the population is associated with observables Z_i and D_i, which respectively denote an indicator for whether the individual received a voucher and the school in $\mathcal{J}_s$ where the individual enrolled. We assume that the observed enrollment choice is the product of an underlying individual-level utility maximization decision. To this end, let Y_{ij} denote the individual's underlying disposable income under the jth school in $\mathcal{J}_g$ or $\mathcal{J}_n$, and let $U_{ij}(Y_{ij})$ denote the corresponding indirect utility under that school. For schools in $\mathcal{J}_v$, we can define similar quantities but we need to explicitly account for the role their prices play as they are altered by the receipt of the voucher. Specifically, let $Y_{ij} - p_j$ denote the individual's underlying disposable income under the jth school in $\mathcal{J}_v$ had the price of that school been set to p_j, and let $U_{ij}(Y_{ij} - p_j)$ denote the corresponding indirect utility under that school given that price. Note that the indirect utilities here can be interpreted as the ex-ante utility of enrolling in that school. Moreover, they can be interpreted as the indirect utility of enrolling in that school net of whether the individual can be admitted there, i.e. it corresponds to $-\infty$ if the individual cannot be admitted to that school.

Using these quantities, we can define the individual's utility maximizing choice had the prices of the schools in $\mathcal{J}_v$ been set to the vector $p = (p_1, \ldots, p_J)$ by

$$D_i(p) = \begin{cases} \arg\max\limits_{j \in \mathcal{J}_g \cup \mathcal{J}_n} U_{ij}(Y_{ij}) & \text{if} \quad \max\limits_{j \in \mathcal{J}_g \cup \mathcal{J}_n} U_{ij}(Y_{ij}) > \max\limits_{j \in \mathcal{J}_v} U_{ij}(Y_{ij} - p_j) \ , \\ \arg\max\limits_{j \in \mathcal{J}_v} U_{ij}(Y_{ij} - p_j) & \text{if} \quad \max\limits_{j \in \mathcal{J}_g \cup \mathcal{J}_n} U_{ij}(Y_{ij}) \leq \max\limits_{j \in \mathcal{J}_v} U_{ij}(Y_{ij} - p_j) \ . \end{cases}$$

The observed enrollment choice is then assumed to be related to the underlying utility maximizing choices and voucher receipt by the relationship

$$D_i = D_i\left(p(\tau_{\text{sq}})\right) \cdot Z_i + D_i\left(p(0)\right) \cdot (1 - Z_i) \ . \tag{1}$$

2.2 Model Discussion

We highlight three features of our model. First, it is entirely nonparametric and imposes no restrictions on the functional form of utilities or on the distribution of unobserved heterogeneity. This is in contrast to the standard approach in discrete choice analysis that fully parameterizes the choice model, with parameterizations chosen such that the model can be point identified. For example, following the practitioner's guide in Nevo (2000), a common parameterization of our setup corresponds to taking

$$U_{ij}(Y_{ij} - p_j) = \gamma_{1i} - \gamma_{2i} p_j + x_j' \gamma_{3i} + \xi_j + \epsilon_{ij} \qquad \text{for} \quad j \in \mathcal{J}_v \ , \tag{2}$$

$$U_{ij}(Y_{ij}) = \gamma_{1i} + x_j' \gamma_{3i} + \xi_j + \epsilon_{ij} \qquad \text{for} \quad j \in \mathcal{J}_g \cup \mathcal{J}_n \ , \tag{3}$$

where x_j and ξ_j denote observed and unobserved school level covariates, respectively, $\gamma_i = (\gamma_{1j}, \gamma_{2j}, \gamma_{3j}')'$ denotes individual level coefficients assumed to be linear in observed individual level covariates X_i and, potentially, a normally distributed stochastic term, and ϵ_{ij} denotes idiosyncratic shocks assumed to have either a Type I extreme value or normal distribution—see Appendix S.4.3 for further details. In comparison to our model, observe that such a model makes a combination of assumptions including specific functional forms on how the utilities vary with individual and school level observed and unobserved covariates, and distributional assumptions on the unobserved heterogeneity across individuals and schools.

Second, we suppose there exists only binary variation in prices, namely the original school prices and these prices under the status-quo voucher. Along with the nonparametric nature of our model, this feature has an important consequence for our subsequent analysis. In particular, it precludes a common approach to identification pursued in nonparametric choice models, which is based on arguing point identification of the parameters of interest by supposing there exists large amounts of variation in the data—see Section 1 for references. The discrete variation, in contrast, will generally imply that our welfare parameters can only be partially identified. A key challenge of our analysis is to show how to characterize what we can learn about the parameters in this case. Our identification analysis in Section 4 captures this point more precisely.

Finally, it is useful to note how we model the variation in prices and its implications on the interpretation of our analysis. We take the voucher to affect utilities only through decreasing school prices and in turn increasing disposable income. A natural implication of this is that our analysis does not capture effects of channels, beyond the change in disposable income, through which the voucher program may affect individual choices. Examples of such channels noted previously in the literature primarily correspond to various general equilibrium type effects such as changes in the set of schools to where an individual can be admitted due to sufficiently large changes in individual application and school admission behavior, and changes in the utilities under different schools due to changes in school incentives to invest in quality (Allende, 2019; Neilson, 2013) or changes in peer composition (Allende, 2019; Gazmuri, 2019). Indeed, capturing such channels requires a richer model that explicitly introduces them in the structure of the model.

As we do not explicitly model these various general equilibrium channels, our resulting analysis should be viewed as taking them as fixed. To this end, our analysis can more appropriately be viewed as a short term partial equilibrium one that takes various general equilibrium channels as fixed. More concretely, it can be viewed as analyzing the welfare effects of a marginal policy that provides a voucher to an additional student who applied to the program but was not admitted.[1] As we highlight in Section 6, extending our analysis to capture general equilibrium effects of vouchers is an interesting direction for future work.

2.3 Average Demand Functions

Our analysis is based on the demand functions for the different schools in the sense that we use them to state our assumptions and define our parameters of interest. These functions correspond to the distribution of enrollment choices across individuals for the different schools at each potential price vector. More formally, let $\mathcal{P} = \prod_{j=1}^{J} [0, p_j(0)] \subset \mathbf{R}_+^J$ denote the domain of price vectors for the schools in $\mathcal{J}_v$ over which we define these functions. Then, for a given $p \in \mathcal{P}$, let

$$\begin{aligned} q_j(p) &= \text{Prob}\{D_i(p) = j\} \ , \\ q_g(p) &= \text{Prob}\{D_i(p) \in \mathcal{J}_g\} \ , \\ q_n(p) &= \text{Prob}\{D_i(p) \in \mathcal{J}_n\} \end{aligned}$$

respectively define the demand for the jth school in $\mathcal{J}_v$, for any school in $\mathcal{J}_g$ and for any school in $\mathcal{J}_n$. Analogously, let $q_j(p|z)$ for $j \in \mathcal{J}_v$, $q_n(p|z)$ and $q_g(p|z)$ respectively define these demand functions conditional on the receipt of the voucher $Z = z \in \{0, 1\}$. Note we only define demand for any school in $\mathcal{J}_g$ and $\mathcal{J}_n$, and not for each specific school in these sets of schools. As we will

[1] It is useful, however, to highlight that such equilibrium channels may potentially be less relevant for many small voucher programs such as the OSP. Nonetheless, they can become relevant here if one is interested in understanding the effects of scaling up such programs.

observe, this is because defining demand over this more parsimonious grouping is sufficient for the definition of our welfare parameters. For notational convenience, let $\mathcal{J} = \{g, n\} \cup \mathcal{J}_v$ denote the set of indices over which the demand functions are defined.

In the following assumption, we state the restrictions we impose on the demand functions under our baseline specification. In particular, note that this specification is entirely nonparametric.

Assumption B. (Baseline)

(i) For each $j \in \mathcal{J}$, $q_j(p|z) = q_j(p)$ for all $p \in \mathcal{P}$ and $z \in \{0, 1\}$.

(ii) For each $j \in \mathcal{J}$, q_j is weakly increasing in p_m for each $m \in \mathcal{J}_v$ such that $m \neq j$.

Assumption B(i) states that the demand functions are invariant to the receipt of the voucher. It follows from this assumption that the underlying demand functions can be uniquely captured by the vector $q \equiv (q_g, q_n, q_1, \ldots, q_J)$ of unconditional demand functions. As a result, in the remainder of our analysis, we focus solely on the unconditional demand; whenever we refer to demand, it is understood we are referring to the unconditional demand. Assumption B(ii) imposes shape restrictions on how demand behaves with the prices of the private schools in the voucher program. In particular, it imposes that for each $p, p' \in \mathcal{P}$ such that $p_j > p'_j$ for $j \in \mathcal{J}' \subseteq \mathcal{J}_v$ and $p_j = p'_j$ for $j \in \mathcal{J}_v \setminus \mathcal{J}'$, we have that

$$q_j(p) \geq q_j(p') \tag{4}$$

for each $j \in \mathcal{J} \setminus \mathcal{J}'$. Since by definition we have that

$$q_j(p) = 1 - \sum_{m \in \mathcal{J} \setminus \{j\}} q_m(p)$$

for each $j \in \mathcal{J}_v$, note that it directly follows from Assumption B(ii) that q_j is weakly decreasing in p_j for $j \in \mathcal{J}_v$, i.e. the standard shape restriction from demand theory that states demand for each good is weakly decreasing with respect to its own price. Moreover, note that Assumption B(ii) corresponds to a version of the weak substitutes assumption from Berry et al. (2013). While Assumption B imposes restrictions directly on the demand functions, note that these restrictions follow from assumptions on the underlying variables of the model. For example, Assumption B(i) follows from assuming the voucher is randomly allocated, i.e. Z_i is statistically independent of the remaining underlying variables of the model. On the other hand, Assumption B(ii) follows from assuming U_{ij} is weakly increasing for each $j \in \mathcal{J}_s$.

As we noted above, a standard approach in the literature on discrete choice analysis is to consider model parameterizations such as those in (2)-(3). Indeed, these parameterizations imply parametric functional form restrictions on the demand functions. Moreover, as they are chosen such that the model is point identified, they also mean that these implied functional forms are point identified. In our analysis, we also consider auxiliary specifications that directly impose parametric restrictions on the demand functions in addition to those in Assumption B, but we do not restrict attention to only those that ensure point identification. In the following assumption, we state the general class of flexible parametric specifications we consider.

Assumption A. (Auxiliary) For each $j \in \mathcal{J}$,

$$q_j(p) = \sum_{k=0}^{K_j} \alpha_{jk} \cdot b_{jk}(p) \tag{5}$$

for some $\{\alpha_{jk} : 0 \leq k \leq K_j\}$, where $\{b_{jk} : 0 \leq k \leq K_j\}$ denote some known functions.

Assumption A states that the demand functions are linear functions of some known functions of prices, where the variable $\alpha \equiv \left(\alpha_g', \alpha_n', \alpha_1', \ldots, \alpha_J'\right)'$, with $\alpha_j = \left(\alpha_{j1}, \ldots, \alpha_{jK_j}\right)'$ for each $j \in \mathcal{J}$, parameterizes the demand functions. As we further discuss in Section 4.3, this assumption allows for several types of parametric specifications. For example, it allows for those that result in point identification of the demand functions such as

$$q_j(p) - q_g(p) = \alpha_{j0} + \alpha_{j1} \cdot p_j \quad \text{for } j \in \mathcal{J}_v \ , \tag{6}$$

$$q_n(p) - q_g(p) = \alpha_n \ , \tag{7}$$

for some $\{\alpha_{jk} : j \in \mathcal{J}_v,\ 0 \leq k \leq 1\}$ and α_n, i.e. the difference in demand for a given school in $\mathcal{J}_v$ and any school in $\mathcal{J}_g$ is a linear functions of that school's own price and the difference in demand for any school in $\mathcal{J}_n$ and any school in $\mathcal{J}_g$ is constant—see Appendix S.2.1 for details on how this specification imposes restrictions similar to a logit specification and, like the logit, achieves point identification. However, Assumption A also allows for more flexible specifications where demand for each school is a polynomial function of own prices as well as prices of all schools, which do not imply point identification.

3 Welfare Effects of Voucher Provision

In the context of our model, the provision of a voucher can make individuals better off by increasing their disposable income when enrolled in schools in the voucher program. In this section, we define the main parameter of interest of our analysis that aims to quantify these potential welfare benefits. We define this parameter for a generic voucher amount $\tau \in \mathbf{R}_+$. As mentioned below, this generality, by choosing different values of τ, allows us to analyze the welfare effects of the status-quo voucher amount as well as alternative counterfactual voucher amounts.

To quantify the benefit for each individual i, we use a money metric for the welfare gains from the receipt of the voucher. Specifically, we use the amount of money that the individual would pay to receive the voucher or, equivalently, the negative of the compensating variation of the reduction in prices induced by the voucher. Formally, the individual's willingness to pay for a voucher of amount τ is defined by the variable $B_i(\tau)$ that solves

$$\begin{gathered}\max\left\{\max_{j\in\mathcal{J}_g\cup\mathcal{J}_n} U_{ij}(Y_{ij}),\ \max_{j\in\mathcal{J}_v} U_{ij}\left(Y_{ij}-p_j(0)\right)\right\}\\ = \\ \max\left\{\max_{j\in\mathcal{J}_g\cup\mathcal{J}_n} U_{ij}(Y_{ij}-B_i(\tau)),\ \max_{j\in\mathcal{J}_v} U_{ij}\left(Y_{ij}-p_j(\tau)-B_i(\tau)\right)\right\}\ ,\end{gathered} \tag{8}$$

i.e. the amount of money to be subtracted from the individual's income under the receipt of the voucher so that they obtain the same utility as in the absence of the voucher. We then quantify the average benefit of a voucher that provides an amount τ by

$$AB(\tau) = E[B_i(\tau)]\ , \tag{9}$$

i.e. the average willingness to pay across individuals to receive that voucher.

As mentioned, our analysis is based on the fact that our parameters of interest can be written as functions of the demand functions introduced in the previous section. In order to show this for the average benefit parameter defined above, we exploit results from Bhattacharya (2015, 2018), who showed that the average value of a variable such as that defined in (8) can be written as a closed form expression of the demand functions. In the following proposition, we formally state this result in terms of our setup and notation. In the statement of this proposition, we use $j(\tau)$ to denote the jth school in $\mathcal{J}_v$ such that $p_{j(\tau)}(0) < \tau$ and $p_{j(\tau)+1}(0) \geq \tau$, i.e. the last school in $\mathcal{J}_v$ for which the voucher amount τ is strictly greater than the tuition amount. In addition, we take $\{a_l(\tau) : 0 \leq l \leq J\}$ to be a set of values such that $a_0(\tau) = 0$, $a_l(\tau) = p_l(0)$ for $1 \leq l \leq j(\tau)$ and $a_l(\tau) = \tau$ for $l > j(\tau)$.

Proposition 3.1. For each individual i, suppose U_{ij} is continuous and strictly increasing for each $j \in \mathcal{J}_s$. Then we have that $B_i(\tau)$ defined in (8) exists and is unique, and that

$$E[B_i(\tau)] = \sum_{l=0}^{j(\tau)} \int_{a_l(\tau)}^{a_{l+1}(\tau)} \left(\sum_{j=l+1}^{J} q_j\left(p_1(0),\ldots,p_l(0),p_{l+1}(\tau)+a,\ldots,p_J(\tau)+a\right)\right) da\ . \tag{10}$$

While voucher provision is beneficial for individuals, it can be costly to the government financing the voucher. To benchmark the benefits and perform a cost-benefit analysis, we also consider parameters that measure these potential costs. To this end, observe that the provision of a voucher introduces costs to the government when individuals enroll in a school in the program, but can also bring about savings depending on the costs the government faces at schools where individuals enroll in the absence of the voucher. To formally capture these net costs, let $c_j(\tau)$ denote the cost that the government associates with the jth demand function in $\mathcal{J}$ under a voucher of amount τ. For example, in our baseline empirical analysis, we take

$$c_j(\tau) = \begin{cases} c_g & \text{for } j = g \ , \\ 0 & \text{for } j = n \ , \\ \min\{p_j(0), \tau\} + \gamma \cdot 1\{\tau > 0\} & \text{for } j \in \mathcal{J}_v \ , \end{cases}$$

i.e. the cost associated with each government-funded school is some known value c_g, the cost associated with each private school not participating in the program is zero, and the cost associated with each private school participating in the program is the voucher amount spent to cover tuition plus some known administrative cost γ of operating the program (i.e. charged only when the voucher amount is positive). We then measure the average net costs from the provision of a voucher of amount τ by

$$AC(\tau) = \sum_{j \in \mathcal{J}} c_j(\tau) \cdot q_j(p(\tau)) - \sum_{j \in \mathcal{J}} c_j(0) \cdot q_j(p(0)) \ , \tag{11}$$

i.e. the average costs the government faces when individuals receive the voucher net of those it faces when individuals do not receive the voucher. Along with the average benefit parameter, we can then also define the average surplus parameter, which can be used to perform a cost-benefit analysis. Specifically, for a voucher of amount τ, let

$$AS(\tau) = AB(\tau) - AC(\tau) \tag{12}$$

denote the average surplus of the voucher, i.e. the average benefit across individuals of receiving the voucher net of the average cost for the government of providing that voucher. Note that the average cost parameter is a function of q and, since the average benefit parameter is a function of q, so is the average surplus parameter.

The benefit, cost and surplus parameters we described above were defined for a generic voucher of amount τ. By taking different values of τ, we can evaluate these parameters for both the status-quo voucher amount as well as alternative counterfactual amounts. More specifically, by taking $\tau = \tau_{\text{sq}}$, we can evaluate these parameters for the status-quo voucher amount, whereas, by taking $\tau = \tau_{\text{c}} \neq \tau_{\text{sq}}$, we can evaluate these parameters for a counterfactual amount of τ_{c}. In our analysis, we also study the difference of the parameters under these amounts, i.e.

$$\Delta AB\left(\tau_{\text{c}}\right) = AB\left(\tau_{\text{c}}\right) - AB\left(\tau_{\text{sq}}\right) \ , \tag{13}$$

$$\Delta AC\left(\tau_{\text{c}}\right) = AC\left(\tau_{\text{c}}\right) - AC\left(\tau_{\text{sq}}\right) \ , \tag{14}$$

$$\Delta AS\left(\tau_{\text{c}}\right) = AS\left(\tau_{\text{c}}\right) - AS\left(\tau_{\text{sq}}\right) \ , \tag{15}$$

which allows us to directly compare the benefit, cost and surplus between the counterfactual and status-quo voucher amounts.

4 Identification Analysis

In the previous section, we described our parameters of interest and noted that each of them was a function of the demand functions. In this section, we study what we can learn about each of these parameters given what we know about the demand functions from the imposed assumptions and the distribution of the data, taken to be known for the purposes of this section.

4.1 General Setup

We begin by formally describing the general setup for the identification analysis we develop below. To this end, let $\theta(q)$ denote a pre-specified parameter of interest from Section 3 that we want to learn about.

Since θ is a known function, it follows that what we can learn about our parameter depends on what we know about the function q. As q is defined to be a function whose image is a vector of probabilities, we know by construction that for each $p \in \mathcal{P}$ we have

$$0 \leq q_j(p) \leq 1 \quad \text{for each } j \in \mathcal{J} \ , \tag{16}$$

$$\sum_{j \in \mathcal{J}} q_j(p) = 1 \ , \tag{17}$$

i.e., for all prices, each demand function lies in the unit interval and their sum together equals one. Under our baseline specification, we know that q satisfies Assumption B(ii), i.e. it satisfies the nonparametric shape restrictions stated in (4). Under our auxiliary specifications, we additionally know that q satisfies Assumption A, i.e. it satisfies the parametric restrictions stated in (5). Finally, under both specifications, the distribution of the data across individuals also restricts the values that q can take. Specifically, it follows from (1) and Assumption B(i) that the distribution of the data reveals

$$q_j(p(0)) = \text{Prob}[D_i = j | Z_i = 0] \equiv P_{j|0} \ , \tag{18}$$

$$q_j\left(p\left(\tau_{\text{sq}}\right)\right) = \text{Prob}[D_i = j | Z_i = 1] \equiv P_{j|1} \tag{19}$$

for $j \in \mathcal{J}_v$, and

$$q_j(p(0)) = \text{Prob}[D_i \in \mathcal{J}_j | Z_i = 0] \equiv P_{j|0} \ , \tag{20}$$

$$q_j\left(p\left(\tau_{\text{sq}}\right)\right) = \text{Prob}[D_i \in \mathcal{J}_j | Z_i = 1] \equiv P_{j|1} \tag{21}$$

for $j \in \{g, n\}$, i.e. the enrollment shares across schools conditional on the receipt of voucher reveal the values the demand functions take at the vector of prices with and without the status-quo voucher amount. To summarize the above information on what we know about q, let $\mathbf{F}$ denote the set of all functions from $\mathcal{P}$ to $\mathbf{R}^{|\mathcal{J}|}$. Then, let

$$\mathbf{Q}_B = \{q \in \mathbf{F} : q \text{ satisfies } (16) - (17),\ (4) \text{ and } (18) - (21)\} \tag{22}$$

denote the admissible set of all demand functions that satisfy the various restrictions imposed by the assumptions and data under our baseline specification, and let

$$\mathbf{Q}_A = \{q \in \mathbf{F} : q \text{ satisfies } (16) - (17),\ (4),\ (5) \text{ and } (18) - (21)\} \tag{23}$$

denote the analogous set of such demand functions under our auxiliary specification.

Given what we know about q, our objective is to characterize what we can then learn about our parameter $\theta(q)$. In some cases, observe that there exists a single admissible value of q under the chosen specification. In such cases, it follows that we can exactly learn the value of $\theta(q)$. For example, as we noted before, this is the case under the specification described in (6)-(7). However, under more flexible parametric specifications as well as the baseline nonparametric specification, there generally exist multiple admissible values of q. In these more general cases, it follows that we can learn a set of values that $\theta(q)$ lies in.

Our analysis aims to show what we can learn across both these two cases. We generally do so by showing how to characterize the identified set. Formally, for a given admissible set of demand functions $\mathbf{Q}$, the identified set is defined by

$$\theta(\mathbf{Q}) = \{\theta_0 \in \mathbf{R} : \theta(q) = \theta_0 \text{ for some } q \in \mathbf{Q}\} \equiv \Theta\ , \tag{24}$$

i.e. the image of the set of admissible functions $\mathbf{Q}$ under the function θ. Intuitively, the identified set corresponds to the set of all parameter values that could have been generated by the admissible values of q. By construction, it sharply captures all that we can learn about the parameter given the data and the chosen specification. Indeed, if the parameter is point identified then the identified set corresponds to a single point. Alternatively, if the parameter is partially identified then the identified set corresponds to the sharpest set of all possible parameter values consistent with the data and specification.

The key challenge of our analysis is how to develop a tractable characterization of the identified set. In what follows, we develop tractable procedures that show how to do so under each of our specifications: first, in Section 4.2, under our baseline specification, i.e. Θ in (24) when $\mathbf{Q} = \mathbf{Q}_B$; and then, in Section 4.3, under our auxiliary specification, i.e. Θ in (24) when $\mathbf{Q} = \mathbf{Q}_A$.

4.2 Identified Set under Baseline Nonparametric Specification

In principle, observe that characterizing the identified set corresponds to searching over the various q in $\mathbf{Q}$ and taking their image under the function θ. Under the baseline specification, this problem can be challenging due to the fact that $\mathbf{Q}_B$ is an infinite-dimensional space. Below, we show how to feasibly proceed in this case. In particular, we exploit the idea that we can replace $\mathbf{Q}_B$ by a finite-dimensional space $\mathbf{Q}_B^{\text{fd}}$ without any loss of information with respect to what we can learn about the parameter in the sense that $\theta(\mathbf{Q}_B) = \theta\left(\mathbf{Q}_B^{\text{fd}}\right)$. This allows us to indirectly characterize the identified set by searching only through q in $\mathbf{Q}_B^{\mathrm{fd}}$, which is a finite-dimensional problem and, hence, potentially feasible in practice.

We begin by defining the finite-dimensional $\mathbf{Q}_B^{\mathrm{fd}}$ we consider. Our choice of $\mathbf{Q}_B^{\mathrm{fd}}$ takes q to be constant on some finite partition of the space of prices. The partition is intuitively chosen such that the resulting q based on it is sufficiently rich to equivalently define the various parameters of interest and the data restrictions in (18)-(21) as well as preserve the information provided by the shape restrictions in (4).

In order to define the partition, we need to first define a collection of sets that play a role in the definition of the parameters and data restrictions, and allow the preservation of the information provided by the shape restrictions. To this end, observe that

$$\mathcal{P}_l(\tau) = \{p \in \mathcal{P} : p_j = \min\{p_j(0), p_j(\tau) + a\} \text{ for } a \in [a_l(\tau), a_{l+1}(\tau)] \text{ for each } j \in \mathcal{J}_v\} \tag{25}$$

for $0 \leq l \leq j(\tau)$ correspond to the various sets of prices that play a role in the definition of the parameter $AB(\tau)$, and

$$\{p(0), p(\tau_{\mathrm{sq}}), p(\tau)\} \tag{26}$$

corresponds to the set of prices that play a role in the definition of the parameter $AC(\tau)$ as well as the data restrictions in (18)-(21). Note it then follows that

$$\mathcal{P}^* = \bigcup_{l=0}^{j(\tau_{\mathrm{sq}})} \mathcal{P}_l(\tau_{\mathrm{sq}}) \bigcup_{l=0}^{j(\tau_c)} \mathcal{P}_l(\tau_{\mathrm{c}}) \bigcup \{p(0), p(\tau_{\mathrm{sq}}), p(\tau_{\mathrm{c}})\} \tag{27}$$

corresponds to the subset of $\mathcal{P}$ that plays a role in the definition of all parameters for the status-quo voucher amount and a counterfactual voucher amount of τ_{c} along with the restrictions imposed by the data. Given this set of prices, we define in the following definition the collection of sets that we later use in the definition of the partition.

Definition 4.1. Let $\mathcal{U}$ denote a finite partition of the set of prices $\mathcal{P}^*$ in (27) such that for all $u \in \mathcal{U}$ we have either

(i) $u = \{p \in \mathcal{P} : p_j = \min\{p_j(0), p_j(\tau) + a\} \text{ for } a \in (\underline{a}_u, \bar{a}_u] \text{ or } (\underline{a}_u, \bar{a}_u) \text{ for each } j \in \mathcal{J}_v\}$ where $\underline{a}_u$ and $\bar{a}_u$ are such that $u \subseteq \mathcal{P}_l(\tau)$ for some $0 \leq l \leq j(\tau)$ and $\tau \in \{\tau_{\mathrm{sq}}, \tau_{\mathrm{c}}\}$; or

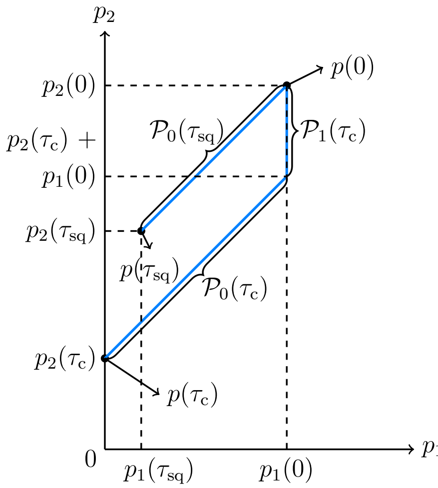
(ii) $u = \{p(\tau)\}$ for some $\tau \in \{0, \tau_{\mathrm{sq}}, \tau_{\mathrm{c}}\}$,

and for all $u, u' \in \mathcal{U}$ we have either

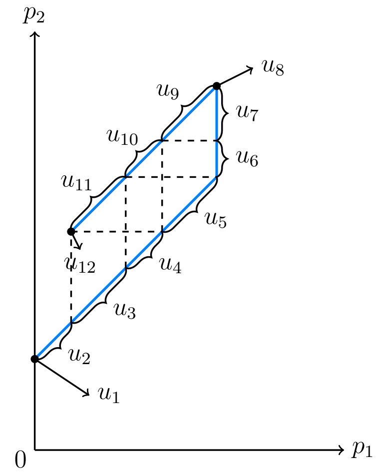
$$u(j) = u'(j) \quad \text{or} \quad u(j) \cap u'(j) = \emptyset \tag{28}$$

for each $j \in \mathcal{J}_v$, where $u(j) = \{t \in \mathbf{R} : p_j = t \text{ for some } p \in u\}$ for each $u \in \mathcal{U}$ and $j \in \mathcal{J}_v$.

Figure 1: Various sets of prices for an example with $J = 2$ and $\tau_{\text{sq}} < p_1(0) < \tau_{\text{c}} < p_2(0)$



(a) Sets that play a role in defining the parameters and data restrictions



(b) A partition $\mathcal{U} \equiv \{u_1, \ldots, u_{12}\}$ of the union of sets in (a) that satisfies Definition 4.1

Definition 4.1 states that $\mathcal{U}$ corresponds to a finite partition of $\mathcal{P}^*$, where each element of the partition satisfies specific properties. In particular, Definition 4.1(i)-(ii) states that each element is a connected subset of that in (25) or (26). In addition, it states in (28) that any pair of sets in this partition are such that they either completely overlap or are disjoint in each price coordinate. Intuitively, note that the first property, as the sets in (25) and (26) are specifically based on the parameters and data restrictions, is what ensures that the finite-dimensional q will be sufficiently rich to define the parameters and data restrictions. On the other hand, note that the latter property, which implies that the sets can be ordered and pairwise compared across each price coordinate, is what ensures that the finite-dimensional q will be able to preserve the information provided by the shape restrictions in (4) that are indeed based on pairwise comparisons of prices.

To better understand these various set of prices, Figure 1(a) first graphically illustrates the sets of prices in (25) and (26) in the context of a simple example with two voucher schools and a specific combination of status-quo and counterfactual voucher amounts. Figure 1(b) then shows how the union of the sets in Figure 1(a) can be partitioned to obtain a collection of sets satisfying Definition 4.1. In particular, it sequentially divides any two sets in Figure 1(a) that partially overlap in a given coordinate until the condition in (28) is satisfied. In Appendix S.2.2, we describe a computational procedure that sequentially divides sets in such a manner to obtain a partition satisfying Definition 4.1 in the case of more than two goods.

Using the above defined collection of sets, we can now define the partition of the space of prices and our choice of $\mathbf{Q}_B^{\text{fd}}$. To define the partition, observe that for each $j \in \mathcal{J}_v$, the collection of sets determined by the prices in $u \in \mathcal{U}$ for the jth school, i.e. $\{u(j) : u \in \mathcal{U}\}$, generates a partition of $[p_j(\max\{\tau_{\text{sq}}, \tau_{\text{c}}\}), p_j(0)] \subseteq [0, p_j(0)]$. Given this implies that $\mathcal{U}_j = \{[0, p_j(\max\{\tau_{\text{sq}}, \tau_{\text{c}}\}))\} \bigcup \{u(j) :$

$u \in \mathcal{U}\}$ corresponds to a partition of $[0, p_j(0)]$ for each $j \in \mathcal{J}_v$, observe that

$$\mathcal{W} = \prod_{j=1}^{J} \mathcal{U}_j \equiv \{w_1, \ldots, w_M\} \ ,$$

denotes a partition of the space of prices $\mathcal{P}$ over which q is defined, where, for each element of the partition, the prices for the jth school in $\mathcal{J}_v$ take values in a set that corresponds to an element of $\mathcal{U}_j$. Then, using this partition, we take

$$\mathbf{Q}_B^{\text{fd}} = \left\{ q \in \mathbf{Q}_B : q_j(p) = \sum_{w \in \mathcal{W}} 1_w(p) \cdot \beta_j(w) \text{ for some } \{\beta_j(w)\}_{w \in \mathcal{W}} \text{ for each } j \in \mathcal{J} \right\} \ , \tag{29}$$

where $1_w(p) \equiv 1\{p \in w\}$, i.e. the space we consider corresponds to a subset of $\mathbf{Q}_B$ such that each q is parameterized to be a constant function over the elements of the partition $\mathcal{W}$.

We next show that replacing $\mathbf{Q}_B$ with this choice of $\mathbf{Q}_B^{\text{fd}}$ leads to no loss of information with respect to what we can learn about the parameter of interest, i.e. $\theta\left(\mathbf{Q}_B\right) = \theta\left(\mathbf{Q}_B^{\text{fd}}\right)$. In addition, we also show that characterizing $\theta\left(\mathbf{Q}_B^{\text{fd}}\right)$, which is a finite-dimensional problem, can be solved using two finite-dimensional optimization problems. In order to state this result, it is useful to first restate $\theta\left(\mathbf{Q}_B^{\text{fd}}\right)$ in terms of the variable $\beta \equiv \left(\beta_g', \beta_n', \beta_1', \ldots, \beta_J'\right)'$, where $\beta_j = (\beta_j(w_1), \ldots, \beta_j(w_M))$ for each $j \in \mathcal{J}$, that parameterizes a given $q \in \mathbf{Q}_B^{\text{fd}}$. To this end, note that given each parameter θ is continuous in q and that q is continuous in β, it follows that θ can be written in terms of a continuous function of β in the sense that there exists a continuous function θ_B of β such that $\theta(q) = \theta_B(\beta)$. Similarly, note that $\mathbf{Q}_B$ can also be written in terms of β by

$$\mathbf{B} = \left\{ \beta \in \mathbf{R}^{d_\beta} : \left(\sum_{w \in \mathcal{W}} 1_w \cdot \beta_j(w) : j \in \mathcal{J} \right) \in \mathbf{Q}_B \right\} \ , \tag{30}$$

where d_β denotes the dimension of β, i.e. the set of values of β that ensure that the corresponding q is in $\mathbf{Q}_B$. Then, we can write $\theta\left(\mathbf{Q}_B^{\text{fd}}\right)$ in terms of β by

$$\{\theta_0 \in \mathbf{R} : \theta_B(\beta) = \theta_0 \text{ for some } \beta \in \mathbf{B}\} \equiv \Theta_B \ . \tag{31}$$

In the following proposition, we state the result that the identified set under the baseline specification, i.e. Θ in (24) when $\mathbf{Q} = \mathbf{Q}_B$, is equal to Θ_B. In addition, the proposition also shows that we can characterize Θ_B by solving two finite-dimensional optimization problems.

Proposition 4.1. Suppose that $\mathbf{Q} = \mathbf{Q}_B$. Then, the identified set in (24) is equal to that in (31), i.e. $\Theta = \Theta_B$. In addition, if $\mathbf{B}$ is empty then by definition Θ_B is empty; whereas, if $\mathbf{B}$ is non-empty then $\Theta_B = [\underline{\theta}_B, \bar{\theta}_B]$, where

$$\underline{\theta}_B = \min_{\beta \in \mathbf{B}} \ \theta_B(\beta) \ \text{ and } \ \bar{\theta}_B = \max_{\beta \in \mathbf{B}} \ \theta_B(\beta) \ . \tag{32}$$

Proposition 4.1 shows that the identified set under the baseline specification when not empty is given by a closed interval, where the endpoints can be obtained by solving the two optimization problems stated in (32). In the proof of the proposition, we explicitly derive $\mathbf{B}$, the constraint set of these optimization problems, and observe that it is determined by constraints that are all linear in β. In addition, we also explicitly derive θ_B, the objectives of these optimization problems, for each of our parameters of interest and observe that they all correspond to linear functions of β. These two observations then imply that these optimization problems are, in fact, linear programming problems, a useful observation in their practical implementation. Lastly, observe that to characterize the identified set using these linear programs, we specifically require that $\mathbf{B}$ is non-empty or, equivalently, that the model is not misspecified. However, when this is not the case, these linear programs automatically terminate, indicating that the model is misspecified.

While the optimization problems in (32) are linear programs, they can nonetheless be computationally expensive in cases where the dimension of the optimizing variable β is large. Such a case arises especially in settings when J is large as in our empirical analysis, where we have that J is equal to 68. To ensure tractability in such cases, it is useful to consider alternative lower-dimensional linear programs that are easier to compute and can continue to allow us to learn about our parameters. To this end, observe that, given how $\mathcal{U}$ captured all sets relevant in defining our parameters, only a restricted subset of $\mathcal{W}$ given by

$$\mathcal{W}^{\mathrm{r}} = \left\{ w \in \mathcal{W} : w = \prod_{j=1}^{J} u(j) \text{ for some } u \in \mathcal{U} \right\} \equiv \{w_1^{\mathrm{r}}, \ldots, w_{M^{\mathrm{r}}}^{\mathrm{r}}\} \ ,$$

corresponds to the sets of prices that play a role in the definition of our parameters. In turn, observe that only a subvector of β defined over these sets given by $\beta^{\mathrm{r}} = \left(\beta_g^{\mathrm{r}\prime}, \beta_n^{\mathrm{r}\prime}, \beta_1^{\mathrm{r}\prime}, \ldots, \beta_J^{\mathrm{r}\prime}\right)' \equiv \phi(\beta)$, where $\beta_j^{\mathrm{r}} = (\beta_j(w_1^{\mathrm{r}}), \ldots, \beta_j(w_{M^{\mathrm{r}}}^{\mathrm{r}}))$ for each $j \in \mathcal{J}$, plays a role in determining θ_B in the sense that there equivalently exists a linear function θ_B^{r} such that $\theta_B^{\mathrm{r}}(\beta^{\mathrm{r}}) = \theta_B(\beta)$. Then, the lower-dimensional linear programs we consider are those in terms of the subvector β^{r} given by

$$\underline{\theta}_B^{\mathrm{r}} = \min_{\beta^{\mathrm{r}} \in \mathbf{B}^{\mathrm{r}}} \ \theta_B^{\mathrm{r}}(\beta^{\mathrm{r}}) \ \text{ and } \ \bar{\theta}_B^{\mathrm{r}} = \max_{\beta^{\mathrm{r}} \in \mathbf{B}^{\mathrm{r}}} \ \theta_B^{\mathrm{r}}(\beta^{\mathrm{r}}) \ , \tag{33}$$

where $\mathbf{B}^{\mathrm{r}}$ denotes a set of β^{r} determined by linear constraints. By an appropriate choice of $\mathbf{B}^{\mathrm{r}}$, these alternative linear programs can continue to allow us to learn about our parameters. To see how, observe first that if $\mathbf{B}^{\mathrm{r}} = \phi(\mathbf{B})$, we have by construction that these programs are equivalent to those in (32). In turn, by taking $\mathbf{B}^{\mathrm{r}}$ to be such that $\phi(\mathbf{B}) \subseteq \mathbf{B}^{\mathrm{r}}$, it follows that we have $\underline{\theta}_B^{\mathrm{r}} \leq \underline{\theta}_B$ and $\bar{\theta}_B^{\mathrm{r}} \geq \bar{\theta}_B$, and can therefore continue to learn about our parameters by obtaining a set that contains the identified set, i.e. $\Theta_B \in \left[\underline{\theta}_B^{\mathrm{r}}, \bar{\theta}_B^{\mathrm{r}}\right]$. In Appendix S.2.3, we provide a natural choice of such a $\mathbf{B}^{\mathrm{r}}$ determined by restrictions on β^{r} implied by those in $\mathbf{B}$, which we find in our empirical analysis can be tractably implemented and also result in informative conclusions.

4.3 Identified Set under Auxiliary Parametric Specifications

We now proceed to show how to characterize the identified set under our auxiliary specification. Under this specification, in contrast to the baseline, note that the problem is finite-dimensional in nature due to the fact that $\mathbf{Q}_A$ is a finite-dimensional parameterized space. As a result, in this case, the identified set can be directly characterized by searching over q in $\mathbf{Q}_A$ and then taking their image under the function θ.

In order to state the result that shows how to do this, it useful to first restate the identified set in terms of the variable α that parameterizes a given $q \in \mathbf{Q}_A$ through (5). Given each parameter θ is continuous in q and that q is continuous in α, note that it follows that θ can be written in terms of a continuous function of α in the sense that there exists a continuous function θ_A of α such that $\theta(q) = \theta_A(\alpha)$. Similarly, note that $\mathbf{Q}_A$ can also be written in terms of α by

$$\mathbf{A} = \left\{ \alpha \in \mathbf{R}^{d_\alpha} : \left(\sum_{k=0}^{K_j} \alpha_{jk} \cdot b_{jk} : j \in \mathcal{J} \right) \in \mathbf{Q}_B \right\} . \tag{34}$$

where d_α denotes the dimension of α, i.e. the set of values of α that ensure that the corresponding q is in $\mathbf{Q}_B$. Then, the identified set under the auxiliary specification, i.e. Θ in (24) when $\mathbf{Q} = \mathbf{Q}_A$, can equivalently be given by

$$\theta_A(\mathbf{A}) = \{\theta_0 \in \mathbf{R} : \theta_A(\alpha) = \theta_0 \text{ for some } \alpha \in \mathbf{A}\} \equiv \Theta_A , \tag{35}$$

i.e. the image of the set $\mathbf{A}$ under the function θ_A. In the following proposition, we show that when $\mathbf{A}$ is connected and non-empty, the closure of this set is equal to an interval, where the endpoints can be characterized as solutions to two finite dimensional optimization problems.

Proposition 4.2. If $\mathbf{A}$ is empty then by definition Θ_A is empty; whereas, if $\mathbf{A}$ is connected and non-empty, then the closure of Θ_A is given by $[\underline{\theta}_A, \bar{\theta}_A]$, where

$$\underline{\theta}_A = \inf_{\alpha \in \mathbf{A}} \ \theta_A(\alpha) \ \text{ and } \ \bar{\theta}_A = \sup_{\alpha \in \mathbf{A}} \ \theta_A(\alpha) . \tag{36}$$

Proposition 4.2 shows how to characterize the identified set under a general class of parametric restrictions. As we mentioned before, this class allows various types of more flexible versions of the parametric specification in (6)-(7) that ensured point identification of the demand functions. We conclude this section by discussing three types of such specifications we later consider in our empirical analysis that can be implemented using Proposition 4.2.

Assumption O. (Own-price) For each $j \in \mathcal{J}_v$,

$$q_j(p) - q_g(p) = \sum_{k=0}^{K} \alpha_{jk} \cdot p_j^k$$

for some $\{\alpha_{jk} : 0 \le k \le K\}$, and $q_n(p) - q_g(p) = \alpha_n$ for some α_n.

Assumption AS. (Additively Separable) For each $j \in \mathcal{J}$,

$$q_j(p) = \sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot p_m^k$$

for some $\{\alpha_{jmk} : m \in \mathcal{J}_v,\ 0 \leq k \leq K\}$.

Assumption NS. (Nonseparable) For each $j \in \mathcal{J}$,

$$q_j(p) = \sum_{\substack{m=1 \\ m \neq j}}^{J} \sum_{k=0}^{K} \sum_{l=0}^{K} \alpha_{jmkl} \cdot p_j^k \cdot p_m^l$$

for some $\{\alpha_{jmkl} : m \in \mathcal{J}_v,\ 0 \leq k, l \leq K\}$, and for each $j \in \{g, n\}$,

$$q_j(p) = \sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot p_m^k$$

for some $\{\alpha_{jmk} : m \in \mathcal{J}_v,\ 0 \leq k \leq K\}$.

Assumption O states that the difference in demand for each $j \in \mathcal{J}_v$ and any $j \in \mathcal{J}_g$ is a function only of its own price, where this function is a polynomial of degree K, and the difference in demand for any $j \in \mathcal{J}_n$ and any $j \in \mathcal{J}_g$ is constant. When K equals one, this corresponds to the linear specification in (6)-(7). However, for larger values of K, it allows for more flexible patterns in prices. Nonetheless, while more flexible, it can still be viewed as restrictive as it assumes the demand for a given school with respect to government-funded schools to be invariant to prices of other schools. To this end, Assumption AS and Assumption NS consider more flexible parametric specifications that allow the demand for each school to depend on the prices of all voucher schools. Assumption AS takes the demand for each $j \in \mathcal{J}$ to be an additively separable function in the prices of each $j \in \mathcal{J}_v$, where these functions are polynomials of degree K. Assumption NS further parsimoniously relaxes the requirement of additive separability by allowing for nonseparability in its own price. In particular, it takes the demand for $j \in \mathcal{J}_v$ to be an additively separable function only in the prices of each $m \in \mathcal{J}_v \setminus \{j\}$, where these bivariate functions are bivariate polynomials of degree K.

While the optimization problems in (36) are finite-dimensional, their computational tractability depends on the structure of the objective θ_A and constraint set $\mathbf{A}$. In Appendix S.2.4, we illustrate that, under each of the different parametric specifications considered above, θ_A for each parameter is a linear function of α and that $\mathbf{A}$ is characterized by linear equality and inequality restrictions on α. However, we observe here that some of the linear restrictions in these cases are evaluated at every possible price vector in $\mathcal{P}$, which implies that the resulting optimization problems in (36) can be generally difficult to compute. To this end, similar in spirit to those in (33), we consider the following alternative optimization problems

$$\underline{\theta}_A^{\mathrm{r}} = \min_{\alpha \in \mathbf{A}^{\mathrm{r}}} \theta_A(\alpha) \quad \text{and} \quad \bar{\theta}_A^{\mathrm{r}} = \max_{\alpha \in \mathbf{A}^{\mathrm{r}}} \theta_A(\alpha) \tag{37}$$

in our empirical analysis, where $\mathbf{A}^{\mathrm{r}}$ corresponds to a subset of $\mathbf{A}$ that evaluates some of the restrictions on only a finite set of prices in $\mathcal{P}$—the exact form of $\mathbf{A}^{\mathrm{r}}$ for each specification is provided in Appendix S.2.4. Indeed, since the objective and the finite number of restrictions determining the constraint sets of these problems are linear in α, they are linear programming problems and hence generally computationally tractable. But, since $\mathbf{A} \subseteq \mathbf{A}^{\mathrm{r}}$, these problems only provide a set that contains the identified set, i.e. $\Theta_A \subseteq [\underline{\theta}^{\mathrm{r}}_A, \bar{\theta}^{\mathrm{r}}_A]$, similar to how those in (33) do so for the identified set under the baseline specification. In our empirical analysis, we nonetheless find that these sets result in informative conclusions.

5 Evaluation of the DC Opportunity Scholarship Program

5.1 Background

The DC Opportunity Scholarship Program (OSP) was a federally-funded school voucher program established by Congress in January 2004, and which started accepting students for the 2004-2005 (henceforth, 2004) school year. The OSP was structured similarly to other voucher programs that existed at the time (Epple et al., 2017). It was open to students residing in Washington, DC, and whose family income was no higher than 185% of the federal poverty line ($18,850 for a family of four in 2004).[2] It could be used only for K-12 education, and at the time of initial receipt was renewable for up to five years. It provided students a voucher worth $7,500 that could be used to offset tuition, fees, and transportation at any private school of their choice participating in the program.

The law that established the program also mandated its evaluation, which culminated with a final report to Congress (Wolf et al., 2010). The report exploited the fact that the OSP randomly allocated vouchers to participating students. In particular, Congress expected the program to be oversubscribed, i.e. the number of applicants would exceed the number of available slots in participating private schools. As a result, it required that vouchers be randomly allocated to applicants through a lottery whenever the program was oversubscribed—see Wolf et al. (2010) for details on the lottery. Wolf et al. (2010) exploited this random allocation by comparing various outcomes of voucher recipients to non-recipients to experimentally evaluate the effect of voucher receipt on these outcomes. The main findings from this report, as listed in its executive summary, can be broadly summarized as follows. First, they find no conclusive evidence that the receipt of the voucher had any significant effects on various outcomes corresponding to student achievement. Second, they find that the receipt of the voucher significantly improved students' chances of graduating from high school. Finally, they find that the receipt of the voucher raised parents' ratings of school safety and satisfaction.

[2] All dollar amounts throughout have been deflated to 2004 dollars.

Table 1: Enrollment shares across school type by voucher receipt

	With voucher	Without voucher	Difference
Government-funded	0.288	0.901	-0.613
	[0.453]	[0.299]	(0.019)
Non-participating private	0.014	0.020	-0.006
	[0.117]	[0.140]	(0.007)
Participating private	0.698	0.079	0.619
	[0.459]	[0.270]	(0.018)
Observations	1,090	730	

Observations rounded to the nearest 10. Standard deviations in square brackets and robust standard errors in parentheses.

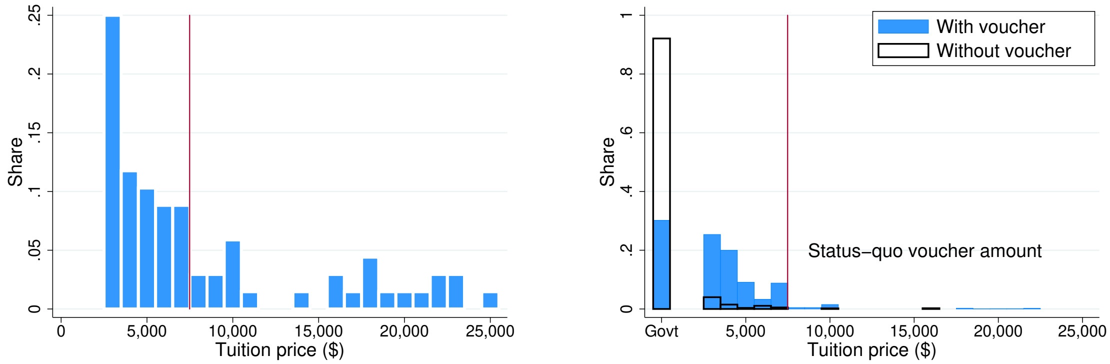
SOURCE: *Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)*, U.S. Department of Education, National Center for Education Statistics previously unpublished tabulations.

In what follows, we use the tools developed in the previous sections to complement these findings by analyzing the welfare effects of providing the status-quo voucher amount as well as alternative counterfactual amounts. Our analysis is based on the premise that while the receipt of the voucher revealed mixed evidence on outcomes in the sense that there are zero as well as some positive effects, parents may nonetheless value the voucher, potentially across dimensions not easily captured by the outcomes. Indeed, as we highlight below, the data from the program reveals that a non-trivial proportion of voucher recipients used the voucher, which, by revealed preference arguments, implies that recipients may value receiving the voucher. Our analysis below estimates these potential welfare benefits using data collected by the OSP.

5.2 Data and Summary Statistics

The OSP collected detailed data for the first two years of the program, 2004 and 2005, and tracked students for at least four years. Across these years, the school settings were different—the composition of applicants and private schools participating in the program changed. Wolf et al. (2010) provide a detailed description on how the data was collected and various summary statistics for the various years. To keep prices and the set of available schools the same for all students, in our analysis we focus on the second year of the program, 2005, which contains around 80% of the entire sample. In addition, we focus on the initial year of the data for students entering the program this year. As we note in Section 6, this avoids complications that arise from the dynamics of the setup. In Appendices S.4.1-S.4.2, we provide details on how our analysis sample was constructed from the original evaluation data and some statistics on the school setting. Below, we present summary statistics for the main variables our analysis exploits, namely the enrollment shares and the prices as measured by the tuition of private schools participating in the program.

Figure 2: Prices and enrollment shares by voucher receipt across participating private schools



(a) Prices of participating private schools

(b) Enrollment shares over prices

SOURCE: *Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)*, U.S. Department of Education, National Center for Education Statistics previously unpublished tabulations.

Table 1 presents the empirical enrollment shares across the three types of schools, i.e. government-funded schools (which includes charter schools) and private schools participating and not participating in the program, by voucher receipt. The proportion of voucher recipients who use the voucher, corresponding to those enrolled in participating private schools, is relatively large (69.8%). By revealed preference, this implies that recipients value the voucher. In addition, the voucher increases the proportion enrolling in private schools by 61.9 percentage points, suggesting that prices play an important role in inducing private school enrollment. Finally, observe the nearly symmetric decline in the proportion enrolled in government-funded schools caused by voucher receipt (-61.3 percentage points), which reveals that nearly all students induced into participating private schools by the voucher would be in government-funded schools absent the voucher.

In 2005, there were 68 private schools participating in the program (out of a total of 109 in DC). Figure 2 presents histograms that summarize the variation in prices across these schools as well as the enrollment shares across these prices. Figure 2(a) reveals that a large number of participating private schools had low prices—around 80% had prices below the status-quo voucher amount. Figure 2(b) reveals that the voucher induced a significant proportion to enroll in these low-price schools—out of the 61.9 percentage point increase in the number of students attending a participating private school, a full 59 percentage points (95%) was into schools with prices less than the status-quo voucher amount. As we highlight below, these observations play an important role in better understanding the welfare effects of the voucher.

To also provide some evidence on why recipients may be choosing participating private schools and, in turn, value the voucher, Table S.2 in Appendix S.4.2 compares characteristics of these schools with those of with government-funded schools, where the majority of students enroll absent the voucher. Private and government-funded schools differ across several attributes. The private schools tend to be more religious and specifically Catholic, have smaller school sizes, are more likely to track students by ability, and are less likely to have programs for students with learning difficulties. This suggests that recipients may value these attributes and, hence, the voucher that makes these schools more affordable.

Recall from Section 3 that our analysis also uses a value of c_g for the costs the government faces when a student enrolls in government-funded schools as well as a value of γ for administrative costs. In our main analysis, we take c_g = \$5,355, which corresponds to the educational expenditure reported by the US Census (2005). This is lower than total per-pupil expenditure (\$12,979, which includes some fixed costs), or educational expenditure as measured in other sources (\$8,105, Sable and Hill (2006)). However, given that our surplus parameters are increasing in c_g, we choose the smaller, more conservative value. On the other hand, we take γ = \$200, which corresponds to cost of administration, adjudication and providing information to families for an alternative school voucher program reported in Levin and Driver (1997).[3] For our baseline government-funded schools cost of \$5,355, Figure 2(b) reveals that a large proportion of recipients (81%) redeem the voucher at schools with prices below this value. Given that Table 1 revealed that the majority of these recipients would have enrolled in government-funded schools absent the voucher, this suggests that the government may face only small net costs or even savings from the provision of a voucher, even accounting for the administrative costs. Our estimates below make this point more precisely.

5.3 Welfare Estimates for the Status-quo Voucher Amount

Table 2 presents the estimates of the welfare effects for the status-quo voucher amount. Each row of the table corresponds to a parameter from (9), (11) or (12), taking $\tau = \tau_{\text{sq}} \equiv \$7,500$. Each column corresponds to a specification of demand, which is either the baseline nonparametric specification defined by Assumption B or an auxiliary parametric specification that additionally imposes either Assumption O, Assumption AS or Assumption NS for some value of K. We consider $K = 1, 2, 3$. The estimates under the nonparametric specification are computed using the optimization problems in (33) with the choice of $\mathbf{B}^{\text{r}}$ described in Appendix S.2.3 and those under the parametric specifications are computed using the optimization problems in (37) with the choices of $\mathbf{A}^{\text{r}}$ described in Appendix S.2.4, where in both cases the enrollment shares in the restrictions in (18)-(21) are replaced by their empirical counterparts. We also report 95% confidence intervals, and specification test p-values in the case of misspecification. These are both constructed using a subsampling procedure from Kalouptsidi et al. (2020), which we describe in Appendix S.3. To compare our results to those obtained using standard discrete choice tools, we also report in Appendix S.4.3, and briefly discuss below, results under common fully-parameterized, point identified versions of our model based on taking a logit specification for the idiosyncratic error terms in (2)-(3).

[3] As a sensitivity analysis, we also present results for a range of other values of c_g and γ in Appendix S.4.4.

Table 2: Estimated welfare effects of the status-quo voucher

	Nonparametric	Own-price			Additively separable			Nonseparable		
		K			K			K		
		1	2	3	1	2	3	1	2	3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	239				1,669	1,000	720	1,669	1,000	720
	364				1,769	1,100	795	1,769	1,100	795
AB(τ_{sq})		$\emptyset$	$\emptyset$	$\emptyset$						
	5,239				1,873	2,518	2,912	1,920	2,608	3,038
	5,414				1,973	2,618	3,037	2,020	2,733	3,163
	-20				0	0	0	0	0	0
AC(τ_{sq})	150	$\emptyset$	$\emptyset$	$\emptyset$	150	150	150	150	150	150
	300				290	290	290	290	290	290
	113				1,519	825	545	1,519	825	545
	213				1,619	950	645	1,619	950	645
AS(τ_{sq})		$\emptyset$	$\emptyset$	$\emptyset$						
	5,088				1,723	2,368	2,762	1,770	2,458	2,887
	5,313				1,848	2,493	2,912	1,870	2,583	3,012
Spec. p-value	-	0.00	0.00	0.00	-	-	-	-	-	-

For each parameter, each panel reports the lower endpoint of the 95% CI, the estimated lower bound, the estimated upper bound, and the upper endpoint of the 95% CI, respectively. Upper and lower bound not repeated if they coincide. The $\emptyset$ denotes the empty set and indicates that the specification was rejected by the data. For rejected specifications, we provide the specification test p-value.

SOURCE: *Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)*, U.S. Department of Education, National Center for Education Statistics previously unpublished tabulations.

The empty sets reveal that some of the specifications may be misspecified. Specifically, the specification in (6)-(7) in Column (2) that implies point identification of the demand function as well as more flexible versions in the form of Assumption O in Columns (3) and (4) may be misspecified. To see why this arises, observe that Assumption O requires the difference between $q_g(p)$ and $q_n(p)$ to be constant for all values of p, which then implies that the difference in enrollment shares for any government-funded and non-participating private school with and without the voucher be equal; however, their empirical counterparts in Table 1 reveal these values are in fact different. The p-values in the final row of Table 2 reveal that this difference is also statistically significant. In contrast, the data do not reject the nonparametric specification in Column (1) or the more flexible parametric specifications in Columns (5)-(10). In these cases, as highlighted in Section 4, there exist multiple demand functions consistent with data and, as a result, we can generally only obtain bounds for the parameters. Nonetheless, as we discuss below, these bounds are quite tight and allow us to reach informative conclusions.

The estimates for $AB(\tau_{\text{sq}})$ under the nonparametric specification in Column (1) reveal that the average benefit from the status-quo voucher is between \$364 and \$5,239. Under auxiliary parametric specifications, the bounds can substantially tighten. For example, under the most informative specification in Column (5), the average benefit is between \$1,769 and \$1,873, whereas, under the most flexible specification in Column (10), it is between \$795 and \$3,038.

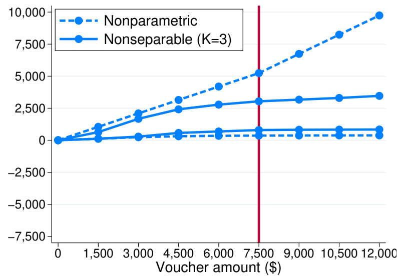
The estimates for $AC(\tau_{\text{sq}})$ reveal that the lower and upper bounds are equal and, in turn, that it is point identified across all specifications. In particular, point identification arises because $AC(\tau_{\text{sq}})$ is a function of demand at values of prices at which the demand is exactly observed in the data, namely the prices with and without the status-quo voucher. The point identified value reveals that the average net cost of providing the status-quo voucher is equal to \$150. While this voucher provides an amount of up to \$7,500, the cost is relatively low due to the fact, as highlighted above, that a large proportion of recipients redeem the voucher at low-cost private schools relative to the government-funded schools they would have enrolled in absent the voucher.

Taking the difference of the average benefit and cost, the estimates for $AS(\tau_{\text{sq}})$ reveal that the average benefit net of costs of the status-quo voucher across all specifications is generally positive. In particular, under the nonparametric specification in Column (1), the average surplus is between \$213 and \$5,088 and, under the most flexible parametric specification in Column (10), between \$645 and \$2,887. Intuitively, the positive net benefit arises due to the relatively low net costs of providing the voucher that we highlighted above. Specifically, the voucher recipients have a high welfare benefit from the low-price private schools at which they redeem the voucher relative to the low net costs the government faces to fund the voucher at these schools, which then implies a positive net benefit.

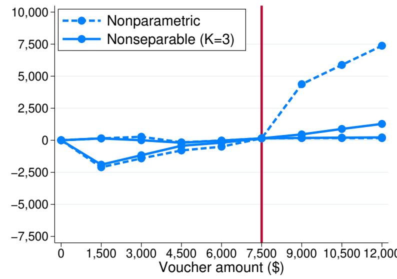
In Appendix S.4.4, we perform several robustness checks on the above conclusion that the provision of the status-quo voucher amount has a positive average surplus. As we noted above, our analysis uses a specific value of c_g for the costs the government faces when a student enrolls in a government-funded school, and a value of γ for the administrative costs of providing a voucher. In addition, while the OSP allowed the voucher to be used to offset additional fees and transportation costs, our analysis implicitly assumed that they could be only used to offset tuition. Our robustness analysis measures the sensitivity of our average surplus estimates to taking different values of c_g and γ, and supposing that the voucher could be used to offset an amount δ in addition to the tuition. We find that our conclusions continue to hold for a range of values of c_g, γ and δ.

In comparison to the above results, Appendix S.4.3 reveals that results based on fully-parameterized specifications of our model can potentially provide a misleading picture of the welfare effects. Specifically, we find that these specifications all provide estimates of the average benefit parameter that are consistently towards the lower bound of our nonparametric specification, and even below those of some of our parametric specifications. In turn, the resulting average surplus parameter may substantially understate the net benefit of the voucher. We find that this pattern holds across different specifications that allow for observed heterogeneity in the price coefficients as well as those that

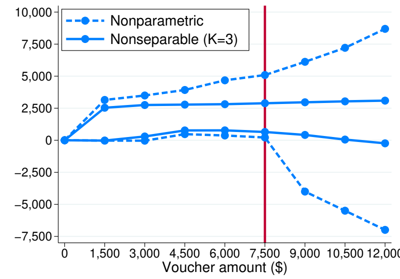
Figure 3: Estimated upper and lower bounds on welfare effects for counterfactual voucher amounts



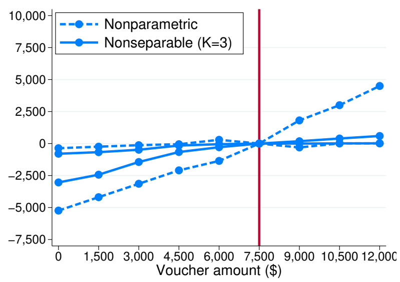
(a) $AB\,(\tau_\mathrm{c})$



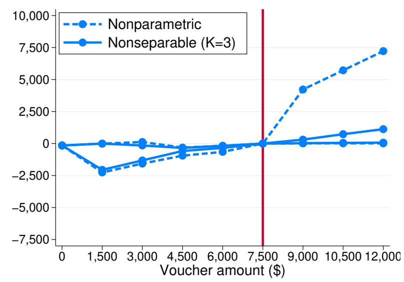
(b) $AC\,(\tau_\mathrm{c})$



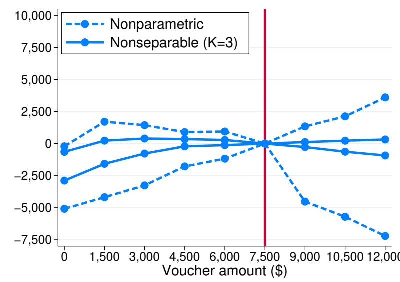
(c) $AS\,(\tau_\mathrm{c})$



(d) $\Delta AB\,(\tau_\mathrm{c})$



(e) $\Delta AC\,(\tau_\mathrm{c})$



(f) $\Delta AS\,(\tau_\mathrm{c})$

SOURCE: *Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)*, U.S. Department of Education, National Center for Education Statistics previously unpublished tabulations.

allow for unobserved heterogeneity, i.e. a mixed logit. This suggests that the parameterizations present in these specifications potentially underestimate the individual valuations for the price decrease induced by the voucher, and in turn how responsive individuals are to this price change. We also find that these patterns continue to hold for parameters measuring the welfare effects of providing counterfactual voucher amounts when comparing the logit estimates of these parameters to those that our tools provide, which we discuss next.

5.4 Welfare Estimates for Counterfactual Voucher Amounts

Figure 3 next presents the estimates of our various parameters measuring the welfare effects of providing counterfactual voucher amounts. These parameters correspond to those illustrated in Table 2 but for a range of values of $\tau = \tau_\mathrm{c}$ not necessarily equal to τ_sq. We also report the differences with the parameter when $\tau = \tau_\mathrm{sq}$, as described in (13)-(15). For conservativeness, we present only results under the nonparametric and the most flexible parametric specifications from Table 2, i.e. Columns (1) and (10), respectively. For expositional reasons, since the plots of the estimated bounds and the 95% confidence intervals are close to each other and hence difficult to visually distinguish, we report only the former. The confidence intervals are available in Appendix S.4.5.

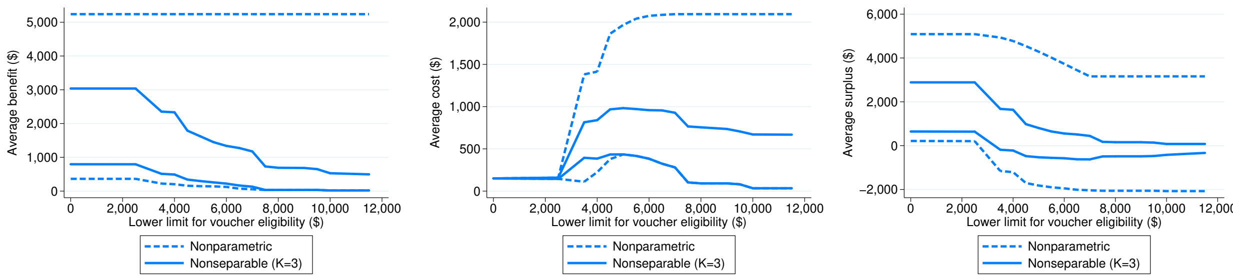
The estimates for $AB(\tau_{\mathrm{c}})$ and $\Delta AB(\tau_{\mathrm{c}})$ reveal, unsurprisingly, that the average benefit increases with the voucher amount. As in Table 2, the bounds under the parametric specification can be considerably tighter than those under the nonparametric specification. Under the parametric specification, we find that the bounds vary more for lower voucher amounts and are more stable for larger amounts. The estimates for $AC(\tau_{\mathrm{c}})$ and $\Delta AC(\tau_{\mathrm{c}})$ reveal, in contrast to the status-quo amount in Table 2, that they are generally not point identified but only bounded. This is because, unlike $AC(\tau_{\mathrm{sq}})$, these parameters are generally functions of demand at values of prices not observed in the data. Unsurprisingly, the bounds under the nonparametric specification vary non-smoothly and those under the parametric specification vary smoothly given that the latter specification imposes a smooth relationship of how demand varies with prices while the former does not. Similar to the average benefit, the average cost also varies more at lower voucher amounts. For some voucher amounts, the estimates are negative, i.e the government has cost savings. This arises because at these values, as before, recipients continue to redeem the voucher and switch to low-price schools from government-funded schools, but now the government actually saves as the costs of funding the voucher at these schools are significantly lower than that of government-funded schools.

Taking the difference of average benefit and cost, the estimates for $AS(\tau_{\mathrm{c}})$ reveal that the provision of counterfactual voucher amounts may have a positive average benefit net of costs. Specifically, under the nonparametric specification, the bounds reveal that we have a positive average surplus for voucher amounts below the status-quo, but potentially not above it. This is because the average costs are low relative to the benefit and potentially even negative at voucher amounts below the status-quo, but drastically increase in a non-smooth manner above the status-quo. Under the parametric specification, the smooth relationship of demand with prices allows the pattern of costs below the status-quo voucher to smoothly extend to voucher amounts above it as well, implying a positive average surplus for all voucher amounts. However, not all counterfactual voucher amounts have the same surplus as the status-quo. Comparing the counterfactual to the status-quo surplus at $\tau_c = \$1{,}500$, the bound for $\Delta AS(\$1{,}500)$ under the parametric specification is $[-\$1{,}578, \$229]$. This suggests that providing vouchers in such low amounts would likely reduce surplus relative to the status-quo amount.

5.5 Role of Low-tuition Schools in the Program

In summary, our welfare estimates reveal that voucher provision has a positive average surplus under both the status-quo and counterfactual voucher amounts. While discussing these results, we specifically noted that they arose in part due to the presence of low-tuition schools in the program that many recipients attend, but that have a small net cost to the government. We conclude our analysis by more directly investigating the importance of these schools in the program when providing the status-quo voucher amount.

Figure 4: Estimated upper and lower bounds on welfare effects of removing schools with tuition at most κ from the program



(a) $AB^{\kappa}(\tau_{\text{sq}})$ (b) $AC^{\kappa}(\tau_{\text{sq}})$ (c) $AS^{\kappa}(\tau_{\text{sq}})$

SOURCE: *Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)*, U.S. Department of Education, National Center for Education Statistics previously unpublished tabulations.

Specifically, we analyze how our estimates change when we remove schools having prices at most a certain amount from the program. To this end, for a given $\kappa \in \mathbf{R}_+$, let $\mathcal{J}^{\kappa} = \{j \in \mathcal{J}_v : p_j(0) \leq \kappa\}$ denote the set of participating private schools with prices no more than κ, and let $j^{\kappa} = \arg\max \mathcal{J}^{\kappa}$ denote the school with the highest price removed from the program. In addition, let $p^{\kappa}(\tau_{\text{sq}}) = (p_1(0), \ldots, p_{j^{\kappa}}(0), p_{j^{\kappa}+1}(\tau_{\text{sq}}), \ldots p_J(\tau_{\text{sq}}))$ denote the prices of the schools in $\mathcal{J}_v$ under the application of the status-quo amount when schools with prices at most κ are removed, i.e. the status-quo voucher amount is applied to only schools with prices above κ. Then, similar to (9), the average benefit of the status-quo voucher amount absent these schools can be defined by

$$AB^{\kappa}(\tau_{\text{sq}}) = E[B_i^{\kappa}(\tau_{\text{sq}})] \; . \tag{38}$$

where $B_i^{\kappa}(\tau_{\text{sq}})$ is given by the variable that solves (8) when replacing $p_j(\tau)$ with $p_j^{\kappa}(\tau_{\text{sq}})$ for $j \in \mathcal{J}_v$. Similarly, the average cost and benefit net of costs can be defined by

$$AC^{\kappa}(\tau_{\text{sq}}) = \sum_{j \in \mathcal{J}} c_j^{\kappa}(\tau_{\text{sq}}) \cdot q_j(p^{\kappa}(\tau_{\text{sq}})) - \sum_{j \in \mathcal{J}} c_j(0) \cdot q_j(p(0)) \; , \tag{39}$$

$$AS^{\kappa}(\tau_{\text{sq}}) = AB^{\kappa}(\tau_{\text{sq}}) - AC^{\kappa}(\tau_{\text{sq}}) \; . \tag{40}$$

where $c_j^{\kappa}(\tau_{\text{sq}}) = c_j(\tau_{\text{sq}})$ for $j \in \mathcal{J} \setminus \mathcal{J}^{\kappa}$ and $c_j^{\kappa}(\tau_{\text{sq}}) = 0$ for $j \in \mathcal{J}^{\kappa}$, i.e. we take the same costs as before except with the difference that we take the schools that are removed from the program to have zero costs. In Appendix S.2.5, we describe how we can continue to use the programs in (33) and (36) to learn about these parameters and, in turn, obtain estimates for these parameters using their empirical counterparts as in Table 2 and Figure 3.

Figure 4 presents the results for the above parameters for a range of values of κ and, as in Figure 3 for conservativeness, for the nonparametric and most flexible parametric specifications from Table 2. Similar to Figure 3, we only report estimates here and present confidence intervals in Appendix S.4.5. The bounds under the nonparametric specification are considerably wider than those under the parametric specification and especially so for the average benefit, where the upper bound stays constant across all values of κ. This is because the data do not provide any cross-price variation and, unlike the parametric specification, the nonparametric specification does not impose any cross-price restrictions. Under the parametric specification, Figure 4(a)-(b) reveal that the average benefits and costs first steeply decrease and increase, respectively, from the removal of low-tuition schools from the program and then become more stable when more expensive schools are removed. This highlights that recipients strongly value the presence of low-tuition schools in the program, and when these schools are removed, switch into relatively expensive government-funded schools.

Taking the difference of the average benefit and costs, Figure 4(c) reveals that the removal of low-tuition schools from the program generally results in the reduction of average surplus. Specifically, we find that absent schools with tuition at most \$3,500 in the program we can potentially have a negative surplus. A closer look at Figure 2(a) reveals that nearly 30% of schools in the program have tuition of at most this value. The estimates from Figure 4(c) highlight that the presence of these low-tuition schools in the program play an essential role in explaining the positive net benefit our analysis finds for the provision of the status-quo voucher amount.

6 Conclusion

In this paper, we develop new discrete choice tools that show how to robustly learn about the welfare effects of providing school vouchers of a given amount in settings where vouchers are randomly allocated. We use our tools to analyze the welfare effects of voucher provision in the DC Opportunity Scholarship Program (OSP). Our estimates reveal that provision of both the status-quo voucher and a wide range of counterfactual vouchers of different amounts have a positive net average benefit, and that these positive results can be explained by the popularity of low-tuition schools in the program.

We conclude by highlighting some fruitful directions for future research, which are beyond scope of this paper. As noted in Section 2, we do not model the potential general equilibrium effects of vouchers. Our analysis should therefore be interpreted as reflecting short-term, partial equilibrium welfare effects, taking long-term general equilibrium responses as fixed. These responses, however, could have potentially ambiguous effects on the welfare of both voucher recipients and non-recipients (Epple et al., 2017). It would hence be interesting to account for these responses and analyze their consequences on the welfare effects of the voucher. For some recent advancements in this direction, see Bhattacharya et al. (2019), who provides welfare results accounting for certain types of equilibrium responses in a different policy setting.

On a related point, note that our welfare analysis is based on a static choice model. In particular, while the OSP provided vouchers valid for at least five years, we model and analyze school choices collected only in the initial year. In this sense, unless individuals do not change their choices across years, our results should be more appropriately interpreted as the welfare effects of a voucher that is to be used in the same year. As noted in Wolf et al. (2010), there is in fact substantial variation in choice across years. It would hence also be interesting to extend our analysis to encompass related welfare parameters in a dynamic discrete choice model, and study if the positive results we find continue to hold.

Finally, and more generally, it would be interesting to generalize our approach to other types of price variation observed in practice. Specifically, we exploit the fact that in our application there was exogenous, discrete variation in prices due to the random allocation of the voucher. However, in many applications, prices could be endogenously related to the underlying variables of the model. A promising avenue for further research would be to extend our tools to accommodate for such endogeneity, for example through instrumental variable-type assumptions as in Berry and Haile (2009, 2014).

References

Abdulkadiroğlu, A., Pathak, P. A. and Walters, C. R. (2018). Free to choose: Can school choice reduce student achievement? *American Economic Journal: Applied Economics*, **10** 175–206.

Allende, C. (2019). Competition under social interactions and the design of education policies. *Job Market Paper*.

Angrist, J., Bettinger, E., Bloom, E., King, E. and Kremer, M. (2002). Vouchers for private schooling in colombia: Evidence from a randomized natural experiment. *American economic review*, **92** 1535–1558.

Arcidiacono, P., Muralidharan, K., Shim, E.-y. and Singleton, J. D. (2016). Valuing school choice: Using a randomized experiment to validate welfare evaluation of private school vouchers.

Berry, S., Gandhi, A. and Haile, P. (2013). Connected substitutes and invertibility of demand. *Econometrica*, **81** 2087–2111.

Berry, S., Levinsohn, J. and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society* 841–890.

Berry, S. T. and Haile, P. A. (2009). Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Tech. rep., National Bureau of Economic Research.

Berry, S. T. and Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, **82** 1749–1797.

Bhattacharya, D. (2015). Nonparametric welfare analysis for discrete choice. *Econometrica*, **83** 617–649.

Bhattacharya, D. (2018). Empirical welfare analysis for discrete choice: Some general results. *Quantitative Economics*, **9** 571–615.

Bhattacharya, D. (2019). The empirical content of binary choice models. *arXiv preprint arXiv:1902.11012*.

Bhattacharya, D., Dupas, P. and Kanaya, S. (2019). Demand and welfare analysis in discrete choice models with social interactions. Tech. rep., National Bureau of Economic Research.

Briesch, R. A., Chintagunta, P. K. and Matzkin, R. L. (2010). Nonparametric discrete choice models with unobserved heterogeneity. *Journal of Business & Economic Statistics*, **28** 291–307.

Carneiro, P. M., Das, J. and Reis, H. (2019). The value of private schools: Evidence from pakistan.

Chesher, A., Rosen, A. M. and Smolinski, K. (2013). An instrumental variable model of multiple discrete choice. *Quantitative Economics*, **4** 157–196.

Chiappori, P. and Komunjer, I. (2009). On the nonparametric identification of multiple choice models.

Compiani, G. (2019). Market counterfactuals and the specification of multi-product demand: a nonparametric approach. Tech. rep.

Dynarski, M., Rui, N., Webber, A. and Gutmann, B. (2018). Evaluation of the dc opportunity scholarship program: Impacts two years after students applied. ncee 2018-4010. *National Center for Education Evaluation and Regional Assistance.*

Epple, D., Romano, R. E. and Urquiola, M. (2017). School vouchers: A survey of the economics literature. *Journal of Economic Literature*, **55** 441–92.

Friedman, M. (1962). *Capitalism and freedom.* University of Chicago press.

Gazmuri, A. (2019). School segregation in the presence of student sorting and cream-skimming: Evidence from a school voucher reform.

Ho, K. and Pakes, A. (2014). Hospital choices, hospital prices, and financial incentives to physicians. *American Economic Review*, **104** 3841–84.

Howell, W. G., Wolf, P. J., Peterson, P. E. and Campbell, D. E. (2000). Test-score effects of school vouchers in dayton, ohio, new york city, and washington, dc: Evidence from randomized field trials.

Kalouptsidi, M., Kitamura, Y., Lima, L. and Souza-Rodrigues, E. A. (2020). Partial identification and inference for dynamic models and counterfactuals. Tech. rep., National Bureau of Economic Research.

Kamat, V. (2019). Identification with latent choice sets. *arXiv preprint arXiv:1711.02048.*

Kitamura, Y. and Stoye, J. (2018). Nonparametric analysis of random utility models. *Econometrica*, **86** 1883–1909.

Krueger, A. B. and Zhu, P. (2004). Another look at the new york city school voucher experiment. *American Behavioral Scientist*, **47** 658–698.

Levin, H. M. and Driver, C. E. (1997). Cost of an educational voucher system. *Education Economics*, **5** 265–283.

Manski, C. F. (2007). Partial identification of counterfactual choice probabilities. *International Economic Review*, **48** 1393–1410.

Matzkin, R. L. (1993). Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics*, **58** 137–168.

Mayer, D. P., Peterson, P. E., Myers, D. E., Tuttle, C. C. and Howell, W. G. (2002). School choice in new york city after three years: An evaluation of the school choice scholarships program. final report.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics* 105–142.

Mills, J. N. and Wolf, P. J. (2017). Vouchers in the bayou: The effects of the louisiana scholarship program on student achievement after 2 years. *Educational Evaluation and Policy Analysis*, **39** 464–484.

Mogstad, M., Santos, A. and Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, **86** 1589–1619.

Muralidharan, K. and Sundararaman, V. (2015). The aggregate effect of school choice: Evidence from a two-stage experiment in india. *The Quarterly Journal of Economics*, **130** 1011–1066.

Neilson, C. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of economics & management strategy*, **9** 513–548.

Sable, J. and Hill, J. (2006). Overview of public elementary and secondary students, staff, schools, school districts, revenues, and expenditures: School year 2004-05 and fiscal year 2004. ed tab. nces 2007-309. *National Center for Education Statistics.*

Tebaldi, P., Torgovitsky, A. and Yang, H. (2019). Nonparametric estimates of demand in the california health insurance exchange. Tech. rep., National Bureau of Economic Research.

Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge university press.

US Census (2005). Public education finances.

Wolf, P., Gutmann, B., Puma, M., Kisida, B., Rizzo, L., Eissa, N., Carr, M. and Silverberg, M. (2010). Evaluation of the dc opportunity scholarship program: Final report. *National Center for Education Evaluation, Institute for Education Sciences Report*, **4018**.

Supplementary Appendix to "Estimating Welfare Effects in a Nonparametric Choice Model: The Case of School Vouchers"

Vishal Kamat
Toulouse School of Economics
University of Toulouse Capitole
vishal.kamat@tse-fr.eu

Samuel Norris
Harris School of Public Policy
University of Chicago
samnorris@uchicago.edu

June 16, 2021

Abstract

This document presents proofs and additional details pertinent to the main analysis for the authors' paper titled "Estimating Welfare Effects in a Nonparametric Choice Model: The Case of School Vouchers." Section S.1 presents proofs of all results. Section S.2 presents additional details pertinent to the identification analysis. Section S.3 presents the procedures used to perform statistical inference in the empirical analysis. Section S.4 presents additional details pertinent to the empirical analysis.

S.1 Proofs of Propositions

S.1.1 Proof of Proposition 3.1

The proof of this proposition follows from Bhattacharya (2018, Proposition 1) and Bhattacharya (2018, Theorem 1). We reproduce these proofs here in the context of our setup and notation for completeness. For convenience, we drop the i sub-index here.

To see why the variable $B(\tau)$ exists and is unique, note first that the right hand side of (8) is continuous in $B(\tau)$ as U_j is a continuous function for each $j \in \mathcal{J}_s$. In addition, since

$$\begin{aligned} Y_j &> Y_j - B(\tau) && \text{for } j \in \mathcal{J}_g \cup \mathcal{J}_n \ , \\ Y_j - p_j(0) &> Y_j - p_j(\tau) - B(\tau) && \text{for } j \in \mathcal{J}_v \ , \end{aligned}$$

for $B(\tau) > \tau$, and

$$\begin{aligned} Y_j &< Y_j - B(\tau) && \text{for } j \in \mathcal{J}_g \cup \mathcal{J}_n \ , \\ Y_j - p_j(0) &< Y_j - p_j(\tau) - B(\tau) && \text{for } j \in \mathcal{J}_v \ , \end{aligned}$$

for $B(\tau) < 0$, note that it follows from the fact that U_j is strictly increasing for each $j \in \mathcal{J}_s$ that if $B(\tau) > \tau$ then the right hand side of (8) is strictly smaller than its left hand side, whereas if $B(\tau) < 0$ then the right hand side will be strictly greater than the left hand side. Then, using these two arguments together, it follows by the intermediate value theorem that there exists a $B(\tau) \in [0, \tau]$ such that the right hand side equals the left hand side, i.e. a solution to (8) exists. Furthermore, given that U_j is strictly increasing for each $j \in \mathcal{J}_s$, it also follows that the solution to (8) must be unique.

To see why the average value of $B(\tau)$ is given by (10), note first from above that $B(\tau) \in [0, \tau]$ and, in turn, that

$$\begin{aligned} \text{Prob}[B(\tau) \leq a] = 0 \quad \text{for } a < 0 \ , \\ \text{Prob}[B(\tau) \leq a] = 1 \quad \text{for } a \geq \tau \ . \end{aligned}$$

Next, to calculate this probability for $a \in [0, \tau)$, note that since U_j is strictly increasing for each $j \in \mathcal{J}_s$, we have that $B(\tau) \leq a$ is equivalent to

$$\max\left\{ \max_{j \in \mathcal{J}_g \cup \mathcal{J}_n} U_j(Y_j) \ , \ \max_{j \in \mathcal{J}_v} U_j\left(Y_j - p_j(0)\right) \right\} \\ \geq \\ \max\left\{ \max_{j \in \mathcal{J}_g \cup \mathcal{J}_n} U_j(Y_j - a) \ , \ \max_{j \in \mathcal{J}_v} U_j\left(Y_j - p_j(\tau) - a\right) \right\} \ .$$

Then, it follows that for $a \in [a_l(\tau), a_{l+1}(\tau))$ for $l = 0, \ldots, j(\tau)$, we have

$$
\begin{aligned}
\text{Prob}[B(\tau) \le a] &= \sum_{j \in \mathcal{J}_g \cup \mathcal{J}_n} \text{Prob}\left[U_j(Y_j) \ge \max \left\{ \begin{array}{c} \max\limits_{m \in \mathcal{J}_g \cup \mathcal{J}_n \setminus \{j\}} U_m(Y_m) \ , \ \max\limits_{m \in \mathcal{J}_v} U_m\left(Y_m - p_m(0)\right), \\ \max\limits_{m \in \mathcal{J}_g \cup \mathcal{J}_n} U_m(Y_m - a) \ , \ \max\limits_{m \in \mathcal{J}_v} U_m\left(Y_m - p_m(\tau) - a\right) \end{array} \right\} \right] \\
&\quad + \sum_{j \in \mathcal{J}_v} \text{Prob}\left[\begin{array}{l} U_j(Y_j - p_j(0)) \\ \ge \max \left\{ \begin{array}{c} \max\limits_{m \in \mathcal{J}_g \cup \mathcal{J}_n} U_m(Y_m) \ , \ \max\limits_{m \in \mathcal{J}_v \setminus \{j\}} U_m\left(Y_m - p_m(0)\right), \\ \max\limits_{m \in \mathcal{J}_g \cup \mathcal{J}_n} U_m(Y_m - a) \ , \ \max\limits_{m \in \mathcal{J}_v} U_m\left(Y_m - p_m(\tau) - a\right) \end{array} \right\} \end{array} \right] \\
&= \sum_{j \in \mathcal{J}_g \cup \mathcal{J}_n} \text{Prob}\left[U_j(Y_j) \ge \max \left\{ \begin{array}{c} \max\limits_{m \in \mathcal{J}_g \cup \mathcal{J}_n \setminus \{j\}} U_m(Y_m) \ , \ \max\limits_{m \in \mathcal{J}_v,\ m \le l} U_m\left(Y_m - p_m(0)\right), \\ \max\limits_{m \in \mathcal{J}_v, m > l} U_m\left(Y_m - p_m(\tau) - a\right) \end{array} \right\} \right] \\
&\quad + \sum_{j=1}^{l} \text{Prob}\left[\begin{array}{l} U_j(Y_j - p_j(0)) \\ \ge \max \left\{ \begin{array}{c} \max\limits_{m \in \mathcal{J}_g \cup \mathcal{J}_n} U_m(Y_m) \ , \ \max\limits_{1 \le m \le l, m \ne j} U_m\left(Y_m - p_m(0)\right), \\ \max\limits_{l < m \le J} U_m\left(Y_m - p_m(\tau) - a\right) \end{array} \right\} \end{array} \right] \\
&= \sum_{\substack{j \in \{g,n\}, \\ 1 \le j \le l}} q_j\left(p_1(0), \ldots, p_l(0), p_{l+1}(\tau) + a, \ldots, p_J(\tau) + a\right) \ ,
\end{aligned}
$$

where the second equality follows from the fact that U_j is strictly increasing for each $j \in \mathcal{J}_s$ along with $Y_j - p_j(\tau) - a \le Y_j - p_j(0)$ for $j \le l$ as $p_j(\tau) = 0$ and $a \ge a_j(\tau) = p_j(0)$, and $Y_j - p_j(\tau) - a \ge Y_j - p_j(0)$ for $j > l$ as $p_j(0) - p_j(\tau) = \min\{\tau, p_j(0)\} \ge a_{l+1}(\tau) > a$, and the final equality follows from the definition of the average demand functions. Finally, given that for a positive random variable X we have that its expectation is given by

$$E[X] = \int_0^\infty \left[1 - \text{Prob}[X \le x]\right] dx \ ,$$

it follows from the above characterization of $\text{Prob}[B(\tau) \le a]$ along with noting that

$$
\begin{aligned}
&\sum_{j=l+1}^{J} q_j(p_1(0), \ldots, p_l(0), p_{l+1}(t) + a, \ldots, p_J(\tau) + a) \\
&\qquad\qquad = 1 - \left(\sum_{\substack{j \in \{g,n\}, \\ 1 \le j \le l}} q_j\left(p_1(0), \ldots, p_l(0), p_{l+1}(\tau) + a, \ldots, p_J(\tau) + a\right) \right) \ ,
\end{aligned}
$$

that the average value of $B(\tau)$ is given by (10). This concludes the proof.

S.1.2 Proof of Proposition 4.1

In order to prove the proposition, we need to show that $\Theta \subseteq \Theta_B$ and $\Theta_B \subseteq \Theta$, and that $\Theta_B = [\underline{\theta}_B, \bar{\theta}_B]$ if $\mathbf{B}$ is non-empty. Below, we divide the proof into three parts respectively showing each of these statements. First, we show $\Theta \subseteq \Theta_B$, i.e for every $\theta_0 \in \Theta$ there exists a $\beta \in \mathbf{B}$ such that $\theta_B(\beta) = \theta_0$. Second, we show $\Theta_B \subseteq \Theta$, i.e. for every $\theta_0 \in \Theta_B$ there exists a $q \in \mathbf{Q}$ such that $\theta(q) = \theta_0$. Third, we show that if $\mathbf{B}$ is non-empty then $\Theta_B = [\underline{\theta}_B, \bar{\theta}_B]$.

Before proceeding, it is useful to first explicitly state the restrictions on β that characterize $\mathbf{B}$ as well as the expression for the function θ_B. To this end, note that $\mathbf{B}$ corresponds to all β such that the corresponding q determined by (29) satisfies the restrictions in (16)-(17), (4) and (18)-(21). In turn, observe that (16) and (17) equivalently correspond to

$$0 \leq \beta_j(w) \leq 1 \quad \text{for each } j \in \mathcal{J} \ , \tag{S.1}$$

$$\sum_{j \in \mathcal{J}} \beta_j(w) = 1 \tag{S.2}$$

for each $w \in \mathcal{W}$. To state the equivalent restriction corresponding to (4), we introduce additional notation where $w(j) = \{t \in \mathbf{R} : p_j = t \text{ for some } p \in w\}$ for each $w \in \mathcal{W}$ and $j \in \mathcal{J}_v$. Then, observe that (4) equivalently corresponds to stating that for each $w, w' \in \mathcal{W}$ such that $t > t'$ for all $t \in w(j),\ t' \in w'(j)$ for $j \in \mathcal{J}' \subseteq \mathcal{J}_v$ and $w(j) = w'(j)$ for $j \in \mathcal{J}_v \setminus \mathcal{J}'$, we have that

$$\beta_j(w) \geq \beta_j(w') \tag{S.3}$$

for each $j \in \mathcal{J} \setminus \mathcal{J}'$. Finally, observe that (18)-(21) equivalently corresponds to

$$\beta_j\left(\{p(0)\}\right) = P_{j|0} \ , \tag{S.4}$$

$$\beta_j\left(\{p(\tau_{\text{sq}})\}\right) = P_{j|1} \ , \tag{S.5}$$

for each $j \in \mathcal{J}$. Then, it follows we can equivalently characterize $\mathbf{B}$ as

$$\mathbf{B} = \left\{ \beta \in \mathbf{R}^{d_\beta} : \beta \text{ satisfies (S.1)} - \text{(S.5)} \right\} \ , \tag{S.6}$$

i.e. the set of all β that satisfy the above restrictions. The expression for θ_B depends on the choice of parameter. As each of our parameters can be characterized by $AB(\tau)$ or $AC(\tau)$ by appropriately choosing $\tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$ along with taking differences, we only derive the expression for θ_B when θ corresponds to either $AB(\tau)$ or $AC(\tau)$ for some generic $\tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$. In the case of $AB(\tau)$, observe that θ and θ_B are given by

$$\theta(q) \equiv \sum_{l=0}^{j(\tau)} \sum_{j=l+1}^{J} \int_{a_l(\tau)}^{a_{l+1}(\tau)} q_j(\min\{p(0), p(\tau) + a\}) \ da \ , \tag{S.7}$$

$$= \sum_{l=0}^{j(\tau)} \sum_{j=l+1}^{J} \sum_{u \in \mathcal{U}(l,\tau)} \int_{\underline{a}_u}^{\bar{a}_u} q_j(\min\{p(0), p(\tau) + a\}) \ da \ , \tag{S.8}$$

$$= \sum_{l=0}^{j(\tau)} \sum_{j=l+1}^{J} \sum_{u \in \mathcal{U}(l,\tau)} (\bar{a}_u - \underline{a}_u) \cdot \beta_j(h(u)) \equiv \theta_B(\beta) \ , \tag{S.9}$$

where $\min\{p(0), p(\tau)+a\} \equiv (\min\{p_1(0), p_1(\tau)+a\}, \ldots, \min\{p_J(0), p_J(\tau)+a\})$, $\mathcal{U}(l,\tau) \equiv \{u \in \mathcal{U} : u \subseteq \mathcal{P}_l(\tau)\}$ and $h(u) \in \mathcal{W}$ for a given $u \in \mathcal{U}$ is such that $h(u) = \prod_{j=1}^{J} u(j)$. In particular, the first line simply recalls rewriting of $AB(t)$ in terms of q from (10), the second line follows from rewriting (10) in terms of the collection of sets $\mathcal{U}$, and the third line then directly follows from substituting the equation in (29). In the case of $AC(\tau)$, observe that θ and θ_B are given by

$$\begin{aligned}\theta(q) &\equiv \sum_{j\in\mathcal{J}} c_j(\tau)\cdot q_j(p(\tau)) - \sum_{j\in\mathcal{J}} c_j(0)\cdot q_j(p(0)) \ , &\text{(S.10)}\\ &= \sum_{j\in\mathcal{J}} c_j(\tau)\cdot \beta_j(\{p(\tau)\}) - \sum_{j\in\mathcal{J}} c_j(0)\cdot \beta_j(\{p(0)\}) \ \equiv \theta_B(\beta) \ , &\text{(S.11)}\end{aligned}$$

where the first line simply recalls the definition of $AC(\tau)$ from (11), and the second line follows directly from substituting the equation in (29).

Given the explicit characterizations of $\mathbf{B}$ and θ_B, we now proceed to presenting the proofs of each of the three parts.

Part 1: Since $\theta_0 \in \Theta$, there exists by definition a $q \in \mathbf{Q}$ such that $\theta(q) = \theta_0$. Using this q, we construct a β such that $\beta \in \mathbf{B}$ and $\theta_B(\beta) = \theta_0$. In particular, we take β to be such that

$$\beta_j(w) = \int_0^1 q_j\left(p(a,w)\right) da \tag{S.12}$$

for each $w \in \mathcal{W}$ and $j \in \mathcal{J}$, where, for each $a \in (0,1)$, $p(a,w) = (p_1(a,w), \ldots, p_J(a,w))$ with $p_j(a,w) = \underline{w}(j) + (\bar{w}(j) - \underline{w}(j))\cdot a$ for each $j \in \mathcal{J}_v$ such that $\bar{w}(j) = \sup\{t : t \in w(j)\}$ and $\underline{w}(j) = \inf\{t : t \in w(j)\}$.

We now show this constructed β is such that $\beta \in \mathbf{B}$ and $\theta_B(\beta) = \theta_0$. First, we show that $\beta \in \mathbf{B}$, i.e. it satisfies the restrictions in (S.1)-(S.5). The restriction in (S.1) is satisfied for each $w \in \mathcal{W}$ and $j \in \mathcal{J}$ as

$$0 \le \beta_j(w) = \int_0^1 q_j\left(p(a,w)\right) da \le \int_0^1 1 da \ ,$$

where the equality follows from (S.12) and the inequalities follow from (16). Similarly, the restriction in (S.2) is satisfied for each $w \in \mathcal{W}$ and $j \in \mathcal{J}$ as

$$\sum_{j\in\mathcal{J}} \beta_j(w) = \int_0^1 \sum_{j\in\mathcal{J}} q_j\left(p(a,w)\right) da = \int_0^1 1 da \ ,$$

where the first equality follows from (S.12) and the second equality from (17). Next, to see why (S.3) is satisfied, take $w, w' \in \mathcal{W}$ such that $t > t'$ for all $t \in w(j)$, $t' \in w'(j)$ for $j \in \mathcal{J}' \subseteq \mathcal{J}_v$ and

$w(j) = w'(j)$ for $j \in \mathcal{J}_v \setminus \mathcal{J}'$. Observe, for each $a \in (0,1)$, it follows that $p_j(a,w) > p_j(a,w)$ for $j \in \mathcal{J}' \subseteq \mathcal{J}_v$ and $p_j(a,w) = p_j(a,w')$ for $j \in \mathcal{J}_v \setminus \mathcal{J}'$. In turn, it directly follows from (4) that

$$q_j(a,w) \geq q_j(a,w')$$

for each $a \in (0,1)$ and $j \in \mathcal{J} \setminus \mathcal{J}'$. Then, taking the integral over $a \in (0,1)$ and using (S.12), it directly follows that (S.3) is satisfied. Finally, we show that the restrictions in (S.4) and (S.5) are satisfied. To this end, observe that this is the case as

$$\beta_j(\{p(0)\}) = q_j(p(0)) = P_{j|0} \ ,$$
$$\beta_j(\{p(\tau_{\text{sq}})\}) = q_j(p(\tau_{\text{sq}})) = P_{j|1}$$

for each $j \in \mathcal{J}$, where in both lines the first equality follows from (S.12) and the second equality follows from (18)-(21).

Next, we show that the constructed β is such that $\theta_B(\beta) = \theta_0$. Since $\theta(q) = \theta_0$, this is equivalent to showing $\theta_B(\beta) = \theta(q)$. As we noted above, each of our parameters can be written in terms of $AB(\tau)$ or $AC(\tau)$ for an appropriate choice of $\tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$. As a result, we only show $\theta_B(\beta) = \theta(q)$ when our parameter is equal to either $AB(\tau)$ or $AC(\tau)$ for a generic $\tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$. In the case of $AB(\tau)$, observe that the various components in $\theta(q)$ in (S.8) can be written in terms of the constructed β as

$$\int_{\underline{a}_u}^{\bar{a}_u} q_j\left(\min\{p(0), p(\tau) + a\}\right) da = (\bar{a}_u - \underline{a}_u) \int_0^1 q_j\left(p(a, h(u))\right) da = \beta_j(h(u))$$

for each $u \in \mathcal{U}(l,\tau)$, $0 \leq l \leq j(\tau)$ and $l+1 \leq j \leq J$, where the first equality follows from the change of variables $a = \underline{a}_u + (\bar{a}_u - \underline{a}_u) \cdot a'$ along with the above definition of $p(a,w)$ with $w = h(u)$, and the second equality from (S.12). Similarly, in the case of $AC(\tau)$, observe the various components of $\theta(q)$ in (S.10) can be written as

$$q_j\left(p(\tau')\right) = \int_0^1 q_j\left(p\left(a, \{p(\tau')\}\right)\right) da = \beta_j(\{p(\tau')\})$$

for each $\tau' \in \{0, \tau\}$ and $j \in \mathcal{J}$, where the first equality follows by construction given that $p\left(a, \{p(\tau')\}\right) = p(\tau')$ and the second equality from (S.12). Then, in both cases, substituting these components rewritten in terms of β in the expressions for $\theta(q)$ in (S.8) and (S.10) results in the corresponding expression for $\theta_B(\beta)$ in (S.9) and (S.11), respectively, and, in turn, that $\theta_B(\beta) = \theta(q)$. This completes the first part of the proof.

Part 2: Since $\theta_0 \in \Theta_B$, there exist by definition a $\beta \in \mathbf{B}$ such that $\theta_B(\beta) = \theta_0$ and, in turn, by how Θ_B and θ_B were constructed, a $q \in \mathbf{Q}_B^{\text{fd}}$ that is related to β by the equation in (29) such that $\theta(q) = \theta_B(\beta) = \theta_0$. Since it holds that $\mathbf{Q}_B^{\text{fd}} \subseteq \mathbf{Q}_B$, it follows that $q \in \mathbf{Q}$. This completes the second part of the proof.

Part 3: Given the various linear restrictions that define $\mathbf{B}$ in (S.6), observe that $\mathbf{B}$ is a convex and compact set. In addition, it is also a non-empty set by assumption. Then, since θ_B is a continuous real-valued scalar function, it follows that the image of this function over $\mathbf{B}$ given by Θ_B is a convex, compact and non-empty set on the real line, i.e. a closed interval with endpoints given by (32). This completes the final part of the proof.

S.1.3 Proof of Proposition 4.2

Note $\mathbf{A}$ is a connected and non-empty set. Then, since θ_A is a continuous real-valued scalar function, it follows that the image of this function over $\mathbf{A}$ given by Θ_A is a connected and non-empty set on the real line, i.e. an interval whose closure has endpoints given by (36).

S.2 Additional Details on Identification Analysis

S.2.1 Point Identification of Demand Functions

In this section, we show the demand functions q can be point identified under the specification in (6)-(7). To this end, observe, under this specification, the data restrictions in (18)-(21) imply

$$\alpha_{j0} + \alpha_{j1} \cdot p_j\left(0\right) = P_{j|0} - P_{g|0} \ , \tag{S.13}$$

$$\alpha_{j0} + \alpha_{j1} \cdot p_j\left(\tau_{\text{sq}}\right) = P_{j|1} - P_{g|1} \tag{S.14}$$

for each $j \in \mathcal{J}_v$, and

$$\alpha_n = P_{n|0} - P_{g|0} = P_{n|1} - P_{g|1} \ . \tag{S.15}$$

From (S.13)-(S.14), given there are two equations and two unknowns, it follows that α_{j0} and α_{j1} are point identified for $j \in \mathcal{J}_v$ by the data provided $p_j\left(\tau_{\text{sq}}\right)$ does not equal $p_j\left(0\right)$. From (S.15), it directly follows that α_n is point identified. In turn, since α is point identified, it follows that the q is point identified under this specification.

It is worth highlighting that the specification in (6)-(7) imposes restrictions conceptually similar to a logit specification, a commonly used parameterization in discrete choice analysis, that also achieves point identification in our setup. This logit specification is given by

$$\log(q_j(p)) - \log(q_g(p)) = \gamma_{j0} + \gamma_{j1} \cdot p_j \quad \text{for } j \in \mathcal{J}_v \ , \tag{S.16}$$

$$\log(q_n(p)) - \log(q_g(p)) = \gamma_n \ , \tag{S.17}$$

for some $\{\gamma_{jk} : j \in \mathcal{J}_v,\ 0 \leq k \leq 1\}$ and $\{\gamma_{j0} : j \in \{g, n\}\}$, i.e. the difference in the log demands for a given school in $\mathcal{J}_v$ and any school in $\mathcal{J}_g$ is a linear function of that school's price and the difference in the log demands for any school in $\mathcal{J}_n$ and any school in $\mathcal{J}_g$ is constant. Often, the logit specification is imposed by making assumptions on the underlying utilities. The one in (S.16)-(S.17) can for example be implied by assuming that

$$
\begin{aligned}
U_{ij} &= \gamma_{j0} + \gamma_{j1} \cdot p_j + \epsilon_{ij} \quad \text{for } j \in \mathcal{J}_v \quad , \\
\max_{j \in \mathcal{J}_n} U_{ij} &= \gamma_n + \epsilon_{in} \ , \\
\max_{j \in \mathcal{J}_g} U_{ij} &= 0 \ ,
\end{aligned}
$$

i.e. the utility of each school in $\mathcal{J}_v$ is linear function of its own price and additively separable in the remaining underlying unobservables and the utility of the maximum government school is normalized to zero, and by also assuming that the components of the remaining underlying unobservables

$$
(\epsilon_{in}, \epsilon_{i1}, \ldots, \epsilon_{iJ})
$$

are independently and identically distributed according to the Type I extreme value distribution.

Importantly, comparing (6)-(7) to (S.16)-(S.17), observe that both specifications impose arguably strong restrictions that do not allow each demand function in $\mathcal{J}$ to flexibly vary with prices of other schools. Similar to how (6)-(7) does not allow $q_j - q_g$ for a given $j \in \mathcal{J}$ to not vary with prices p_m for $m \neq j \in \mathcal{J}_v$, (S.16)-(S.17) also does not allow $\log(q_j) - \log(q_g)$ to vary in such a manner. As a result, similar to how the data in our empirical analysis in Section 5 reveals that (6)-(7) is misspecified, the data will also imply that this is the case with (S.16)-(S.17).

S.2.2 Procedure to Compute $\mathcal{U}$

In this section, we describe a procedure that can be used to obtain a collection of sets $\mathcal{U}$ that partition $\mathcal{P}^*$ in (27) as in Definition 4.1. Before proceeding, note, in the formal sense, a partition ensures that each element of $\mathcal{P}^*$ is only in one set of $\mathcal{U}$. As a result, this requires carefully defining the boundaries of the set $u \in \mathcal{U}$ satisfying Definition 4.1(i) to be either closed or open. However, in a practical sense, this distinction is not required in our analysis as our parameters only take Lebesgue integrals over these sets. To this end, in this section, we instead consider the following alternative set of prices to that in (27):

$$
\mathcal{P}^{**} = \bigcup_{l=0}^{j(\tau_{\text{sq}})} \mathcal{P}'_l\left(\tau_{\text{sq}}\right) \bigcup_{l=0}^{j(\tau_{\text{c}})} \mathcal{P}'_l\left(\tau_{\text{c}}\right) \bigcup \left\{p(0), p\left(\tau_{\text{sq}}\right), p\left(\tau_{\text{c}}\right)\right\}
$$

where $\mathcal{P}'_l(\tau) = \{p \in \mathcal{P} : p_j = \min\{p_j(0), p_j(\tau) + a\} \text{ for } a \in (a_l(\tau), a_{l+1}(\tau)) \text{ for each } j \in \mathcal{J}_v\}$ for $0 \leq l \leq j(\tau)$ for each $\tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$, and describe a procedure to show how to obtain a collection of sets $\mathcal{U}$ that partition this set as in Definition 4.1. We emphasize, however, by using more notation to carefully adjust the endpoints of the intervals, the collection of sets we obtain will also correspond to a partition of $\mathcal{P}^*$.

In order to understand the main idea behind our procedure, observe first that

$$\mathcal{U}(\tau_{\text{sq}}) \cup \mathcal{U}(\tau_{\text{c}})$$

equals $\mathcal{P}^{**}$, where

$$\mathcal{U}(\tau) = \left\{\mathcal{P}'_0(\tau), \ldots, \mathcal{P}'_{j(\tau)}(\tau), \{p(0)\}, \{p(\tau)\}\right\} \ ,$$

For each $\tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$, observe that $\mathcal{U}(\tau)$ corresponds to a partition of the union of the sets in it and that and also that each $u \in \mathcal{U}(t)$ satisfies Definition 4.1(i) or (ii) along with each $u, u' \in \mathcal{U}(t)$ satisfying (28). However, for a given $u \in \mathcal{U}(\tau_{\text{sq}})$ and $u' \in \mathcal{U}(\tau_{\text{c}})$, it may possibly be that (28) is not satisfied. In particular, it may be the case that $u(j) \neq u'(j)$ for some $j \in \mathcal{J}_v$, but that $u(j) \cap u'(j)$ is a non-empty set equal to some interval with end points corresponding to those of either $u(j)$ or $u'(j)$. For example, if $u = \mathcal{P}'_l(\tau_{\text{sq}})$ for some $l \in \{0, \ldots, j(\tau_{\text{sq}})\}$ and $u' = \mathcal{P}'_l(\tau_{\text{c}})$ for some $l' \in \{0, \ldots, j(\tau_{\text{c}})\}$, we can have for some $j \in \mathcal{J}_v$ that

$$u(j) \cap u'(j) = \left(\min\left\{p_j(0), p_j(\tau_{\text{c}}) + a_{l'}(\tau_{\text{c}})\right\}, \min\left\{p_j(0), p_j(\tau_{\text{sq}}) + a_{(l+1)}(\tau_{\text{sq}})\right\}\right)$$

or

$$u(j) \cap u'(j) = \left(\min\left\{p_j(0), p_j(\tau_{\text{sq}}) + a_l(\tau_{\text{sq}})\right\}, \min\left\{p_j(0), p_j(\tau_{\text{c}}) + a_{(l'+1)}(\tau_{\text{c}})\right\}\right)$$

depending on whether the lower (upper) end point of $u'(j)$ is greater or lower than the lower (upper) end point of $u'(j)$, and, alternatively, if $u' = \{p(\tau_{\text{c}})\}$ instead, we can have for some $j \in \mathcal{J}_v$ that

$$u(j) \cap u'(j) = \{p_j(\tau_{\text{c}})\} \ .$$

Our procedure is based on the idea that in such cases we can further partition the elements of $\mathcal{U}(\tau_{\text{sq}})$ and $\mathcal{U}(\tau_{\text{c}})$ to alternatively obtain $\mathcal{U}_1(\tau_{\text{sq}})$ and $\mathcal{U}_1(\tau_{\text{c}})$ such that, for each $\tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$, we continue to have that each $u \in \mathcal{U}_1(\tau)$ satisfies Definition 4.1(i) or (ii) and that each $u, u' \in \mathcal{U}_1(\tau)$ satisfies (28), but, in addition, we also have that each $u \in \mathcal{U}_1(\tau_{\text{sq}})$ and $u' \in \mathcal{U}_1(\tau_{\text{c}})$ satisfy (28) for a given $j^* \in \mathcal{J}_v$. For a given $j^* \in \mathcal{J}_v$ and each $\tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$, denoting by

$$\mathcal{A}_1(\tau) = \{a_0(\tau), \ldots, a_{j(\tau)+1}(\tau)\}$$

the set of points used to define the sets in (25) and by

$$\mathcal{A}_2(\tau) = \mathcal{A}_1(\tau) \cup \{\min\{p_{j^*}(0), a + p_{j^*}(\tau')\} - p_{j^*}(\tau) : a \in \mathcal{A}_1(\tau')\}$$

the resulting set that includes all the points in $\mathcal{A}_1(\tau)$ in addition to those that play a role, when we sum it with $p_{j^*}(\tau)$, in defining the end points of the interval $u(j^*)$ for each $u \in \mathcal{U}(\tau')$ for $\tau' \neq \tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$, this alternative set can be given by

$$\mathcal{U}_1(\tau) = \left\{\mathcal{P}''_0(\tau), \ldots, \mathcal{P}''_{L(\tau)-1}(\tau), \{p(0)\}, \{p(\tau)\}\right\} \tag{S.18}$$

where $a^*_0(\tau) < \ldots < a^*_{L(\tau)}(\tau)$ denotes the ordered values of the set $\mathcal{A}_2(\tau)$ and

$$\mathcal{P}''_l(\tau) = \left\{p \in \mathcal{P} : p_j = \min\{p_j(0), p_j(\tau) + a\},\ a \in \left(a^*_l(\tau), a^*_{l+1}(\tau)\right) \text{ for each } j \in \mathcal{J}_v\right\}$$

for $0 \leq l \leq L(\tau)-1$. As before, observe that each $u \in \mathcal{U}_1(\tau)$ satisfies Definition 4.1(i) or (ii) and that each $u, u' \in \mathcal{U}_1(\tau)$ satisfies (28). To see why, in addition, for each $u \in \mathcal{U}_1(\tau_{\text{sq}})$ and $u' \in \mathcal{U}_1(\tau_{\text{c}})$ we have that (28) is satisfied for $j^* \in \mathcal{J}_v$, suppose that it wasn't the case. In this case, we would then instead have that $u(j^*) \cap u'(j^*) = [\underline{a}, \bar{a}]$ (or, alternatively, $[\bar{a}, \underline{a}]$), where $\underline{a} = \min\{p_{j^*}(0), p_{j^*}(\tau_{\text{sq}}) + a\}$ for some $a \in \mathcal{A}_2(\tau_{\text{sq}})$ and $\bar{a} = \min\{p_{j^*}(0), p_{j^*}(\tau_{\text{c}}) + a'\}$ for some $a' \in \mathcal{A}_2(\tau_{\text{c}})$. However, observe since, by construction, we have that $\bar{a} - p_{j^*}(\tau_{\text{sq}}) \in \mathcal{A}_2(\tau_{\text{sq}})$ and $\underline{a} - p_{j^*}(\tau_{\text{c}}) \in \mathcal{A}_2(\tau_{\text{c}})$, it follows that there also exists a $\tilde{u} \in \mathcal{U}_1(\tau_{\text{sq}})$ and $\tilde{u}' \in \mathcal{U}_1(\tau_{\text{c}})$ such that $\tilde{u}(j^*), \tilde{u}'(j^*) \subseteq [\underline{a}, \bar{a}]$. As a result, since it holds that $u, \tilde{u} \in \mathcal{U}_1(\tau_{\text{sq}})$ and $u', \tilde{u}' \in \mathcal{U}_1(\tau_{\text{c}})$ both satisfy (28), it must in fact be the case that $u(j^*) = u'(j^*) = [\bar{a}, \underline{a}]$ or that $u(j^*) \cap u'(j^*) = \emptyset$.

Our procedure is based on applying this same idea simultaneously to all $j \in \mathcal{J}_v$ with the aim of ensuring that (28) is satisfied for all $j \in \mathcal{J}_v$. However, unlike doing it for a given j^*, doing so simultaneously may generate sets that do not necessarily satisfy (28) for all $j \in \mathcal{J}_v$. As a result, our procedure then continues to iterate through simultaneous applications of the idea until it generates sets that satisfy (28) for all $j \in \mathcal{J}_v$.

To summarize, we present the above described procedure in terms of the following step-wise algorithm:

Step 1: For each $\tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$, take $\mathcal{A}^*_0(\tau) = \emptyset$ and

$$\mathcal{A}^*_1(\tau) = \{a_0(\tau), \ldots, a_{j(\tau)+1}(\tau)\} ,$$

which corresponds to the set of points used to define the sets in (25).

$\vdots$

Step $s+1$: If $\mathcal{A}^*_s(\tau) = \mathcal{A}^*_{s-1}(\tau)$, then stop and take

$$\mathcal{U} = \mathcal{U}(\tau_{\text{sq}}) \cup \mathcal{U}(\tau_{\text{c}}) ,$$

where, for each $\tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$, we have that

$$\mathcal{U}(\tau) = \left\{\mathcal{P}^*_0(\tau), \ldots, \mathcal{P}^*_{L^*(\tau)-1}(\tau), \{p(0)\}, \{p(\tau)\}\right\} ,$$
$$\mathcal{P}^*_l(\tau) = \left\{p \in \mathcal{P} : p_j = \min\{p_j(0), p_j(\tau) + a\},\ a \in \left(a^*_l(\tau), a^*_{l+1}(\tau)\right) \text{ for each } j \in \mathcal{J}_v\right\}$$

$0 \leq l \leq L^*(\tau)-1$, and $a^*_0(\tau) < \ldots < a^*_{L(\tau)}(\tau)$ denotes the ordered values of $\mathcal{A}^*_s(\tau)$. Otherwise, i.e. if $\mathcal{A}^*_s(\tau) \neq \mathcal{A}^*_{s-1}(\tau)$, then for each $\tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$ with $\tau' \neq \tau \in \{\tau_{\text{sq}}, \tau_{\text{c}}\}$, take

$$\mathcal{A}^*_{s+1}(\tau) = \mathcal{A}^*_s(\tau) \bigcup_{j=1}^{J} \left\{\min\{p_j(0), a + p_j(\tau')\} - p_j(\tau) : a \in \mathcal{A}^*_s(\tau') \setminus \mathcal{A}^*_{s-1}(\tau')\right\}$$

which correspond to the set of all points that simultaneously define the end points of the intervals for all $j \in \mathcal{J}_v$ not included in the previous step.

$\vdots$

S.2.3 Example of a Set $\mathbf{B}^{\mathrm{r}}$

In this section, we describe an example of a set $\mathbf{B}^{\mathrm{r}}$ such that $\phi(\mathbf{B}) \subseteq \mathbf{B}^{\mathrm{r}}$ that we use in our empirical analysis when implementing the linear programs in (33).

To this end, it is first useful to consider an equivalent representation of the restrictions in (S.3) written in terms of pairs $w, w' \in \mathcal{W}$. Specifically, consider the following restrictions

$$\sum_{j \in \mathcal{J}^{>}_{w',w} \cup \mathcal{J}^{\dagger}} \beta_j(w) \geq \sum_{j \in \mathcal{J}^{>}_{w',w} \cup \mathcal{J}^{\dagger}} \beta_j(w') \tag{S.19}$$

for each $\mathcal{J}^{\dagger} \subseteq \mathcal{J}^{=}_{w,w'}$ and $w, w' \in \mathcal{W}$, where

$$\begin{aligned} \mathcal{J}^{>}_{w,w'} &= \left\{ j \in \mathcal{J}_v : t > t' \text{ for all } t \in w(j) \ , \ t' \in w'(j) \right\} \ , \\ \mathcal{J}^{=}_{w,w'} &= \mathcal{J} \setminus \left(\mathcal{J}^{>}_{w,w'} \cup \mathcal{J}^{>}_{w',w} \right) \ . \end{aligned}$$

and, as before, $w(j) = \{t \in \mathbf{R} : p_j = t \text{ for some } p \in w\}$ for each $w \in \mathcal{W}$ and $j \in \mathcal{J}_v$. To see why these restrictions imply those in (S.3), consider $w, w' \in \mathcal{W}$ such that $t > t'$ for all $t \in w(j)$, $t' \in w'(j)$ for $j \in \mathcal{J}' \subseteq \mathcal{J}_v$ and $w(j) = w'(j)$ for $j \in \mathcal{J}_v \setminus \mathcal{J}'$. In this case, note that $\mathcal{J}^{>}_{w,w'} = \mathcal{J}'$, $\mathcal{J}^{>}_{w',w} = \emptyset$, and $\mathcal{J}^{=}_{w,w'} = \mathcal{J} \setminus \mathcal{J}'$. Then, taking $\mathcal{J}^{\dagger} = \{j\}$ for each $j \in \mathcal{J}^{=}_{w,w'}$ in (S.19) implies that (S.3) holds for each $j \in \mathcal{J}'$. To see why the restrictions in (S.3) imply those in (S.19), consider $w, w' \in \mathcal{W}$ as well as a $w'' \in \mathcal{W}$ such that we have $w''(j) = w'(j) = w(j)$ for $j \in \mathcal{J}_v \setminus \left(\mathcal{J}^{>}_{w,w'} \cup \mathcal{J}^{>}_{w',w} \right)$, $w''(j) = w(j)$ for $j \in \mathcal{J}^{>}_{w,w'}$, and $w''(j) = w'(j)$ for $j \in \mathcal{J}^{>}_{w',w}$. Since this implies that $t'' > t$ for all $t \in w(j)$, $t'' \in w''(j)$ for $j \in \mathcal{J}^{>}_{w',w}$ and $w(j) = w''(j)$ for $j \in \mathcal{J}_v \setminus \mathcal{J}^{>}_{w',w}$, it follows from (S.3) that

$$\beta_j(w'') \geq \beta_j(w) \tag{S.20}$$

for each $j \in \mathcal{J}^{=}_{w,w'} \cup \mathcal{J}^{>}_{w,w'}$. Similarly, since it also implies that $t'' > t'$ for all $t \in w'(j)$, $t'' \in w''(j)$ for $j \in \mathcal{J}^{>}_{w,w'}$ and $w'(j) = w''(j)$ for $j \in \mathcal{J}_v \setminus \mathcal{J}^{>}_{w,w'}$, it also follows from (S.3) that

$$\beta_j(w'') \geq \beta_j(w') \tag{S.21}$$

for each $j \in \mathcal{J}^{=}_{w,w'} \cup \mathcal{J}^{>}_{w',w}$. Then, for each $\mathcal{J}^{\dagger} \subseteq \mathcal{J}^{=}_{w,w'}$, this implies that (S.19) holds as

$$\begin{aligned} \sum_{j \in \mathcal{J}^{>}_{w',w} \cup \mathcal{J}^{\dagger}} \beta_j(w') \leq \sum_{j \in \mathcal{J}^{>}_{w',w} \cup \mathcal{J}^{\dagger}} \beta_j(w'') &= 1 - \sum_{j \in \mathcal{J} \setminus \left(\mathcal{J}^{>}_{w',w} \cup \mathcal{J}^{\dagger} \right)} \beta_j(w'') \\ &\leq 1 - \sum_{j \in \mathcal{J} \setminus \left(\mathcal{J}^{>}_{w',w} \cup \mathcal{J}^{\dagger} \right)} \beta_j(w) = \sum_{j \in \mathcal{J}^{>}_{w',w} \cup \mathcal{J}^{\dagger}} \beta_j(w) \end{aligned}$$

where the first inequality follows from (S.21), the second equality follows from (S.2), the third inequality from (S.20), and the final equality from (S.2).

Given the equivalence between the restrictions in (S.3) and (S.19), we can alternatively write $\mathbf{B}$ in (S.6) as

$$\mathbf{B} = \left\{ \beta \in \mathbf{R}^{d_\beta} : \beta \text{ satisfies (S.1)} - \text{(S.2)}, \text{ (S.19)}, \text{ and (S.4)} - \text{(S.5)} \right\} .$$

In this set, observe that each restriction on β is for a given $w \in \mathcal{W}$ or for a pair of $w, w' \in \mathcal{W}$. Our choice of $\mathbf{B}^{\mathrm{r}}$ corresponds to the subset of these restrictions on β for $w \in \mathcal{W}^{\mathrm{r}}$ or for pairs of $w, w' \in \mathcal{W}^{\mathrm{r}}$, i.e. the subset of restrictions that directly correspond those that are in terms of β^{r}. More specifically, these restrictions correspond to the following

$$0 \leq \beta^{\mathrm{r}}_j(w) \leq 1 \quad \text{for each } j \in \mathcal{J} \text{ and } w \in \mathcal{W}^r , \tag{S.22}$$

$$\sum_{j \in \mathcal{J}} \beta^{\mathrm{r}}_j(w) = 1 \quad \text{for each } w \in \mathcal{W}^r , \tag{S.23}$$

$$\sum_{j \in \mathcal{J}^{>}_{w',w} \cup \mathcal{J}^\dagger} \beta^{\mathrm{r}}_j(w) \geq \sum_{j \in \mathcal{J}^{>}_{w',w} \cup \mathcal{J}^\dagger} \beta^{\mathrm{r}}_j(w') \quad \text{for each } \mathcal{J}^\dagger \subseteq \mathcal{J}^{=}_{w,w'} \text{ and } w, w' \in \mathcal{W}^{\mathrm{r}} , \tag{S.24}$$

$$\beta^{\mathrm{r}}_j(\{p(0)\}) = P_{j|0} \quad \text{for each } j \in \mathcal{J} , \tag{S.25}$$

$$\beta^{\mathrm{r}}_j(\{p(\tau_{\mathrm{sq}})\}) = P_{j|1} \quad \text{for each } j \in \mathcal{J} . \tag{S.26}$$

Then, denoting by $d_{\beta^{\mathrm{r}}}$ the dimension of β^{r}, the set we consider is given by

$$\mathbf{B}^{\mathrm{r}} = \left\{ \beta^{\mathrm{r}} \in \mathbf{R}^{d_{\beta^{\mathrm{r}}}} : \beta^{\mathrm{r}} \text{ satisfies (S.22)} - \text{(S.26)} \right\} .$$

S.2.4 Implementation Details for Auxiliary Parametric Assumptions

In this section, we characterize θ_A and $\mathbf{A}$ for the optimization problems in (36), for each of our parameters under each specification in Assumption O, Assumption AS and Assumption NS. As highlighted in Section 4.3, we compute bounds for our parameters under these specifications using the linear programs in (37), which are based on an alternative constraint set $\mathbf{A}^{\mathrm{r}}$ under each specification. As we will observe more precisely below, this is because the corresponding sets $\mathbf{A}$ are based on restrictions evaluated on all possible prices in $\mathcal{P}$, which makes the problems in (36) difficult to compute. The set $\mathbf{A}^{\mathrm{r}}$ we consider is based on taking these same restrictions but only evaluated on a given, finite set of prices in $\mathcal{P}$. The set of prices $\mathcal{P}^{\mathrm{r}}$ we consider is given by $\prod_{j=1}^{J} \mathcal{P}^{\mathrm{r}}_j$, where $\mathcal{P}^{\mathrm{r}}_j = \{0, p_j(0)/L, p_j(0) \cdot 2/L, \ldots, p_j(0) \cdot (L-1)/L, p_j(0)\}$ for some pre-specified value of L, i.e. a set of $(L+1)$ equidistant points in $[0, p_j(0)]$. In our empirical results, we take $L = 4$. In unreported results, we find that increasing the value of L generally tightens the bounds in a gradual manner, but at the cost of increased computational time, specifically for the confidence intervals.

Below, for completeness, we also present the set of restrictions that determine $\mathbf{A}^{\mathrm{r}}$ for each of the specifications. For these purposes, let $\{p_{0,j}, \ldots, p_{L,j}\}$ denote the set of ordered values of $\mathcal{P}_j^{\mathrm{r}}$ for each $j \in \mathcal{J}_v$ and let $\Delta p_{l,j}^k = p_{l,j}^k - p_{l-1,j}^k$ denote the difference in two consecutive prices in this set, where each of these two prices is raised to the power of k.

S.2.4.1 Under Assumption AS

We start with Assumption AS as Assumption O can be shown to be a special case of Assumption AS. Similar to Section S.1.2, we only derive the expression for θ_A for the parameters $AB(\tau)$ and $AC(\tau)$ for a generic value of τ as the expressions for the remainder of the parameters can then be straightforwardly derived from them. In the case of $AB(\tau)$, observe that we have

$$\begin{aligned}\theta_A(\alpha) &\equiv \sum_{l=0}^{j(\tau)} \int_{a_l(\tau)}^{a_{l+1}(\tau)} \sum_{j=l+1}^{J} \left(\sum_{m=1}^{l} \sum_{k=0}^{K} \alpha_{jmk} \cdot (p_m(0))^k + \sum_{m=l+1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot (p_m(\tau) + a)^k \right) da \\ &= \sum_{l=0}^{j(\tau)} \sum_{j=l+1}^{J} \left(\sum_{m=1}^{l} \sum_{k=0}^{K} \alpha_{jmk} \cdot (p_m(0))^k \cdot (a_{l+1}(\tau) - a_l(\tau)) + \right. \\ &\qquad \left. \sum_{m=l+1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot \left(\frac{(p_m(\tau) + a)^{k+1}}{k+1} \bigg|_{a_l(\tau)}^{a_{l+1}(\tau)} \right) \right) , \end{aligned} \tag{S.27}$$

where the first line follows from directly substituting the relation between q and α from Assumption AS in (10), and the second line from evaluating the integrals. Similarly, by substituting the relation between q and α in (11), observe that in the case of $AC(\tau)$ we have

$$\theta_A(\alpha) \equiv \sum_{j \in \mathcal{J}} \sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot \left(c_j(\tau) \cdot (p_m(\tau))^k - c_j(0) \cdot (p_m(0))^k \right) . \tag{S.28}$$

Next, to see the linear restrictions that determine $\mathbf{A}$, observe that by substituting the relation between q and α into the various restrictions in (16)-(17), (4) and (18)-(21) we obtain

$$\sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot p_m^k \geq 0 \ \text{ for each } j \in \mathcal{J} , \tag{S.29}$$

$$\sum_{j \in \mathcal{J}} \sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot p_m^k = 1 \tag{S.30}$$

for all $p \in \mathcal{P}$,

$$\sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot \left((p_m)^k - (p'_m)^k \right) \geq 0 \tag{S.31}$$

for each $j \in \mathcal{J} \setminus \mathcal{J}'$ and $p, p' \in \mathcal{P}$ such that $p_j > p'_j$ for $j \in \mathcal{J}' \subseteq \mathcal{J}_v$ and $p_j = p'_j$ for $j \in \mathcal{J}_v \setminus \mathcal{J}'$, and

$$\sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot (p_m(0))^k = P_{j|0} \ , \tag{S.32}$$

$$\sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot (p_m(\tau_{\text{sq}}))^k = P_{j|1} \tag{S.33}$$

for each $j \in \mathcal{J}$. In turn, we have that $\mathbf{A} = \{\alpha \in \mathbf{R}^{d_\alpha} : \alpha \text{ satisfies (S.29)} - \text{(S.33)}\}$, i.e. the set of α satisfying the above restrictions.

As mentioned above, considering the above restrictions at all prices in $\mathcal{P}$ can be generally difficult. We therefore only consider a subset of these restrictions that are evaluated at prices in $\mathcal{P}^{\text{r}}$. Simplifying and removing some jointly redundant restrictions using algebra, these restrictions can be given by

$$\sum_{k=0}^{K} \alpha_{jjk} \cdot p_{l,j}^k + \sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot p_{1,m}^k \geq 0 \ \text{ for each } j \in \mathcal{J}_v, \ 0 \leq l \leq L \ , \tag{S.34}$$

$$\sum_{m=i}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot p_{1,m}^k \geq 0 \ \text{ for } j \in \{g, n\} \ , \tag{S.35}$$

$$\sum_{j \in \mathcal{J}} \sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot p_{1,m}^k = 1 \ , \tag{S.36}$$

$$\sum_{j \in \mathcal{J}} \sum_{k=0}^{K} \alpha_{jmk} \cdot \Delta p_{l+1,m}^k = 0 \ \text{ for } m \in \mathcal{J}_v, \ 0 \leq l \leq L-1 \ , \tag{S.37}$$

$$\sum_{k=0}^{K} \alpha_{jmk} \cdot \Delta p_{l+1,m}^k \geq 0 \ \text{ for each } j \in \mathcal{J}, \ m \neq j \in \mathcal{J}_v, \ 0 \leq l \leq L-1 \ , \tag{S.38}$$

and (S.32)-(S.33), and, in turn, we have that $\mathbf{A}^{\text{r}} = \{\alpha \in \mathbf{R}^{d_\alpha} : \alpha \text{ satisfies (S.34)} - \text{(S.38) and (S.32)} - \text{(S.33)}\}$.

S.2.4.2 Under Assumption O

We can show that Assumption O corresponds to a special case of Assumption AS. In particular, summing the equations in the statement of Assumption O across $j \in \mathcal{J}_v \cup \{n\}$ and then using the restriction in (17), we can show that

$$q_g(p) = \frac{1}{J+2} \left(1 - \alpha_n - \sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{mk} \cdot p_m^k \right) \ .$$

Along with relation between q_g and the other demand functions under Assumption O, this then implies that Assumption O corresponds to a special case of Assumption AS.

As a result, the implementation of the bounds under Assumption O is the same as that under Assumption AS but with a different constraint set that includes the additional restrictions imposed by Assumption O. To this end, note that, in the presence of Assumption AS, imposing Assumption O is equivalent to requiring that $q_j(p) - q_g(p) = q_j(p') - q_g(p')$ for all $j \in \mathcal{J}_v$ and all $p, p' \in \mathcal{P}$ such $p_j = p'_j$, and $q_n(p) - q_g(p) = q_n(p') - q_g(p')$ for all $p, p' \in \mathcal{P}$, i.e. the difference in demand between a given private voucher school or any non-participating private school and any government-funded school is invariant to changes in the prices of the other voucher schools. Substituting the relation between q and α from Assumption AS in these restrictions allows us to rewrite them in terms of α, and $\mathbf{A}$ in this case can be given by the same one as that under Assumption AS above but with the addition of these restrictions.

Considering these restriction only on the prices in $\mathcal{P}^{\mathrm{r}}$ and simplifying and removing some jointly redundant restrictions, the additional restrictions correspond to

$$\sum_{k=0}^{K} (\alpha_{jmk} - \alpha_{gmk}) \cdot \Delta p^k_{l+1,m} = 0 \tag{S.39}$$

for each $j \in \mathcal{J}$, $m \neq j \in \mathcal{J}_v$ and $0 \leq l \leq L-1$. In turn, we have that $\mathbf{A}^{\mathrm{r}} = \{\alpha \in \mathbf{R}^{d_\alpha} : \alpha$ satisfies (S.34) – (S.38), (S.32) – (S.33) and (S.39)$\}$, i.e. the set of restrictions under Assumption AS along with the additional restriction in (S.39).

S.2.4.3 Under Assumption NS

As in Assumption AS, we derive the expressions for θ_A for the parameters $AB(\tau)$ and $AC(\tau)$ for some generic value of τ. In the case of $AB(\tau)$, observe that we have

$$\begin{aligned}
\theta_A(\alpha) \equiv & \sum_{l=0}^{j(\tau)} \int_{a_l(\tau)}^{a_{l+1}(\tau)} \sum_{j=l+1}^{J} \Bigg(\sum_{m=1}^{l} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot (p_j(\tau)+a)^{k_j} \cdot (p_m(0))^{k_m} da \, + \\
& \qquad \sum_{m=l+1, m\neq j}^{J} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot (p_j(\tau)+a)^{k_j} \cdot (p_m(\tau)+a)^{k_m} da \Bigg) \\
= & \sum_{l=0}^{j(\tau)} \sum_{j=l+1}^{J} \Bigg(\sum_{m=1, m\neq j}^{l} \sum_{k_j,k_m=0}^{K} \alpha_{jik_jk_m} \cdot (p_m(0))^{k_m} \left(\frac{(p_j(\tau)+a)^{k_j+1}}{k_j+1} \Bigg|_{a_l(\tau)}^{a_{l+1}(\tau)} \right) \\
& \qquad \sum_{m=l+1, m\neq j}^{J} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot \int_{a_l(\tau)}^{a_{l+1}(\tau)} (p_j(\tau)+a)^{k_j} \cdot (p_m(\tau)+a)^{k_m} da \,,
\end{aligned} \tag{S.40}$$

where the first line follows from directly substituting the relation between q and α from Assumption NS in (10), and second line follows from evaluating one of the integrals. Moreover, using the binomial theorem, the remaining integral can also be analytically given by

$$\int_{a_l(\tau)}^{a_{l+1}(\tau)} (p_j(\tau)+a)^{k_j} \cdot (p_m(\tau)+a)^{k_m} da = \sum_{l_j=0}^{k_j} \sum_{l_m=0}^{k_m} \left(\binom{k_j}{l_j} \binom{k_m}{l_m} (p_j(\tau))^{k_j-l_j} \cdot (p_m(\tau))^{k_m-l_m} \cdot \left(\frac{a^{l_j+l_m+1}}{l_j+l_m+1} \Bigg|_{a_l(\tau)}^{a_{l+1}(\tau)} \right) \right) .$$

Similarly, substituting the relation between q and α in (11), observe that in the case of $AC(\tau)$ we have

$$\theta_A(\alpha) \equiv \sum_{j=1}^{J} \sum_{m=1,m\neq j}^{J} \sum_{k_j,k_m=0}^{K} \alpha_{jik} \cdot \Big(c_j(\tau) \cdot (p_j(\tau))^{k_j} \cdot (p_m(\tau))^{k_m} - c_j(0) \cdot (p_j(0))^{k_j} \cdot (p_m(0))^{k_m} \Big) + \sum_{j\in\{g,n\}} \sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jik} \cdot \Big(c_j(\tau) \cdot (p_m(\tau))^{k_j} - c_j(0) \cdot (p_m(0))^{k_m} \Big) . \tag{S.41}$$

Similar to (S.29)-(S.33) under Assumption AS, we can substitute the relation between q and α from Assumption NS in (16)-(17), (4) and (18)-(21) to derive analogous restrictions that determine $\mathbf{A}$ in this case. Below we state $\mathbf{A}^{\mathrm{r}}$ that evaluates these restrictions at prices in $\mathcal{P}^{\mathrm{r}}$. Simplifying and removing some jointly redundant restrictions, these restrictions can be stated as

$$\sum_{m=1,m\neq j}^{J} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot p_{l,j}^{k_j} \cdot p_{1,m}^{k_m} \geq 0 \quad \text{for } j \in \mathcal{J}_v,\ 0 \leq l \leq L , \tag{S.42}$$

$$\sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot p_{1,m}^{k} \geq 0 \quad \text{for } j \in \{g,n\} , \tag{S.43}$$

$$\sum_{j=1}^{J} \sum_{m=1,m\neq j}^{J} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot p_{1,j}^{k_j} \cdot p_{1,m}^{k_m} + \sum_{j\in\{g,n\}} \sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot p_{1,m}^{k} = 1 , \tag{S.44}$$

$$\sum_{m=1,m\neq j}^{J} \sum_{k_j,k_m=0}^{K} (\alpha_{jmk_jk_m} + \alpha_{mjk_mk_j}) \cdot \Delta p_{l+1,j}^{k_j} \cdot p_{1,m}^{k_m} + \sum_{m\in\{g,n\}} \sum_{k=0}^{K} \alpha_{mjk} \cdot \Delta p_{l+1,j}^{k} = 0 \quad \text{for } j \in \mathcal{J}_v,\ 0 \leq l \leq L-1 , \tag{S.45}$$

$$\sum_{k_j,k_m=0}^{K} (\alpha_{jmk_jk_m} + \alpha_{mjk_mk_j}) \cdot \Delta p_{l_j+1,j}^{k_j} \cdot \Delta p_{l_m+1,m}^{k_m} = 0 \quad \text{for } j \in \mathcal{J}_v,\ m \neq j \in \mathcal{J}_v,\ 0 \leq l_j, l_m \leq L-1, \tag{S.46}$$

$$\sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot p_{l_j,j}^{k_j} \cdot \Delta p_{l_m+1,m}^{k_m} \geq 0 \quad \text{for } j \in \mathcal{J}_v,\ m \neq j \in \mathcal{J}_v , \tag{S.47}$$

$$0 \leq l_j \leq L,\ 0 \leq l_m \leq L-1\ ,$$

$$\sum_{k=0}^{K} \alpha_{jmk} \cdot \Delta p_{l+1,m}^{k} \geq 0 \quad \text{for } j \in \{g,n\},\ m \in \mathcal{J}_v\ , \tag{S.48}$$

$$0 \leq l \leq L-1\ ,$$

$$\sum_{m=1, m\neq j}^{J} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot (p_j(0))^{k_j} \cdot (p_m(0))^{k_m} = P_{j|0} \quad \text{for } j \in \mathcal{J}_v\ , \tag{S.49}$$

$$\sum_{m=1, m\neq j}^{J} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot (p_j(\tau_{\text{sq}}))^{k_j} \cdot (p_m(\tau_{\text{sq}}))^{k_m} = P_{j|1} \quad \text{for } j \in \mathcal{J}_v\ , \tag{S.50}$$

$$\sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot (p_j(0))^{k} = P_{j|0} \quad \text{for } j \in \{g,n\}\ , \tag{S.51}$$

$$\sum_{m=1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot (p_j(\tau_{\text{sq}}))^{k} = P_{j|1} \quad \text{for } j \in \{g,n\}\ , \tag{S.52}$$

and, in turn, we have that $\mathbf{A}^{\text{r}} = \{\alpha \in \mathbf{R}^{d_\alpha} : \alpha \text{ satisfies (S.42)} - \text{(S.52)}\}$.

S.2.5 Extension for the Role of Low-tuition Schools in the Program

In the empirical analysis in Section 5.5, we noted that our identification analysis can be straightforwardly extended to evaluate what we can learn for the parameters defined in (38)-(40) for a given value of $\kappa \in \mathbf{R}_+$. In this section, we describe this extension.

To begin, observe, similar to the parameters in Section 3, each of these parameters can be written as functions of q. Indeed, this is true by definition for $AC^\kappa(\tau_{\text{sq}})$. For $AB^\kappa(\tau_{\text{sq}})$, one can modify the arguments in Proposition 3.1 in a straightforward manner to obtain that

$$\begin{aligned} AB^\kappa(\tau_{\text{sq}}) = & \int_0^{a_{j^\kappa+1}(\tau_{\text{sq}})} \left(\sum_{j=j^\kappa+1}^{J} q_j(p_1(0), \ldots, p_{j^\kappa}(0), p_{j^\kappa+1}(\tau_{\text{sq}}) + a, \ldots, p_J(\tau_{\text{sq}}) + a) \right) da \\ & + \sum_{l=j^\kappa+1}^{j(\tau_{\text{sq}})} \int_{a_l(\tau_{\text{sq}})}^{a_{l+1}(\tau_{\text{sq}})} \left(\sum_{j=l+1}^{J} q_j(p_1(0), \ldots, p_l(0), p_{l+1}(\tau_{\text{sq}}) + a, \ldots, p_J(\tau_{\text{sq}}) + a) \right) da\ . \end{aligned} \tag{S.53}$$

Then, given that both $AB^\kappa(\tau_{\text{sq}})$ and $AC^\kappa(\tau_{\text{sq}})$ can be written in terms of q, it also directly follows that $AS^\kappa(\tau_{\text{sq}})$ can as well. Below, we show how we can continue to apply Proposition 4.1 and Proposition 4.2 to characterize what we can learn for each of these parameters under the baseline nonparametric specification and under the auxiliary parametric specifications, respectively.

S.2.5.1 Baseline Nonparametric Specification

We start with the baseline nonparametric specification. Recall from the proof of Proposition 4.1 that it is useful to first derive the expression for θ_B for each of the parameters. In the case of $AB^\kappa(\tau_{\text{sq}})$, similar to $AB(\tau_{\text{sq}})$ in (S.7)-(S.9), observe that θ and θ_B can be shown to be given by

$$
\begin{aligned}
\theta(q) &\equiv \sum_{l=0}^{j^\kappa+1} \sum_{j=j^\kappa+1}^{J} \sum_{u \in \mathcal{U}(l,\tau_{\text{sq}})} \int_{\underline{a}_u}^{\bar{a}_u} q_j(p_1(0), \ldots, p_{j^\kappa}(0), p_{j^\kappa+1}(\tau_{\text{sq}}) + a, \ldots, p_J(\tau_{\text{sq}}) + a) \, da \\
&+ \sum_{l=j^\kappa+1}^{j(\tau_{\text{sq}})} \sum_{j=l+1}^{J} \sum_{u \in \mathcal{U}(l,\tau_{\text{sq}})} \int_{\underline{a}_u}^{\bar{a}_u} q_j(\min\{p(0), p(\tau_{\text{sq}}) + a\}) \, da \\
&= \sum_{l=0}^{j^\kappa+1} \sum_{j=j^\kappa+1}^{J} \sum_{u \in \mathcal{U}(l,\tau_{\text{sq}})} (\bar{a}_u - \underline{a}_u) \cdot \beta_j(h^\kappa(u)) + \sum_{l=j^\kappa+1}^{j(\tau_{\text{sq}})} \sum_{j=l+1}^{J} \sum_{u \in \mathcal{U}(l,\tau_{\text{sq}})} (\bar{a}_u - \underline{a}_u) \cdot \beta_j(h(u)) \\
&\equiv \theta_B(\beta) \ ,
\end{aligned}
$$

where, similar to before, $\min\{p(0), p(\tau_{\text{sq}}) + a\} \equiv (\min\{p_1(0), p_1(\tau_{\text{sq}}) + a\}, \ldots, \min\{p_J(0), p_J(\tau_{\text{sq}}) + a\})$, $\mathcal{U}(l, \tau_{\text{sq}}) \equiv \{u \in \mathcal{U} : u \subseteq \mathcal{P}_l(\tau_{\text{sq}})\}$ and $h(u) \in \mathcal{W}$ for a given $u \in \mathcal{U}$ is such that $h(u) = \prod_{j=1}^{J} u(j)$, and, in addition, $h^\kappa(u) \in \mathcal{W}$ for a given $u \in \mathcal{U}$ is such that $h^\kappa(u)(j) = \{p_j(0)\}$ for $j \le j^\kappa(\tau_{\text{sq}})$ and $h^\kappa(u)(j) = u(j)$ for $j > j^\kappa(\tau_{\text{sq}})$. In the case of $AC^\kappa(\tau_{\text{sq}})$, similar to $AC(\tau)$ in (S.10)-(S.11), observe that θ and θ_B can be shown to be given by

$$
\begin{aligned}
\theta(q) &\equiv \sum_{j \in \mathcal{J}} c_j^\kappa(\tau_{\text{sq}}) \cdot q_j(p^\kappa(\tau_{\text{sq}})) - \sum_{j \in \mathcal{J}} c_j(0) \cdot q_j(p(0)) \\
&= \sum_{j \in \mathcal{J}} c_j^\kappa(\tau_{\text{sq}}) \cdot \beta_j(\{p^\kappa(\tau_{\text{sq}})\}) - \sum_{j \in \mathcal{J}} c_j(0) \cdot \beta_j(\{p(0)\}) \ \equiv \theta_B(\beta) \ .
\end{aligned}
$$

The expression from $AS^\kappa(\tau_{\text{sq}})$ can similarly be derived from the difference of the above two expressions. Then, recall again from the proof of Proposition 4.1 that we can show that the proposition applies to these parameters if in Part 1 of the proof we can show that every for every $q \in \mathbf{Q}_B$ with $\theta(q) = \theta_0$, the β constructed from this q through (S.12) satisfies $\theta_B(\beta) = \theta_0$. This can be shown for these parameters using arguments analogous to those used to show this was the case for $AB(\tau_{\text{sq}})$ and $AC(\tau_{\text{sq}})$ in Part 1 of the proof, from which it then follows that Proposition 4.1 applies to these parameters.

Recall in our empirical analysis for the parameters in Section 3 we don't directly apply optimization problems from Proposition 4.1, but the more computationally tractable alternatives from (33). However, for the above parameters, the constructed $\mathcal{W}^{\text{r}}$ is slightly different as they are defined over a slightly different sets of prices. For these parameters, observe the following larger set of prices given by

$$
\mathcal{W}^{\kappa,\text{r}} = \mathcal{W}^{\text{r}} \cup \{w \in \mathcal{W} : w(j) = \{p_j(0)\} \text{ for } j \le j^\kappa,
$$

$$w(j) \in (p_j(\tau_{\text{sq}}), p_j(\tau_{\text{sq}}) + a_{j^\kappa+1}(\tau_{\text{sq}})) \text{ or } \{p_j(\tau_{\text{sq}})\} \text{ for } j > j^\kappa\}$$

corresponds to a set of prices that are sufficient in the definition of these additional parameters. Then, replacing $\mathcal{W}^{\text{r}}$ by $\mathcal{W}^{\kappa,\text{r}}$ in the arguments leading to (33), we can similarly only consider a subvector of β defined over this set sufficient to define the parameters, and similarly obtain linear programs as in (33) giving outer sets containing the identified set for these parameters with constraints corresponding to those on this subvector determined from every pair of $w, w' \in \mathcal{W}^{\kappa,\text{r}}$ as in Section S.2.3.

S.2.5.2 Auxiliary Parametric Specifications

We next proceed to the auxiliary parametric specifications. In this case, recall from the proof of Proposition 4.2 that the proof, and hence the proposition, directly applies since each of these parameters, analogous to the parameters in Section 3, can be rewritten as continuous functions of α. However, as observed in Section S.2.4, recall that the expressions for these functions play an important role in the implementation of the problems in (36) and the more practical versions in (37). Similar to the derivation of the expressions for $AB(\tau_{\text{sq}})$ and $AC(\tau_{\text{sq}})$ in Section S.2.4, one can derive these expressions for $AB^\kappa(\tau_{\text{sq}})$ and $AC^\kappa(\tau_{\text{sq}})$, and in turn $AS^\kappa(\tau_{\text{sq}})$, for the specifications in Assumption O, Assumption AS and Assumption NS. For completeness, we present these expressions below.

We start with Assumption AS, as Assumption O corresponds to a special case of it. In the case of $AB^\kappa(\tau_{\text{sq}})$, similar to (S.27), observe that by directly substituting the relation between q and α from Assumption AS in (S.53) we have

$$\begin{aligned}
\theta_A(\alpha) \equiv & \sum_{j=j^\kappa+1}^{J} \sum_{k=0}^{K} \left(\sum_{m=1}^{j^\kappa} \alpha_{jmk} \cdot (p_m(0))^k \cdot a_{j^\kappa}(\tau_{\text{sq}}) + \sum_{m=j^\kappa+1}^{J} \alpha_{jmk} \cdot \frac{(p_m(\tau_{\text{sq}}) + a_{j^\kappa+1}(\tau))^{k+1}}{k+1} \, da \right) \\
& + \sum_{l=j^\kappa+1}^{j(\tau_{\text{sq}})} \sum_{j=l+1}^{J} \left(\sum_{m=1}^{l} \sum_{k=0}^{K} \alpha_{jmk} \cdot (p_m(0))^k \cdot (a_{l+1}(\tau_{\text{sq}}) - a_l(\tau_{\text{sq}})) + \right. \\
& \qquad \left. \sum_{m=l+1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot \left(\frac{(p_m(\tau_{\text{sq}}) + a)^{k+1}}{k+1} \bigg|_{a_l(\tau_{\text{sq}})}^{a_{l+1}(\tau_{\text{sq}})} \right) \right)
\end{aligned}$$

where the integrals can be explicitly characterized as in (S.27). In the case of $AC^\kappa(\tau_{\text{sq}})$, similar to (S.28), observe that by directly substituting the relation between q and α from Assumption AS in (39) we have

$$\begin{aligned}
\theta_A(\alpha) \equiv & \sum_{j\in\mathcal{J}} \sum_{m=1}^{j^\kappa} \sum_{k=0}^{K} \alpha_{jmk} \cdot \left(c_j^\kappa(\tau_{\text{sq}}) - c_j(0)\right) \cdot (p_m(0))^k \\
& + \sum_{j\in\mathcal{J}} \sum_{m=j^\kappa+1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot \left(c_j^\kappa(\tau_{\text{sq}}) \cdot (p_m(\tau_{\text{sq}}))^k - c_j(0) \cdot (p_m(0))^k\right) \; .
\end{aligned}$$

Next, we present the expressions for Assumption NS. In the case of $AB^{\kappa}(\tau_{\text{sq}})$, similar to (S.40), observe that by directly substituting the relation between q and α from Assumption NS in (S.53) we have

$$
\begin{aligned}
\theta_A(\alpha) \equiv \sum_{j=j^{\kappa}+1}^{J} \Bigg[& \sum_{m=1}^{j^{\kappa}} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot (p_m(0))^{k_m} \cdot \left(\int_{0}^{a_{j^{\kappa}+1}(\tau_{\text{sq}})} (p_j(\tau_{\text{sq}})+a)^{k_j} \, da \right) \\
& + \sum_{\substack{m=j^{\kappa}+1 \\ m\neq j}}^{J} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot \left(\int_{0}^{a_{j^{\kappa}+1}(\tau)} (p_j(\tau_{\text{sq}})+a)^{k_j} \cdot (p_m(\tau_{\text{sq}})+a)^{k_m} \, da \right) \Bigg] \\
+ \sum_{l=j^{\kappa}+1}^{j(\tau_{\text{sq}})} \sum_{j=l+1}^{J} \Bigg[& \sum_{m=1}^{l} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot (p_m(0))^{k_m} \cdot \left(\int_{a_l(t)}^{a_{l+1}(\tau_{\text{sq}})} (p_j(\tau_{\text{sq}})+a)^{k_j} \, da \right) \\
& + \sum_{\substack{m=l+1 \\ m\neq j}}^{J} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot \left(\int_{a_l(\tau_{\text{sq}})}^{a_{l+1}(\tau_{\text{sq}})} (p_j(\tau_{\text{sq}})+a)^{k_j} \cdot (p_m(\tau_{\text{sq}})+a)^{k_m} \, da \right) \Bigg] ,
\end{aligned}
$$

where the integrals can be explicitly characterized (S.40). In the case of $AC^{\kappa}(\tau_{\text{sq}})$, similar to (S.41), observe that by directly substituting the relation between q and α from Assumption AS in (39) we have

$$
\begin{aligned}
\theta_A(\alpha) \equiv & \sum_{j=1}^{j^{\kappa}} \sum_{\substack{m=1 \\ m\neq j}}^{j^{\kappa}} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot (p_j(0))^{k_j} \cdot (p_m(0))^{k_m} \cdot \left(c_j^{\kappa}(\tau_{\text{sq}}) - c_j(0) \right) \\
& + \sum_{j=1}^{j^{\kappa}} \sum_{\substack{m=j^{\kappa}+1 \\ m\neq j}}^{J} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot (p_j(0))^{k_j} \cdot \Big((p_m(\tau_{\text{sq}}))^{k_m} \cdot c_j^{\kappa}(\tau_{\text{sq}}) - (p_m(0))^{k_m} \cdot c_j(0) \Big) \\
& + \sum_{j=j^{\kappa}+1}^{J} \sum_{\substack{m=1 \\ m\neq j}}^{j^{\kappa}} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot (p_m(0))^{k_m} \cdot \Big((p_j(\tau_{\text{sq}}))^{k_j} \cdot c_j^{\kappa}(\tau_{\text{sq}}) - (p_j(0))^{k_j} \cdot c_j(0) \Big) \\
& + \sum_{j=j^{\kappa}+1}^{J} \sum_{\substack{m=j^{\kappa}+1 \\ m\neq j}}^{J} \sum_{k_j,k_m=0}^{K} \alpha_{jmk_jk_m} \cdot \Big((p_j(\tau_{\text{sq}}))^{k_j} \cdot (p_m(\tau_{\text{sq}}))^{k_m} \cdot c_j^{\kappa}(\tau_{\text{sq}}) \\
& \qquad\qquad - (p_j(0))^{k_j} \cdot (p_m(0))^{k_m} \cdot c_j(0) \Big) \\
& + \sum_{j\in\{g,n\}} \sum_{m=1}^{j^{\kappa}} \sum_{k=0}^{K} \alpha_{jmk} \cdot (p_m(0))^{k} \cdot \left(c_j^{\kappa}(\tau_{\text{sq}}) - c_j(0) \right) \\
& + \sum_{j\in\{g,n\}} \sum_{m=j^{\kappa}+1}^{J} \sum_{k=0}^{K} \alpha_{jmk} \cdot \Big((p_m(\tau_{\text{sq}}))^{k} \cdot c_j^{\kappa}(\tau_{\text{sq}}) - (p_m(0))^{k} \cdot c_j(0) \Big) .
\end{aligned}
$$

S.3 Statistical Inference

In this section, we describe the procedures used to construct confidence intervals for our parameters of interest and p-values for the specification tests presented in our empirical analysis in Section 5. To this end, let

$$\{(D_i, Z_i) : 1 \leq i \leq N\} \tag{S.54}$$

denote our sample of N observations, assumed to be independently and identically distributed, on which our statistical tests are based. In Section S.3.1 below, we first describe the procedure for constructing confidence intervals; and, in Section S.3.2, we then describe the related procedure for constructing p-values for testing the specification.

S.3.1 Confidence Interval

Recall that our parameters of interest are generally bounded across our various specifications, where the lower and upper bounds are given by minimization and maximization problems, respectively. We construct confidence intervals such that each point in these bounds lies in the interval with probability at least $(1-\alpha)\%$ for some pre-specified value of $\alpha \in (0,1)$. In order to describe the common procedure that we use across all the parameters and specifications, it is useful to first define the common structure present in all these cases. To this end, note that each point in the bounds across these cases can be written as $c'x$ for some vector $x \in \mathbf{R}^{d_x}$ of dimension d_x that satisfies

$$A_1 x = b_1 \ , \tag{S.55}$$

$$A_2 x \leq b_2 \ , \tag{S.56}$$

where x corresponds to the optimizing variable in the minimization and maximization problems, c corresponds to the vector defining the objective in these problems that depends on the choice of parameter, the restrictions in (S.55) capture the restriction imposed on the optimizing variable by the observed enrollment shares through (18)-(21), and the restrictions in (S.56) capture the restriction imposed by the shape restrictions through (16)-(17) and (4). Note that across these cases the values c, A_1, A_2 and b_2 are known and deterministic, and only b_1 needs to be estimated as it corresponds to the observed enrollment shares.

We construct confidence intervals for the various parameters across the various specifications by test inversion. In particular, we test the null hypothesis at level α that there exists a x satisfying the restrictions in (S.55) and (S.56) such that $c'x = \theta_0$ for some given value of $\theta_0 \in \mathbf{R}$. Confidence intervals are then constructed by collecting the set of values of θ_0 that are not rejected.

We test this null hypothesis using a recentered subsampling procedure from Kalouptsidi et al. (2020), who show it can have several desirable theoretical properties. In our above setup, the test procedure can be described in the following steps:

1. Compute the test statistic
$$TS_N(\theta_0) = N \cdot \min_x (\hat{b}_1 - A_1 x)'(\hat{b}_1 - A_1 x) \tag{S.57}$$
subject to x satisfying $c'x = \theta_0$ and the restrictions in (S.56), where $\hat{b}_1$ corresponds to the estimated counterpart of b_1 using the data in (S.54), namely where the enrollment shares are replaced by the empirical counterparts using the empirical distribution of the data.

2. For $l = 1, \ldots, B$, compute the recentered subsampling test statistics
$$TS_{l,N}(\theta_0) = N_s \cdot \min_x (\hat{b}_{l,1} - \hat{b}_1 + A_1\hat{\nu} - A_1 x)'(\hat{b}_{l,1} - \hat{b}_1 + A_1\hat{\nu} - A_1 x) \tag{S.58}$$
subject to x satisfying $c'x = \theta_0$ and the restrictions in (S.56), where $\hat{\nu}$ corresponds to the argmin of the minimization problem in (S.57), and $\hat{b}_{l,1}$ corresponds to the analogue of $\hat{b}_1$ using the lth subsample of size N_s drawn without replacement from the data in (S.54). Here note that N_s, the subsample size, is a tuning parameter that is required to satisfy $N_s < N$, $N_s \to \infty$ and $N_s/N \to 0$ as $N \to \infty$. For our empirical results, following Kalouptsidi et al. (2020), we take $N_s \approx 8N^{1/2}$. Moreover, we take $B = 200$.

3. Compute the critical value of the test $\hat{c}(1-\alpha, \theta_0)$ by taking the $(1-\alpha)$-quantile of the distribution of computed subsample test statistics
$$H_N(t, \theta_0) = \frac{1}{B}\sum_{l=1}^{B} 1\{TS_{l,N}(\theta_0) \leq t\} \ .$$

4. The test procedure is given by $\phi_N(\theta_0) = 1\{TS_N(\theta_0) > \hat{c}(1-\alpha, \theta_0)\}$, i.e we reject if the test statistic is greater than the critical value, and do not reject if it is equal to it or below.

Given the above test procedure for a given value of θ_0, we can then use it construct confidence intervals by collecting the set all points at which we don't reject, namely $\{\theta_0 \in \mathbf{R} : \phi_N(\theta_0) = 0\}$.

S.3.2 Specification Test

We next describe how we construct p-values for testing whether a chosen specification is correctly specified. To this end, note that testing whether this is the case is equivalent to testing whether there exists an admissible demand function that can generate the observed enrollment shares as well as satisfy the various imposed restrictions. Alternatively, given the common structure present across all our specifications noted above, this corresponds to testing the null hypothesis that there exists a x such that it satisfies the restrictions in (S.55) and (S.56), where the x and the restrictions, as before, depend on the choice of specification.

Indeed, this null hypothesis is the same as that on which our confidence intervals were based except for the absence of the additional restriction that $c'x = \theta_0$ for some value of θ_0. As a result, we can use the same stepwise test procedure described above except for removing this restriction in the computation of the test statistics. Specifically, in this case, the test statistic in (S.57) becomes

$$TS_N = N \cdot \min_x (\hat{b}_1 - A_1 x)'(\hat{b}_1 - A_1 x)$$

subject to x just satisfying the restrictions in (S.56); whereas the test statistic in each subsample in (S.58) becomes

$$TS_{l,N} = N_s \cdot \min_x (\hat{b}_{l,1} - \hat{b}_1 + A_1\hat{\nu} - A_1 x)'(\hat{b}_{l,1} - \hat{b}_1 + A_1\hat{\nu} - A_1 x)$$

subject to x just satisfying the restrictions in (S.56) and $\hat{\nu}$ corresponding to the argmin of the minimization problem in computing TS_N above. Using these test statistics, the test can as before be performed by $\phi_N = 1\{TS_N > \hat{c}(1-\alpha)\}$, where $\hat{c}(1-\alpha)$ denotes the $(1-\alpha)$-quantile of the distribution of the computed subsample test statistics. Given these test statistics, we can also compute a p-value for the null hypothesis by

$$p_{value} = \frac{1}{B}\sum_{l=1}^{B} 1\{TS_{l,N} \geq TS_N\} \ .$$

S.4 Additional Details on Empirical Analysis

S.4.1 Data Construction

In this section, we describe how we construct the data used in our empirical analysis in Section 5. The original data sample comes from the replication files for the evaluation of OSP, which are available from the US Department of Education (Wolf et al., 2010). Recall that our analysis focuses on the initial school choice for students who entered the experiment in 2005. Beginning with this subsample, we make the following data-cleaning choices to reach our final analysis data.

Our analysis requires only the prices (as measured by the tuition) of the participating private schools and the school choices of the students (to compute their enrollment shares). For all participating private schools, we observe tuition in either the first or second year of the study, but not necessarily both. If we observe tuition only in the first year, we assume that it was unchanged between the first and second year (recall we use only the second year of data). Under these tuition values, we find that the dimension of the program in (33) used to compute bounds under the baseline specification can be quite large for parameters measuring welfare effects under some counterfactual voucher amounts. This specifically makes computing the confidence intervals for these parameters computationally intensive. However, these computations become very tractable when we round tuition to the closest \$500. To maintain consistency across all results, we therefore round the tuition to the closest \$500 in our analysis sample. In unreported results, we also compute the results in Table 2 without any rounding, and find that the numerical results are very similar and the qualitative conclusions remain unchanged.

The second requirement is school choices. In our data, these are missing for 36% of the students. By a fortunate quirk of the research design, however, participating private schools reported all voucher students to the researchers. Unobserved school choices must therefore be either in non-participating private schools, or government-funded schools. For these students, we assume they enroll in these two groups at the same relative rate that students with observed choices enroll in non-participating private and government-funded schools. Once we obtain these school choices, we weight these observed choices using the baseline weights of the original evaluation—see Wolf et al. (2010, Appendix A.7) for details on how these weights were constructed.

S.4.2 Summary Statistics on School Setting

In this section, we describe some additional information on the students and schools in our analysis data that we did not present in Section 5.2.

Table S.1 reports mean characteristics of students and their families. Only families making less than 185% of the federal poverty line were eligible for the program, and so unsurprisingly the students are relatively disadvantaged. Approximately 50% of the children's mothers were married, and fewer than 50% were employed at baseline. Family income was slightly less than \$17,000. Baseline achievement reflects both positive and negative selection: families selected into participation in the experiment, but they also had to be relatively poor to qualify. The table also reveals that voucher recipients and non-recipients are balanced in terms of the various predetermined characteristics. This suggests that the receipt of the voucher was random, in line with Assumption B(i).

Table S.2 reports characteristics of the private and government-funded schools in the sample, both unweighted and weighted by attendance. Panel A reveals that the private schools are substantially whiter, have smaller student/teacher ratios, and are more likely to track students by ability. Most strikingly, many of the private schools are religious—35% of them are Catholic, and an additional 20% another religion. In addition, private schools tend to have lower share of minorities, lower share of student/teacher ratio, lower school sizes, have more students tracked by ability and have lower learning difficulties program. Comparing Panel A and Panel B reveals that among the schools that students actually attended (as reported in the attendance-weighted results in Panel B), there are smaller differences between private and government-funded schools. For example, while the average private school is only 73% minority (relative to 96% for the government schools),

Table S.1: Student and family characteristics by voucher receipt

	With voucher	Without voucher	Difference
Mother married (=1)	0.52	0.55	-0.034
Mother years education	12.20	12.25	-0.057
Mother works full time (=1)	0.35	0.38	-0.028
Mother works part time (=1)	0.11	0.11	0.007
Family income ($)	16,725	17,372	-647
HH receives govt transfers (=1)	0.03	0.01	0.016
Household size	4.11	4.14	-0.030
Black (=1)	1.00	1.00	-0.001
Male (=1)	0.50	0.47	0.030
Grade $\leq$ 5 (=1)	0.65	0.65	0.000
Grade 6-8 (=1)	0.21	0.21	-0.000
Grade $\geq$ 9 (=1)	0.14	0.14	0.000
Child learning disabilities (=1)	0.09	0.09	-0.004
Observations	1,090	730	

Table shows mean student and family characteristics by treatment group, weighted using the baseline weights. Observations rounded to the nearest 10.

SOURCE: *Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)*, U.S. Department of Education, National Center for Education Statistics previously unpublished tabulations.

the average private school attended by a voucher student is 96% minority.

S.4.3 Details and Empirical Results on Parameterized Models

In this section, we describe the various fully parameterized models we consider and how we estimate them, and also report the estimates under them. Denoting by $P_i = p(\tau_{\text{sq}}) \cdot Z_i + p(0) \cdot (1 - Z_i)$ the realized price vector that an individual faces, the general model in (2)-(3) under the realized price in the absence of school level covariates can be given by

$$\begin{aligned} U_{ij} &= -\gamma_{2i} P_{ij} + \xi_j + \epsilon_{ij} && \text{for} \quad j \in \mathcal{J}_v \ , \\ U_{in} &= \xi_n + \epsilon_{in} && \text{for} \quad j \in \mathcal{J}_n \ , \\ U_{ig} &= 0 && \text{for} \quad j \in \mathcal{J}_g \ . \end{aligned}$$

Note that we do not include school level covariates because we do not observe them for all the schools in our data, and also because they are constant within each school and hence their coefficients cannot be separately identified from the school fixed effects unless additional assumptions are made. Moreover, note here that we have normalized the utility of the group of government schools to be 0 and hence do not have γ_{1i} in the equations.

Table S.2: Characteristics of sample schools

	Private	Government-funded	Difference
Panel A: Unweighted characteristics			
Share minority	0.73	0.96	-0.227
School size	222.97	325.90	-102.924
Student/teacher ratio	8.92	12.82	-3.897
Catholic (=1)	0.35	0.00	0.354
Other religious (=1)	0.20	0.00	0.200
Secular (=1)	0.45	1.00	-0.554
Gifted program (=1)	0.35	0.39	-0.040
Learning difficulties program (=1)	0.48	0.93	-0.447
Individual tutors available (=1)	0.64	0.69	-0.052
Students tracked by ability (=1)	0.79	0.60	0.192
Remedial classes available (=1)	0.61	0.68	-0.070
Panel B: Attendance-weighted characteristics			
Share minority	0.96	0.98	-0.017
School size	205.86	419.46	-213.605
Student/teacher ratio	13.17	13.72	-0.551
Catholic (=1)	0.53	0.00	0.534
Other religious (=1)	0.25	0.00	0.252
Secular (=1)	0.21	1.00	-0.786
Gifted program (=1)	0.34	0.34	-0.000
Learning difficulties program (=1)	0.45	0.96	-0.518
Individual tutors available (=1)	0.80	0.77	0.026
Students tracked by ability (=1)	0.70	0.55	0.157
Remedial classes available (=1)	0.68	0.73	-0.048
Observations	60	160	

Displays school characteristics for private and government-run schools. We do not break out the private schools by participation status because the non-participating schools almost never responded. Observations rounded to the nearest 10.

SOURCE: *Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)*, U.S. Department of Education, National Center for Education Statistics previously unpublished tabulations.

The various specifications we consider all assume that ϵ_{ij} are independent and identically distributed with a Type I extreme value distribution. The difference across the various specifications is how they treat γ_{2i}, i.e. the individual level coefficient on prices. They are given by the following:

L1 (Logit 1): $\gamma_{2i} = \bar{\gamma}_0$, i.e. the price coefficient is constant across individuals.

ML1 (Mixed Logit 1): $\gamma_{2i} = \bar{\gamma}_0 + v_i$, where v_i is normally distributed with mean 0 and variance σ^2, i.e. there is unobserved heterogeneity in the price coefficient across individuals.

L2 (Logit 2): $\gamma_{2i} = \bar{\gamma}_0 + \bar{\gamma}_1 X_i$, i.e. there is observed heterogeneity in the price coefficient across individuals.

ML2 (Mixed Logit 2): $\gamma_{2i} = \bar{\gamma}_0 + \bar{\gamma}_1 X_i + v_i$, where v_i is normally distributed with mean 0 and variance σ^2, i.e. there is both observed and unobserved heterogeneity in the price coefficient across individuals.

Since $P_i = p(\tau_{\text{sq}}) \cdot Z_i + p(0) \cdot (1 - Z_i)$, where Z_i is statistically independent of all the unobservables, we can estimate the coefficients in the utilities and distributions of the unobserved terms in the various specifications above using standard logit and mixed logit methods based on maximum likelihood estimation. In our implementation, the individual level covariates correspond to: an indicator whether the mother is married, an indicator for the mother is working full time, and indicators for which bin the family income lies in, where there are four bins determined by quartiles of the empirical distribution of family income.

Once we have the coefficient estimates, we can estimate the demand for $j \in \mathcal{J}_v \cup \{g, n\}$ at price p, potentially using simulation methods. For example, for **ML2**, we have that

$$\hat{q}_j(p) = \sum_{i=1}^{n} \int \frac{e^{-\gamma_2^* p_j + \hat{\xi}_j}}{1 + e^{\hat{\xi}_n} + \sum\limits_{l \in \mathcal{J}_v} e^{-\gamma_2^* p_l + \hat{\xi}_j}} \phi(\gamma^* | \hat{\bar{\gamma}}, \hat{\sigma}, X_i) d\gamma^* \text{ for } j \in \mathcal{J}_v \ ,$$

$$\hat{q}_n(p) = \sum_{i=1}^{n} \int \frac{e^{\hat{\xi}_n}}{1 + e^{\hat{\xi}_n} + \sum\limits_{l \in \mathcal{J}_v} e^{-\gamma_2^* p_l + \hat{\xi}_j}} \phi(\gamma^* | \hat{\bar{\gamma}}, \hat{\sigma}, X_i) d\gamma^* \ ,$$

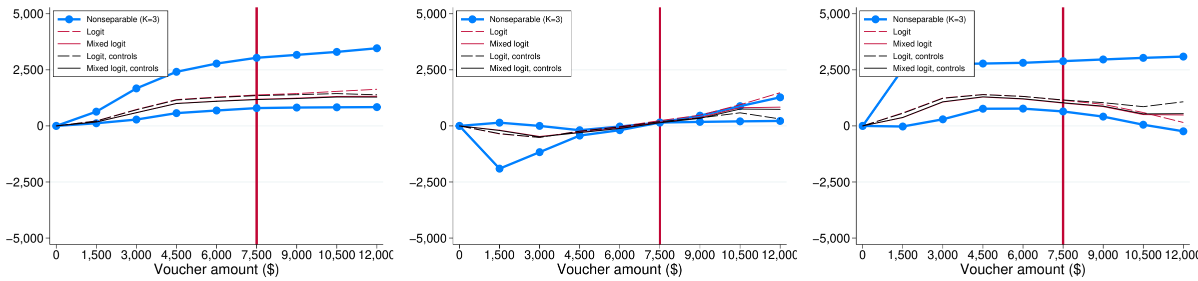
$$\hat{q}_g(p) = \sum_{i=1}^{n} \int \frac{1}{1 + e^{\hat{\xi}_n} + \sum\limits_{l \in \mathcal{J}_v} e^{-\gamma_2^* p_l + \hat{\xi}_j}} \phi(\gamma^* | \hat{\bar{\gamma}}, \hat{\sigma}, X_i) d\gamma^* \ ,$$

where $\hat{\xi}_j$ denotes the estimated version of ξ_j, and the integrals are estimated by simulating a large number of draws from ϕ which corresponds to the density function of a normal distribution with mean $\hat{\bar{\gamma}}_0 + \hat{\bar{\gamma}}_1 X_i$ and variance $\hat{\sigma}^2$, where $\hat{\bar{\gamma}}$ and $\hat{\sigma}$ are the estimated versions of $\bar{\gamma} = (\bar{\gamma}_0, \bar{\gamma}_1)$ and σ, respectively. Given these estimates, we can then compute $AB(\tau)$ using the expression in (10) by numerical integration, and $AC(\tau)$ using the expression in (11) and the estimated values of the demands at $p(\tau_{\text{sq}})$ and $p(0)$. In particular, for $AB(\tau)$, we take an equally spaced fine grid $\{a_1, \ldots a_M\}$ with spacings Δ on $[0, \tau]$ and then take

$$\widehat{AB}(\tau) = \sum_{m=1}^{M} \sum_{l=0}^{j(\tau)} \sum_{j=l+1}^{J} 1\{a_m \in (a_l(\tau), a_{l+1}(\tau))\} \cdot \Delta \cdot \hat{q}_j\left(p_1(0), \ldots, p_l(0), p_{l+1}(\tau) + a_m, \ldots, p_J(\tau) + a_m\right) \ .$$

Figure S.1 reports the estimates of our various parameters measuring the welfare effects of providing various voucher amounts. These are the counterparts of Figure 3(a)-(c). For comparison, we also plot the bounds under the parametric specification from these figures here. To make the figures visually clearer, we do not plot the bounds under the nonparametric specification.

Figure S.1: Parametric bounds and logit point estimates on welfare effects for counterfactual voucher amounts



(a) $AB\,(\tau_c)$ (b) $AC\,(\tau_c)$ (c) $AS\,(\tau_c)$

SOURCE: *Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)*, U.S. Department of Education, National Center for Education Statistics previously unpublished tabulations.

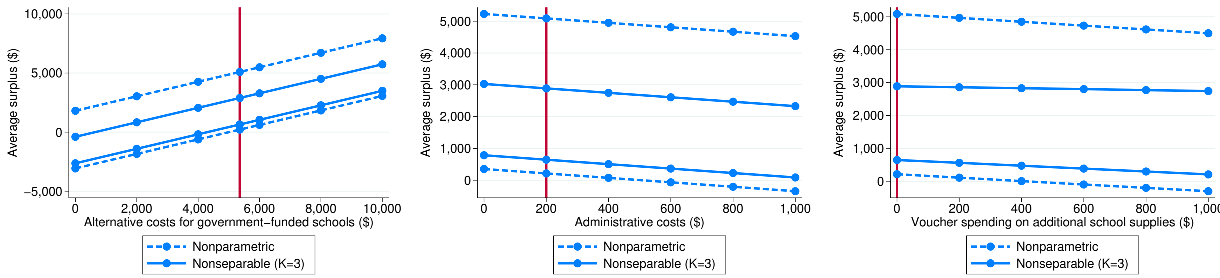
As briefly discussed in Section 5.3, Figure S.1(a) reveals that the various logit specifications all produce estimates of the welfare benefits that are towards our lower bounds. For example, the average benefit of the status-quo voucher is \$1,173 under the most flexible mixed logit model, **ML2**, relative to a bound of [\$795, \$3,038] under our most flexible parametric specification and a bound of [\$364, \$5,239] under our nonparametric specification. The estimates are also sometimes below the lower bounds of some of our stronger parametric specifications, specifically the AS and NS specifications with K=1. We see a similar pattern at all counterfactual voucher amounts; if the voucher were worth \$12,000 then **ML2** would predict an average benefit of \$1,285, relative to a bound of [\$839, \$3,463] under our most flexible parametric specification. As a result, given that the estimated average costs are near the middle of the bounds (Figure S.1(b)), the logit specifications produce estimates of the average surplus that are also towards our lower bounds, and hence may understate the net benefits of the voucher (Figure S.1(c)). For example, the average benefit of the status quo voucher is \$1,033 under the most flexible mixed logit model, **ML2**, relative to a bound of [\$645, \$2,887] under our most flexible parametric specification. More broadly, this suggests that the logit specifications potentially underestimate the individual valuations for the price decrease induced by the voucher, and in turn how responsive individuals are to this price change.

S.4.4 Sensitivity Analysis

In this section, we perform several sensitivity analyses to analyze the robustness of our empirical results in Section 5.3 that revealed that the provision of the status-quo voucher amount results in a positive net average benefit.

First, recall that our analysis used a pre-specified value for c_g, the cost the government faces when a student enrolls in a government-funded school. Our baseline value was \$5,355. Figure

Figure S.2: Estimated upper and lower bounds of average surplus for status-quo voucher amount under alternative pre-specified values



(a) Government school cost c_g (b) Administrative costs γ (c) Additional spending δ

SOURCE: *Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)*, U.S. Department of Education, National Center for Education Statistics previously unpublished tabulations.

S.2(a) presents estimates for the parameter $AS(\tau_{sq})$, estimated in the same way as in Table 2, for a range of values of c_g. For conservativeness, we focus on the two most flexible specifications in Table 2, i.e. the nonparametric specification in Column (1) and the most flexible auxiliary parametric specification in Column (10). The estimates reveal the average surplus increases with c_g. This arises simply due to the fact that the voucher induces recipients away from government-funded schools, which implies that a higher c_g results in higher net cost savings and hence a higher average surplus. Observe that as long as we assume that c_g is at least slightly above \$5,000, the conclusion that there is positive net average benefit continues to hold under both specifications.

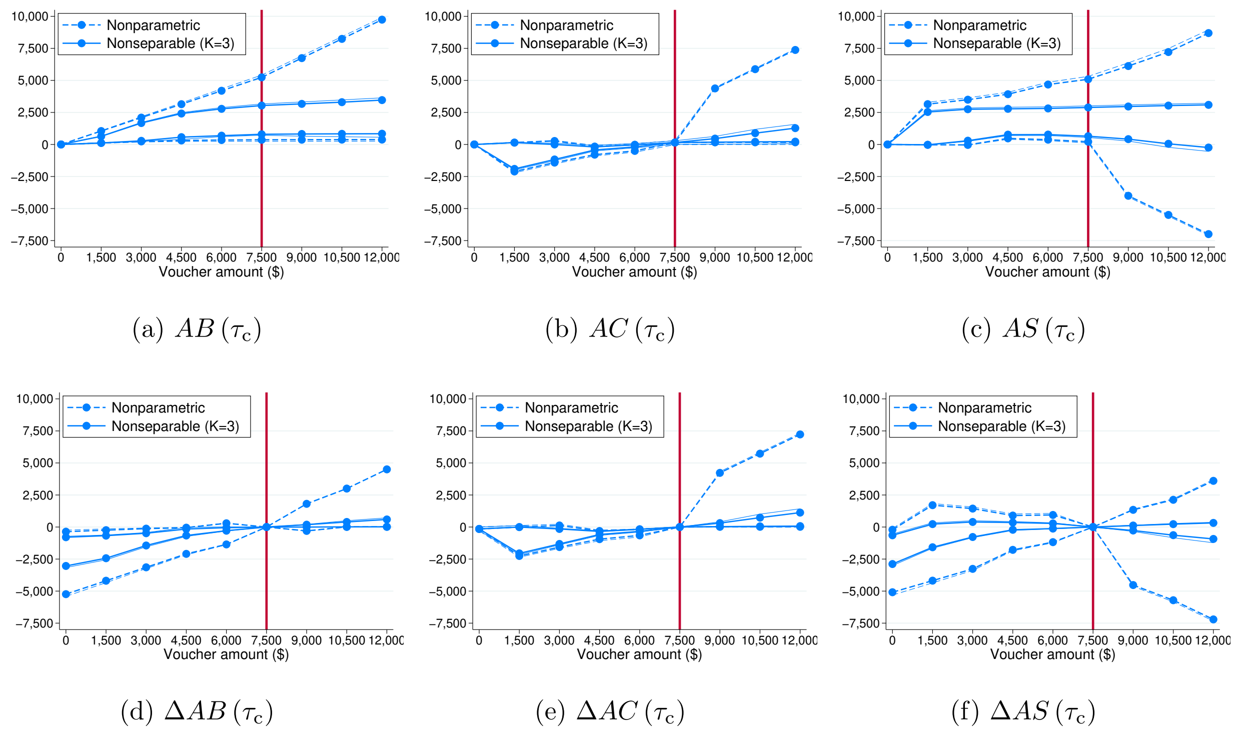
Next, recall that our analysis used a pre-specified value for γ, the possible administrative costs of providing a voucher. Our baseline value is \$200. Figure S.2(b) presents estimates for $AS(\tau_{sq})$ for a range of values of γ. The estimates reveal that average surplus decreases with γ. This arises simply because it makes providing the voucher more costly. For values of γ lower than \$400, we continue to find positive net average benefits under both specifications.

Finally, recall that our analysis presumed that the price of private schools was only the tuition amount and, in turn, implicitly assumed that the voucher could be used to only offset tuition. However, in practice, the voucher could also be used to offset other costs such as fees and transportation costs. Unfortunately, it is difficult to fully account for these other additional costs in our analysis due to the fact that they are unobserved for each individual. Nonetheless, to analyze the sensitivity of our conclusions to this implicit assumption, we suppose that each individual has an additional homogenous cost δ for each private participating school j in $\mathcal{J}_v$ that the voucher can be used to offset. In this case, observe that the price of that school corresponds to $p_j(0) + \delta$ as opposed to $p_j(0)$, the case in our analysis. While higher values of δ weakly increase the average benefit of the voucher as the voucher potentially offsets a higher amount, it can also increase the net costs of the voucher. Figure S.2(c) presents the estimates for the parameter $AS(\tau_{sq})$ for a range of values of δ when the price for the jth school in $\mathcal{J}_v$ in our main analysis is redefined from $p_j(0)$ to $p_j(0) + \delta$. The estimates reveal that average surplus decreases with δ, implying the increase in costs is greater than that of benefits. Observe that as long as we assume that the costs are not more than around \$400, the conclusion that there is positive net average benefit continues to hold under both specifications.

S.4.5 Confidence Intervals for Main Results

Given that the estimated confidence intervals nearly coincide with the estimated bounds, for clarity we did not report them in Figures 3 or 4. For completeness, we do so here; our substantive conclusions are unchanged.

Figure S.3: Estimated upper and lower bounds on counterfactual voucher amounts



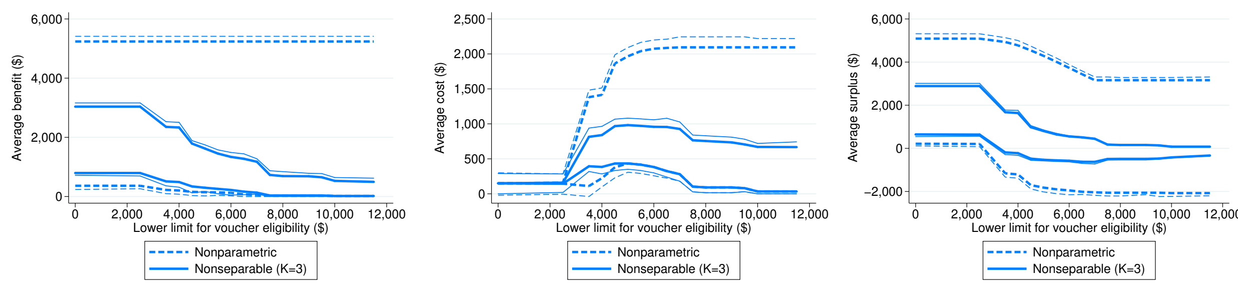
(a) $AB\,(\tau_c)$ (b) $AC\,(\tau_c)$ (c) $AS\,(\tau_c)$

(d) $\Delta AB\,(\tau_c)$ (e) $\Delta AC\,(\tau_c)$ (f) $\Delta AS\,(\tau_c)$

Panels (a)-(c) respectively show the upper and lower bounds on estimated average benefit, average cost, and average surplus for each possible value of the voucher. Panels (d)-(f) show the upper and lower bounds on the difference between the parameter at a given level of voucher generosity and the parameter for the status-quo voucher generosity. In each figure, point estimates are represented by the thicker lines and 95% confidence intervals by the thinner lines.

SOURCE: *Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)*, U.S. Department of Education, National Center for Education Statistics previously unpublished tabulations.

Figure S.4: Estimated upper and lower bounds on welfare effects of removing schools with tuition below κ from the program



(a) $AB^{\kappa}\left(\tau_{\text{sq}}\right)$ (b) $AC^{\kappa}\left(\tau_{\text{sq}}\right)$ (c) $AS^{\kappa}\left(\tau_{\text{sq}}\right)$

Panels (a)-(c) respectively show the upper and lower bounds on the estimated average benefit, average cost, and average surplus for a voucher of the status quo amount ($7,500), but excluding schools below the tuition level given on the x axis. In each figure, point estimates are represented by the thicker lines and 95% confidence intervals by the thinner lines.

SOURCE: *Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)*, U.S. Department of Education, National Center for Education Statistics previously unpublished tabulations.

References

Bhattacharya, D. (2018). Empirical welfare analysis for discrete choice: Some general results. *Quantitative Economics*, **9** 571–615.

Kalouptsidi, M., Kitamura, Y., Lima, L. and Souza-Rodrigues, E. A. (2020). Partial identification and inference for dynamic models and counterfactuals. Tech. rep., National Bureau of Economic Research.

Wolf, P., Gutmann, B., Puma, M., Kisida, B., Rizzo, L., Eissa, N., Carr, M. and Silverberg, M. (2010). Evaluation of the dc opportunity scholarship program: Final report. *National Center for Education Evaluation, Institute for Education Sciences Report*, **4018**.