

Identification and Estimation of Generalized Selection Models, with an Application to Texas Direct

Yujin Kwon[‡]

This version: October 22, 2026

Abstract

This paper studies models in which the observability of people's outcomes depends on their economic choices. In these models, the outcomes resulting from a given treatment are observed only for those who chose to participate in that treatment through utility maximization. The proposed method enables the identification and estimation of general models with such choice-dependent outcomes while relaxing some of the identifying restrictions commonly used by existing tools, providing a simple framework for studying counterfactual choice-specific outcomes for the population in a way that reflects the available information. The generality of our method allows it to be applied even when data limitations, such as limited or censored instrumental variables, preclude researchers from fully determining how people's decisions affect the observability of their outcomes. The methods are applied to evaluate the effect of Texas Direct, a major community college transfer initiative. Using Texas administrative data, we find that the benefits of completing the Texas Direct curricula components for students' post-transfer outcomes depend on how students accumulate and apply credits from each curriculum.

1 Introduction

Models reflecting people's choices, in which an outcome associated with a given treatment is observed only for a self-selected subsample of people who choose to engage in that treatment, are prevalent in economics research. People's rational decision-making process causes some outcomes of interest to never be realized and observed for members of the population who never select certain choices

*The University of Texas at Austin, 2225 Speedway, BRB 4.132, C3100, Austin, TX 78712, USA (email: kwonyujin@utexas.edu).

[‡]I am deeply grateful to Brendan Kline, Jason Abrevaya, Lauren Schudde, and Xinyue Bei for their invaluable guidance and support. I also thank participants at the 2026 Midwest Econometrics Group Annual Meeting, the 2026 Texas Camp Econometrics, and the UT Austin Econometrics Writing Seminar for helpful comments and suggestions. All errors are my own.

that lead to them, making it infeasible to rely solely on observed information to evaluate outcomes pertaining to those unchosen alternatives. A classic example is a model of occupational choice, in which the returns to the sector each person chose are observed, but the returns they would have received in the sectors they did not choose can never be observed. Because people tend to choose sectors based on their skills, the returns they would have received in unchosen sectors may differ from those observed among people working in those sectors, making it hard to learn the effect of sectoral choice on people's returns. Another example is our analysis of Texas Direct, a community college transfer policy in Texas. Students participate in the policy based on their knowledge of transferring to a four-year university and their academic goals, both of which are likely associated with their academic outcomes. Unobserved outcomes for those who never choose to participate make it hard to draw conclusions about the effect of the policy. Motivated by this inherent limitation of observed data, this paper studies the model defined in equation (1), which encompasses a wide range of selection models with additively separable errors that have been studied in econometrics for decades (Heckman, 1976; Ahn and Powell, 1993; Das et al., 2003; Newey, 2009; Eisenhauer et al., 2015).

$$\begin{aligned} Y &= Y(D), Y(t) = m_t(X) + e_t, \\ D &= \arg \max_{t \in \mathcal{T}} U_t, t \in \mathcal{T}. \end{aligned} \tag{1}$$

Y denotes the outcome available in the observed data, whereas $Y(t)$ denotes a counterfactual outcome under treatment t , which belongs to the set of treatments $\mathcal{T}$. X is a vector of covariates. e_t denotes the error term, and U_t denotes utility from choice t . People choose D to maximize utility; the observed outcome reflects the counterfactual outcome after receiving the utility-maximizing treatment, so $Y = Y(D)$. The utility values are assumed to have the following functional form:

$$U_t = g_t(Z_t) + v_t, t \in \mathcal{T}, \tag{2}$$

where Z_t is a vector of variables related to utility from treatment t among Z , which collects Z_t across all treatments for each individual. v_t are unobserved error terms. The primary parameter of interest for researchers is usually the population conditional mean parameter, $m_t(X)$, for a specific treatment. However, the potential outcome for the treatment of interest is only observed for individuals who select it through utility maximization. The unobserved heterogeneity affecting individuals' utility, $V = (v_0, \dots, v_M)$, is often correlated with the outcome errors, e_t , in unknown ways. This creates a

discrepancy between the observed conditional average of outcomes, $E[Y|D = t, X]$, and the true parameter of interest, $E[Y(t)|X]$.

The key contribution of this paper is to provide researchers with a toolkit to address this identification problem arising from “missing data” on unchosen outcomes in selection models, particularly in settings that cannot be accommodated by existing tools, which are often based on strict data requirements or strong statistical assumptions. Information about the data-generating process of the observed sample plays a key role in resolving the discrepancy between the population and the self-selected subsample. For example, with the matching method, also used by Spencer (2021) to evaluate the effect of Ohio’s statewide transfer policy similar in spirit to Texas Direct, identifiability of the model is established after assuming that conditional on observed covariates, treatment assignment is effectively random. That is, individuals sharing the same observed characteristics can be compared as if treatment had been randomly allocated among them. Such an assumption can be restrictive, since the unobserved errors in the observed outcomes are not allowed to depend on self-selection after conditioning on the covariates (Heckman and Navarro-Lozano, 2004). To allow for more flexibility, we make a different assumption: the instrumental variables, Z , shift individuals’ utilities in a way that allows for the identification of the treatment-choice-generating process. By explicitly modeling how the unobserved variables affecting the outcomes are related to those that generate the observed choices, we account for the discrepancy between the observed data and what we would observe under random allocation.

Central to this approach of controlling for self-selection, known as the “control function approach” (Heckman and Robb, 1985; Ahn and Powell, 1993; Han and Kaido, 2026), are the instrumental variables Z and the properties they are assumed to have. In practice, those assumptions can be restrictive and challenging to satisfy. Moreover, they often become stronger as the treatment structure becomes more complicated, for example, when moving from binary to multinomial choice. One way that the methods proposed in this paper improve upon existing methods is that we do not require the available dataset to satisfy those properties, allowing the methods to be applicable in various empirical settings.

In general, such restrictions used in the existing literature are made to ensure that the mapping from the model parameters to conditional choices, given the observed variables, is one-to-one. They ensure that the choice-generating process can be fully characterized, thereby allowing us to identify and estimate the “control function” that takes into account the gap between observed outcomes and unobserved population-level outcomes. For example, Heckman and Sedlacek (1985) assume that there are sector-specific attributes in their study of an extended Roy model. Lewbel (2007) has

shown that identification of the selection model can be achieved if Z is a “special regressor” with large support that can shift individuals’ selection probabilities to 1 or 0. In many cases, however, these conditions are not satisfied. For instance, as in our study of Texas Direct, choice-specific utility shifters may be unavailable because the research question concerns participation in a bundle of treatments relative to participation in each treatment separately. In studies using administrative datasets, key utility-shifting instrumental variables may be interval-censored due to confidentiality restrictions, resulting in more than one utility-maximizing choice prediction from a given utility function according to the true, unobserved value of the instrument the person actually received. The proposed methods accommodate cases where the data-generating process is partially identified, that is, cases in which the mapping from the observed information to the models that can rationalize the data is a correspondence. They also allow for the choices to be generated by incomplete models (Tamer, 2003), where the model and the observed variables map to more than one prediction of utility-maximizing choice. By allowing for such cases, the proposed tools can be used under much relaxed conditions.

An alternative or additional assumption that has frequently been used in the existing literature is that V follows a specific distribution, such as an independent Gumbel distribution or a multivariate normal distribution. For instance, Dahl (2002) studies a multi-market Roy model while assuming that the selection unobservables are independent and identically distributed with the extreme value distribution. D’Haultfœuille and Maurel (2013) study the identification of an extended Roy model under more general restrictions that leverage a specific dependence structure among the selection unobservables. A limitation of these approaches is that although they may be too restrictive, these assumptions cannot be empirically tested in general cases. In reality, the population may consist of groups of individuals who share patterns in how their unobserved heterogeneity affects their choices, and each individual’s unobserved determinants of utility are often correlated across choices in complicated ways. To allow for more flexibility and a more realistic reflection of the real world, we do not make such assumptions. Instead, we collect all distributions of V that may underlie the observed choices based on the observed variables and then take into account all possible choice-generating processes when identifying the outcome model parameters. By doing so, our identification strategy accommodates various empirical settings while enabling a more credible study of the model of interest.

Our approach of allowing the data to map onto the identified set of parameters, which is key to relaxing restrictive assumptions required to establish a one-to-one mapping from the data to the model, is built upon the literature on partial identification. On the one hand, because of the

importance of optimizing behavior in economics and because the observed choice often provides a set of inequality constraints on the underlying utility function, the choice model has been a source of partial identification results across a series of settings (Kline and Tamer, 2023; examples for general choice problems include Chesher, Rosen, and Smolinski, 2013, and Barseghyan et al., 2021, among many others). On the other hand, due to the nature of the observed data, which is subject to the missing data problem (Manski, 2005), partial identification results have been established in the potential outcomes framework (Mullahy, 2018; Mogstad and Torgovitsky, 2018; Mourifié, Henry, and Méango, 2020). However, the relationship between the identified set of parameters for models with selection bias and the choice model that characterizes their sample selection process has not been thoroughly discussed, and tools are limited for incorporating the rich set of implications garnered from the revealed preference model into identification and estimation of parameters related to potential outcomes that are observed conditional on the utility-maximizing behavior. This paper, by bridging the two strands of literature to propose a practical tool for studying generalized selection models, makes both theoretical and practical contributions.

This paper builds on recent developments in partial identification using random set theory (Beresteanu, Molchanov, and Molinari, 2012; Molchanov and Molinari, 2018). An advantage of these developments is that they enable researchers to characterize the sharp identified set of potentially partially identified model parameters in a tractable way. For models defined by their distributional features, such as our choice model in equation (2), Artstein’s inequality enables characterization of the sharp identified set. On the other hand, for models characterized by conditional expectations, such as our outcome model in equation (1), Aumann expectations over random closed sets provide a straightforward characterization. This difference reflects the strengths and limitations of the two approaches. Artstein’s inequality approach is often challenging to implement for models such as the model in equation (1), where taking conditional expectations is crucial for estimation and identification. Conversely, although assumptions such as statistical independence of certain variables from the unobserved error terms are often crucial for identifying choice models such as the model in equation (2), incorporating such assumptions into the Aumann expectation approach is not straightforward (Beresteanu, Molchanov, and Molinari, 2012). The key challenge our paper addresses is that both aspects, taking conditional expectations and learning how observed choices are generated by leveraging instrumental variables to restrict the distribution of unobservables conditional on those variables, are essential for identifying the parameters of interest. In addressing that key challenge, this paper bridges the theoretical gap between the identified sets characterized via the “Aumann expectation approach” and the “capacity functional approach” in the random set literature.

We complement and extend the literature on applying random set-based tools, including generalized instrumental variables proposed by Chesher and Rosen (2017) and set-valued control functions proposed by Han and Kaido (2026). Our approach provides a natural link between the two approaches, allowing researchers to efficiently leverage the advantages of both tools. Specifically, the method introduced by Han and Kaido (2026) provides a convenient characterization of the identified set of outcomes using set-valued control functions, leveraging the convexification property of Aumann expectations. However, constructing valid set-valued control functions incorporating distributional assumptions is nontrivial when the structural model underlying the source of bias is itself set-identified. In our setting, this requires first characterizing the set of distributions of unobserved values that enter the choice-specific utilities that the observed data can rationalize. We show that, due to the unique structure of the selection model, this characterization can be carried out in a two-step procedure analogous to the classical control function approach of Heckman (1976). Specifically, we first apply the findings of Chesher, Rosen, and Smolinski (2013) on revealed-preference models, and Chesher and Rosen (2017)’s generalized instrumental variables framework, to characterize the set of data-generating processes consistent with the observed data. Our results show that characterizing this set is an essential first step in determining the identified set of outcome-model parameters, and it enables us to incorporate distributional restrictions into models characterized by taking conditional expectations, thereby overcoming the limitations of relying solely on Aumann-expectation-based approaches.

The estimator we introduce is a convenient two-step estimator for which consistency is established by applying our identification results to the theoretical findings of Chernozhukov, Hong, and Tamer (2007) and Kaido and White (2014). This estimator can be viewed as a set-valued analog of the two-stage estimator proposed by Heckman (1976). An estimator for set-identified triangular models with instrumental variables, such as ours, has also been proposed by Chesher, Kim, and Rosen (2023). Although their estimator can be applied to our case by allowing the endogenous selection rule to be deterministic, implementing their method for our model, which deals with polychotomous choices without distributional assumptions on the selection unobservable, may easily become intractable. The method proposed in this paper provides an alternative that can yield substantial computational gains and is especially useful when the researcher is only interested in the conditional mean of the outcomes, as is often the case in various empirical studies. The proposed estimator includes a bootstrap procedure for subsampling-based inference that accounts for uncertainty from using the estimated first-stage model to estimate the second-stage parameters, providing a comprehensive set of tools for practical application.

We apply these insights to evaluate the effects of two central policy levers of Texas Direct, a major initiative of the Texas Higher Education Coordinating Board (THECB) designed to improve the efficiency of transfers between two-year and four-year institutions in the state. Specifically, we examine the role of the Texas Core Curriculum (TCC; Texas Education Code, Chapter 61, Subchapter S), a 42-semester-credit-hour core curriculum for all undergraduate students in Texas, and the Field of Study Curriculum (FOSC; Texas Education Code, Chapter 61, Section 61.823), which is a set of lower-division courses that transfer and count toward a degree program, as required by state law (THECB, 2026). Although the Texas Direct initiative and the Texas Direct Transfer Degree were launched in 2021, joining the FOSC and the TCC together, the FOSC and the TCC have existed for decades. FOSC was first mandated by the Texas Legislature in 1997, and the TCC has been required for all Texas undergraduates since 1987. Throughout this paper, the term “Texas Direct” is used to refer to these two underlying policies rather than to the 2021 initiative itself.

Using a statewide administrative dataset containing individual-level records on student demographics, course enrollment, credit loss, and degree attainment for the population of students in the Texas higher education system, we examine how these transfer policies influence the loss of pre-transfer credits upon transfer and ultimately, bachelor’s degree attainment. Credit portability is one of the key challenges that community college students intending to transfer to a university face; only 58% of community college transfers were able to bring over 90% of their college credits to the four-year institution, while students who transfer most of their credits are more likely to complete their bachelor’s degree (Monaghan and Attewell, 2015). Given the large role of community colleges in providing a pathway to universities in Texas, with 70% of all Texas bachelor’s degree graduates having some transfer credits from a two-year college in 2019 (THECB Almanac, 2020), Texas Direct is a significant initiative for THECB.

Even with a comprehensive population-level dataset, it is challenging to study causal effects of the TCC and the FOSC on student outcomes and how the two interact. As in the model of sectoral choice, students’ degree attainment outcomes in our dataset may not reflect what Texas community college students would achieve if all students completed the TCC and FOSC requirements. Completion of each curriculum sequence is inherently endogenous: students choose coursework based on its perceived future value (Galotti and Umscheid, 2019), which varies greatly among students in ways that are also associated with their academic outcomes. Challenges in disentangling the complicated correlation between students’ course-taking decisions and their outcomes have impeded research efforts.

As discussed further in Section 6, students who complete the lower-division major coursework

of the Field of Study Curriculum before vertical transfer likely differ from those who complete only the general education requirements of the Texas Core Curriculum, in part because the former are already progressing toward degrees in specific majors. Students who complete both curricula likely also differ from both groups because they choose to accumulate many credits before transfer, a choice that can reflect the goal of earning a community college credential before transfer, limited access to transfer information, or, conversely, an awareness of different transfer policies. The differences in academic goals and transfer information are likely to be associated with their outcomes, creating a gap between the overall student population and the subgroups whose outcomes under each Texas Direct policy component are observed. By addressing this identification challenge while accounting for the highly heterogeneous nature of the Texas transfer student population, our analysis provides unique insights.

As in Texas, several states have statewide articulation agreements that include a set of core courses guaranteed to transfer, major-specific pathways, or an associate's degree that guarantees transfer of a full set of course credits and entry with junior standing (Whinnery and Peisach, 2022). Previous research on such policies has found positive effects on bachelor's degree attainment (Stern, 2016) and credit accumulation (Worsham et al., 2021). However, a thorough investigation in the Texas context is critical, as heterogeneity in the intent and implementation of statewide policies may limit understanding of the extent to which articulation agreements affect students when these policies are treated as a homogeneous construct (Worsham et al., 2021).

Almost 40 percent of community college students in the U.S. switch among broad major categories between their first and third years of college (Schudde, Ryu, and Brown, 2020), potentially limiting the effectiveness of a degree-specific pathway like the FOSC (Hodara et al., 2017). We show that students who complete the lower-division major coursework of FOSC lose a positive proportion of their pre-transfer credits upon transfer, regardless of their core completion status, providing the first evidence on the effect of the FOSC. Furthermore, we add to the findings on the effect of the core curriculum by showing that although students who complete the TCC or both curriculum components of Texas Direct can increase their chances of earning a four-year degree compared with students who complete neither curriculum, these gains may be masked because each additional pre-transfer credit may contribute less to degree attainment for curriculum completers than for students who complete neither curriculum. Our results add to the existing evidence that pre-transfer core credit accumulation can benefit bachelor's degree attainment only up to about 44 credits, which is 2 credit hours more than the minimum needed to be considered core complete (Schudde, Bicak, and Shea, 2023). By capturing both distinct and overlapping effects of the TCC and the FOSC, we

illustrate the mechanism underlying the effect of pre-transfer core credits. Our findings offer several important implications for policymakers and students aspiring to transfer while also contributing to the broader literature on strengthening community colleges as a pathway to four-year degree attainment.

The remainder of this paper is structured as follows. Section 2 introduces the setup. Section 3 presents the identification strategy. Section 4 describes a computationally simple two-step estimation procedure for computing the identified set of parameters. Section 5 presents the results of Monte Carlo simulations illustrating the proposed procedures. Section 6 applies the proposed procedures to analyze the effects of Texas Direct. Finally, Section 7 concludes.

2 Setup

The dataset $(Y, D, X, Z) \sim P_0$ consists of a scalar outcome Y reflecting people's choices D among the treatments in $\mathcal{T}$, covariates X , and instruments Z that affect utility but not the outcome. $\mathcal{Y}$, $\mathcal{X}$ and $\mathcal{Z}$ denote their respective supports. All random variables, (Y, D, X, Z) , are defined on a complete, nonatomic probability space $(\Omega, \mathfrak{F}, P)$. When a subset of covariates or instruments is only relevant to a certain treatment t , we denote them using subscripts as X_t and Z_t , and their respective supports as $\mathcal{X}_t$ and $\mathcal{Z}_t$.

The potential outcomes for each $t \in \mathcal{T}$ are generated by the following model, where e_t , $t \in \mathcal{T}$ denotes individuals' unobserved characteristics that affect their potential outcomes:

$$Y(t) = m_t(X) + e_t, (e, v) \sim F_{e|V, X, Z}(\cdot, \cdot \mid X, Z). \quad (3)$$

We denote individuals' unobserved characteristics affecting their utility from each choice alternative as v_t , $t \in \mathcal{T}$, where $|\mathcal{T}| = T = M + 1$. Choices are made by maximizing the utilities over the treatment options in $\mathcal{T}$, as follows:

$$D = \arg \max_{t \in \mathcal{T}} U_t, \text{ where } U_t = g_t(Z_t) + v_t, v|X, Z \sim F_{V|X, Z}(\cdot \mid X, Z). \quad (4)$$

This model can be connected to various treatment evaluation problems. To illustrate, suppose that the outcome model has a linear form:

$$Y(t) = \beta_{0t} + X\beta_{1t} + e_t.$$

The most commonly used specification assumes that β_{1t} is the same across all t , and focuses on the difference in β_{0t} among different treatments (Heckman, 1990). The problem posed by self-selection, when we are interested in the effect of treatment t compared to the “default” of no treatment, can then be written as follows:

$$Y = \beta_{00} + \mathbf{1}(D = t)(\beta_{0t} - \beta_{00}) + X\beta + (1 - \mathbf{1}(D = t))e_0 + \mathbf{1}(D = t)e_t. \quad (5)$$

The confounding from self-selection can be seen as the endogeneity problem due to nonzero $E[\mathbf{1}(D = t)(e_t - e_0)|X, D]$, caused by the correlation between D and e_t , where the former is a function of V , and the latter is correlated with V .

In empirical studies, researchers are often interested in the average effect of particular treatments. Although the average effect can be defined in various ways, as Heckman (1990) has summarized, two parameters have been frequently used. First, as in equation (5), one may be interested in $\beta_{0t} - \beta_{0t'}$, where t, t' are different treatments within $\mathcal{T}$. Alternatively, the parameter of interest may be $\bar{X}_t\beta_{1t} - \bar{X}_{t'}\beta_{1t'}$, where $\bar{X}_t$ is the average value among individuals who choose t (i.e., $E[X|D = t]$).

To illustrate the use of the methods introduced in this paper within this setting, the following three running examples will be discussed throughout the paper. Note that the identification results for multinomial models in settings like Example 2 and 3 have been studied, but neither has been studied in the context of selection models in connection to the conditionally observed outcomes.

Example 1: Heckman (1976)

$\mathcal{T} = \{0, 1\}$, $g_1(Z) = Z\alpha$, $m_1(X) = X\beta$; $e \sim N(0, \sigma)$, $V \sim N(0, 1)$, and $\text{corr}(e, v) = \rho$.

Example 1 is a classic model that leverages distributional assumptions to guarantee point-identification of the choice model. In this approach, the parameters of the choice model can then be conveniently applied to characterize the observed, conditional-on-treatment outcomes. It can be compared with Examples 2 and 3, where the data conditions are nonstandard and existing tools are often challenging to apply.

Example 2: Non-exclusive targeting

$\mathcal{T} = \{0, 1, 2, 3\}$, instruments for treatment 3 are a subset of those for choices 1 and 2.

In general, nonparametric identification of $F_{V|X, Z}$ in Example 2 is challenging without additional assumptions. Such settings are especially common when treatment 3 is a “bundle” that combines treatments 1 and 2, incorporating aspects of both. Understanding these treatment settings is important for learning whether multiple treatments have synergistic effects. That is, whether their

combined effect differs from what would be expected from each treatment individually. Our empirical problem of identifying the effects of TCC and FOSC completion individually, as well as their combined use through Texas Direct, fits this framework due to the structural challenge of targeting full completion of both components without also targeting the individual policy levers.

Example 3: Interval-valued instruments

For some $\tilde{Z}_t \subseteq Z_t$, the observed data includes $([z_{Lt}, z_{Ut}])$, where $P(z_{Lt} \leq \tilde{z}_t \leq z_{Ut}) = 1$.

Example 3 differs from the aforementioned examples because, even when the choice model is point identified, its predicted effects on observed outcomes may remain set identified due to the use of set-valued instruments. The tools proposed in this paper provide a convenient way to handle such settings, including cases where the choice model itself is set identified.

Throughout the paper, we will maintain the regularity conditions in Assumption R1, most of which are commonly used in the random set literature (e.g., Chesher, Rosen, and Smolinski, 2013; Aristodemou and Rosen, 2022).

Assumption R1 (i) The functions $g_t(\cdot) : Z \rightarrow \mathbb{R}$ and $m_t(\cdot) : X \rightarrow \mathbb{R}$ are measurable and continuous in their arguments for all $t \in \mathcal{T}$. (ii) The conditional distribution of e given (D, X, Z) , and the conditional distribution of V given (X, Z) , has continuous and strictly positive density with respect to Lebesgue measure on $\mathbb{R}^M$ almost surely. (iii) $E[|e_t|] < \infty$ for all $t \in \mathcal{T}$.

Notations We sometimes suppress the arguments of set-valued functions $\mathbf{A}(x)$ and denote $\mathbf{A}$ for convenience. We denote the set of all measurable selections of a random set $\mathbf{A}$ as $Sel(\mathbf{A})$. $E[A]$ denotes expectation of a random vector A , while $\mathbb{E}[\cdot]$ denotes Aumann expectation over a random set A . The conditional expectations are denoted similarly. We sometimes deal with parameters that enter the choice or the outcome model specifically; in such cases, we denote them using superscripts, i.e., θ^D and $\theta^{Y(t)}$. We also use $g(\cdot)$, in place of $g_t(\cdot)$, $\forall t \in \mathcal{T}$, for notational simplicity.

3 Identification

3.1 Framework for Identification

We begin by considering a normalized version of the choice model in equation (4), where we normalize the utility from one alternative to zero. Specifically, we normalize the unobserved heterogeneity of a particular alternative, v_{t_n} , to zero by redefining $v_t = v_t - v_{t_n}$. We also do the same for the observed terms as $g_t(Z_t) - g_{t_n}(Z_{t_n})$. In this paper, we normalize with respect to the utility from no treatment,

setting $t_n = 0$. Therefore, among $(v_0, v_1, \dots, v_M) \in \mathbb{R}^T$, one of them is fixed to 0, and the other M variables are normalized with respect to that variable.

Assumption I1 serves as a baseline condition for our analysis. It implies that the econometrician knows the parameter space that aligns with the structure of the model, while allowing for the possibility that the parameter space is defined based on their knowledge about the model of interest.

Assumption I1 *There exists a collection of functions and conditional distributions of unobserved variables $(m, g, F_{e|V,X,Z}, F_{V|X,Z})$ in a parameter space Θ that is known to the econometrician.*

The choice of the parameter space in Assumption I1 can have a direct impact on the identified set. For instance, in Example 3, Manski and Tamer (2002) show that the model is point identified when $(g(\cdot), F_{V|X,Z})$ are in a parametric family and the model has a monotone-index form, when specific conditions on the data are satisfied. The dimensions of the parameters of the functions $m(\cdot)$ and $g(\cdot)$, and the conditional distributions $F_{V|X,Z}$ and $F_{e|V,X,Z}$, are allowed to be finite or infinite. Therefore, they may include any of the parametric, nonparametric, or semiparametric classes.

The potential outcome available from the dataset contains only the one given the selected choice, $Y(D = d)$. The conditional mean observed in the data can therefore be written as $E[Y(t)|D = d, X] = m_d(X) + E[e_d|D = d, X]$, the sum of the conditional mean function and the conditional mean of unobservables. The latter term is not necessarily 0 even when its population counterpart is 0. Our goal is to model this “selection bias” explicitly to learn the average effect of X on the outcomes for the choice t , $m_t(X)$, which is our parameter of interest. Given our choice model, we can consider a level set that collects V underlying the observed outcome data, which is observed conditional on self-selection. We define $\mathbf{V}(D, X, Z; g)$ as follows:

$$\mathbf{V}(D = d, X, Z; g) = \left\{ v \in \mathbb{R}^T \text{ s.t. } d = \arg \max_{t \in \mathcal{T}} (g(Z_t) + v_t) \right\}. \quad (6)$$

$\mathbf{V}(D, X, Z; g)$ collects all unobserved errors that will generate the choice D given the variables X, Z and the utility function $g(\cdot)$. In some cases, the assumptions imposed by the econometrician allow $\mathbf{V}(D, X, Z; g)$ to be a singleton set for every observation. We see this in Example 1: we have $\mathcal{T} = \{0, 1\}$, V is known to be Gaussian conditional on X and Z , and $g(Z) = Z\alpha$. This model is generally point identified, and $\mathbf{V}(D, X, Z; g)$ is a singleton set including a standard normal random variable. In various settings, especially when the data on hand lacks precision, such as in Examples 2 and 3, the assumptions that guarantee a singleton $\mathbf{V}(D, X, Z; g)$ are not satisfied by the data or are not desirable. Not only is point identification of the choice model often challenging, but

even when the underlying choice model is point identified, the prediction of the unobservables for each observation can also be set-valued in settings such as Example 3, leading to non-singleton $\mathbf{V}(D, X, Z; g)$. In order to consider all possibilities underlying the generation of choices, we need to consider the set-valued $\mathbf{V}(D, X, Z; g)$.

We incorporate the generation of the observed outcomes by using the set of underlying choice model unobservables, $\mathbf{V}(D, X, Z; g)$, as a set-valued control function (Han and Kaido, 2026) in the outcome model. Taking conditional expectation over them and denoting the residual $e_D - E[e_D|X, V]$ using η_D , the set-valued outcome predictions can be represented as equation (7).

$$\mathbf{Y}(\eta, D, X, Z; \theta) = \text{cl} \left\{ y \in \mathcal{Y} : y = m_D(X) + E_{\theta^D}[e_D|V, X, Z] + \eta_D, V \in \text{Sel}(\mathbf{V}(D, X, Z; g)) \right\} \quad (7)$$

Observe that the outcome model predictions can be constructed for any selection process, depending on the choice model that is used to construct the prediction set for some parameters of the outcome model. In order for us to overcome the identification challenge from self-selection, which is that we do not observe the counterfactual outcomes given treatments other than the one people self-selected into, we need to ensure that the outcome predictions are constructed in a way that reflects all information about their unobserved counterfactual outcomes, conveyed in their decision-making process. In other words, we first need to learn the decision-making process that aligns with how people in the sample made their choices to ensure that the outcome model parameters reflect our population. Theorem 1 enables this process by showing that we can gather all information about the choice model without relying on the observed outcomes, allowing us to use the decision-making process to learn the outcome model parameters.

Theorem 1 *Suppose that Assumption I1 holds, and that $F_{e|V, X, Z}$ is known to the econometrician for every V in its support conditional on X and Z , for all $X \in \mathcal{X}$, and $Z \in \mathcal{Z}$. Then, any θ^D is in the sharp identified set of parameters given the dataset $(D, X, Z) \sim P_0$, denoted $\Theta_I^D(P_0)$, if and only if it is in the sharp identified set $\Theta_I^{D*}(P_0)$ from the full dataset $(Y, D, X, Z) \sim P_0$.*

Theorem 1 implies that even when the econometrician can map the unobserved variables in the outcome model directly to the choice model unobservables, including the conditional outcome observations do not add any more information about the generation of choices. The logic behind this result is similar to the well-known non-identifiability of survival distributions shown in Tsiatis (1975), or the non-identifiability of Roy models shown by Heckman and Honoré (1990); because we only observe a single cross-section of data conditional on choices, we can construct a theoretical

joint distribution of outcomes that has the empirically observed outcome when $t = d$, and has the counterfactual outcomes that align with the choice model otherwise. Because we only observe outcomes conditional on D , the observed data cannot reject this model.

Going back to the choice model, we know that for individuals with $D = d$ and covariates (X, Z) , their selection unobservable V is in the level set, $\mathbf{V}(D, X, Z; g)$, with probability 1. In other words, V is a measurable selection of $\mathbf{V}$. Therefore, if we learn the conditional distribution $F_{V|X, Z}$ and the parameters of the corresponding structural function $g(\cdot)$ for all V that are in this set, among them would be the parameters of the choice model underlying our observed subsample of the population. Therefore, we consider all distributions $F_{V|x, z}$ that are selectionable with respect to $\mathbf{V}$.

To characterize such parameters, we consider the dual set of $\mathbf{V}(D, X, Z; g)$, which we denote using $\mathbf{D}(V, X, Z; g)$, that will contain the choice of D for each V and g given the covariates. Specifically, we define $\mathbf{D}(V, X, Z; g)$ as follows:

$$\mathbf{D}(V, X, Z; g) = \{d \in \mathcal{T} : \arg \max_{d \in \mathcal{T}} (g_d(Z_d) + v_d), v_d \sim F_{V|X, Z}\}.$$

If the functions g were the true functions underlying our dataset, and if the unobserved terms V that are used as inputs were drawn from the distribution that underlies our dataset, the distribution of choices included in $\mathbf{D}$ will coincide with the conditional choice distribution in the dataset.

We use the capacity functional, $T_{\theta^D}(A|X, Z) = P(\mathbf{D}(V, X, Z; g) \cap A \neq \emptyset | X, Z)$, where $V \sim F_{V|X, Z}$, to characterize the distribution of the random closed set $\mathbf{D}$. The intuition behind the capacity dominance condition is simple: the data are compatible with coarsely predicted $\mathbf{D}$ if and only if the data do not put more mass in any region than $\mathbf{D}$ intersects it. The information about the parameters of the selection process is summarized by Theorem 2.

Theorem 2 *Under Assumptions R1 and I1, the sharp identified set $\Theta_I^D(P_0)$ of $(g, F_{V|X, Z})$ from $(D, X, Z) \sim P_0$ can be characterized as follows:*

$$\Theta_I^D(P_0) = \{(g(\cdot), F_{V|X, Z}) \in \Theta^D : T_{\theta^D}(A|X, Z) \geq P_0(D \in A | X, Z), \text{ a.e., } \forall \mathcal{X}, \mathcal{Z}, \forall A \subseteq \mathcal{T}\}.$$

Having established Theorem 2, it is convenient to incorporate the set of choice model parameters in the notation for the outcome model as $E_{\theta^D}[e_d | X, Z, V]$, using the fact that the outcomes depend on self-selection only through e_d . The value of the added first-step procedure and the instrumental variables excluded from the outcome model is that we can restrict $\theta^D \in \Theta^D$ to be in the choice model identified set from the dataset, $\Theta_I^D(P_0)$. This ensures that we can use the knowledge from the

choice model, and the model assumptions, to impose identifying restrictions on the outcome model.

To incorporate the identified set of choice-model parameters to characterize the set of outcomes that can result from the possible underlying decision-making processes, we apply the method of set-valued control functions (Han and Kaido, 2026). To facilitate this, we first introduce Corollary (1).

Corollary 1 *Let a random closed set $\mathbf{V}(D, X, Z; g)$ be constructed from $(g, F_{V|X,Z})$ such that $F_{V|X,Z}$ is selectionable with respect to the conditional distribution of $\mathbf{D}(V, X, Z; g)$, for all $V \in \mathbf{V}(D, X, Z; g)$.*

- *Suppose that $F_{e_d|D,X,Z} = F_{e_d|D,X}$, $\forall d \in \mathcal{T}$. Then, $\mathbf{V}(D, X, Z; g)$ is a set-valued control function such that for all $V \in \mathbf{V}(D, X, Z; g)$, $F_{e|D,X,Z,V} = F_{e|X,Z,V}$.*
- *Suppose that $E[e_d|D, X, Z] = E[e_d|D, X]$, $\forall d \in \mathcal{T}$. Then, $\mathbf{V}(D, X, Z; g)$ is a set-valued control function such that for all $V \in \mathbf{V}(D, X, Z; g)$, $E[e_d|D, X, Z, V] = E[e_d|X, V]$.*

We then add requirements in Assumption I2, which will be used to establish Theorem 3.

Assumption I2 *(i) For all $t \in \mathcal{T}$, there is at least one element of $\mathcal{Z}_t$ that is excluded from $\mathcal{X}$; that is, $\mathcal{Z}_t \not\subseteq \mathcal{X}$. (ii) For all $t \in \mathcal{T}$, $E[e_t|D, X, Z] = E[e_t|D, X]$.*

The two conditions in Assumption I2 allow us to vary the correction terms separately from the conditional means and identify the outcome-model parameters using observed outcomes. The first requirement of Assumption I2 is a necessary condition to use the variation in the observed conditional expectations given X and Z to learn about the parameters of the outcome model, and it is weaker than the second requirement, in the sense that it is not possible to satisfy the second requirement while violating the first. When this condition is not satisfied, identification of the conditional mean parameter would solely rely on the shape restrictions imposed by the parameter space. The second condition is a mean independence condition that is necessary to ensure that the choices affect the observed outcomes only through V , which can be varied with Z . This is generally weak, and can be useful especially since stochastic independence of V from Z conditional on X does not imply stochastic independence of e from Z conditional on X and V . For example, in our empirical model predicting students' credit loss and four-year degree attainment outcomes for our study of Texas Direct, we assume that unobserved factors affecting students' outcomes are mean independent of course flexibility conditional on pre-transfer credits earned and unobserved characteristics affecting their decision to complete the curricula. This allows for the possibility that, for instance, a wide variation in instruction quality from increased course availability and remote sections in lower-division coursework causes a large variation in four-year degree attainment following transfer among

those who complete the field of study curricula, which would not be accommodated by the stochastic independence assumption.

Maintaining the assumptions of Theorem 2 and Assumption I2, we first construct the set of outcome predictions corresponding to each observation with covariates X and Z in the data:

$$\mathbf{Y}(D, X, Z; \theta) = cl\{y \in \mathcal{Y} : y = m_D(X) + E_{\theta^D}[e_D|X, V] + \eta_D, V \in Sel(\mathbf{V}(D, X, Z; g))\}.$$

where $E_{\theta^D}[e_D|D, X, V] = E_{\theta^D}[e_D|X, V]$ by definition of V , and $\eta_D = e_D - E[e_D|X, V]$.

To aggregate the individual predictions and obtain their expectation, we take the conditional Aumann expectation over the set-valued predictions $\mathbf{Y}(D, X, Z; \theta)$. Aumann expectation is the closure of the set of expectations of all integrable selections of $\mathbf{Y}$. Under our regularity conditions, it can be understood as a closed set that contains all conditional expectations of predicted outcomes. Because we cannot pinpoint the true data-generating process underlying our observed data, we construct this set of expected predictions of outcomes for all data-generating process that may be underlying our population of interest, which is conveyed by $\Theta_I^D(P_0)$. The joint conditional distribution of unobserved terms affecting people's choices can then be used to take into account the gap between the population-level conditional expectation of counterfactual outcomes given a particular choice and the observed data to identify the parameter of interest.

A characteristic of Aumann expectations that is especially favorable in practice is that it is always a convex set, thereby providing a convenient moment inequality specification. Leveraging this feature, we seek outcome model parameters satisfying the following inequality constraint:

$$m_t(X) + \inf_{V \in \mathbf{V}} E_{\theta^D}[e_t|t, X, Z] \leq E_{P_0}[Y|D, X, Z] \leq m_t(X) + \sup_{V \in \mathbf{V}} E_{\theta^D}[e_t|t, X, Z], \theta^D \in \Theta_I^D(P_0). \quad (8)$$

The identified set of parameters of the conditional mean model is the set of parameters that satisfy equation (8). Theorem 3 formalizes this.

Theorem 3 (Identification of the Outcome Model) *Suppose that the conditions for Theorem 2 and Assumption I2 are satisfied, and $E[|Y|] < \infty$. For all $t \in \mathcal{T}$, $\Theta_I^{Y(t)}(P_0)$ defined as*

$$\begin{aligned} \Theta_I^{Y(t)}(P_0) = \{ & (m_t, F_{e_t|V, X, Z}) \in \Theta^{Y(t)} : m_t(X) + \inf_{V \in \mathbf{V}(t, X, Z; g)} E_{\theta^D}[e_t|X, V] \leq E[Y|D = t, X, Z] \\ & \leq m_t(X) + \sup_{V \in \mathbf{V}(t, X, Z; g)} E_{\theta^D}[e_t|X, V], \theta^D \in \Theta_I^D(P_0), \forall \mathcal{X}, \forall \mathcal{Z} \}, \end{aligned} \quad (9)$$

characterizes the sharp identified set of $\theta^{Y(t)} = (m_t(\cdot), F_{e_t|V, X, Z})$ from $(Y, D, X, Z) \sim P_0$.

Generating the set of outcome predictions using the set-valued control functions, and then applying the convexification property of Aumann expectations, is a straightforward application of the method developed in Han and Kaido (2026). The novel aspect of our result is connecting the model to the set of choice model parameters, based on Theorem 1. This connection ensures that the control functions provide meaningful restrictions consistent with the generation of the observed outcomes, while guaranteeing that we use the full identifying power of the instruments. In particular, settings with set-valued selection unobservables fall into two cases. First, for a given set of parameters, only one choice will be predicted by the data; yet, multiple pairs of parameters may be admissible by the dataset on hand. In other words, the model is partially identified, but complete in the sense of Tamer (2003). This is the case for our Example 2; even when Assumption I3 is satisfied, the distribution of instruments for the third choice is degenerate conditional on the instruments for the other choices. Therefore, without additional restrictions, point identification of the choice model cannot be guaranteed even for a fixed scale. On the other hand, the mapping from a unique vector of parameters to choices may be a correspondence. In this case, even when the choice model is point-identified, it provides set-valued predictions because we only observe z_{Lt} or z_{Ut} instead of $\tilde{z}_t$, with $P(z_{Lt} \leq \tilde{z}_t \leq z_{Ut}) = 1$. For example, observe that the model provides the following set of choice predictions for each individual:

$$\begin{aligned} \mathbf{D}(V, X, Z; g) = \{d \in \mathcal{T} : \exists \tilde{z}_d \in [z_{Ld}, z_{Ud}] \text{ s.t. } z_d \alpha_d + v_d \geq z_t \alpha_t + v_t, \\ \forall t \in \mathcal{T}, t \neq d, \text{ and some } \tilde{z}_t \in [z_{Lt}, z_{Ut}]\}. \end{aligned} \quad (10)$$

The model parameters can be learned by considering the capacity functionals for these sets. By connecting the method of set-valued control functions (Han and Kaido, 2026) to identification results for partially identified, revealed-preference models (e.g., Chesher, Rosen, and Smolinski, 2013), we enable identification and estimation under both circumstances.

Remark 1 *Note that the sharpness result in Theorem 3 is only with respect to $F_{e_t|V,X,Z}$, and not with respect to the full joint distribution of e . Additional restrictions on the joint distribution can further sharpen the identified sets by using the information gathered in the identified sets for each treatment.*

Remark 2 *The results from Theorems 2 and 3 can be applied to learn various parameters of interest. In particular, one can construct the set of predicted outcomes given treatment t for those with $D = d$*

as follows:

$$\{y \in \mathcal{Y} : y = m_t(X) + E_{\theta^D}[e_t|X, V] + \eta_t, V \in \text{Sel}(\mathbf{V}(d, X, Z; g)), (\theta^D, \theta^{Y(t)}) \in \Theta_I(P_0)\} \quad (11)$$

Taking Aumann expectations over the set in equation (11) and taking the supremum and infimum over the set yields the set of counterfactual expected outcomes, which can then be used to learn parameters such as the average treatment effect on the treated (ATT).

3.2 Implications of Identifying Assumptions

In practice, it is often necessary to use additional identifying assumptions, such as stochastic independence of instrumental variables, to derive meaningful implications from the available data. For instance, stochastic independence of instruments, Z , from the selection unobservables, V , has been commonly used for identification in models with selection bias (e.g., Powell, 1994; French and Taber, 2011; D'Haultfœuille and Maurel, 2013). The results under Assumptions I3 and I4 illustrate the role each plays in identifying the model parameters. In particular, Assumption I3 aids in identifying the choice model parameters, which in turn affect the identified set of outcome model parameters. Assumption I4, by contrast, applies only to the outcome model parameters. Since the identification of the outcome model builds on the identification of the choice model, we first introduce Assumption I3 and establish the corresponding results before considering Assumption I4.

Assumption I3 V is stochastically independent of Z conditional on X .

Assumption I3 implies that the joint distribution of unobserved heterogeneity V is stochastically independent of Z . This assumption is plausible when Z is randomly assigned to individuals, either by the researcher or by some natural experiment.

Corollary 2 Suppose that Assumptions R1, I1 and I3 hold. Then, the sharp identified set of θ^D , denoted Θ_I^D , from the dataset $(D, X, Z) \sim P_0$ is as follows:

$$\Theta_I^D(P_0) = \{\theta^D \in \Theta^D : P(\mathbf{D}(V, X, Z; g) \cap A \neq \emptyset | X, Z) \geq P_0(D \in A | X, Z), a.e., V \sim F_{V|X}, \forall X, \forall Z, \forall A \subseteq \mathcal{T}\}.$$

In Corollary 2, $F_{V|X, Z}$ is restricted to be a singleton set across Z by Assumption I3. This allows the instruments to only affect the choice probability through the function $g(\cdot)$ and not the unobserved components, therefore constituting the necessary condition to let the effect of $g(\cdot)$ be distinguished from the role of $F_{V|X, Z}$. As a consequence, it also enables the identification of $F_{V|X}$

from the variation in the observed probabilities according to the observed components of the utility functions.

To see this more clearly, observe that the identified set in Corollary 2 is the result of taking the essential infimum across Z for each $F_{V|X,Z}$ that generates $\mathbf{D}(V, X, Z; g)$ to obtain $T_{\theta^D}(A|X, Z)$ in Theorem 2. Unless this condition is guaranteed, there can be many different pairs of $(g(\cdot), F_{V|X,Z})$ that satisfy the inequality condition for each Z . In the absence of an additional condition that allows for comparison of these parameters across different values of Z , all such pairs can be included in the identified set.

One may wonder whether we can reduce the capacity dominance condition into a probit equation, for instance, when each of $(g(\cdot), F_{V|X,Z})$ predicts a unique choice. While the relationship itself is true, the identification strategy remains different. The usual identification argument for a multinomial probit model leverages the functional form of the normal distributions, while we verify whether the inequalities hold for all subsets of $\mathcal{T}$. Therefore, although the probit parameter estimates are always point identified, the identified set satisfying the dominance condition in Theorem 1 or Corollary 2 may be empty or contain many different forms of g . In cases where the identified set is empty under maintained assumptions, it may indicate model misspecification.

Corollary 2 showed the role of Assumption I3 in sharpening the identified set. While having a sharper set of choice model parameters generally leads to a sharper identified set of outcome model parameters, without specific restrictions on the joint distribution of e and V , the restrictions on the outcomes from the choice model are unclear (Puhani, 2000). In general, the outcome model may align with different joint relationships of e and V , leading to a one-to-many relationship between $F_{e|V,X,Z}$ for a given conditional distribution of V . Restrictions on the conditional distribution of outcome model unobservables, such as Assumption I4, can restrict the extent to which they can affect the outcome model identified set.

Assumption I4 $F_{e|V=v,X,Z}$ is known to the econometrician for every V in its support conditional on X and Z , for all $X \in \mathcal{X}$, and for all $Z \in \mathcal{Z}$.

Assumption I4 says that $F_{e|V,X,Z}$ is known to the econometrician as a singleton for each X , Z and V , establishing a one-to-one mapping from the choice model to the outcome. This assumption has been used extensively in the literature. For instance, Heckman (1979) assumed joint normality of the unobserved terms in the outcome model and the choice model, while Hay (1980) and Dubin and McFadden (1984) extended Heckman (1979) to accommodate multinomial choices while assuming a specific relationship between the outcome and the choice models. On the other hand, Dahl (2002)

proposes to estimate this semiparametrically as a function of the cumulative distribution function of a number of choice probabilities, which he estimates using bins of discrete variables, and Newey (2009) considers nonparametric estimation of the correction term constructed using a known transformation of a semiparametric choice model.

In the point identification literature, one of the key conditions for the identification of the conditional mean parameters, $m_t(X)$, is that the control functions can be varied independently from them. This requires that there is variation in Z for each X . The role of such a condition is closely related to Assumption I4. Without Assumption I4, there may be cases where features of $F_{e_t|V,X,Z}$ can be identified from the outcome model. In particular, when the conditional distribution is known only up to a finite-dimensional parameter, the variation in Z may allow the parameters of the conditional expectation of e_d to be identified separately from $m_t(\cdot)$. Having sufficient variation in Z can allow for the identification of such parameters without imposing Assumption I4, as the outcome model for each choice-generating process can be made explicit by representing the outcome model as intersection bounds, as in equation (12) using Corollary 2 of Han and Kaido (2026):

$$\sup_{z \in \mathcal{Z}} \left(E[Y|D=t, X, Z] - \sup_{V \in \mathbf{V}} E_{\theta^D}[e_d | X, V] \right) \leq m_t(X) \leq \inf_{z \in \mathcal{Z}} \left(E[Y|D=t, X, Z] - \inf_{V \in \mathbf{V}} E_{\theta^D}[e_d | X, V] \right). \quad (12)$$

These discussions highlight the importance of restrictions placed on the conditional distribution of outcome model error terms for deriving meaningful information about the model. Having a closed-form representation that can be computed easily in standard software, while incorporating the choice model unobservables V , can be useful for applications. In general, the outcome model unobserved error terms can be characterized using the Knothe-Rosenblatt transformation (Rosenblatt, 1952; Knothe, 1957), as in Han and Kaido (2026), where, for $\eta_0, \eta_1, \dots, \eta_T \stackrel{iid}{\sim} U(0, 1)$,

$$\begin{aligned} e_0 &= T_0(\eta_0; v, x, z) = F_{e_0|V,X,Z}^{-1}(\eta_0 | v, x, z), e_1 = T_1(\eta_1; e_0, v, x, z) = F_{e_1|e_0,V,X,Z}^{-1}(\eta_1 | e_0, v, x, z), \\ e_T &= T_T(\eta_T; e_0, \dots, e_{T-1}, v, x, z) = F_{e_T|e_0, \dots, e_{T-1}, V, X, Z}^{-1}(\eta_T | e_0, \dots, e_{T-1}, v, x, z). \end{aligned}$$

Lee (1983) used the maximum order statistic of a polychotomous selection model to transform the two marginal distributions of e_D and $v_D - v_{D-1}$, where v_{D-1} denotes the unobserved term entering the utility from the next-best choice. This is a specific example of the Knothe-Rosenblatt procedure, based on the assumption that the selection unobservables are independent across choices. By relaxing this assumption and considering a general multivariate distribution of V that underlies the observed choices, we can also learn about the outcome-model parameters without imposing specific restrictions

on the relationship between e and v .

4 Estimation and Inference

4.1 Estimation

In this section, we propose an estimator for computing the identified set of parameters introduced in Section 3. In particular, it consists of two steps that align with our identification results. We first obtain a consistent estimate of the set of choice model parameters, $\Theta_I^D(P_0)$. Then, using them as inputs to construct the set of correction terms, we evaluate the moment inequalities for the outcome equation. Consistency of the second-stage estimates is established by applying the results of Kaido and White (2014), who provide conditions under which the two-step estimator is a set-valued random element that is measurable in an appropriate sense and is consistent.

Because the vector of parameters used to construct the outcome model predictions includes the subvector of parameters for the choice model, one may consider estimating the parameters by evaluating the moment conditions appearing in the choice model and the outcome model simultaneously, as is common for the likelihood estimation of the point-identified model in Example 1. In general, not only is considering the outcome model parameters for each candidate choice model parameter in the parameter space highly inefficient in our setting, but it also comes with several critical drawbacks. Theoretically, the estimator suffers from poor power properties. In Section 3, we showed that any choice model can rationalize the outcome model parameters; accordingly, we established the result in Theorem 3 conditional on the availability of the identified set of choice model parameters. Without incorporating the restrictions generated by the set-identified choice model into the moment inequalities used to estimate the outcome model parameters, the objective function will often be very small (for minimization problems), even when the candidate choice model parameters are not in the identified set. Computationally, evaluating all parameters simultaneously will likely be highly burdensome due to the large dimensionality of the problem. The proposed estimator deals with these challenges.

To simplify the exposition, we maintain Assumptions R1, I1, I2, and I3 throughout this section. They are standard and are frequently used in practice to ensure that the estimated parameters yield meaningful implications.

For both steps, the identified sets are computed as contour sets of level c , defined in equation (13). We define $C_N^D(c)$ as the estimates of Θ_I^D , computed using the criterion function $Q^D(\theta)$. Similarly,

for the parameters of the outcome model, $\Theta_I^{Y(t)}$, we use $C_{N_d}^{Y(t)}$ as the estimates of $\Theta_I^{Y(t)}$, where N_d denotes the number of individuals with selected choice d .

$$C_N^D(c) := \{\theta \in \Theta : Q^D(\theta) \leq c/a_N\}, C_{N_d}^{Y(t)}(c) := \{\theta \in \Theta : Q^{Y(t)}(\theta) \leq c/a_{N_d}\} \quad (13)$$

a_N and a_{N_d} are normalizing sequences, corresponding to sample sizes N and N_d . $Q^D(\theta)$ and $Q^{Y(t)}(\theta)$ are objective functions.

We supplement the maintained assumptions with the following regularity conditions to guarantee consistency of our estimators. Let $F(A|X, Z) = [P(d \in A|X, Z), A \subseteq \mathcal{T}/\emptyset]$ denote the vector of conditional choice probabilities, and let $\Sigma_{X,Z} \equiv \text{diag}(F(A|X, Z)) - F(A|X, Z)F(A|X, Z)'$

Assumption E1 1. The data (Y, D, X, Z) are i.i.d.

2. $\mathcal{X}$ and $\mathcal{Z}$ consist of a finite number of discrete elements; the vector of support points for each $\mathcal{X}$ and $\mathcal{Z}$ has full rank.

3. The parameter space Θ is a compact subset of a finite-dimensional Euclidean space.

4. $\Sigma_{X,Z}$ is nonsingular with $a < \|\Sigma_{X,Z}\| < b$, $X, Z - a.s.$ for some constants $0 < a < b < \infty$, where $\|\Sigma_{X,Z}\|$ is a matrix norm for $\Sigma_{X,Z}$ compatible with a Euclidean norm.

5. $E[|\eta_t|^2] < \infty$, $\forall V \in \mathbf{V}(D, X, Z; g)$ and for all $t \in \mathcal{T}$.

6. The moment functions $E[m^c(A, X, Z; \theta)|X, Z]$ and $E[m^{Y(t)}(Y, X, Z; \theta)|X, Z]$ are nondegenerate and have continuous Jacobian at each $\theta \in \Theta$ and for all $X \in \mathcal{X}, Z \in \mathcal{Z}$.

The conditions in Assumption E1 are standard for estimation in the partial identification literature. Assumption E1 - 2 accounts for discretization of the support points used to aggregate across conditional moment conditions, while ruling out redundant parameterization. The discreteness can be relaxed by using nonparametric methods to deal with conditional moments, such as the tools introduced in Andrews and Shi (2013). Assumption E1-3 restricts our estimator to be parametric. Flexible possibilities of distributions of unobserved terms can be incorporated by using finite mixtures of distribution functions known up to finite-dimensional parameters, and including these parameters as a part of the parameter vector to be estimated.

We denote the total number of (X, Z) cells using $\kappa_{XZ} = |\mathcal{X} \times \mathcal{Z}|$. Let w_{xz} denote the weights for each cell of (x, z) , that accounts for the number of individuals with the same values of X and Z , which sum up to 1 across different values of X and Z .

For the observed conditional probabilities, we use a standard frequency estimator:

$$\hat{P}(d \in A|X, Z) = \frac{\sum_{i=1}^N \mathbf{1}(d_i \in A) \mathbf{1}(x_i = X, z_i = Z)}{\sum_{i=1}^N \mathbf{1}(x_i = X, z_i = Z)}.$$

The capacity functionals $\hat{T}_{\theta^D}(\mathbf{D}(V, X, Z; g) \cap A \neq \emptyset | X, Z)$ are simulated for each candidate parameter vector. To do this, we first generate S i.i.d. draws of V from $F_{V|X,Z}$, computing set-valued prediction $\mathbf{D}(V, X, Z; \theta)$ using these simulated draws for each X, Z . Then, we aggregate them by computing $\frac{1}{S} \sum_{s=1}^S \mathbf{1}(\mathbf{D}(V_s, X, Z; \theta) \cap A \neq \emptyset)$ for each subset $A \subseteq \mathcal{T}$. We assume that the number of simulation draws grows to infinity at a rate $S = O(N^{2+\alpha})$, where $\alpha > 0$. This construction guarantees that they do not affect the asymptotic distribution (Ciliberto and Tamer, 2009).

Denote the vector of simulated quantities by $Q(X, Z; \theta^D)$, where each row corresponds to a hitting functional evaluated on the same subset used to obtain the observed probability in the corresponding row of $F(A|X, Z)$.

The objective function $Q(\theta^D)$, displayed in equation (14), is used to compute the choice model identified set. It can be seen as a classic inequality estimator with each non-empty subset of $\mathcal{T}$ as a moment inequality constraint.

$$\begin{aligned} Q(\theta^D) &= \int \|(F(A|X, Z) - Q(X, Z; \theta^D))_+\|^2 dF_{x,z} \\ &\simeq \sum_{x,z} w_{x,z} \|(F(A|X, Z) - Q(X, Z; \theta^D))_+\|^2 \end{aligned} \tag{14}$$

We use the sample equivalent of $Q(\theta^D)$, $Q_N(\theta^D)$, in practice:

$$Q_N(\hat{\theta}^D) = \sum_{i=1}^N \|\hat{F}(A|X, Z) - Q(X, Z; \theta^D)\|_+^2 w_{xz}.$$

Lemma 1 shows that the outlined procedure provides a consistent estimator of $\Theta_I^D(P_0)$.

Lemma 1 *Suppose that Assumptions R1, I1, E1-1,2,3,4 and 6 are satisfied, and the number of simulation draws used to compute $\hat{T}_{\theta^D}(A|X, Z)$ grows to infinity at the rate $S = O(N^{2+\alpha})$, $\alpha > 0$. Then, $\Theta_I^D \subseteq \hat{\Theta}_I^D$ w.p.a 1 and $d_H(\hat{\Theta}_I^D, \Theta_I^D) = o_p(1)$.*

For the outcome model parameters, which are the key parameters of our interest, we first define a residual set $\mathbf{R}^{Y(t)}(Y, X, Z; \theta)$:

$$\mathbf{R}^{Y(t)}(Y, X, Z; \theta) = \text{cl}\{r : r = Y - m_t(X) - E_{\theta^D}[e_t|X, V] : V \in \text{Sel}(\mathbf{V}(D = t, X, Z; g))\}.$$

$\mathbf{R}^{Y(t)}(Y, X, Z; \theta)$ is a random closed set in $\mathbb{R}$, from the same logic applied to the random closed set $\mathbf{Y}(D, X, Z; \theta)$. Observe that we can characterize the identified set of $Y(t)$ parameters using the support function of $\mathbf{R}^{Y(t)}(Y, X, Z; \theta)$, as follows:

$$\left\{ \theta \in \Theta_I^{Y(t)} : E[h(\mathbf{R}^{Y(t)}(Y, X, Z; \theta), u) \mid D = t, X, Z] \leq 0, \|u\| \leq 1 \right\}. \quad (15)$$

The specification in (15) again makes use of the equivalence of the support function of Aumann expectations and the expected value of the support function of a random closed set. This provides the following moment conditions for each (X, Z) pair:

$$\begin{aligned} \hat{r}1^{Y(t)}(Y, X, Z; \theta) &= ((X\beta_t - (\hat{E}[Y|D = t, X, Z] - \inf_{V \in \mathbf{V}} E_{\theta^D}[e_t|X, V])), \\ \hat{r}2^{Y(t)}(Y, X, Z; \theta) &= ((\hat{E}[Y|D = t, X, Z] - \sup_{V \in \mathbf{V}} E_{\theta^D}[e_t|X, V]) - X\beta_t). \end{aligned} \quad (16)$$

Aggregating the sample residuals from equation (16) provides the following sample moments:

$$\hat{m}(Y, X, Z; \theta^{Y(t)}) = \frac{1}{\kappa_{XZ}} \sum_{\kappa_{XZ}} \begin{bmatrix} \hat{r}1^{Y(t)} \\ \hat{r}2^{Y(t)} \end{bmatrix},$$

which can be used to construct the following test statistic:

$$\hat{Q}^{Y(t)}(\theta) = \left\| \left| \hat{m}(Y, X, Z; \theta^{Y(t)}) \right|_+ \right\|^2,$$

where $\|\cdot\|$ is the Euclidean norm.

Observe that once we obtain $\hat{\Theta}_I^D$, we can conveniently plug in the estimated identified set of choice model parameters and include the correction terms $E_{\theta^D}[e_t|X, V]$, $\theta^D \in \Theta_I^D(P_0)$, in the moment function to impose restrictions from the data generating process for the outcome model parameters. That will allow us to forego estimating $(g(\cdot), F_{V|X, Z})$ jointly with the outcome model parameters to ensure that the correction terms are created using all choice model parameters that align with the observed data. That is, we can consider $\mathbf{R}^{Y(t)}(\theta^{Y(t)}, \hat{\theta}^D)$, where $\hat{\theta}^D \in \hat{\Theta}_I^D(P_0)$, to incorporate the a priori restrictions from the choice model when estimating the outcome model parameters.

In cases such as Example 2, each pair $(g(\cdot), F_{V|X, Z})$ predicts a unique correction term for each observation. On the other hand, in Example 3, multiple predictions of correction terms may arise according to the specific value of the instruments within the interval under consideration, even for a single pair of the choice model parameters. In both cases, the outcome model parameters are easy

to estimate using standard computing software once the choice model estimates have been obtained.

Theorem 4 *Suppose that the conditions of Lemma 1, Assumptions I2, and E1 5-6 hold. Then $d_H(\hat{\Theta}^{Y(t)}, \Theta_I^{Y(t)}) = o_p(1)$.*

The two-step procedure provides sizable gains in computational efficiency. This can be especially valuable in the set estimation setting, which often entails heavy computational burden. Notably, the choice model identified set can be computed once and reused across all outcomes, provided the same selection rule applies to each. This is a valuable feature for applied researchers who often wish to examine the effects of covariates on multiple outcomes while controlling for selection bias. Moreover, when the researcher is interested only in specific subgroups among self-selected choices, evaluating joint probabilities of the outcome model and the choice model would require processing all observations across every selection group. Our method avoids this, thereby reducing both memory usage and computation time. Finally, our procedure for the outcome parameters does not require dividing the continuous outcomes into discrete cells, which not only adds to the computational burden, particularly when conditioning on a large number of covariates, but also tends to obscure model interpretability.

4.2 Inference

Most of the time, the parameters of interest for empirical researchers are those of the outcome model, while those of the choice model only serve the purpose of constructing the correction terms. In this context, it can be particularly useful to obtain the confidence set $C_{N_d}(c(\alpha))$ for the outcome model parameters associated with treatment d that covers the identified set $\Theta_I^{Y(d)}$ with α probability. The most common way of conducting inference for set-identified parameters is to construct the confidence set as a collection of parameters θ that satisfy $a_N Q_N(\theta) \leq c_N(\alpha)$: that is, by constructing the level sets in equation (13), using a particular critical value, $c_N(\alpha)$. A well-known way of computing the critical value $c_N(\alpha)$, which constitutes the key step for inference, is to obtain the α -quantile of $\{\hat{C}_{j,b} = \sup_{\theta \in C_N(c_0)} a_b Q_{j,b}(\theta), j = 1, 2, \dots, B_{N_d}\}$, where $Q_{j,b}$ denotes the criterion function evaluated using the j -th subsample of size b , and c_0 is an initial value that is chosen efficiently¹. For the estimated identified sets from our first-stage computation of the choice model, Theorem 5 provides validity of such subsampling-based critical values for obtaining the confidence sets.

Theorem 5 *Suppose that the conditions of Lemma 1 are satisfied, $b \rightarrow \infty$, and $b/N \rightarrow 0$ at polynomial rates as $N \rightarrow \infty$. Let $\alpha \in (0, 1)$ denote the desired coverage level. Then, the following are true*

¹See Chernozhukov, Hong, and Tamer (2007) for additional details

for $C_N^D(\hat{c}(\alpha))$ generated using the sample critical value $\hat{c}(\alpha)$ from a generic subsampling algorithm for Θ_I^D and an estimated identified set $C_N^D(\cdot)$:

1. $\hat{c}(\alpha) \rightarrow c(\alpha) := \inf\{c : P(C^D \leq c \leq \alpha)\}$,
2. $\lim_{N \rightarrow \infty} P\{\Theta_I^D \subseteq C_N^D(\hat{c}(\alpha))\} = \alpha$ if $c(\alpha) > 0$, and $\lim_{N \rightarrow \infty} \inf P(\Theta_I^D \subseteq C_N^D(\hat{c}(\alpha))) \geq \alpha$ if $c(\alpha) = 0$.

Subsampling-based inference on the estimated identified set of outcome model parameters, on the other hand, can be particularly complicated. In our two-step procedure, the critical values depend on the first-stage identified set Θ_I^D , which is usually estimated a priori from the data. Although inference for two-step procedures with point-identified first-stage parameters has been discussed in Kaido and White (2014), much less is known about inference when the first-stage input is itself set identified.

To be specific, for a treatment of interest $d \in \mathcal{T}$, suppose that we have a consistent estimate of $\Theta_I^{Y(d)}$, that is, $\hat{\Theta}_I^{Y(d)}$, which can be guaranteed by Theorem 4. Moreover, assume that we know a priori the true first-stage identified set, Θ_I^D . In that counterfactual scenario, it is straightforward to establish the validity of the α -level confidence set obtained in a similar way as the confidence set for the choice model parameters.

In practice, the prior knowledge of Θ_I^D is often not guaranteed. Instead, the researcher would have the estimated $\hat{\Theta}_I^D$ on hand. Because the bootstrap procedure depends on the $\hat{\Theta}_I^D$ -dependent initial estimate of $\Theta_I^{Y(d)}$, we cannot ensure that the critical values are obtained by evaluating all parameters in the true identified set, $\Theta_I^{Y(d)}$. Moreover, it is challenging to establish the conditions for which the $\hat{\Theta}_I^D$ -dependent objective functions evaluated by the bootstrap subsamples, $\sup_{\theta \in C_N(c_0)} a_b Q_{j,b}(\theta)$, will provide a valid estimate of the critical values covering $\hat{\Theta}_I^D$ -dependent $\hat{\Theta}_I^{Y(t)}(\hat{\Theta}_I^D)$ with a given probability, accounting for the first-stage estimation uncertainty. This would require characterizing the asymptotic behavior of the second-stage estimator, which depends on two non-independent samples of different sizes: one used for first-stage estimation and the other for which the outcome of interest is observed. Tackling this challenge is beyond the scope of this paper, as our primary contribution is the identification analysis, development of the estimation procedure, and the empirical application. To this end, we leave this as a possibility for future research.

Still, it can be useful to have a tool on hand that accounts for the added uncertainty from first-stage estimation and is feasible to compute in standard computing environments. Building on the fact that first-stage identified sets have the confidence set property and that we can apply a generic subsampling procedure to compute the critical values, we provide a bootstrap procedure for the second-stage identified sets. It foregoes the need to estimate the first-stage identified sets as many

times as the number of bootstrap subsamples by using a confidence set of choice model parameters containing the true identified set, Θ_I^D , with at least $1 - \beta$ probability in place of the set of choice model parameters. We denote this $C_N^D(\hat{c}^D(1 - \beta))$.

Consider the following inverse quantile function:

$$J_t(u, \Theta^D, \theta^Y) = \inf \left\{ a \in \mathbf{R}^{Y(t)} : P \left(\sup_{\theta^Y \in \hat{\Theta}^Y} Q_{N_d}(\theta^Y(\Theta^D)) \leq a \right) \geq u \right\},$$

where we use notation $\theta^Y(\Theta^D)$ to make explicit its dependence on Θ^D . We propose the following algorithm:

Algorithm for Obtaining the confidence set of outcome model parameters:

1. For each treatment $t \in \mathcal{T}$ of interest, subsample B_t groups of size N_{B_t} .
2. Compute the test statistic for each outcome model parameter in $\hat{\Theta}_I^{Y(t)}$, $\sup_{\theta^Y \in \hat{\Theta}^Y} Q_{N_{B_t}}^b(\hat{\Theta}^Y)$, $b = 1, 2, \dots, B_t$, for each $\hat{\Theta}_I^{Y(t)}$ computed using $C_N^D(\hat{c}^D(1 - \beta))$.
3. Take the $1 - (\alpha - \beta)$ quantile critical value across B_t values, $J_t(1 - (\alpha - \beta), \theta^D, \theta^Y)$; report this value as $\hat{c}^Y(1 - \alpha)$.

In the above algorithm, we control for the extra β - overrejection probability embedded in the first stage by being more conservative in the bootstrap procedure in the second stage, while using the confidence set of choice model parameters, $C_N^D(\hat{c}^D(1 - \beta))$, as inputs to account for the possible small-sample bias in the first-stage estimates entering the model. The following rationale justifies the procedure. Observe that the correction terms $E_{\theta^D}[e_d|X, V]$ are images of parameters $\theta^D \in \Theta_I^D$, and Θ_I^D maps all selectable choice-model unobservables, identified by the observed data, to the outcome model. Because images of sets preserve set inclusion, $\mathbb{E}[e_d|X, V] = \{E_{\theta^D}[e_d|X, V], \theta^D \in C(\hat{c})\}$, where $P(\Theta_I^D \subseteq C(\hat{c})) = \alpha$, will also include $\{E_{\theta^D}[e_d|X, V], \theta^D \in \Theta_I^D\}$ with at least $1 - \beta$ probability. Then, in the absence of finite-sample bias specific to the sample used to compute the outcome model, the set of predicted outcomes, $\mathbf{Y}(D, X, Z; \theta)$ obtained using the $1 - \beta$ level confidence set for the choice model parameters would include its population-level counterpart with the same confidence.

In practice, additional finite-sample bias specific to the second-stage estimation is inevitable, and we do not know how it interacts with the finite-sample bias caused by the sample used for estimating the first-stage identified set. Instead, leveraging the fact that the expectation operator and the support function operator are both additive, we assume that the $1 - \beta$ level confidence

set that treats the estimated set of choice-model parameters as if they are the true choice-model identified set will not include the true outcome-model parameters due to the first-stage finite-sample bias at most β percent, so when we use it to obtain the $1 - \alpha$ confidence set for the outcome model parameter identified set, it will cover the true outcome model identified set with at least $1 - \alpha - \beta$ probability. In that case, by taking the $1 - \alpha + \beta$ quantile across the subsample test statistic rather than the $1 - \alpha$ quantile, we can fully take into account the added uncertainty from the estimated first stage. Our back-of-the-envelope simulations suggest that they provide good coverage, offering a stepping stone toward future research on consistent finite-sample inference in settings such as ours.

4.3 Computational Aspects

Existing research has highlighted the computational challenges of estimating identified sets under partial identification. However, the computational requirements of an estimator depend heavily on the specific model under consideration. The model studied in this paper differs substantially from those considered in the existing set estimation literature, most of which involve variants of choice models with multiple equilibria. We therefore discuss the computational tasks associated with our model, the methods considered and proposed in this paper, and practical approaches for addressing these tasks.

Our proposed procedure requires computing the identified set of choice model parameters in the first stage. In general, estimating identified sets requires searching for the parameter values over a grid of candidate values. In any grid-search procedure, computational cost increases exponentially with the dimensionality of the model parameters. Our procedure tackles this computational challenge by computing the identified sets for the choice and outcome models separately, dramatically reducing the computational burden relative to dealing with all model parameters simultaneously. For example, if a researcher were to evaluate the objective function over a closed interval for each parameter split into 10 equal points, and estimate 2 choice model parameters and 2 outcome model parameters, it would involve 10^4 evaluations when done simultaneously, relative to $10^2 \times 2$ evaluations when done separately. Regardless, because estimating the choice model identified set requires searching over a grid for the areas corresponding to the full parameter vector, using simulated error terms that reflect the candidate distribution for choice model error terms, it inevitably requires more computational resources than classical point-identified models.

Still, the proposed method is practical to implement using commonly available computing resources for problems with a moderate number of parameters. The baseline choice model in the

simulation of Subsection 5.1 required approximately three hours to evaluate 90,720 grid points, computed as four concurrent jobs of four cores each (16 cores in total) on AMD EPYC 7763 hardware (2.45 GHz, 32 GB RAM)². This is comparable to a current high-end laptop or entry-level workstation.

Alternatives to the grid search procedure include global optimization algorithms, such as simulated annealing (see, for instance, Ciliberto and Tamer, 2009) or differential evolution, which we use for our empirical application. In those cases, the criteria for updating the status quo using the candidate parameter vector are used as the criteria for acceptance into the identified set. An advantage of using such algorithms in place of grid search is that it can be a more efficient way to explore the parameter region compared to grid search. A downside may be that it is not as straightforward as grid search in ensuring sufficient coverage of the identified set, as it is not guaranteed that the proposed iterations of objective function evaluations will include evaluations of all parameters that could be included in the identified set. Moreover, most of them involve judicious choice of tuning parameters to guarantee optimal performance, which may be challenging to achieve in practice. Although specific guidance would differ from case to case, these algorithms could be a convenient tool when the parameter dimension is very large or when evaluating the objective function at each parameter grid point is infeasible or expensive.

Once the choice model identified set has been computed, the researcher should construct the set of control functions, or correction terms, to be used in the outcome model criterion function. This can become computationally burdensome if the identified set of choice model parameters is large. Some models have analytic expressions for the expected value of truncated random variables that can be conveniently calculated, such as the well-known inverse Mills ratio. However, in complicated, multinomial choice models with flexible distributional features, such as the models the current paper focuses on, closed-form formulations are often not readily available, and the control functions need to be computed via numerical integration and simulation.

Finally, the outcome model parameters need to be computed. This step can be done considerably faster and with minimal computational burden relative to the preceding steps. The convex set of outcome model predictions, together with the inequality constraints, implies that the researcher can simply compute the minimum and the maximum of the control functions for each covariate across the choice model identified set to evaluate the inequality constraints and obtain the objective function.

²Specifically, the computation was carried out on the vm-small queue of Lonestar6 at the Texas Advanced Computing Center, University of Texas at Austin. The same computation could be completed considerably faster on other queues, at higher service unit (SU) cost and typically longer queue wait times. We report the vm-small timing because it provides the environment most similar to commonly available hardware.

Unlike the choice model parameters, the outcome model parameters are evaluated separately for each choice rather than simultaneously across all choices, which helps keep the dimensionality of the problem within a reasonable range.

For researchers using grid search to obtain the identified set for the outcome-model parameters, it can be especially useful to have an approximate range that contains the identified set. One way to construct such a range is to first form a point-valued control function for each parameter value in the choice-model identified set. One can then estimate the outcome-model parameter using each of these control functions, as in the standard control-function approach. The resulting collection of parameter estimates can serve as starting points for the grid search.

The justification for this approach follows from the convexification property of Aumann expectations (Molinari, 2020), which implies that the set-valued correction term, $\mathbb{E}[e_t|V, X, Z] = cl\{\int e_t dF_{e|V, X, Z}(e | V, X, Z), V \in Sel(\mathbf{V}(D, X, Z; g))\}$, is equal to the expectation of the closed convex hull of the set of e conditional on V, X , and Z . Since the convex hull contains every element of the original set, each point-valued control function constructed using an element of the choice-model identified set is an element of the set-valued correction term. Consequently, the outcome-model parameters obtained from these point-valued control functions are included in the identified set for the outcome-model parameters. The range spanned by these parameter estimates may, however, be considerably narrower than the true identified set. Thus, the grid should extend over a reasonably wide range surrounding this region to ensure that the identified set is adequately covered.

5 Simulation

This section provides results from Monte Carlo simulations illustrating the proposed estimation procedure. In Subsection 5.1, we estimate the parameters of a model designed similarly to Example 2. In Subsection 5.2, we estimate the parameters of a model with interval-valued observations, as in Example 3.

5.1 Selection Models with Choice-Specific Instrumental Variables

5.1.1 Data-Generating Process

We consider a selection model with four alternatives, $j \in \{0, 1, 2, 3\}$, where alternative 0 (getting no treatment) is the baseline. The variables in the model, (x, z_1, z_2) , are drawn from the following

distributions:

$$x \sim \mathcal{N}(5, 3), z_1 \sim \mathcal{N}(-1, \sigma), z_2 \sim \mathcal{N}(1, \sigma),$$

while varying σ to examine the power of utility shifters with large support. Unobserved determinants of choices (v_0, v_1, v_2, v_3) are drawn from a mixture of two multivariate normal distributions, $u_j \sim N(\mu_j, \Sigma_j), j = 1, 2$, where $\mu_1 = (1, 3, 3, 1)$, and $\mu_2 = (5, -2, -2, 5)$. Σ_j are equicorrelated with $\rho_1 = 0.8$ and $\rho_2 = 0.1$, and all variables have unit variance. We normalized V to $\bar{v}$ by subtracting v_0 from all other elements for each individual.

Latent utilities are specified as follows:

$$\begin{aligned} U_0 &= 0, U_1 = x\beta_1 + \alpha_1 z_1 + \bar{v}_1, \\ U_2 &= x\beta_2 + \alpha_2 z_2 + \bar{v}_2, U_3 = x\beta_3 + \alpha_1 z_1 + \alpha_2 z_2 + \bar{v}_3, \end{aligned}$$

and the observed treatment is $D = \arg \max_{j \in \{0, 1, 2, 3\}} U_j$. Potential outcomes are generated by the following equation:

$$Y(t) = x\theta_t + e_t, e_t = v_t + \varepsilon, \varepsilon \sim \mathcal{N}(0, 0.1), t = 0, 1, 2, 3, \quad (17)$$

The observed outcome is $Y = Y(D)$.

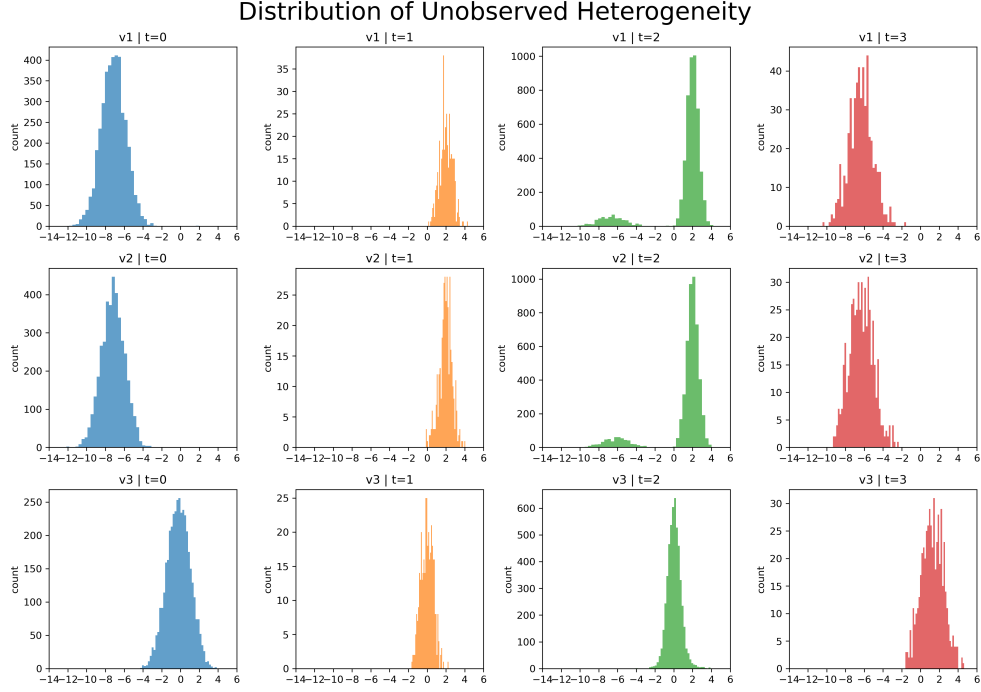
We generate a large sample of $N = 20,000$. Sampling imprecision should be a minor concern with such a large sample size, so the choice model parameter estimates should approximate their identification region, i.e., the “identified set”.

Figure 1 displays the distribution of unobserved heterogeneity by treatment group. Clearly, the utility-maximizing choices closely align with the choice where they maximize their unobserved utility components. For example, the unobserved components v_2 for those who selected treatment 3 are on average much smaller than the majority of those who selected treatment 2 or treatment 1. If we assume that the true data generating process is known, we can obtain the estimates from the third row, which is very close to the true parameters. In reality, because our choice model is not nonparametrically identified, we cannot obtain these quantities.

5.1.2 Implementation

To approximate the conditional choice probabilities, we discretized all variables into equally sized intervals. We divided x into 35 bins, whereas z_1 and z_2 were both divided into 20 bins. The bin widths were not very wide, ranging from approximately 0.714 (for x) to 0.804 (for z_1). For our

Figure 1: Distribution of Unobserved Heterogeneity by Treatment Choice



“baseline” model, we used $\sigma = 2$, and for our “large support” model, we used $\sigma = 4$. As a result, the supports of z_1 and z_2 were $[-8.745, 6.534]$ and $[-6.487, 7.690]$, respectively, in our baseline model, whereas for our large support model, they were $[-17.170, 16.905]$ and $[-15.292, 14.718]$, respectively.

To estimate the identified sets for both the choice model and outcome model parameters, we performed a grid search over Cartesian products of one-dimensional grids defined on closed intervals. For the choice model parameters, we searched over the interval $[-1, 1]$, partitioned into 8 equally spaced bins. For the outcome model parameters, we searched over the interval $[\theta_t - 5, \theta_t + 5]$, partitioned into 401 equally spaced points. Our estimator for both the first and the second stage requires choosing an appropriate value of c for constructing the level sets $C_N^D(c)$ and $C_{N_d}^{Y(t)}(c)$. For all our estimates presented throughout Sections 5 and 6, we used $1/\sqrt{N}$ and $1/\sqrt{N_d}$ for each of the choice model and the outcome model estimates.

5.1.3 Results

Before presenting our main results, to establish a benchmark, we compare estimates from several alternative models. Table 1 reports the parameter estimates from an ordinary least squares (OLS) model, as well as estimates obtained using the “true” control variables and the “default” control variables. The “true” control variables correspond to those that would be available if the choice model

were point identified, whereas the “default” control variables are constructed under the conventional assumption that the choice model unobservables are normally distributed with unit variance and are uncorrelated across choices.

Table 1: Baseline Outcomes

	θ_0	θ_1	θ_2	θ_3
True	5	3	−5	−1
OLS	5.844	1.758	−4.680	0.049
True Control Var	5.000	3.000	−5.000	−1.000
Default Control Var	6.017	1.940	−4.684	−1.543
N_d	8,191	532	10,467	810

As expected, the OLS estimates show substantial selection bias, while the estimates obtained using the “true” control variables recover the data-generating parameters almost perfectly. Given our data conditions, the latter estimates are likely unattainable. The estimates based on the “default” control variables, which use a choice model specification commonly available in standard statistical software, are generally closer to the true parameter values than the OLS estimates. Nevertheless, they remain biased relative to the one reflecting the true data-generating process. In Table 7 in

Table 2: Outcome Parameter Estimates

Parameter	$\theta_0 = 5$	$\theta_1 = 3$	$\theta_2 = -5$	$\theta_3 = -1$
$\hat{\theta}_{t,Baseline}$	[4.600, 5.325]	[2.775, 3.900]	[−5.175, −4.850]	[−3.000, −0.825]
N_d	8,191	532	10,467	810
$\hat{\theta}_{t,Largesupp}$	[4.600, 5.325]	[2.8, 3.875]	[−5.175, −4.850]	[−2.975, −0.825]
N_d	7,837	572	10,553	1,038

Appendix A, we display the identified set of choice model parameters, obtained using the procedure outlined in Section 4. We see that V can be mapped to multiple distributions with various mixture weight distributions. In line with the findings from point identification literature that relies on instruments with large support, the identified set of parameters is narrower when the instruments have larger support and more variation. Because the instruments enter the outcome model through the set-valued control function that reflects the choice model identified sets, we can expect the instruments with large support to provide more identifying power relative to those with more limited support.

Even with a wide range of processes that may underlie the observed choices and the outcomes conditional on them, our data, together with the modeling assumptions, provide meaningful informa-

tion about the model parameters of interest. Table 2 reports the outcome parameter identified sets, obtained using the set-valued control functions constructed using our first-stage identified set. Overall, the restrictions implied by the choice model are informative about our parameters of interest. In particular, the identified set for θ_1 excludes both the OLS estimate and the “default control variable” estimate. This finding highlights the usefulness of our approach; by incorporating all selection mechanisms that may underlie the data, the control variables can generate identifying restrictions that are always reflective of reality, which may have been missed by conventional estimators, which often leverage auxiliary assumptions when the data on hand are insufficient to pinpoint the data generating process.

5.2 Selection Models with Interval-Censored Instrumental Variables

We consider a trinomial choice model with three alternatives $j \in \{0, 1, 2\}$, where alternative 0 is the baseline. Observed covariates are drawn as

$$x \sim \mathcal{N}(-5, 3), \quad z_1 \sim \mathcal{N}(2, 1), \quad z_2 \sim \mathcal{N}(1, 1).$$

Although the assigned instruments z_1 and z_2 are continuous for each individual, in our observed dataset, they are interval-valued. Each observation is assigned to one of three bins according to the empirical quantile of its true value in the observed dataset, using the 25th and 50th percentiles as cutpoints. The observed data therefore include $[z_1^\ell, z_1^u]$ and $[z_2^\ell, z_2^u]$ rather than the true values, as in Example 3.

Unobserved heterogeneity (v_0, v_1, v_2) is drawn from a mixture of two multivariate normal distributions with weights 0.5, for each of $U_1 \sim N(-3 \cdot \mathbf{1}_3, \Sigma_1)$ and $U_2 \sim N(3 \cdot \mathbf{1}_3, \Sigma_2)$, where Σ_1 is equicorrelated with $\rho_1 = 0.3$ and variance 1. Similarly, Σ_2 is equicorrelated with $\rho_2 = -0.3$.

We normalized the choice unobservables to $\bar{v}_t = v_t - v_0$, and the outcome unobservables retained the undifferenced draws v_t . Therefore, latent utilities are

$$U_0 = 0, \quad U_1 = \alpha_1 z_1 + \alpha_2 x - \bar{v}_1, \quad U_2 = \alpha_3 z_2 + \alpha_4 x - \bar{v}_2,$$

and the observed treatment is $D = \arg \max_{j \in \{0, 1, 2\}} U_j$. Potential outcomes are

$$Y(t) = x\theta_t + v_t, \quad t = 0, 1, 2,$$

and the observed outcome is $Y = Y(D)$.

5.2.1 Implementation

We discretized all intervals into 500 equally spaced points to simulate the capacity functionals, where the choice model predictions are generated as in equation (10). To compute the conditional probabilities, x was divided into 15 points.

Similarly to the procedure from Subsection 5.1, to estimate the identified sets for both the choice model and outcome model parameters, we performed a grid search over Cartesian products of one-dimensional grids defined on closed intervals. For the choice model parameters, we searched the true value ± 1 in steps of 0.25, restricted to $[-1, 1]$. The mixture weights were searched over $\{0, 0.25, 0.5, 0.75, 1\}$. For the outcome model parameters, we searched over the interval $[\theta_t - 2, \theta_t + 2]$, partitioned into 401 equally spaced points.

5.2.2 Results

In Table 8 of Appendix A, we display our estimates from the selection model. Similarly to the previous section, the data can map to various selection processes. Because the interval-censored instruments lead to a set of control functions even for a point-identified choice model, we can expect the outcome model parameters to be relatively wide.

Table 3: Outcome Parameter Estimates

	θ_0	θ_1	θ_2
True	1.5	0.5	1
$\hat{\theta}_t$	[1.08, 2.38]	[-1.5, 2.5]	[-0.71, 3]
N	18,328	583	1,089

Table 3 displays the estimates of the parameters of our outcome model. The bounds on the parameters are wide; notably, the identified set for θ_1 encompasses the entire search space, partly due to the small sample size. Still, they always include the true parameter. To date, there is no tool for incorporating interval-censored instruments to identify parameters in self-selection models. In practice, it is common to assume that the choice predictions are point-identified. Our simulation results highlight the perils of such restrictions, while the tools proposed in previous sections enable researchers to verify the true identifying power of the data at hand when such assumptions are relaxed.

6 Empirical Application

In this section, we study how participation in Texas Direct, a major policy initiative designed to facilitate seamless transfer degree pathways between two- and four-year institutions in Texas, alters students’ academic trajectories following transfer from a community college to a four-year institution (vertical transfer). Specifically, focusing on the role of its two key components, the Texas Core Curriculum (TCC) and the Texas Field of Study Curriculum (FOSC), we ask whether completion of a Texas Direct curricula component leads to better outcomes for students. We evaluate two academic measures as outcomes: the total number of courses with lost credits and bachelor’s degree attainment. For degree attainment, we track students over two overlapping time horizons— within 3 years of transfer and within 6 years of initial college enrollment.

Different transfer policies play a central role in shaping how students’ pre-transfer credits affect their post-transfer trajectories. To evaluate whether Texas Direct is reaching its intended goal, it is crucial to study its causal effects by capturing how the policy would alter a student’s outcomes if we assigned a randomly selected student with a particular number of pre-transfer credits to each curriculum completion status. Learning this effect would let us disentangle the policy effect from the effects of unobserved characteristics of students who chose to complete those curricula, such as academic preparedness, knowledge of transfer policies, and socioeconomic resources. Because students’ outcomes are only observed conditional on their completion of Texas Direct curricula, and students who complete any of them are likely different from students who transferred without completing them, such causal effects are challenging to study in the absence of a randomized experiment.

One key reason the observed effects of the policy likely include self-selection is that choosing how many courses to take and which ones to take before transfer requires students to navigate complex transfer policies, including local articulation agreements and institution-specific transfer requirements. Many community colleges provide limited advising and support (Bailey et al., 2017; Schudde, Bradley, and Absher, 2020), often causing students to rely not only on institutional guidance but also on family members and peer networks for information (Jabbar et al., 2022). Differences in their access to information, academic goals, and socioeconomic resources create substantial heterogeneity in pre-transfer coursetaking decisions. Consequently, observed post-transfer outcomes reflect not only the effect of pre-transfer course enrollment, but also the unobserved heterogeneity that caused it. Unfortunately, for each student, we observe outcomes only conditional on having taken specific sets of courses for each curriculum; outcomes under alternative curricula remain unobserved.

Ideally, we would learn from the data about the unobserved heterogeneity in students’ preferences

that shapes their decisions to complete different curriculum components, without imposing restrictive assumptions on students’ decision-making processes. As one might expect, achieving this requires sufficiently informative variation in the data. In particular, it requires instrumental variables that either uniquely shift the utility of a particular choice, thereby providing choice-specific variation that can be used to isolate the determinants of that choice, or have infinite support, allowing the instrument to take arbitrarily large values and, at extreme values, effectively induce students to select a given choice with probability approaching one. In our setting, students may complete both the TCC and the FOSC, so no obvious natural instrument uniquely targets either completion choice. Nor do we have a variable with sufficiently large support to induce students to complete either, or both, curricula with probability approaching one. Our method is therefore especially valuable in this context, as it extracts information about unobserved heterogeneity from observed choices without requiring the data to contain either form of instrumental variation. We leverage the availability and flexibility of TCC and FOSC courses for each student as instrumental variables, and use the identified set of choice model parameters to account for the likely lack of point identification. Then, we consider the identified sets of outcome model parameters for each curriculum completion status— which measures how the pre-transfer academic attainment translates to post-transfer outcomes conditional on the completion of each curriculum for the population of transfer students — corresponding to the identified set of unobserved heterogeneity underlying the dataset. Comparing these outcome model parameters across different “treatment” or completion statuses provides information about how the effect of completing one or both of these curricula compares. For this purpose, it is convenient to fix the terms used to refer to the completion of each, both, or neither of the curricula. Accordingly, students who complete neither curriculum prior to transfer are referred to as the “Neither” group, while those who complete both curricula are referred to as the “Both” group. Students who complete only one of the two curricula are classified as either the “FOSC only” or “TCC only” group, depending on which curriculum they complete.

6.1 Policy Context

Community colleges play a vital role in providing access to higher education for students pursuing a bachelor’s degree. Fifty community college districts in Texas enrolled about 700,000 students in 2017, or 46 percent of all Texas higher education students, and 34 percent of four-year college graduates had transferred at least 30 credit hours from a community college that same year (Texas Comptroller of Public Accounts, n.d.). Credit portability is one of the main challenges that students

aspiring to transfer from community colleges to four-year institutions encounter. In efforts to resolve this issue, Texas launched Texas Direct in 2021, strengthening the roles of two prior transfer policies: the Texas Core Curriculum (TCC), which is a 42 semester credit hour (SCH) core curriculum for all undergraduate students in Texas (Texas Education Code, Chapter 61, Subchapter S), and the Field of Study Curriculum (FOSC), which is a set of lower-division courses that transfer and apply to a degree program, as required by state law (Texas Education Code, Chapter 61, Section 61.823)³. Even after the launch of Texas Direct, students continue to face challenges in selecting courses that transfer efficiently. The FOSC does not exist for all majors, and students and advisors generally lack awareness about them (Schudde, Jabbar, and Chevalier, 2018). Completing the TCC does not always guarantee that students have fulfilled all the prerequisites for their intended majors at their designated institutions (Bailey et al., 2017). In many cases, core credits are accepted as electives rather than counting toward major-specific requirements.

The FOSC and the TCC serve as the state’s primary policy levers in Texas’s ongoing efforts to improve the effectiveness of vertical transfer. Under Texas Direct, existing FOSC pathways continue to be revised, while new FOSC pathways for other majors in popular transfer disciplines are being developed. The Alternative Field of Study Curriculum, approved in July 2024, further expands them by providing greater flexibility for disciplines that require fewer than the full 42 semester credit hours of the TCC but more than 18 semester credit hours of lower-division coursework to complete an associate degree at a community college. Evaluating how these two components of Texas Direct influence student outcomes, independently and in combination, will be critical for informing ongoing policy development and strengthening transfer pathways in Texas.

6.2 Empirical Strategy

6.2.1 Data

We used Texas statewide administrative data, which contain student- and school-level information from all public and private higher education institutions in Texas, collected by the Texas Higher Education Coordinating Board (THECB) and provided through the Texas Education Research Center (ERC) at the University of Texas at Austin. To construct a full longitudinal dataset, we used the admissions, coursework, financial aid, enrollment, graduation, and credit loss information included in the THECB dataset. To construct variables for the availability and the remote coursework for the TCC and the FOSC courses, the coursework data was supplemented with information from the

³Coordinating Board rules regarding the TCC and FOSC are in the Texas Administrative Code, Title 19, Part 1, Chapter 4, Subchapter B.

Texas General Education Core Curriculum Dashboard and the THECB website’s list of coursework included in legacy FOSC. We additionally used the Texas Education Agency dataset, which contains all public education information from K-12 schools in Texas, to capture students whose dual enrollment records were not captured by the THECB data.

The credit loss data (CBM00T) within the THECB dataset, collected since Fall 2020, is novel in that course-level data on credit loss are rarely collected by states (Giani, Schudde, and Sultana, 2026a). The lack of such comprehensive data had made it challenging for researchers to compute credit loss for each student according to their institution and conduct analysis. However, it is not without shortcomings. Notably, the data includes whether a student lost credits from specific courses when at least 50% of the awarded credit hours for the course did not transfer and apply to the student’s major at the receiving institution, rather than the “credits lost upon transfer”. Since we do not know how many credit hours each student was able to transfer from each course with credit loss, we constructed our measure of credit loss using the number of courses with at least 50% of their credits lost. Throughout this section, we will use the term “number of courses with lost credit”, or “post-transfer credit loss”, to refer to this variable. We further discuss the limitations of the data and their implications for our analysis in the subsection discussing the limitations of our study (Subsection 6.4).

6.2.2 Analytic Sample

We focus on the analytic sample of students constructed using the population of first-time-in-college students who entered a Texas community college in Fall 2018 or Fall 2019 ($N_{2018} = 99,055$, $N_{2019} = 99,963$) and transferred to a Texas four-year university after two ($N_{2018} = 7,284$, $N_{2019} = 7,507$) or three ($N_{2018} = 4,418$, $N_{2019} = 4,271$) years. We restricted our sample to those with no prior enrollment records, including dual enrollment, in the Texas community college system before they graduated high school. Once we restrict our attention to those with non-missing observations for the variables used in our model, there are $N = 20,969$ students in our sample, with $N_{2018} = 10,599$, $N_{2019} = 10,370$.

Using a year-based criterion for inclusion in our analytic sample, rather than credit accumulation-based criteria, offered a number of advantages. From the policy standpoint, both cohorts were subject to the same FOSC policy (2017–2019 enactment), allowing for a cleaner evaluation of a specific set of FOSC guidelines. From the analysis standpoint, they provided sufficient follow-up after transfer to earn a degree (3 years, and 6 years since first college entrance), and ensured that all students were subject to inclusion in the THECB transfer report. Finally, we could include a wide

range of students with varying degrees of pre-transfer credit accrual, allowing us to produce more generalizable findings.

Tables 10 and 11 of Appendix B provide the descriptive statistics of the demographic and academic characteristics of students in our analytic sample, and Table 9 includes the definitions of the variables we used for our analysis, together with sources of the data used to construct each variable. Clearly, students in each group differ in their demographic and academic characteristics. Demographically, students who completed at least one of the FOSC or the TCC prior to transfer were more likely to be Hispanic, be a Pell Grant recipient, and report lower adjusted gross income of parents relative to those who did not. Academically, students observed in the Neither group had the lowest GPA both prior to, and after, vertical transfer. Students in the FOSC only group were at least 3 percentage points more likely to reverse transfer than students in the other groups, and had the highest number of courses with credit loss. Although the students in our analytic sample were those who have transferred to a four-year university within two to three years of initial community college enrollment, more than half of the students in FOSC only, TCC only, and Both group had the educational goal of a two-year degree upon entrance. In particular, 64.4% of the Both group students had the goal of two-year degree attainment. Their completion of both curricula may reflect this goal, as completion of the TCC is required for graduation with any academic associate degree in Texas, and the FOSC consists of lower-division courses in major fields, some of which may have been mandatory for earning an associate degree. Finally, while students observed in the Both group on average took the longest time until vertical transfer, the total number of years until obtaining a four-year bachelor’s degree, among those who did, was the shortest for them.

6.2.3 Empirical Model

Students’ utility functions are modeled as follows:

$$U_t = X\alpha_{t0} + \sum_{k=1}^K \alpha_{tk}Z_{tk} + \bar{v}_t = U_{x,t} + U_{z,t} + \bar{v}_t$$

where X denotes the pre-transfer total credits earned, and Z_{t1} denotes the average number of course sections available for curriculum t , and Z_{t2} denotes the average number of remote sections available for curriculum t , which are used as instrumental variables. We normalized the utility from completing none of the curricula to 0, and let $\bar{v}_t = v_t - v_0$. We assume that $U_{z,Both} = U_{z,FOSC} + U_{z,TCC}$. This restricts any synergistic effect of the two curricula on students’ utility to occur only from unobserved

differences among students or from the number of pre-transfer credits a student is willing to earn before transferring. Such a modeling choice is standard in models of “bundles” (e.g., Fox and Lazzati, 2017). The model includes 8 parameters to be estimated: one for the distributional features of v_t , which will be described later in this section, and 7 coefficients for the variables in the model. Of these 7 coefficients, 3 correspond to the coefficient on X entering the utilities, and 4 correspond to α_{t1} and α_{t2} , for $t = 1, 2$.

In our model, pre-transfer credit accumulation captures the mechanical relationship between the chances of curricula completion and pre-transfer credit accumulation. It can be seen as measuring the “cost” of curricula completion: more pre-transfer credit makes curricula completion “cheaper”, increasing the chances of curricula completion. Although not necessary for our identification strategy, in implementing our model, we restricted $F_{V|x,z}$ to be equal not only across different instruments but also across different numbers of pre-transfer credits. In other words, our analysis is based on the assumption that V is stochastically independent of X . This assumption may be violated if students made their decision for curricula completion before taking community college coursework, and their pre-transfer credits earned reflect this strategic process, rather than students being more or less likely to complete each curriculum based on their credit accrual during community college enrollment. We believe this is unlikely to be the case, since most community college transfer-intending students consider transfer policies as providing them a pathway towards transfer, rather than as a plan to be strategically executed during enrollment (Schudde et al., 2021), such as by deliberately completing a curriculum⁴.

The average number of course sections and the remote sections available serve as our instrumental variables and measure course flexibility and accessibility. Based on findings that course flexibility may shift community college students’ enrollment decisions (Fox, 2017; McKinney et al., 2022; Grosz, Kurlaender, and Stevens, 2022), we assume that these instruments satisfy the relevance condition. That is, we assume that the variation in these factors shifts the utility of course completion for students with different preferences, providing information about the unobserved heterogeneity within the student population. For example, McKinney et al. (2022), in their study of Texas community college students, find that the notion of flexibility, in terms of both course formats and meeting times, was an important determinant of course enrollment. Ortagus, Tyler, and Dirghalli (2026), in their study of historically underserved students in community colleges, find that students often struggle to balance optimal course scheduling with work and financial obligations, which can

⁴This assumption was made mostly due to computational reasons, with the possibility of relaxing it and exploring the robustness of it in the future

lead to suboptimal course enrollment decisions. They also find that the propensity to use online coursework as a scheduling tool is especially prominent among working students and those with familial responsibilities, whose optimal course sequences can only be met through online coursework, while others with a stronger preference for high levels of peer-to-peer and faculty interaction have a lower preference for those options. Moreover, we assume that they satisfy the exclusion restriction, varying in statistically independent ways from the students’ unobserved determinants of selection into completing each curriculum. We believe this condition is also plausible: many community colleges create term-by-term schedules that can be unpredictable for students (Community College Research Center, 2026), making students’ self-selection into semesters with more flexible coursework difficult and therefore non-endogenous.

The outcome model is specified as follows:

$$Y(t) = \theta_{t1} + X\theta_{t2} + e_t.$$

The treatment-specific intercept, θ_{t1} is fixed to 0 when evaluating credit loss, since it is not possible for students to lose credits when they have not earned any credits from a community college. For our degree attainment outcomes, if we assume that the marginal effect of pre-transfer credits on degree attainment does not vary with curriculum completion status, we can compare θ_{t1} across different t to learn about the average treatment effect. We assume that $e_t = \rho v_t$, where ρ is a scale parameter that takes into account how much e_t changes with the choice model unobserved terms. We report our results after fixing $\rho = 0.5$ in this section, and report an alternative specification with $\rho = 0.2$ in Appendix B. Although the predictions of outcomes change slightly in levels, we find that the conclusion of this paper is robust to the parameterization of ρ ⁵.

We allow students’ unobserved heterogeneity that shifts their propensities to complete either or both of the FOSC and the TCC to follow a mixture of two multivariate normal distributions with mean 0 and unit variance, accommodating the possibility that transfer students consist of two separate subpopulations. In one component, the unobserved determinants of utility are positively correlated across choices; in the other, they are negatively correlated. Both components are equicorrelated, with the absolute value of the correlation fixed to 0.33. The mixture weight, which we treat as an unknown parameter in $[0, 1]$, determines the share of each type, so the student population may consist of either type alone (a weight of 0 or 1) or of both. Therefore, the transfer student population

⁵Due to our data usage limitations, we found it challenging to incorporate an additional dimension of parameter to be estimated jointly with other outcome model parameters. More details on the implementation of our procedure are in Appendix B.

can consist of different subgroups: for one group, students who are more likely to complete the TCC will also be more likely to complete FOSC, and also more likely to complete both; for the other group, students who are more likely to complete the TCC will be less likely to complete FOSC, and students who complete either one will be less likely to complete both. We posit that students who are generally constrained in financial resources or are certain about their field of study and the institution to which they will transfer are more likely to fall into the former category. In contrast, students who benefit from transferring as soon as possible because they have sufficient financial capital to complete most of their coursework at a four-year institution, as well as students who are unsure about their field of study or have been advised to transfer sooner for increased success, are more likely to fall into the latter category. By accounting for heterogeneity among students who complete different pre-transfer course sequences before ultimately transferring, this approach allows evaluation of the effect of pre-transfer course credit accumulation on post-transfer outcomes without relying on restrictive assumptions that may not reflect the composition of transfer students in Texas.

6.3 Results

Table 4: Bounds on Predicted Outcomes when $X = 60$

Number of Courses with 50% or more Credits Lost				
Treatment	Neither	FOSC only	TCC only	Both
$E[\hat{Y}(t) X = 60]$	[0, 0.410]	[0.024, 0.603]	[0, 0.458]	[0.121, 0.603]
Bachelor's Degree Attainment (6 Years)				
Treatment	Neither	FOSC only	TCC only	Both
$E[\hat{Y}(t) X = 60]$	[0.335, 0.455]	[0.454, 0.665]	[0.398, 0.640]	[0.423, 0.768]
Bachelor's Degree Attainment (3 Years, Post-Transfer)				
Treatment	Neither	FOSC only	TCC only	Both
$E[\hat{Y}(t) X = 60]$	[0.226, 0.485]	[0.274, 0.676]	[0.368, 0.621]	[0.411, 0.744]
Observations	10,773	800	7,779	1,617

Table 4 displays the bounds on the predicted outcomes when the pre-transfer credits are fixed to 60. That is, they reflect the predicted number of courses with half or more credits denied if an average transfer student with 60 pre-transfer credits were to transfer with each curriculum completion status. The bounds on predicted number of courses with lost credit overlap across different curricula completion statuses, making it challenging to compare the effect of pre-transfer credit accumulation across them. It is still interesting to observe that conditional on completing the FOSC, the predicted number of credit loss is always positive, implying that an average transfer

student is likely to lose more than 0 pre-transfer credits upon transferring after completing their lower-division major coursework. In reality, students who complete each of the curricula components complete a different number of pre-transfer credit hours on average. For instance, students in the “Neither” group, who did not complete TCC or FOSC, transfer with about 55 credit hours on average, while those who complete both curricula earn about 69 pre-transfer credits. Table 5 displays the predicted loss of credits once this difference is taken into account.

Table 5: Bounds on Predicted Outcomes when $X = \bar{X}_t$

Number of Courses with 50% or more Credits Lost				
Treatment	Neither	FOSC only	TCC only	Both
$E[\hat{Y}(t) X = \bar{X}_t]$	[0, 0.377]	[0.027, 0.669]	[0, 0.508]	[0.139, 0.693]
Observations	10,773	800	7,779	1,617

For bachelor’s degree attainment, we observe that the predicted probabilities also overlap across treatments. It is informative to focus on the intercept estimates, which represent the predicted treatment effect for a student who transfers without any pre-transfer credit accrual, or the treatment effect when the marginal effect of pre-transfer credits on degree attainment is assumed to be homogeneous regardless of curricula completion status. This assumption is standard in empirical studies of treatment effects (Heckman, 1990).

Table 6: Bounds on θ_{t1}

Bachelor’s Degree Attainment (6 Years)				
Treatment	Neither	FOSC only	TCC only	Both
$\hat{\theta}_{t1}$	[0.395, 0.395]	[0.395, 0.605]	[0.395, 0.605]	[0.395, 0.816]
Bachelor’s Degree Attainment (3 Years, Post-Transfer)				
Treatment	Neither	FOSC only	TCC only	Both
$\hat{\theta}_{t1}$	[0.184, 0.395]	[0.184, 0.605]	[0.395, 0.605]	[0.395, 0.816]
Observations	10,773	800	7,779	1,617

Table 6 indicates that, when the marginal contribution of pre-transfer credits is assumed to be the same across curriculum completion statuses, the effect of completing the core curriculum alone, or both the core and the field of study curricula, relative to completing neither, is bounded below by zero, pointing to a nonnegative and possibly positive effect of core completion. The increase in the chances of degree attainment can be as large as 42.1 percentage points, although the interpretation should be made with caution, given the large widths of the bounds and the coarseness of the grid. Interestingly, we do not observe the same guarantee of nonnegative gains from core completion

when students have already completed the FOSC, suggesting a possible negative synergistic effect, although the imprecision makes it hard to conclude.

If we restrict our attention to the bounds on θ_{t1} after restricting θ_{t2} to take the same values within the vector, we can further sharpen the identified sets. Because it is more standard to report projections of the identified set of parameters, as we do here, we do not pursue this possibility.

Because each treatment’s intercept and slope jointly form a point in the identified set, comparing either parameter in isolation can be misleading, since the largest intercept in the identified set is generally paired with smaller slopes, whereas the smallest intercepts tend to be paired with larger slopes. In Table 15 of Appendix B, we report, for the identified sets of parameters for each curriculum-completion status, the smallest slope paired with the smallest admissible intercept and the largest slope paired with the largest. For both the six-year and three-year degree attainment outcomes, the smallest slope paired with the smallest intercept for the Both group is below the largest slope paired with the largest intercept for the Neither group. Thus, the data allow for the possibility that each additional pre-transfer credit contributes less to degree attainment for students who complete both curricula than for students who complete neither, offsetting the gains implied by the intercept comparison in Table 6, preventing us from guaranteeing that curricula completion will lead to increased likelihood of degree attainment. In other words, Texas Direct participation has a heterogeneous effect according to the number of pre-transfer credits earned: because the positive effect of curricula completion is decreasing in the number of pre-transfer credits earned, at any given nonzero level of pre-transfer credits, those credits contribute less toward degree attainment once they are in the curricula-completion status.

6.4 Limitations

Our observation period overlaps with the onset of the COVID-19 pandemic. Specifically, our analytic sample consists of students who transferred from community colleges to a four-year university between Fall 2020 and Fall 2022 and were subject to the pandemic’s impact during both their pre- and post-transfer periods. Our focus on this period may raise several concerns, one regarding the generalizability of our results and the other regarding our instrumental variables that capture course availability and the number of remote course sections available.

Texas community colleges responded flexibly and promptly during the pandemic, and persistence and vertical transfer rates remained stable after the pandemic began for cohorts who matriculated before its onset (Hughes et al., 2025). This suggests that our analytic sample of students, which

includes first-time-in-college students who entered in Fall 2018 and Fall 2019, is likely similar to a sample selected under the same criteria without the pandemic’s impact. However, these students may be a more selective group that was more resilient to pandemic-related disruptions and therefore demonstrated academic patterns different from the general population of Texas community college students under similar conditions.

Another pandemic-related factor that may affect the generalizability of our results is that many community colleges adopted more lenient grading practices (Bird, Castleman, and Lohner, 2022), and average grades increased during the pandemic (Bulman and Fairlie, 2022), whereas course completion rates decreased. This may have led to a temporary change in students’ coursetaking behavior if the leniency of the grading criteria caused them to take more or fewer “challenging courses” before applying and transferring to a four-year university. For example, a student who planned to earn 60 credit hours during community college enrollment and would have completed only a core curriculum may end up completing both curricula and taking more courses because it became relatively “easy” to pass lower-division major coursework under a more lenient grading policy. In such cases, the composition of students in each curriculum-completion group in our sample may differ from what is typically expected, making our results hard to generalize to other periods.

In Table 12 of Appendix B, we show differences in pre-transfer academic measures for each of the 2018 and 2019 entrants. Although the 2019 cohort was subject to the pandemic-induced changes for a longer period of time, we see that the cohorts do not demonstrate significant differences in their pre-transfer GPAs, credits enrolled, and credits earned. However, we do notice an increase in the average number of remote TCC and FOSC sections during students’ enrollment, as well as an increase in the FOSC only group students in the 2019 cohort. Although these differences do not threaten the robustness of our analysis to pandemic-induced confounders, other factors not captured by these pre-transfer academic attainment measures may bias our results in unpredictable ways. We therefore note this as a limitation.

The credit loss data (CBM00T) component of the THECB dataset, which we used to construct our outcome variable, has several limitations that may have impacted our analysis. First, it began collecting in Fall 2020, when the earliest students in our analytic sample transferred to a four-year university. Because this was the first time universities reported these data, the first data collection faced challenges, such as institutions interpreting some report parameters differently (THECB, 2021). In the next round of data collection, institutions provided information on courses that were denied transfer credit in Spring 2021 and re-reported denied course information for Fall 2020 (THECB, 2022). Therefore, our results should be interpreted with this imperfect data collection process in

consideration.

Second and more importantly, these data are only available for students whose major area did not change between the time of application and the census date. Therefore, credit loss information included in them may be downward biased if students who changed majors lost many credits but universities did not report them. Credits lost for the FOSC only group and the Both group may have been reported more frequently relative to the other two groups if they were less likely to change majors, given the major-specific nature of the FOSC. Because the THECB datasets do not include the exact number of credit hours transferred or the major to which a student applied, this limitation cannot be addressed and should be acknowledged.

6.5 Discussion

Since the launch of Texas Direct in March 2021, its two key policy levers, the Texas Core Curriculum and the Field of Study Curriculum, have played key roles in improving the transfer function of Texas community colleges. However, their roles in shaping students' academic pathways after transfer have largely remained unknown, mainly due to the difficulty of conducting causal analyses amid substantial heterogeneity among students who select into each curriculum within such a complicated policy structure. Accounting for uncertainty from the wide range of unobserved heterogeneity that may underlie the Texas community college student population, our results provide a multifaceted view of its effect.

Our findings show that students are likely to lose at least some proportion of pre-transfer credits when transferring after completing a lower-division degree-specific coursework of FOSC. Completing the FOSC does not mean students always apply to and enroll in a major aligned with the FOSC they have completed. Students are largely unaware of FOSC (Schudde, Jabbar, and Chevalier, 2018), and many change their mind about their major (Hodara et al., 2017). They may end up completing FOSC without accurate information about which majors the curriculum applies to at the destination institution, apply to an institution that does not offer the corresponding degree program, or apply to another program because they changed their mind about the major they want to pursue. In such cases, because FOSC transfer is subject to completion of the set of courses adopted by the institution for that particular field of study (Texas Education Code, Chapter 61, Section 61.823), the credits earned toward FOSC completion can be lost due to them falling outside the degree program requirements. Given that the most prevalent cause for credit loss was credits being outside the student's major at the time of matriculation (Giani, Schudde, and Sultana, 2026b),

it is plausible that completing the degree-specific transfer pathway, such as the FOSC, may not be the most efficient strategy for preventing credit loss for all students with a transfer goal.

Our result that TCC completion can be guaranteed to have a nonnegative effect on the likelihood of four-year degree attainment only when the marginal effect of pre-transfer credits is the same across curriculum completion statuses provides a more nuanced assessment of the core curriculum. It builds on the finding by Schudde, Bicak, and Shea (2023) that pre-transfer credit accumulation can improve the likelihood of four-year degree attainment by showing how it shapes students' outcomes after they reach core completion status, while pointing to its underlying mechanism: completing the TCC lowers the contribution of pre-transfer credits towards degree attainment, while raising the baseline likelihood of it. The possibly smaller contribution of each additional pre-transfer credit for TCC completers, relative to non-completers, makes it difficult to say with certainty that students will do better than in the counterfactual scenario where they transfer without completing the TCC.

Before students attain core-complete status, additional pre-transfer credits are more likely to count toward components of the core, which are guaranteed to transfer. Once the core is complete, however, additional credits carry no such guarantee, so they may contribute less to degree attainment for TCC completers than for non-completers, particularly when students lack information about transfer requirements, which may mute the positive effect of core completion.

Specifically, our descriptive analysis shows that students in the TCC only and Both groups attempted and earned more credits than those in the Neither group. Although attaining core completion status requires 42 credit hours, students in the TCC only group earned around 67 credits, and those in the Both group earned as many as 69 credits, compared with an average of 55 credit hours completed by students in the Neither group. Excess credits may indicate that students lack information about the requirements for their academic goals (Zeidenberg, 2012). Our results show that expected credit loss is bounded away from zero for FOSC completers, including those who also complete the TCC. Many Texas public universities cap the number of community college credits that can be applied toward a bachelor's degree, commonly at 66 semester credit hours (e.g., The University of Texas at San Antonio 2026-28 Undergraduate Catalog). Together with evidence of a general lack of awareness about the transferability of coursework among transfer students (Allen, Smith, and Muehleck, 2013; Schudde, Bradley, and Absher, 2020), this suggests that those additional credits may not be effectively directed toward students' degree programs.

Several policy implications follow from our results. First, completing lower-division degree coursework may not always prevent credit loss. Because not all students are certain about the field of study they plan to pursue, or know whether their FOSC credits will be applicable in the destination in-

stitution, it may not be worth the additional credit hours for every student to complete the FOSC. To make Texas Direct and the FOSC an effective policy for improving credit portability, it would be critical to ensure that the students and college advisors are equipped with accurate knowledge about those policies.

Moreover, although TCC completion can have positive effects on degree attainment, its effect may be muted because additional pre-transfer credits contribute less to degree attainment for TCC completers than for non-completers. Recall that our predicted outcomes pertain to the one that a randomly selected student would attain. Therefore, for students who take additional transferable coursework that helps them progress toward their academic goals, additional credit accumulation may not necessarily diminish their gains from completing the TCC. Regardless, for an average transfer student, our results show that preventing excess credit accumulation should be prioritized to reap the full benefit of the TCC.

Finally, the uncertain effects of Texas Direct do not imply that students will not benefit in other ways. Notably, the initiative also grants an associate degree upon the completion of the FOSC and the TCC. Two-year degrees have been shown to have significant labor market returns (Jepsen, Troske, and Coomes, 2014; Stevens, Kurlaender, and Grosz, 2019). Although additional efforts in improving the efficiency of transfers, as well as further investigation into the effects of Texas Direct, are needed, the initiative’s promise to provide students with a tangible return for their pre-transfer coursework can be a positive step towards improving students’ future trajectories.

7 Conclusion

This paper studies self-selection models, in which the observability of individuals’ outcomes depends on their economic choices. Focusing on cases where the underlying selection process is not guaranteed to be point identified or the model does not necessarily predict a single utility-maximizing choice, we provide a straightforward characterization of the identified sets of parameters of interest by applying random set-based methods from Chesher and Rosen (2017) and Han and Kaido (2026). In particular, the identified sets for the parameters of the choice-conditional outcome models include parameters that are consistent with the choice model’s set-valued restrictions, which encompass all possible distributions of unobservables affecting the utilities associated with each choice that are consistent with the data. The generality of our results lets researchers study models with choice-dependent outcomes without the strong restrictions required by existing tools. Based on our identification results, we introduce a computationally convenient two-step estimator for the identified sets. Our estimator lets

practitioners harness the full identifying power of instrumental variables by incorporating stochastically independent instruments into models characterized by conditional moment restrictions, while improving accuracy and reducing computational burden in this set-estimation setting.

The proposed tools are applied to study the effects of the two curriculum levers of Texas Direct, the Texas Core Curriculum and the Texas Field of Study Curriculum, using an administrative student-level longitudinal dataset from Texas that includes credit loss information. Our findings suggest that Texas Direct's benefits for students' post-transfer outcomes depend on how students accumulate and apply pre-transfer credits. Specifically, completing the Field of Study Curriculum before transferring to a four-year university may lead to credit loss, and the benefits of completing the Core Curriculum before transfer depend on whether pre-transfer credits contribute less to degree attainment for students who complete it than for students who complete neither curriculum. Together, these findings highlight the need to improve how accurately transfer information reaches students and college advisors and to help students plan their transfer pathways early in their academic careers for Texas Direct to achieve its full potential.

References

- Ahn, H., & Powell, J. L. (1993). Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism. *Journal of Econometrics*, 58(1–2), 3–29.
- Allen, J. M., Smith, C. L., & Muehleck, J. K. (2013). What Kinds of Advising Are Important to Community College Pre- and Posttransfer Students? *Community College Review*, 41(4), 330–345.
- Andrews, D. W. K., & Shi, X. (2013). Inference Based on Conditional Moment Inequalities. *Econometrica*, 81(2), 609–666. <https://doi.org/10.3982/ECTA9370>
- Aristodemou, E., & Rosen, A. M. (2022). A Discrete Choice Model for Partially Ordered Alternatives. *Quantitative Economics*, 13(3).
- Bailey, T., Jenkins, D., Fink, J., Cullinane, J., & Schudde, L. (2017, January). *Policy Levers to Strengthen Community College Transfer Student Success in Texas* (tech. rep.). Community College Research Center, Teachers College, Columbia University. <https://doi.org/10.7916/D86W9KR5>
- Barseghyan, L., Coughlin, M., Molinari, F., & Teitelbaum, J. C. (2021). Heterogeneous Choice Sets and Preferences. *Econometrica*, 89(5), 2015–2048.
- Beresteanu, A., Molchanov, I., & Molinari, F. (2008). *Sharp Identification Regions in Games* (CEMMAP Working Paper No. CWP15/08). Centre for Microdata Methods and Practice (CEMMAP), Institute for Fiscal Studies. https://econpapers.repec.org/paper/ifscemmap/15_2F08.htm
- Beresteanu, A., Molchanov, I., & Molinari, F. (2011). Sharp Identification Regions in Models with Convex Moment Predictions. *Econometrica*, 79(6), 1785–1821. <https://doi.org/10.3982/ECTA8680>
- Beresteanu, A., Molchanov, I., & Molinari, F. (2012). Partial Identification Using Random Set Theory. *Journal of Econometrics*, 166(1), 17–32.
- Beresteanu, A., & Molinari, F. (2008). Asymptotic Properties for a Class of Partially Identified Models. *Econometrica*, 76(4), 763–814.
- Bird, K. A., Castleman, B. L., & Lohner, G. (2022). Negative Impacts from the Shift to Online Learning during the COVID-19 Crisis: Evidence from a Statewide Community College System. *AERA Open*, 8.
- Bulman, G., & Fairlie, R. (2022). The Impact of COVID-19 on Community College Enrollment and Student Success: Evidence from California Administrative Data. *Education Finance and Policy*, 17(4), 745–764.

- Chernozhukov, V., Hong, H., & Tamer, E. (2007). Estimation and Confidence Regions for Parameter Sets in Econometric Models. *Econometrica*, 75(5), 1243–1284. <https://doi.org/10.1111/j.1468-0262.2007.00794.x>
- Chesher, A., Kim, D., & Rosen, A. M. (2023). IV Methods for Tobit Models. *Journal of Econometrics*, 235(2).
- Chesher, A., & Rosen, A. M. (2017). Generalized Instrumental Variable Models. *Econometrica*, 85(3), 959–989. <https://doi.org/10.3982/ECTA12223>
- Chesher, A., Rosen, A. M., & Smolinski, K. (2013). An Instrumental Variable Model of Multiple Discrete Choice. *Quantitative Economics*, 4(2), 157–196.
- Ciliberto, F., & Tamer, E. (2009). Market Structure and Multiple Equilibria in Airline Markets. *Econometrica*, 77(6), 1791–1828. <https://doi.org/10.3982/ECTA5368>
- Community College Research Center. (2026). Optimizing Course Availability and Delivery for On-Time Completion [Published: Research project description, Teachers College, Columbia University]. <https://ccrc.tc.columbia.edu/research-project/optimizing-course-availability.html>
- Dahl, G. B. (2002). Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica*, 70(6), 2367–2420.
- Das, M., Newey, W. K., & Vella, F. (2003). Nonparametric Estimation of Sample Selection Models. *Review of Economic Studies*, 70(1), 33–58.
- D’Haultfœuille, X., & Maurel, A. (2013). Inference on an Extended Roy Model, with an Application to Schooling Decisions in France. *Journal of Econometrics*, 174(2), 95–106.
- Dubin, J. A., & McFadden, D. L. (1984). An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. *Econometrica*, 52(2), 345–362.
- Eisenhauer, P., Heckman, J. J., & Vytlacil, E. (2015). The Generalized Roy Model and the Cost-Benefit Analysis of Social Programs. *Journal of Political Economy*, 123(2), 413–443.
- Fox, H. L. (2017). What Motivates Community College Students to Enroll Online and Why It Matters. Insights on Equity and Outcomes. Issue 19. *Office of Community College Research and Leadership*.
- Fox, J. T., & Lazzati, N. (2017). A Note on Identification of Discrete Choice Models for Bundles and Binary Games. *Quantitative Economics*, 8(3).
- French, E., & Taber, C. (2011). Identification of Models of the Labor Market. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (pp. 537–617, Vol. 4A). Elsevier.
- Galotti, K. M., & Umscheid, V. A. (2019). Students Choosing Courses: Real-Life Academic Decision Making. *American Journal of Psychology*, 132(2), 149–159.

- Giani, M. S., Schudde, L., & Sultana, T. (2026a). New Insights on Sources of Credit Loss. *Educational Researcher*, 55(4), 281–285. <https://doi.org/10.3102/0013189X251410173>
- Giani, M. S., Schudde, L., & Sultana, T. (2026b). Toward a Comprehensive Model Predicting Credit Loss in Vertical Transfer. *American Educational Research Journal*. <https://doi.org/10.3102/00028312251409063>
- Grosz, M., Kurlaender, M., & Stevens, A. (2022). Capacity and Flexibility in Community College CTE Programs: Program Offerings and Student Success. *Research in Higher Education*, 63. <https://doi.org/10.1007/s11162-021-09645-9>
- Han, S., & Kaido, H. (2026). Set-Valued Control Functions [Published: arXiv:2403.00347].
- Hay, J. W. (1980). *Occupational Choice and Occupational Earnings: Selectivity Bias in a Simultaneous Logit-OLS Model* [PhD Thesis]. Yale University.
- Heckman, J. J. (1974). Shadow Prices, Market Wages, and Labor Supply. *Econometrica*, 42(4), 679–694.
- Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, 5(4), 475–492.
- Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153–161. <https://doi.org/10.2307/1912352>
- Heckman, J. J. (1990). Varieties of Selection Bias. *American Economic Review*, 80(2), 313–318.
- Heckman, J. J., & Honoré, B. E. (1990). The Empirical Content of the Roy Model. *Econometrica*, 58(5), 1121–1149.
- Heckman, J. J., & Navarro-Lozano, S. (2004). Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models. *Review of Economics and Statistics*, 86(1), 30–57.
- Heckman, J. J., & Robb, R. (1985). Alternative Methods for Evaluating the Impact of Interventions. In J. J. Heckman & B. Singer (Eds.), *Longitudinal Analysis of Labor Market Data* (pp. 156–245). Cambridge University Press.
- Heckman, J. J., & Sedlacek, G. (1985). Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market. *Journal of Political Economy*, 93(6), 1077–1125.
- Hodara, M., Martinez-Wenzl, M., Stevens, D., & Mazzeo, C. (2017). Exploring Credit Mobility and Major-Specific Pathways: A Policy Analysis and Student Perspective on Community

- College to University Transfer. *Community College Review*, 45(4), 331–349. <https://doi.org/10.1177/0091552117724197>
- Hughes, K., Kosiewicz, H., Miller, T., Khalid, A., & Vargas Castaño, D. (2025). *Resilience and Transformation: The Pandemic's Effects on Texas Community Colleges* (EdWorkingPaper No. 25-1309). Annenberg Institute at Brown University. <https://doi.org/10.26300/30mv-h525>
- Jabbar, H., Schudde, L., Garza, M., & McKinnon-Crowley, S. (2022). Bridges or Barriers? How Interactions between Individuals and Institutions Condition Community College Transfer. *Journal of Higher Education*, 93(3), 375–398. <https://doi.org/10.1080/00221546.2021.1953339>
- Jepsen, C., Troske, K., & Coomes, P. (2014). The Labor-Market Returns to Community College Degrees, Diplomas, and Certificates. *Journal of Labor Economics*, 32(1), 95–121.
- Kaido, H., & White, H. (2014). A Two-Stage Procedure for Partially Identified Models. *Journal of Econometrics*, 182(1).
- Kline, B., & Tamer, E. (2023). Recent Developments in Partial Identification. *Annual Review of Economics*, 15.
- Knothe, H. (1957). Contributions to the Theory of Convex Bodies. *Michigan Mathematical Journal*, 4(1), 39–52.
- Lee, L.-F. (1983). Generalized Econometric Models with Selectivity. *Econometrica*, 51(2), 507–512.
- Lewbel, A. (2007). Endogenous Selection or Treatment Model Estimation. *Journal of Econometrics*, 141(2), 777–806.
- Manski, C. F. (2005). Partial Identification with Missing Data: Concepts and Findings. *International Journal of Approximate Reasoning*, 39(2–3), 151–165. <https://doi.org/10.1016/j.ijar.2004.10.006>
- Manski, C. F., & Tamer, E. (2002). Inference on Regressions with Interval Data on a Regressor or Outcome. *Econometrica*, 70(2), 519–546.
- McKinney, L., Burrige, A. B., Lee, M. M., Bourdeau, G. V., & Miller-Waters, M. (2022). Incentivizing Full-Time Enrollment at Community Colleges: What Influences Students' Decision to Take More Courses? *Community College Review*, 50(2), 144–170. <https://doi.org/10.1177/00915521211061416>
- Mogstad, M., & Torgovitsky, A. (2018). Identification and Extrapolation of Causal Effects with Instrumental Variables. *Annual Review of Economics*, 10, 577–613.
- Molchanov, I. (2017). *Theory of Random Sets* (2nd ed.). Springer.

- Molchanov, I., & Molinari, F. (2018). *Random Sets in Econometrics*. Cambridge University Press.
- Molinari, F. (2020). Microeconometrics with Partial Identification. In S. N. Durlauf, L. P. Hansen, J. J. Heckman, & R. L. Matzkin (Eds.), *Handbook of Econometrics* (pp. 355–486, Vol. 7A). Elsevier.
- Monaghan, D. B., & Attewell, P. (2015). The Community College Route to the Bachelor’s Degree. *Educational Evaluation and Policy Analysis*, 37(1), 70–91.
- Mourifié, I., Henry, M., & Méango, R. (2020). Sharp Bounds and Testability of a Roy Model of STEM Major Choices. *Journal of Political Economy*, 128(8), 3220–3283.
- Mullahy, J. (2018). Individual Results May Vary: Inequality-Probability Bounds for Some Health-Outcome Treatment Effects. *Journal of Health Economics*, 61, 151–162.
- Newey, W. K. (2009). Two-Step Series Estimation of Sample Selection Models. *Econometrica Journal*, 12(s1), S217–S229.
- Ortagus, J. C., Tyler, O., & Dirghalli, J. (2026). Navigating Course-Sequencing Decisions and Community College Pathways for Historically Underserved Student Populations [ISBN: 0022-1546]. *The Journal of Higher Education*, 97(4), 675–709.
- Pakes, A., & Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, 57(5), 1027–1057.
- Pollard, D. (1989). Asymptotics via Empirical Processes. *Statistical Science*, 4(4), 341–354.
- Powell, J. L. (1994). Estimation of Semiparametric Models. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of Econometrics* (pp. 2443–2521, Vol. 4). Elsevier.
- Puhani, P. A. (2000). The Heckman Correction for Sample Selection and Its Critique. *Journal of Economic Surveys*, 14(1), 53–68.
- Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23(3), 470–472.
- Schudde, L., Bicak, I., & Shea, M. (2023). Getting to the Core of Credit Transfer: How Do Pre-Transfer Core Credits Predict Baccalaureate Attainment for Community College Transfer Students? *Educational Policy*, 37(4), 1014–1043. <https://doi.org/10.1177/08959048211049415>
- Schudde, L., Bradley, D., & Absher, C. (2020). Navigating Vertical Transfer Online: Access to and Usefulness of Transfer Information on Community College Websites. *Community College Review*, 48(1), 3–30. <https://doi.org/10.1177/0091552119874500>
- Schudde, L., Jabbar, H., & Chevalier, A. (2018). *To Improve Community College Transfer in Texas, Students Need Better Information* (PRC Research Brief No. 3(17)). Population Research Center, The University of Texas at Austin.

- Schudde, L., Jabbar, H., Epstein, E., & Yucel, E. (2021). Students' Sense Making of Higher Education Policies During the Vertical Transfer Process. *American Educational Research Journal*. <https://doi.org/10.3102/00028312211003050>
- Schudde, L., Ryu, W., & Brown, R. S. (2020). Major Movement: Examining Meta-Major Switching at Community Colleges. *Review of Higher Education*, 44(2), 189–235. <https://doi.org/10.1353/rhe.2020.0044>
- Spencer, G. (2021). Off the Beaten Path: Can Statewide Articulation Support Students Transferring in Nonlinear Directions? *American Educational Research Journal*, 58(5), 1070–1102. <https://doi.org/10.3102/0002831221999782>
- Stern, J. M. B. (2016). The Effect of Articulation Agreements on Community College Transfers and Bachelor's Degree Attainment. *Community College Journal of Research and Practice*, 40(5), 355–369. <https://doi.org/10.1080/10668926.2015.1065209>
- Stevens, A. H., Kurlaender, M., & Grosz, M. (2019). Career Technical Education and Labor Market Outcomes: Evidence from California Community Colleges. *Journal of Human Resources*, 54(4), 986–1036.
- Storn, R., & Price, K. (1997). Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11(4), 341–359.
- Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, 70(1), 147–165.
- Tamer, E. (2010). Partial Identification in Econometrics. *Annual Review of Economics*, 2, 167–195.
- Texas Comptroller of Public Accounts. (n.d.). Texas' Community Colleges: Statewide Overview [Published: Web page]. <https://comptroller.texas.gov/economy/economic-data/colleges/texas.php>
- Texas Education Code. (n.d.). Chapter 61, Subchapter S and Section 61.823 [Published: Texas Statutes].
- Texas Higher Education Coordinating Board. (2020). *2020 Texas Public Higher Education Almanac* (tech. rep.). Texas Higher Education Coordinating Board. Austin, TX.
- Texas Higher Education Coordinating Board. (2021, February). *Report on Non-Transferable Credit and Courses Taken at Public Two-Year Colleges* (Report). Texas Higher Education Coordinating Board. Austin, TX. Retrieved September 30, 2026, from <https://reportcenter.highered.texas.gov/reports/data/sb-25-report-on-non-transferable-credit-february-2021/>

- Texas Higher Education Coordinating Board. (2022). *Transfer Report 2022* (Legislative Report).
Texas Higher Education Coordinating Board. Austin, TX. Retrieved September 30, 2026,
from <https://reportcenter.highered.texas.gov/reports/legislative/transfer-report-2022/>
- Texas Higher Education Coordinating Board. (2026, August). Texas Direct. Retrieved September
30, 2026, from <https://www.highered.texas.gov/texas-direct/>
- The University of Texas at San Antonio. (n.d.). Transferring Courses [Published: Undergraduate Cat-
alog]. <https://catalog.utsa.edu/undergraduate/bachelorsdegreeregulations/transferringcourses/>
- Tsiatis, A. (1975). A Nonidentifiability Aspect of the Problem of Competing Risks. *Proceedings of
the National Academy of Sciences*, 72(1), 20–22.
- Whinnery, E., & Peisach, L. (2022, July). *50-State Comparison: Transfer and Articulation Poli-
cies* (tech. rep.). Education Commission of the States. [https://www.ecs.org/50-state-
comparison-transfer-and-articulation/](https://www.ecs.org/50-state-comparison-transfer-and-articulation/)
- Worsham, R., DeSantis, A. L., Whatley, M., Johnson, K. R., & Jaeger, A. J. (2021). Early Effects of
North Carolina’s Comprehensive Articulation Agreement on Credit Accumulation Among
Community College Transfer Students. *Research in Higher Education*, 62(7). [https://doi.
org/10.1007/s11162-021-09626-y](https://doi.org/10.1007/s11162-021-09626-y)
- Zeidenberg, M. (2012). *Valuable Learning or “Spinning Their Wheels”? Understanding Excess Credits
Earned by Community College Associate Degree Completers* (CCRC Working Paper No. 44).
Community College Research Center, Teachers College, Columbia University.

A Appendix A: Additional Results for Section 5

Table 7: Choice Parameter Estimates for Subsection 5.1

	α_1	α_2	β_1	β_2	β_3	λ
DGP	0.2	-0.25	-0.75	0.75	-0.5	0.5
Overall	[-0.3, 0.45]	[-0.25, 0.5]	[-1, 0.25]	[-0.25, 1]	[-1, 0.5]	{0.25,0.5,0.75}
$\lambda = 0.25$	[-0.3,0.45]	[-0.25, 0.25]	[-1, 0.25]	[-0.25, 1]	[-1, 0.5]	
$\lambda = 0.5$	[-0.3,0.45]	[-0.25, 0.5]	[-1, 0.25]	[-0.25, 1]	[-1, 0.5]	
$\lambda = 0.75$	[-0.3,0.45]	[-0.25, 0.5]	[-1, 0.25]	[-0.25, 1]	[-1, 0.5]	
Large Support						
Overall	[-0.3, 0.2]	[-0.25, 0.25]	[-1, 0.25]	[-0.25, 1]	[-1, 0.5]	{0.25,0.5,0.75}
$\lambda = 0.25$	[-0.3,0.2]	[-0.25, 0.25]	[-1, 0.25]	[-0.25, 1]	[-1, 0.5]	
$\lambda = 0.5$	[-0.3,0.2]	[-0.25, 0.25]	[-1, 0.25]	[-0.25, 1]	[-1, 0.5]	
$\lambda = 0.75$	[-0.3,0.2]	[-0.25, 0.25]	[-1, 0.25]	[-0.25, 1]	[-1, 0.5]	
N	20,000					

Table 8: Choice Parameter Estimates for Subsection 5.2

	α_1	α_2	α_3	α_4	λ
DGP	-0.5	1	-1	0.5	0.5
Overall	[-1, 0.25]	[0, 1]	[-1, 0]	[-0.5, 1]	{0,0.25,0.5,0.75,1}
N	20,000				

B Appendix B: Details for Empirical Application

B.1 Variable Descriptions

Table 9: Variable Definitions

Variable	Definition
<i>Treatment Group</i>	
Neither	Have not completed either the TCC or the FOSC at the time of vertical transfer, obtained from THECB graduation data and enrollment data.
FOSC only	Have completed the FOSC, but not the TCC, at the time of vertical transfer. Obtained from the THECB graduation data and enrollment data.

Continued on next page

Table 9 continued from previous page

Variable	Definition
TCC only	Have completed the TCC, but not the FOSC, at the time of vertical transfer. Obtained from the THECB graduation data and enrollment data.
Both	Have completed both the TCC and the FOSC at the time of vertical transfer, obtained from THECB graduation data and enrollment data.
<i>Demographic Characteristics</i>	
Female	Identified as female, obtained from THECB enrollment data.
Pell Grant Recipient	Ever received Pell grant, derived using THECB financial aid data.
First Year Age	Student's age in their first year of college enrollment, derived from THECB enrollment data.
Parent AGI	Parent's adjusted gross income, obtained from THECB financial aid data.
Hispanic	Identified as Hispanic, obtained from THECB enrollment data.
White	Identified as white, obtained from THECB enrollment data.
Black	Identified as Black, obtained from THECB enrollment data.
Asian	Identified as Asian, obtained from THECB enrollment data.
Native American	Identified as Native American, obtained from THECB enrollment data.
Pacific Islander	Identified as Pacific Islander, obtained from THECB enrollment data.
Unknown	Student's race was unknown or unreported.
<i>Educational Goal upon College Entry</i>	
Associate	Reported earning a 2-year associate degree as their enrollment intent in the first year of college entrance, obtained from THECB enrollment data.
Certificate	Reported earning a certificate as their enrollment intent in the first year of college entrance, obtained from THECB enrollment data.
Transfer	Reported earning credits or to transfer as their enrollment intent in the first year of college entrance, obtained from THECB enrollment data.

Continued on next page

Table 9 continued from previous page

Variable	Definition
Personal enrichment	Reported taking courses for personal enrichment as their enrollment intent in the first year of college entrance, obtained from THECB enrollment data.
Job skills	Reported getting a new or better job, or improving skills for current job, as their enrollment intent in the first year of college entrance, obtained from THECB enrollment data.
BAT	Reported earning a Bachelor of Applied Technology (BAT) degree as their enrollment intent in the first year of college entrance, obtained from THECB enrollment data.
Missing	Unreported or unknown.
<i>Instruments</i>	
Average TCC section count	Average number of sections among TCC courses offered at the student's community college in each semester, averaged over the college-semesters in which the student was enrolled. Constructed from THECB course schedule data and the Texas General Education Core Curriculum Dashboard.
Average FOSC section count	Average number of sections among FOSC courses offered at the student's community college in each semester, averaged over the college-semesters in which the student was enrolled. Constructed from THECB course schedule data and the THECB list of legacy Field of Study Curriculum.
Average remote TCC sections	Average number of remote sections among TCC courses offered at the student's community college in each semester, averaged over the college-semesters in which the student was enrolled. Constructed from THECB course schedule data and the Texas General Education Core Curriculum Dashboard.
Average remote FOSC sections	Average number of remote sections among FOSC courses offered at the student's community college in each semester, averaged over the college-semesters in which the student was enrolled. Constructed from THECB course schedule data and the THECB list of legacy Field of Study Curriculum.
<i>Pre-Transfer Characteristics</i>	

Continued on next page

Table 9 continued from previous page

Variable	Definition
First-year credits enrolled	Number of course credits enrolled in the first year of college; derived from THECB schedule data.
First-year credits earned	Number of course credits earned in the first year of college; derived from THECB schedule data.
GPA	Student's GPA, averaged across semesters enrolled before vertical transfer, derived from THECB schedule data.
Time to 4-year (years)	Number of years from the time of community college entrance to first enrollment in a four-year university, derived from THECB enrollment data.
Credits enrolled	Number of course credits enrolled before vertical transfer; derived from THECB schedule data.
Credits earned	Number of course credits earned before vertical transfer; derived from THECB schedule data.
Core credits enrolled	Number of TCC course credits enrolled before vertical transfer. Constructed from THECB course schedule data and the Texas General Education Core Curriculum Dashboard.
Core credits earned	Number of TCC course credits earned before vertical transfer. Constructed from THECB course schedule data and the Texas General Education Core Curriculum Dashboard.
FOSC credits enrolled	Number of FOSC course credits enrolled before vertical transfer. Constructed from THECB course schedule data and the THECB list of legacy Field of Study Curriculum.
FOSC credits earned	Number of FOSC course credits earned before vertical transfer. Constructed from THECB course schedule data and the THECB list of legacy Field of Study Curriculum.
<i>Post-Transfer Characteristics</i>	
GPA	Average GPA throughout the semesters enrolled at a four-year university after vertical transfer, derived from THECB schedule data.
Credits enrolled	Number of course credits enrolled at a four-year university after vertical transfer; derived from THECB schedule data.
Credits earned	Number of course credits earned at a four-year university after vertical transfer; derived from THECB schedule data.

Continued on next page

Table 9 continued from previous page

Variable	Definition
Reverse transfer	Ever transferred to a two-year community college after vertical transfer, derived from THECB enrollment data.
Credit loss	Total number of courses completed at a community college before vertical transfer for which at least 50 % of awarded credit hours did not transfer and apply to the student's major at the receiving institution, constructed from THECB course schedule data and transfer report data.
<i>Degree Outcomes</i>	
BA within 2 years, post-transfer	Earned a bachelor's degree within 2 years following vertical transfer, constructed using THECB enrollment data and graduation data.
BA within 3 years, post-transfer	Earned a bachelor's degree within 3 years following vertical transfer, constructed using THECB enrollment data and graduation data.
BA within 4 years, total	Earned a bachelor's degree within 4 years following college entry, constructed using THECB enrollment data and graduation data.
BA within 6 years, total	Earned a bachelor's degree within 6 years following college entry, constructed using THECB enrollment data and graduation data.
Time to BA, post-transfer (years)	Number of years from vertical transfer to bachelor's degree attainment, derived from THECB graduation and enrollment data. Defined only for BA recipients.
Time to BA, total (years)	Number of years from college entrance to bachelor's degree attainment, derived from THECB graduation and enrollment data. Defined only for BA recipients.

B.2 Descriptive Statistics of Analytic Sample

Table 10: Demographic Characteristics by Treatment Group

	Neither	FOSC only	TCC only	Both	Total
Female	0.525 (0.499)	0.576 (0.494)	0.640 (0.480)	0.605 (0.489)	0.576 (0.494)
Pell Grant Recipient	0.596 (0.491)	0.649 (0.478)	0.671 (0.470)	0.720 (0.449)	0.635 (0.481)
First Year Age	18.708 (2.866)	18.661 (2.916)	18.935 (3.577)	18.897 (3.391)	18.805 (3.192)
Parent AGI	\$ 77,714.85 (95,307.33)	\$ 64,058.80 (62,308.69)	\$ 66,654.37 (79,117.32)	\$ 60,201.83 (66,493.81)	\$ 71,608.81 (86,508.45)
Hispanic	0.369 (0.482)	0.445 (0.497)	0.442 (0.497)	0.518 (0.500)	0.410 (0.492)
White	0.554 (0.497)	0.531 (0.499)	0.540 (0.498)	0.560 (0.497)	0.548 (0.498)
Black	0.117 (0.321)	0.090 (0.286)	0.098 (0.297)	0.066 (0.249)	0.105 (0.306)
Asian	0.089 (0.284)	0.080 (0.271)	0.078 (0.268)	0.062 (0.241)	0.082 (0.275)
Native American	0.067 (0.249)	0.061 (0.240)	0.072 (0.259)	0.059 (0.236)	0.068 (0.252)
Pacific Islander	<0.01	<0.01	<0.01	<0.01	<0.01
Unknown	0.134 (0.341)	0.175 (0.380)	0.164 (0.370)	0.168 (0.374)	0.149 (0.356)
N	10,773	800	7,779	1,617	20,969

Notes: Standard deviations in parentheses. AGI: Adjusted Gross Income. Parent AGI $N_{Neither} = 9,629$; $N_{FOSC} = 730$; $N_{TCC} = 7,262$; $N_{Both} = 1,517$. Age is as recorded in students' first year of enrollment.

Table 11: Academic Characteristics by Treatment Group

	Neither	FOSC only	TCC only	Both	Total
<i>Educational Goal upon College Entry</i>					
Associate	0.475 (0.499)	0.531 (0.499)	0.579 (0.494)	0.644 (0.479)	0.529 (0.499)
Certificate	0.019 (0.136)	0.021 (0.144)	0.018 (0.133)	0.015 (0.121)	0.018 (0.134)
Transfer	0.399 (0.490)	0.334 (0.472)	0.331 (0.471)	0.279 (0.449)	0.362 (0.481)

Continued on next page

Table 11 continued from previous page

	Neither	FOSC only	TCC only	Both	Total
Personal enrichment	0.020 (0.139)	0.021 (0.144)	0.020 (0.139)	0.017 (0.128)	0.020 (0.138)
Missing	0.108 (0.310)	0.102 (0.303)	0.058 (0.234)	0.042 (0.201)	0.084 (0.278)
Job skills	<0.01	<0.01	<0.01	<0.01	<0.01
BAT	<0.01	<0.01	<0.01	<0.01	<0.01
<i>Pre-Transfer Characteristics</i>					
First-year credits enrolled	27.014 (6.415)	27.962 (6.364)	28.708 (6.087)	28.548 (5.998)	27.797 (6.315)
First-year credits earned	23.932 (7.331)	25.005 (7.283)	26.912 (6.622)	26.564 (6.762)	25.282 (7.174)
Average TCC section count	17.433 (17.183)	17.317 (15.714)	16.682 (15.841)	19.505 (14.709)	17.310 (16.473)
Average FOSC section count	14.123 (9.272)	12.254 (7.503)	12.270 (9.228)	13.144 (8.229)	13.289 (9.160)
Average Remote TCC Sections	0.050 (0.048)	0.065 (0.052)	0.060 (0.050)	0.065 (0.050)	0.055 (0.049)
Average Remote FOSC section count	0.032 (0.024)	0.039 (0.024)	0.032 (0.021)	0.031 (0.020)	0.032 (0.023)
GPA	2.743 (0.691)	2.845 (0.627)	3.077 (0.540)	3.032 (0.555)	2.893 (0.647)
Time to 4-year (years)	2.322 (0.467)	2.449 (0.498)	2.435 (0.496)	2.479 (0.500)	2.381 (0.486)
Credits enrolled	62.557 (16.379)	73.465 (14.508)	71.169 (12.020)	73.553 (12.752)	67.016 (15.279)
Credits earned	55.210 (15.646)	66.547 (11.859)	66.571 (10.140)	68.979 (9.999)	60.919 (14.553)
Core credits enrolled	35.901 (25.490)	47.553 (29.099)	45.058 (26.171)	51.506 (25.076)	40.946 (26.423)
Core credits earned	31.569 (23.003)	43.171 (26.361)	42.161 (24.335)	48.516 (23.109)	37.248 (24.409)
FOSC credits enrolled	23.513 (10.536)	33.008 (11.447)	25.874 (9.538)	31.416 (9.676)	25.360 (10.483)
FOSC credits earned	20.806 (9.594)	30.441 (10.123)	24.135 (8.582)	29.724 (8.368)	23.096 (9.615)
Credit Loss	0.594 (1.832)	0.927 (2.315)	0.705 (2.016)	0.854 (1.987)	0.668 (1.936)
<i>Post-Transfer Characteristics</i>					

Continued on next page

Table 11 continued from previous page

	Neither	FOSC only	TCC only	Both	Total
GPA	2.557 (0.981)	2.705 (0.920)	2.861 (0.891)	2.878 (0.863)	2.700 (0.949)
Credits enrolled	68.000 (30.036)	61.733 (24.929)	64.678 (25.070)	61.909 (22.256)	66.059 (27.625)
Credits earned	60.173 (30.459)	55.761 (25.352)	59.323 (25.468)	57.059 (22.938)	59.449 (27.991)
Reverse transfer	0.283 (0.451)	0.329 (0.470)	0.270 (0.444)	0.230 (0.421)	0.276 (0.447)
<i>Degree outcomes</i>					
BA within 2 years, post-transfer	0.408 (0.491)	0.532 (0.499)	0.537 (0.499)	0.640 (0.480)	0.479 (0.500)
BA within 3 years, post-transfer	0.587 (0.492)	0.681 (0.466)	0.697 (0.460)	0.748 (0.434)	0.644 (0.479)
BA within 4 years, total	0.322 (0.467)	0.396 (0.489)	0.395 (0.489)	0.469 (0.499)	0.363 (0.481)
BA within 6 years, total	0.635 (0.481)	0.698 (0.460)	0.726 (0.446)	0.769 (0.421)	0.682 (0.466)
Time to BA, post-transfer (years)	2.352 (0.898)	2.040 (0.824)	2.138 (0.876)	1.903 (0.822)	2.216 (0.892)
Time to BA, total (years)	4.623 (0.961)	4.454 (0.935)	4.533 (0.982)	4.361 (0.960)	4.558 (0.971)
<i>N</i>	10,773	800	7,779	1,617	20,969

Notes: Standard deviations in parentheses. Post-transfer GPA N: $N_{Neither} = 10,763$; $N_{FOSC} = 800$; $N_{TCC} = 7,773$; $N_{Both} = 1,613$. Time to BA is defined only for BA recipients; N: $N_{Neither} = 7,070$; $N_{FOSC} = 570$; $N_{TCC} = 5,830$; $N_{Both} = 1,270$. Credit loss is defined as the number of courses with at least 50% of credits denied from the receiving institution.

Table 12: Summary Statistics by Entry Cohort

	2018	2019	Total
<i>Treatment Group</i>			
Neither	0.510 (0.500)	0.518 (0.500)	0.514 (0.500)
FOSC only	0.027 (0.161)	0.050 (0.217)	0.038 (0.192)
TCC only	0.388 (0.487)	0.354 (0.478)	0.371 (0.483)

Continued on next page

Table 12 continued from previous page

	2018	2019	Total
Both	0.075 (0.264)	0.079 (0.270)	0.077 (0.267)
Time to 4-year (years)	2.383 (0.486)	2.378 (0.485)	2.381 (0.486)
<i>Instruments</i>			
Average TCC section count	14.274 (13.263)	20.413 (18.706)	17.310 (16.473)
Average FOSC section count	14.379 (10.014)	12.175 (8.044)	13.289 (9.160)
Average remote TCC sections	0.037 (0.036)	0.074 (0.054)	0.055 (0.049)
Average remote FOSC sections	0.023 (0.018)	0.041 (0.024)	0.032 (0.023)
<i>Pre-transfer</i>			
GPA	2.879 (0.633)	2.908 (0.660)	2.893 (0.647)
Credits enrolled	67.240 (15.219)	66.787 (15.338)	67.016 (15.279)
Credits earned	61.099 (14.500)	60.735 (14.605)	60.919 (14.553)
<i>N</i>	10,599	10,370	20,969

Notes: Standard deviations in parentheses. Credit loss is defined as the number of courses with at least 50% of credits denied from the receiving institution.

B.3 Implementation Details

Computing the identified sets from level sets is intensive (Tamer, 2010). The recent increase in access to high-performance computing (HPC) servers has made it feasible for researchers to leverage their computational power to conduct analysis. Yet, due to tight security requirements by THECB and TEA, the Texas ERC data are only allowed to be accessed within the ERC-provided secure servers and cannot be directly accessed from any other computer or server. Because these servers are shared across users and not intended for heavy computational procedures, it was challenging to incorporate a rule-of-thumb grid search procedure. To address this challenge, we used Differential Evolution

(DE; Storn and Price, 1997), an evolutionary optimization algorithm, as in Beresteanu, Molchanov, and Molinari (2011). To summarize, in the algorithm, a candidate to replace a current member of the population is created by combining information from members of the current population; it is accepted into the population as a replacement if it satisfies a certain criterion. This feature could be leveraged for set estimation, treating the criterion for admission to the new generation as the criterion for inclusion in the identified set. Then, the points admitted could be considered as those that belong to the region of interest. For each of the choice model parameters α_{tk} , $k = 0, 1, 2$, we set the parameter space to be a closed interval of $[-1, 1]$. For the mixture weights, we searched over a closed interval of $[0, 1]$.

For the outcome model parameters, we conducted a grid search over the closed, outcome-specific interval after dividing them into equally spaced bins. For credit loss, we searched over $\theta_{t2} \in [-0.8, 0.8]$, split into 200 bins. For degree attainment 3 years within transfer, we searched over $\theta_{t1} \in [-1.5, 2.5]$, and $\theta_{t2} \in [-0.0015, 0.0015]$. For degree attainment 6 years since enrollment, we searched over $\theta_{t1} \in [-1.5, 2.5]$, and $\theta_{t2} \in [-0.001, 0.001]$. All of them were split into 20 bins. As mentioned in the main text, we estimated the identified sets for $\rho = 0.2$ and $\rho = 0.5$. The next subsection provides the results with $\rho = 0.2$.

B.4 Alternative Parameterization

We provide the results from our analysis using an alternative specification of $F_{e|V,X,Z}$. The conclusions and the discussions from Section 6 do not change.

Table 13: Bounds on Predicted Outcomes when $X = 60$, $\rho = 0.2$

Number of Courses with 50% or more Credits Lost				
Treatment	Neither	FOSC only	TCC only	Both
$E[\hat{Y}(t) X = 60]$	[0.265, 0.651]	[0.410, 0.844]	[0.361, 0.603]	[0.410, 0.748]
Bachelor's Degree Attainment (6 Years)				
Treatment	Neither	FOSC only	TCC only	Both
$E[\hat{Y}(t) X = 60]$	[0.404, 0.621]	[0.423, 0.813]	[0.545, 0.762]	[0.552, 0.844]
Bachelor's Degree Attainment (3 Years, Post-Transfer)				
Treatment	Neither	FOSC only	TCC only	Both
$E[\hat{Y}(t) X = 60]$	[0.362, 0.572]	[0.418, 0.792]	[0.515, 0.735]	[0.534, 0.830]
Observations	10,773	800	7,779	1,617

Table 14: Bounds on θ_{t1} , $\rho = 0.2$

Bachelor's Degree Attainment (6 Years)				
Treatment	Neither	FOSC only	TCC only	Both
$\hat{\theta}_{t1}$	[0.395, 0.395]	[0.395, 0.605]	[0.395, 0.605]	[0.395, 0.816]
Bachelor's Degree Attainment (3 Years, Post-Transfer)				
Treatment	Neither	FOSC only	TCC only	Both
$\hat{\theta}_{t1}$	[0.395, 0.605]	[0.395, 0.816]	[0.605, 0.816]	[0.605, 0.816]
Observations	10,773	800	7,779	1,617

B.5 Additional Results

Table 15: Bounds on θ_{t2} for Degree Attainment Outcomes

Bachelor's Degree Attainment (6 Years)				
Treatment	Neither	FOSC only	TCC only	Both
$\{\hat{\theta}_{t2,min}, \hat{\theta}_{t2,max}\}$	{0.0010, 0.0015}	{0.0015, 0.0010}	{0, 0}	{0, -0.0010}
Bachelor's Degree Attainment (3 Years, Post-Transfer)				
Treatment	Neither	FOSC only	TCC only	Both
$\{\hat{\theta}_{t2,min}, \hat{\theta}_{t2,max}\}$	{-0.0010, 0.0010}	{-0.0010, 0.0010}	{0, 0.0006}	{0.0004, -0.0008}

Notes: $\hat{\theta}_{t2,min}$ denotes the smallest value of $\hat{\theta}_{t2}$ for the smallest value of $\hat{\theta}_{t1}$ in the identified set; $\hat{\theta}_{t2,max}$ denotes the largest value of $\hat{\theta}_{t2}$ for the largest value of $\hat{\theta}_{t1}$ in the identified set.

Table 16: Bounds on θ_{t2} for Degree Attainment Outcomes, $\rho = 0.2$

Bachelor's Degree Attainment (6 Years)				
Treatment	Neither	FOSC only	TCC only	Both
$\{\hat{\theta}_{t2,min}, \hat{\theta}_{t2,max}\}$	{0.0001, 0.0003}	{0.0005, 0}	{-0.0010, -0.0009}	{-0.0009, 0.0005}
Bachelor's Degree Attainment (3 Years, Post-Transfer)				
Treatment	Neither	FOSC only	TCC only	Both
$\{\hat{\theta}_{t2,min}, \hat{\theta}_{t2,max}\}$	{-0.0005, -0.0010}	{0.0004, -0.0004}	{-0.0015, -0.0013}	{0.0015, 0.0002}

Notes: $\hat{\theta}_{t2,min}$ denotes the smallest value of $\hat{\theta}_{t2}$ for the smallest value of $\hat{\theta}_{t1}$ in the identified set; $\hat{\theta}_{t2,max}$ denotes the largest value of $\hat{\theta}_{t2}$ for the largest value of $\hat{\theta}_{t1}$ in the identified set.

C Appendix C: Technical Proofs

For convenience, we first present the proof of Theorem 2, then show the proof of Theorem 1.

Proof of Theorem 2 Observe that $\mathbf{V}(D, X, Z; g)$ is always a convex polytope, since they are a form of:

$$\{v \in \mathbb{R}^T : (v_d - v_0 \leq g_1(Z_1) - g_d(Z_d) \wedge v_d - v_1 \leq g_2(Z_2) - g_d(Z_d) \wedge \dots)\}.$$

Moreover, $\mathbf{D}(V, X, Z; g)$ is closed since it is a collection of finite discrete values. Therefore, they are random closed sets.⁶ The distribution of V conditional on X and Z that generated the observed conditional choice probabilities in the dataset is that of a measurable selection of $\mathbf{V}(D, X, Z; g)$ if and only if the observed conditional distribution of choices are a measurable selection of $\mathbf{D}(V, X, Z; g)$. By Artstein's inequality, the distribution of D conditional on X, Z is a measurable selection of the conditional law generated by X, Z and θ^D if and only if $T_{\theta^D}(A|X, Z) \geq P(d \in A|X, Z)$ for all $A \subseteq \mathcal{T}$, and for all X and Z .⁷ Because $\Theta_I^D(P_0)$ includes all such parameters satisfying this inequality given the dataset $(Y, D, X, Z) \sim P_0$, and because $\Theta_I^{D*}(P_0) = \Theta_I^D(P_0)$, this completes the proof.

Proof of Theorem 1 This proof is logically similar to Tsiatis (1975), Theorem 1. Without loss of generality, we consider $F_{e|X,Z} = F_{V|X,Z}$, since this is the case in which the observed outcomes can provide the most information about $(F_{V|X,Z}, g(\cdot))$. For the “if” argument, Y is observed conditional on D , so since the sharp identified set, by definition, must include all parameters $\theta = (g(\cdot), F_{V|X,Z})$ that could rationalize the distribution of D conditional on X and Z in the data, any $\theta \in \Theta_I^*(P_0)$ must be in $\Theta_I(P_0)$. For the “only if” argument, we show that any $\theta \in \Theta_I^D(P_0)$ cannot be rejected by additionally observing the outcomes. Without the knowledge of $m(\cdot)$, for each X, Z and $t \in \mathcal{T}$, the outcome predictions are equivalent to the support of $\mathcal{Y}$. Denote the joint prediction of choices and outcomes given x and z as $\mathcal{Y}^d(x, z; \theta)$. Because the probability that $\mathcal{Y} \cap A \neq \emptyset$ for any $A \subseteq \mathcal{Y}$ is 1, for any $K \subseteq \mathcal{Y} \times \mathcal{T}$, the capacity functional of the joint prediction of outcomes and choices, $P(\mathcal{Y}^d(X, Z; \theta) \cap K|X, Z)$, is equal to the capacity functional for the choice predictions, $P(\mathbf{D}(V, X, Z; g) \cap A^d \neq \emptyset|X, Z)$, where $A^d \subseteq \mathcal{T}$. Because $P_0((y, D) \in K|X, Z)$ is no larger than $P_0(D \in A|X, Z)$, for any $(F_{V|X,Z}, g(\cdot)) \in \Theta_I^D(P_0)$, any $\theta \in \Theta_I^D(P_0)$ must be in $\Theta_I^{D*}(P_0)$, as it is

⁶Molchanov and Molinari (2018), Example 1.5, shows that a set of finite random points is a random closed set. Example 1.6 shows that random polytopes that appear as convex hulls of any (fixed or random) number of random points in the Euclidean space are random closed sets.

⁷Another way to show this result is that by Theorem 1 of Chesher and Rosen (2017), $P_0(D = d|X, Z)$ is selectionable with respect to conditional law induced by $\mathbf{D}(V, X, Z; g)$ if and only if $F_{V|X,Z}$ is selectionable with respect to the conditional distribution of $\mathbf{V}(D, X, Z; g)$ when $D \sim P_0(D = d|X, Z)$; by their Theorem 2 (ii), Θ_I^D is the identified set of θ^D .

characterized as follows by Artstein's inequality:

$$\Theta_I^*(P_0) = \left\{ \theta^D \in \Theta^D : P(\mathcal{Y}^d(x, z; \theta) \cap K \neq \emptyset | X, Z) \geq P_0(K | X, Z), a.e., \forall \text{Supp}(Z), \forall \text{Supp}(X), \forall K \subseteq \mathcal{F}(\mathcal{Y}) \times \mathcal{T} \right\}.$$

This completes the proof.

Proof of Corollary 1

1. By our models (3) - (4), and stochastic independence of e_d from Z conditional on D, X , $F_{e_d|D, X, Z}$ depends on D only through $F_{V|X, Z}$. Because $\mathbf{V}(D, X, Z; g)$ is constructed from $(g, F_{V|X, Z})$ such that $F_{V|X, Z}$ is selectable with respect to the conditional distribution of $\mathbf{D}(V, X, Z; g)$, for all $V \in \mathbf{V}(D, X, Z; g)$, $\forall d \in \mathcal{T}$, any pair $(g, F_{V|X, Z})$ satisfies $D = \underset{d \in \mathcal{T}}{\operatorname{argmax}} g_d(Z_d) + v_d$. By stochastic independence, $F_{e_d|D, X, V, Z} = F_{e_d|X, Z, V}$, for all $d \in \mathcal{T}$.
2. The argument follows that for the proof of part (1), but by mean independence, we have $E[e_d|D, V, X, Z] = E[e_d|V, X, Z]$, for all $d \in \mathcal{T}$.

Proof of Corollary 2 Similarly to the proof of Theorem 2, the distribution of V conditional on X and Z that generated D is a measurable selection of $\mathbf{V}(D, X, Z; g)$ such that V satisfies Assumption I3, which holds if and only if the observed conditional distribution of choices are a measurable selection of $\mathbf{V}(D, X, Z; g)$ that is stochastically independent of Z . This implies that the distribution of its measurable selections does not depend on Z conditional on X . This restricts $F_{V|X, Z}$ in the parameter space to those that are stochastically independent of Z . The rest of the proof follows from the proof of Theorem 2.

Proof of Theorem 3 This result is a direct application of Han and Kaido (2026), Theorem 2. Therefore, we verify that the conditions required for that result are satisfied. From the discussion in the proof of Theorem 2, $\mathbf{V}(D, X, Z; g)$ is a random closed set. By construction, the random variables V underlying the dataset for an individual who selected D given X, Z and the model parameters are in $\mathbf{V}(D, X, Z; g)$ w.p.1. provided that $\mathbf{V}(D, X, Z; g)$ is constructed using $\theta^D \in \Theta_I^D(P_0)$. The conditions in their Assumption 3 are satisfied by Assumption R1 (iii) and because $D = d$ given X, Z and $\theta^D \in \Theta_I^D$ if and only if $V \in \mathbf{V}(D, X, Z; g)$, providing $E[e_d|D, X, Z, V] = E[e_d|X, V, Z]$, and by Assumption I2, we have $E[e_d|X, V, Z] = E[e_d|X, V]$. Finally, their Assumption 4 is satisfied by Assumption R1-(ii). This completes the proof.

Before proving Lemma 1, we establish the following Lemma, which is an adaptation of Beresteanu, Molchanov, and Molinari (2008), Theorem 4.1. First define $Q(\mathbf{D}(X, Z; \theta^D))$ as follows:

$$Q(\mathbf{D}(X, Z; \theta^D)) = \{q : [\mathbf{1}(d = t_0), 0, 0, 0, \dots], [0, \mathbf{1}(d = t_1), 0, 0, 0, \dots], \dots, [0, 0, 0, \mathbf{1}(d = t_T)], \\ d \in \text{Sel}(\mathbf{D}(V, X, Z; g))\},$$

and let $S = \{u = [u_1, \dots, u_{|\mathcal{T}|}]' : u_i \in \{0, 1\}, i = 1, 2, \dots, |\mathcal{T}|\}$.

Lemma C1 *The following are equivalent:*

- $\{\theta \in \Theta^D : P_0(D|X, Z) \in \mathbb{E}[Q(\mathbf{D}(X, Z; \theta^D))|X, Z], \forall X, Z\}$.
- $\{\theta \in \Theta^D : P_0(D|X, Z) \leq T_{\theta^D}(A|X, Z), a.e., \forall A \subseteq \mathcal{T}, X, Z\}$.
- $\{\theta \in \Theta^D : \max_{u \in B} \left(u'F(D|X, Z) - E[h(Q(\mathbf{D}(X, Z; \theta^D)), u)|X, Z] \right) = 0, \forall X, Z\}$

where $T_{\theta^D}(A|X, Z) = P(\mathbf{D}(V, X, Z; g) \cap A \neq \emptyset | X, Z), V \sim F_{V|X, Z}$.

Proof of Lemma C1 The proof adapts the one for Theorem 4.1 in Beresteanu, Molchanov, and Molinari (2008). By Theorem 2.1 of Beresteanu, Molchanov, and Molinari, 2011, the first set is equivalent to:

$$\{\theta \in \Theta^D : u'F(D|X, Z) \leq E[h(Q(\mathbf{D}(X, Z; \theta^D)), u)|X, Z], \forall u \in B = \{u : \|u\| \leq 1\}, \forall X, Z\}, \quad (18)$$

where their Assumptions 2.1 and 2.2 are satisfied by our setup, and Assumption 2.3 is satisfied because the set $\mathbf{D}$ contains a finite number of vectors of ones and zeros and therefore is integrably bounded. The equivalence of the set in equation (18) and the one in the third condition has been shown in Beresteanu, Molchanov and Molinari, 2011, Theorem C.1. To show that the first condition and the second condition are equivalent, observe that from the equation in equation (18), because the support function is positively homogeneous, this condition is equivalent to $u'P(D = d|X, Z) \leq E[h(Q(\mathbf{D}(X, Z; \theta^D)), u)|X, Z], \forall u \in \mathbf{R}^T$. Take any $\bar{u} \in \mathbf{R}^T$ such that its entries are zeros and ones. The coordinates of $\bar{u}$ correspond to the subsets of $\mathcal{T}$. By definition, the scalar product $\bar{u}'P(D = d|X, Z)$ equals $P(d \in K_{\bar{u}}|X, Z)$, where we use K_u to denote the subsets of $\mathcal{T}$ that corresponds to $\bar{u}$. Moreover,

$$h(\mathbb{E}(Q(\mathbf{D}(X, Z; \theta^D)), \bar{u}) = E[h(Q(\mathbf{D}(V, X, Z; g))|X, Z), \bar{u})|X, Z] = E\left[\sup_{V \in \mathbf{V}(D, X, Z; g)} \bar{u}'d(V)|X, Z\right],$$

where we use $d(V)$ to denote $d \in \text{Sel}(\mathbf{D}(V, X, Z; g))$ while making explicit the dependence of d on V . The realizations of any $d \in \mathbf{D}(V, X, Z; g)$ are vectors of zeros and ones. Therefore, given the choice of $\bar{u}$, the value of $\bar{u}'1(d=1)$ equals one if $d \in \mathbf{D}(V, X, Z; g)$ and 0 otherwise. Therefore, the first condition reduces to:

$$\begin{aligned} P(d \in K_{\bar{u}}|X, Z) &= \bar{u}'F(d|X, Z) \leq E[h(Q(\mathbf{D}(X, Z; \theta^D), \bar{u})|X, Z)] = E\left[\sup_{V \in \mathbf{V}(D, X, Z; g)} \bar{u}'d(V)|X, Z\right] \\ &= E[1(\mathbf{D}(V, X, Z; g) \cap K_{\bar{u}} \neq \emptyset)|X, Z] = P(\mathbf{D}(V, X, Z; g) \cap K_{\bar{u}} \neq \emptyset|X, Z). \end{aligned} \tag{19}$$

Suppose now that the second inequality holds. The second inequality implies $d \in \text{Sel}(\mathbf{D}(V, X, Z; g))$ by Theorem 2. This implies that there exists $V \in \text{Sel}(\mathbf{V}(D, X, Z; g))$ such that $d \in Q(\mathbf{D}(X, Z; \theta^D))$ with probability 1, and the same relationship as displayed in (19) applies.

Proof of Lemma 1 We verify the conditions of Chernozhukov, Hong and Tamer (2007), Theorem 3.1. For Condition C.1., Condition (a) is satisfied by our assumption. For condition (b), by the aforementioned assumption, $g(\cdot)$ is continuous, and the parameter space is compact. Then, by the theorem of maximum, $\mathbf{D}$ is upper-hemicontinuous with nonempty and compact values and therefore has a closed graph. This implies that the population objective function has a closed graph and therefore, $\sup_{\theta_N^D \rightarrow \theta} \mathbb{E}(Q_{\theta_N^D}) \subseteq \mathbb{E}(Q_{\theta^D})$. Then we have $\liminf_{\theta_n \rightarrow \theta} s(\theta_n) = s(\theta)$, since, for all $u \in S$ corresponding to subset K_u ,

$$\begin{aligned} \liminf_{\theta_n \rightarrow \theta} s(\theta_n) &\geq \liminf_{\theta_n \rightarrow \theta} \left(u'P(D = d|X, Z) - \mathbb{E}[h(Q_{\theta_N^D}(D(V, X, Z; g), u|X, Z)] \right) \\ &= \left(u'P(D = d|X, Z) - \liminf_{\theta_n \rightarrow \theta} \mathbb{E}[h(Q_{\theta_N^D}(D(V, X, Z; g), u|X, Z)] \right) \\ &= \left(u'P(D = d|X, Z) - \mathbb{E}[h(Q_{\theta^D}(D(V, X, Z; g), u|X, Z)] \right) \\ &= \left((P(d \in K_u|X, Z) - P(\mathbf{D}(V, X, Z; g) \cap K_u|X, Z)) \right) = s(\theta), \end{aligned}$$

where the last equality is from Lemma C1. Because the rest of operations to construct $Q_N(\theta^D)$ – taking the larger value between $s(\theta)$ and 0, squaring, and taking the weighted average – preserve lower semicontinuity, our objective function is lower semicontinuous in θ^D . By Theorem 2, the set of minimizers of $Q(\theta^D)$ is the identified set. The sample criterion function $Q_N(\theta^D)$ is jointly measurable in the parameters $\theta^D \in \Theta^D$ and the data defined on a complete probability space by our baseline conditions. By Assumption E1 - 1, the sample frequency estimator satisfies $\sup_{x \in \mathcal{X}, z \in \mathcal{Z}} |\hat{P}(d \in K|X, Z) -$

$P(d \in K|X, Z) = o_p(1)$ by the law of large numbers, and the simulated capacity functionals' growth rate ensures that, $\sup_{\theta^D \in \Theta^D} |Q_N(\theta^D) - Q(\theta^D)| = O_p(1/\sqrt{N})$. Therefore the sample criterion function is uniformly not smaller than the population function in large samples since $Q(\theta^D) \leq |Q_N(\theta^D) - Q(\theta^D)| + Q_N(\theta^D)$, and their condition (d) is satisfied. The sample criterion function $Q_N(\theta^D)$ satisfies $\sup_{\theta^D \in \Theta_I^D} |Q(\theta^D) - Q_N(\theta^D)| = O_p(1/N)$, satisfying their condition (e). Therefore, Condition C.1 of Chernozhukov, Hong, and Tamer (2007) is satisfied. Then, by the result of their Theorem 3.1 $\Theta_I^D \subseteq \hat{\Theta}_I^D$ w.p.a 1 and $d_H(\hat{\Theta}_I^D, \Theta_I^D) = o_p(1)$.

Proof of Theorem 4 We verify that the conditions of Corollary 4.3 of Kaïdo and White (2014) are satisfied. Recall that $\mathbf{R}^{Y(t)}(Y, X, Z; \theta)$ is defined as:

$$\mathbf{R}^{Y(t)}(Y, X, Z; \theta) = [r = [y - m_t(X) - \mathbb{E}[e_D|X, Z, V]] : V \in \text{Sel}(\mathbf{V}(D = t, X, Z; \Theta^D))],$$

Since the above expression is a closed interval, we obtain the following restrictions:

$$bE[r|D, X, Z] \leq h(\mathbb{E}[\mathbf{R}^{Y(t)}(Y, X, Z; \theta)|D, X, Z], u), \quad b \in \{-1, 1\}$$

which, by the definition of support functions, provide us with the moment conditions $r1^{Y(t)}$ and $r2^{Y(t)}$:

$$\begin{aligned} r1^{Y(t)}(Y, X, Z; \theta) &= ((X\beta_t - (E[Y|D, X, Z] - \inf_{V \in \mathbf{V}} \mathbb{E}_{\theta^D}[e_t|X, V])), \\ r2^{Y(t)}(Y, X, Z; \theta) &= ((E[Y|D, X, Z] - \sup_{V \in \mathbf{V}} \mathbb{E}_{\theta^D}[e_t|X, V]) - X\beta_t). \end{aligned}$$

Condition 2.1 (i) of the theorem is satisfied because, for all (Y, X, Z) and θ , $\mathbf{R}^{Y(t)}(D, X, Z; \theta)$ consists of continuous functions where θ is mapped from a compact set, so it is compact; support function of a closed compact set is continuous, so $r1^{Y(t)}$ and $r2^{Y(t)}$ are continuous. Since averaging, taking the maximum with zero, squaring, and summing are continuous operations, the resulting objective function $Q^{Y(t)}(\theta)$ is continuous. All our variables are defined on a complete, nonatomic probability space; measurability of the functions are satisfied from Assumption R1. Their assumption 2.3 is satisfied because, by Proposition 5.3.9 of Molchanov (2017), $Q(\theta^D)$ is a normal integrand; therefore, by Proposition 5.3.13 (i) of Molchanov (2017), our choice model identified set, $\hat{\Theta}_I^D$ generated by a level set estimator, is Effros-measurable. To verify their Assumption 4.1., observe that by the strong law of large numbers for set-valued random variables (Beresteanu

and Molinari, 2008, Theorem A.1), $d_H(\frac{1}{N_d} \bigoplus_{i=1}^{N_d} \mathbf{R}_i^{Y(t)}(Y, X, Z; \theta), \mathbb{E}[\mathbf{R}^{Y(t)}(\theta)|D, X, Z]) = o_p(1)$. Denote $\widehat{\mathbb{E}}[\mathbf{R}_\theta^{Y(t)}|D, X, Z] = \frac{1}{N_d} \bigoplus_{i=1}^{N_d} \mathbf{R}_i^{Y(t)}(Y, X, Z; \theta)$. By Lemma A.1 of Beresteanu and Molinari (2008), $d_H(\widehat{\mathbb{E}}[\mathbf{R}_\theta^{Y(t)}|D, X, Z], \mathbb{E}[\mathbf{R}^{Y(t)}(\theta)|D, X, Z]) = o_p(1)$. Then, because $\max_{u \in \{-1, 1\}} (-E[h(\mathbb{E}[\mathbf{R}^{Y(t)}(\theta)|D, X, Z], u)) \leq 0$ if and only if $\theta \in \Theta_I$, and when $u = -1$ the maximum is obtained by $(-E[Y|D, X, V] + m_d(X) + \inf_{V \in \mathbf{V}} \mathbb{E}[e_d|d, X, V])$, and when $u = 1$ the max is obtained by $(E[Y|D, X, Z] - m_d(X) - \sup_{V \in \mathbf{V}} \mathbb{E}[e_d|D, X, V])$, $\sup_{\Theta} \{Q^{Y(t)}(\theta) - \widehat{Q}_N^{Y(t)}(\theta)\}_+ = o_p(1)$ by the continuous mapping theorem. Lemma 1 provides $\widehat{\Theta}_I^D$ is such that $d_H(\widehat{\Theta}_I^D, \Theta_I^D) = o_p(1)$, and $P(\Theta_I^D \subseteq \widehat{\Theta}_I^D) \rightarrow 1$. This suffices to prove the theorem, since our estimator is defined using the level set estimator; for some $\hat{\epsilon}_n$ such that $\hat{\epsilon}_n/a_{N_d} = o_p(1)$ and $\lim_{N_d \rightarrow \infty} P(\sup_{\theta \in \Theta_I} Q^{Y(t)}(\theta) \leq \hat{\epsilon}_n/a_{N_d}) = 1$. Finally, our second-stage estimator satisfies the following definition of the two-stage set estimator:

1. The parameter vector consists of two sub-vectors, a “first-stage parameter” and a “second-stage” parameter
2. The identification of the first-stage parameter depends on neither the identification nor the value of the second-stage parameter
3. The identified set for the second-stage parameter depends on the first-stage parameter through the criterion function.

This completes the proof.

Proof of Theorem 5 We verify the conditions of Chernozhukov, Hong, and Tamer (2007), Theorem 4.2, as our estimator is defined by moment inequality restrictions. In particular, the weights for each cell can be seen as part of the weighting matrix for the moment inequality restrictions pertaining to each cell. Therefore, verifying this condition suffices to prove Theorem 5. Their condition M.2 (a) holds under Assumption E1, our baseline setup and our definition of the criterion function. To see that Condition M.2 (c) is satisfied, the first part of Lemma C1 shows that the moment functions are given by the sum of a linear function of u and a Lipschitz function of u , with Lipschitz constant equal to 1. The processes $\{u'P(d_i|X, Z), u \in \mathbf{S}^T, i \leq N, N \geq 1\}$, where $\mathbf{S}^T = \{u \in \mathbb{R}^T : \|u\| = 1\}$, are manageable with a constant envelope function by Lemma 3 of Andrews and Shi (2013), since they are constructed in the same way as their $\mathcal{G}_{box}$. $E[h(Q_\theta^D, u)|X, Z]$ is manageable because they are Lipschitz in u , with Lipschitz constant equal to 1. By Lemma 2.13 of Pakes and Pollard (1989), the class of functions $\{E[h(Q_\theta^D, u)|X, Z], u \in \mathbf{S}^T\}$ is Euclidean with envelope equal to a constant and is therefore manageable. By Lemma E1 of Andrews and Shi (2013),

$\{u'P(d_i|X, Z) - E[h(Q_\theta^D, u)|X, Z], u \in \mathbf{S}^T, i \leq N, N \geq 1\}$ is manageable with a constant envelope. A manageable class for a constant envelope is a universal Donsker class (Pollard, 1989), so they satisfy the P-Donsker condition. To see that their condition (d) holds, observe that:

$$\begin{aligned} C \left(\sum_{u \in S} u'P(D|X, Z) - E[h(Q_\theta^D, u)|X, Z] \right) &\geq \max_{u \in S} (u'P(D|X, Z) - \mathbb{E}[Q_\theta^D|X, Z]) \\ &= d_H(P(D|X, Z), \mathbb{E}[Q_\theta^D|X, Z]), \end{aligned}$$

where C is some constant larger than 1. We have:

$$d_H(P(D|X, Z), \mathbb{E}[Q_\theta^D|X, Z]) + d_H(P(D|X, Z), \mathbb{E}[Q_{\theta^*}^D|X, Z]) \geq d_H(\mathbb{E}[Q_\theta^D|X, Z], \mathbb{E}[Q_{\theta^*}^D|X, Z]),$$

by the triangle inequality, so for $\theta \in \Theta_I^D$ and $\theta^* \in \Theta^D \setminus \Theta_I^D$,

$$\begin{aligned} d_H(\mathbb{E}[Q_\theta^D|X, Z], \mathbb{E}[Q_{\theta^*}^D|X, Z]) &\leq d_H(P(D|X, Z), \mathbb{E}[Q_{\theta^*}^D|X, Z]) \leq C' \sum_{u \in S} \left[u'P(D|X, Z) - E[h(Q_{\theta^*}^D, u)|X, Z] \right]_+ \\ &= C \sum_{A \in \mathcal{T}/\emptyset} \|m(A, X, Z; \theta^{*D})\|_+, \end{aligned}$$

where C and C' are universal constants. Because $d_H(\mathbb{E}[Q_\theta^D|X, Z], \mathbb{E}[Q_{\theta^*}^D|X, Z])$, for $\theta \in \Theta_I^D$ and all $\theta^* \in \Theta^D$ is proportional to $d_H(\Theta_I^D, \theta^*)$ by construction, once θ^* is bounded away from Θ_I^D , each of the moment equations are bounded below by a positive constant proportional to the distance away from the identified set so long as conditions 1, 2, 3, and 6 of Assumption E1 are satisfied. Condition (e) holds because the cell weights are nonzero constants under Assumption E1-2, and converge to w as $N \rightarrow \infty$ by the law of large numbers. These conditions together establish the necessary conditions for Chernozhukov, Hong, and Tamer (2007) Theorem 4.2. Then, by that theorem, their conditions C.2, C.4, and C.5 hold. Because we have proven their condition C.1 in our proof of Lemma 1, the conditions for their Theorem 3.3 are satisfied. This completes the proof.