Robust Counterfactual Analysis for Nonlinear Panel Data Models[*]

Yan Liu[†]

September 2, 2024

Abstract

This paper studies robust counterfactual identification in a wide variety of nonlinear panel data models. We impose only mild assumptions, including time homogeneity on the distribution of unobserved heterogeneity and index separability on the structural function. We derive the sharp identified set for the distribution of the counterfactual outcome, noting that point identification is impossible in general. We also provide tractable implementation procedures that circumvent the need to directly search over latent distributions. We propose estimating sharp bounds on counterfactual probabilities based on aggregate intersection bounds and conducting inference using Bonferroni confidence intervals. We apply our approach to empirical data to predict female labor force participation rates under counterfactual values of husband's income, as well as market shares of different saltine cracker brands under counterfactual pricing schemes.

Keywords: Nonlinear panel, discrete choice, counterfactual analysis, sharp partial identification

1 Introduction

A frequent goal in empirical research is to predict the counterfactual behavior of an outcome variable under ceteris paribus manipulations of endogenous explanatory variables. Panel data offers the possibility of dealing with endogeneity due to individual-specific unobserved heterogeneity (or

[*] This paper supersedes my earlier paper "Counterfactual Analysis for Semiparametric Panel Multinomial Choice Models". I am indebted to Hiroaki Kaido for his guidance and encouragement throughout this project. For helpful comments and discussions, I thank Rubaiyat Alam, Eli Ben-Michael, Xu Cheng, Iván Fernández-Val, Jean-Jacques Forneron, Shakeeb Khan, Aureo de Paula, Kirill Ponomarev, Marc Rysman, Davide Viviano, and participants at the BU Econometric Seminar, Greenline Workshop in Econometrics 2023, and AMES2024-China. All errors are my own.

[†] Department of Economics, Boston University. Email: yanliu@bu.edu.

fixed effects) by utilizing multiple observations for a single economic unit over time. In nonlinear models that naturally arise in the context of discrete outcomes, unobserved heterogeneity enters in a nonadditive manner. As a result, counterfactual predictions involve information about the distribution of unobserved heterogeneity. It is desirable to extract this information from data rather than rely on parametric distributional assumptions, but the method for doing so with panel data is not yet fully established.

In this paper, we develop a robust method of counterfactual identification in nonlinear panel data models with minimal assumptions on the distribution of unobserved heterogeneity. The only restriction we impose on the distribution of unobserved heterogeneity is time homogeneity, which can be interpreted as "time is randomly assigned" or "time is an instrument" (Chernozhukov, Fernández-Val, Hahn, and Newey, 2013). We note that when the outcome distribution exhibits mass points (e.g., discrete or mixed), it is generally impossible to point identify both structural parameters and the distribution of the counterfactual outcome. Instead, we derive sharp identified sets for the latter. A crucial tool for this purpose is the set of unobserved heterogeneity that produces the same outcome value, which we refer to as the "U-level set". Time homogeneity simplifies the sharp identified set by allowing us to take intersections of restrictions from observed marginal distributions of the outcome variable across time periods. When it comes to implementation, we further exploit the index separability of the structural function and focus on two important classes of models: monotone transformation models (e.g., binary choice, ordered choice, censored regression, etc.) and multinomial choice models. We provide tractable implementation procedures based on set inclusion relationships of U-level sets and bypass the need to directly search over latent distributions.

We investigate how identifying power varies with the length of time periods and the cardinality of outcome support through numerical experiments. We propose estimating sharp bounds on counterfactual probabilities based on aggregate intersection bounds and performing inference using Bonferroni confidence intervals. We apply our approach to empirical data to predict female labor force participation rates under counterfactual values of husband's income, as well as market shares of different saltine cracker brands under counterfactual pricing schemes. We also consider an extension of our identification strategy to dynamic binary choice models.

This paper is related to three strands of literature. First, there is a growing literature on semiparametric identification of nonlinear panel data models, including Manski (1987), Khan, Ponomareva, and Tamer (2016, 2023), Shi, Shum, and Song (2018), Gao and Li (2020), Khan, Ouyang, and Tamer (2021), Botosaru, Muris, and Pendakur (2023), Gao and Wang (2024), Pakes and Porter (2024). Nonetheless, these works exclusively focus on learning the finite-dimensional structural parameters. We take a step forward to facilitate counterfactual analysis. Second, this paper complements the literature on counterfactual identification in discrete outcome models, including Manski (2007), Chiong, Hsieh, and Shum (2021), Gu, Russell, and Stringham (2024), Tebaldi, Torgovitsky, and Yang (2023). Manski (2007) focused on counterfactual scenarios concerning unrealized choice sets. Chiong et al. (2021) assumed exogeneity of product-specific attributes and proposed using *cyclic monotonicity* to bound counterfactual market shares under changes in these attributes. Tebaldi et al. (2023) and Gu et al. (2024) also consider counterfactuals that manipulate explanatory variables. Tebaldi et al. (2023) restricted explanatory variables to be finitely supported. In this case, searching over latent distributions reduces to a finite-dimensional problem characterized by a finite partition of the space of unobserved heterogeneity, termed the *minimal relevant partition*. Gu et al. (2024) extended this insight to account for model misspecification and model incompleteness. An obvious feature of our approach is that we exploit the panel data structure. Moreover, we allow explanatory variables to be both endogenous and continuous. Third, this paper contributes to the literature on counterfactual identification in nonlinear panel data models, including Hoderlein and White (2012), Chernozhukov et al. (2013), Chernozhukov, Fernández-Val, and Newey (2019), Liu, Poirier, and Shiu (2021), Botosaru and Muris (2024). The identification results of Hoderlein and White (2012) and Chernozhukov et al. (2019) are confined to the subpopulation of “stayer”, i.e., the population for which explanatory variables do not change over time. Chernozhukov et al. (2013) only considered finitely supported explanatory variables. Liu et al. (2021) restricted attention to binary choice models and achieved point identification of average partial effects by imposing index sufficiency on the distribution of fixed effects. Botosaru and Muris (2024) derived bounds on counterfactual survival probabilities in monotone transformation models. Our results differ in that we handle continuous explanatory variables and deliver sharp identified sets of counterfactual distributions for a relatively wide variety of nonlinear models.

The remainder of this paper is organized as follows. Section 2 outlines the setup and specifies the type of counterfactuals under our consideration. Section 3 presents the sharp identified set for the distribution of the counterfactual outcome. Section 4 discusses tractable implementation of the sharp identified set. Section 5 documents the results of numerical experiments. Section 6 addresses estimation and inference. Section 7 contains empirical illustrations using data on female labor force participation and purchases of saltine crackers. Section 8 explores the extension to dynamic binary choice models. Section 9 concludes. All the proofs are collected in Appendix A.

2 Setup

Consider panel data structures of the form

$$Y_t = g(X_t, U_t; \theta_0), \ t = 1, \ldots, T,$$

where, for $i = 1, \ldots, N$, $Y_t = Y_{it} \in \mathcal{Y} \subseteq \mathbb{R}$ denotes an observed scalar outcome, $X_t = X_{it} \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$ denotes explanatory variables, $U_t = U_{it} \in \mathbb{R}^{d_u}$ denotes unobserved heterogeneity, and g is a function known up to a finite-dimensional parameter θ_0. Write $X = (X_1, \ldots, X_T)$.

Example 1 (Monotone Transformation Models)**.** Consider the model

$$Y_t = h(X_t^\top \beta_0 + U_t; \gamma_0),$$

where $\beta_0 \in \mathbb{R}^{d_x}$ is a vector of unknown coefficients, and h is a transformation function that is weakly increasing, right-continuous, and known up to a finite-dimensional parameter γ_0. Here $\theta_0 = (\beta_0, \gamma_0)$. For binary response models, $\mathcal{Y} = \{0, 1\}$ and $h(v; \gamma) = 1\{v \geq 0\}$. For ordered choice models, $\mathcal{Y} = \{0, 1, \ldots, J\}$, $\gamma_0 = (\gamma_0^0, \gamma_0^1, \ldots, \gamma_0^J)^\top$, and $h(v; \gamma) = \sum_{j=0}^{J} 1\{v \geq \gamma^j\}$, where γ_0^j are unknown thresholds satisfying $\gamma_0^j > \gamma_0^{j-1}$ and $\gamma_0^0 = -\infty$. For censored regression, $\mathcal{Y} = [0, \infty)$ and $h(v; \gamma) = \max\{0, v\}$.

Example 2 (Multinomial choice models)**.** Suppose that $\mathcal{Y} = \{0, 1, \ldots, J\}$, and X_t and U_t consist of alternative-specific components:

$$X_t = (X_{0t}, X_{1t}, \ldots, X_{Jt}), \ U_t = (U_{0t}, U_{1t}, \ldots, U_{Jt}),$$

where for each j, $X_{jt} \in \mathbb{R}^k$ and $U_{jt} \in \mathbb{R}$. Consider the model

$$Y_t = \max \arg \max_j X_{jt}^\top \beta_0 + U_{jt},$$

where $\beta_0 \in \mathbb{R}^k$ is a vector of unknown coefficients. Here $\theta_0 = \beta_0$. Note that the normalization $\tilde{X}_{jt} = X_{jt} - X_{0t}, \tilde{U}_{jt} = U_{jt} - U_{0t} \ \forall j$ leads to the same choice.

Assumption 1. *$U_t \overset{d}{=} U_1 | X$ for all t.*

Assumption 1 is a time-homogeneity condition commonly imposed for semiparametric or nonparametric identification of nonlinear panel data models.[1]. It requires the conditional distribution of U_t given X to be the same in each time period. A sufficient condition is that U_t has an error component structure: $U_t = A + V_t$, where $V_t \overset{d}{=} V_1 | X, A$ for all t. It is worth noting that Assumption 1 excludes lagged Y_t from X_t and focuses on static models. On the other hand, Assumption 1 allows U_t to be correlated with X and dependent over time. Moreover, it places no parametric distributional restriction on U_t.

Assumption 2. *θ_0 is known or point-identified.*

Assumption 2 is satisfied for a broad class of structural functions g under Assumption 1 and rich support conditions for U_t and X.

Example 1 (continued)**.** Botosaru et al. (2023) converted monotone transformation models into a collection of binary choice models via time-varying binarization. Then, point identification of θ_0 can be shown by invoking Manski (1987).

Example 2 (continued)**.** Point identification of θ_0 is established in Shi et al. (2018) and Khan et al. (2021). Shi et al. (2018) exploited the cyclic monotonicity property of the choice probability vector. Khan et al. (2021) utilized the subsample of observations in which covariates for all alternatives but one are fixed over time to construct a localized rank-based objective function analogous to Manski (1987).

Fixing a counterfactual value $\underline{x}$ for X_t, we are interested in the distribution of the counterfactual outcome $Y_t(\underline{x})$ that satisfies $Y_t(\underline{x}) = g(\underline{x}, U_t; \theta_0)$. This can be understood as the result of an intervention that exogenously sets the value of X_t to $\underline{x}$, without altering the structural function $g(\cdot; \theta_0)$ or the distribution of U_t. We can form summary measures of the distribution of $Y_t(\underline{x})$ in the spirit of the *average structural function* introduced in Blundell and Powell (2003). In Example 1, we consider the counterfactual survival probability $\Pr(Y_t(\underline{x}) \geq y)$ for $y \in \mathcal{Y} \setminus \inf \mathcal{Y}$. In Example 2, we consider the counterfactual choice probability $\Pr(Y_t(\underline{x}) = y)$ for $y \in \mathcal{Y}$. These counterfactual probabilities

[1] See, e.g., Manski (1987), Abrevaya (2000), Graham and Powell (2012), Hoderlein and White (2012), Chernozhukov et al. (2013), Chernozhukov, Fernández-Val, Hoderlein, Holzmann, and Newey (2015), Khan et al. (2016, 2023), Shi et al. (2018), Gao and Li (2020), Khan et al. (2021), Botosaru et al. (2023), Ouyang and Yang (2023), Wang (2023), Botosaru and Muris (2024), Gao and Wang (2024), Ouyang and Yang (2024), Pakes and Porter (2024).

are important parameters per se in evaluating the impact of counterfactual interventions. Moreover, they can serve as building blocks for various welfare measures. For example, Bhattacharya (2015, 2018) showed that in binary and multinomial choice models, the distribution of compensating and equivalent variation under a range of economic changes can be expressed as closed-form functionals of choice probabilities.

Remark 1. *The framework of Chesher, Rosen, and Zhang (2024) also permits counterfactual analysis. They impose a fixed effect structure on unobserved heterogeneity while leaving the distribution of the fixed effect completely unrestricted. As a result, their approach cannot say much about the counterfactual probability in a single period, which is our focus, because the fixed effect can be arbitrarily moved to justify any outcome. Instead, their approach may potentially bound the probability of switching across multiple periods.*

3 Identification

Notation: For a generic random vector W, let $\mathcal{F}_{W|X} = \{F_{W|X=x} : x \in \mathrm{Supp}(X)\}$ denote the collection of conditional distributions of W given X, where for all $\mathcal{S} \subseteq \mathrm{Supp}(W|X=x)$, $F_{W|X=x} = \Pr(W \in \mathcal{S}|X=x)$.

It is convenient to define the U-level sets as

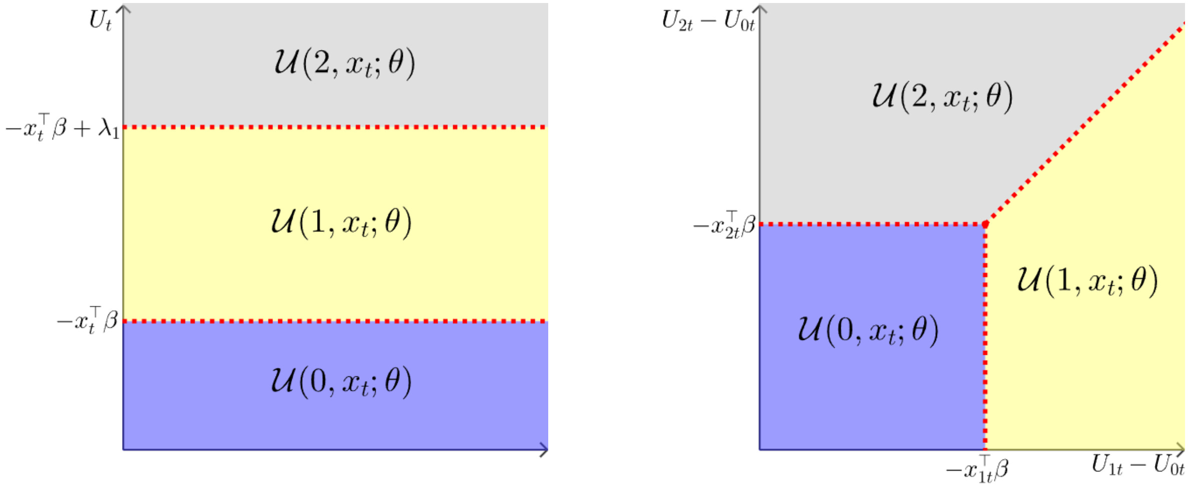
$$\mathcal{U}(y_t, x_t; \theta) = \{u_t : y_t = g(x_t, u_t; \theta)\}$$

so that

$$u_t \in \mathcal{U}(y_t, x_t; \theta) \iff y_t = g(x_t, u_t; \theta).$$

In words, $\mathcal{U}(y_t, x_t; \theta)$ denotes the set of values of U_t that solves $Y_t = g(X_t, U_t; \theta)$ with structural function $g(\cdot; \theta)$ when $Y_t = y_t$ and $X_t = x_t$. Figure 1 contains stylized depictions of U-level sets in Examples 1 and 2 with $\mathcal{Y} = \{0, 1, 2\}$. For any closed subset $\mathcal{T}$ of $\mathcal{Y}$, let

$$\mathcal{U}(\mathcal{T}, x_t; \theta) = \bigcup_{y_t \in \mathcal{T}} \mathcal{U}(y_t, x_t; \theta).$$


U_t
$\mathcal{U}(2, x_t; \theta)$
$-x_t^\top \beta + \lambda_1$
$\mathcal{U}(1, x_t; \theta)$
$-x_t^\top \beta$
$\mathcal{U}(0, x_t; \theta)$
$U_{2t} - U_{0t}$
$\mathcal{U}(2, x_t; \theta)$
$-x_{2t}^\top \beta$
$\mathcal{U}(1, x_t; \theta)$
$\mathcal{U}(0, x_t; \theta)$
$-x_{1t}^\top \beta$
$U_{1t} - U_{0t}$

(a) Example 1: Ordered Choice Model

(b) Example 2: Multinomial Choice Model

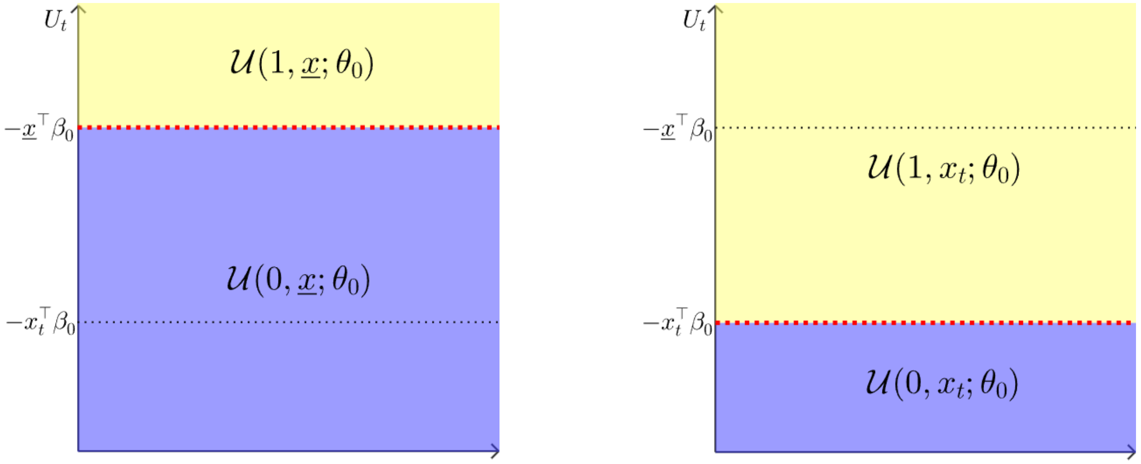
Figure 1: Stylized Depictions of U-Level Sets

We can characterize the distribution of the counterfactual outcome $Y_t(\underline{x})$ as

$$\forall \mathcal{T} \in \mathsf{F}(\mathcal{Y}), F_{Y_t(\underline{x})|X=x}(\mathcal{T}) = F_{U_t|X=x}(\mathcal{U}(\mathcal{T}, \underline{x}; \theta_0)) \text{ } a.e. \text{ } x \in \operatorname{Supp}(X),$$

where $\mathsf{F}(\mathcal{Y})$ denotes the collection of all closed subsets of $\mathcal{Y}$. Therefore, to identify the distribution of $Y_t(\underline{x})$, we need to identify θ_0 and the distribution of $U_t|X = x$ over $\mathcal{U}(\mathcal{T}, \underline{x}; \theta_0)$ for each $\mathcal{T} \in \mathsf{F}(\mathcal{Y})$. The former, as discussed in Section 2, has been studied in the literature for a broad class of nonlinear panel data models. The latter is a new element that emerges in the analysis of counterfactuals. When the outcome distribution exhibits mass points, such as in discrete or mixed distributions, point identification of both elements is impossible. We give a heuristic explanation of why in the binary choice model

$$Y_t = 1\{X_t^\top \beta_0 + U_t \geq 0\}. \tag{1}$$


U_t
$\mathcal{U}(1,\underline{x};\theta_0)$
$-\underline{x}^\top\beta_0$
$\mathcal{U}(0,\underline{x};\theta_0)$
$-x_t^\top\beta_0$
U_t
$-\underline{x}^\top\beta_0$
$\mathcal{U}(1,x_t;\theta_0)$
$-x_t^\top\beta_0$
$\mathcal{U}(0,x_t;\theta_0)$

(a) U-Level Sets under the Counterfactual

(b) U-Level Sets for the Observed Data

Figure 2: Discrepancy of U-Level Sets: Binary Choice Model

As shown in Figure 2, for each $x \in \mathrm{Supp}(X)$, we want to learn how $F_{U_t|X=x}$ allocates probability across $\mathcal{U}(1,\underline{x};\theta_0)$ and $\mathcal{U}(0,\underline{x};\theta_0)$. However, what we observe, $\Pr(Y_t = 1|X = x) = F_{U_t|X=x}(\mathcal{U}(1,x_t;\theta_0))$, only tells us how probability is allocated across $\mathcal{U}(1,x_t;\theta_0)$ and $\mathcal{U}(0,x_t;\theta_0)$, which differ from $\mathcal{U}(1,\underline{x};\theta_0)$ and $\mathcal{U}(0,\underline{x};\theta_0)$ unless $\underline{x} = x_t$. Assumption 1 allows us to also learn from $\Pr(Y_{t'} = 1|X = x)$ for $t' \neq t$, but they may still lead to different U-level sets from what we want. This discrepancy occurs for almost every $x \in \mathrm{Supp}(X)$ if X_t contains at least one continuous component, which is typically required for the point identification of θ_0. As a result, we cannot uniquely determine the distribution of U_t across $\mathcal{U}(1,\underline{x};\theta_0)$ and $\mathcal{U}(0,\underline{x};\theta_0)$

Given the impossibility of point identification, we provide the sharp identified set of the distribution of $Y_t(\underline{x})$ in Theorem 1. The proof is in Appendix A. The sharp identified set relies on the standard definition of *observational equivalence*, that is, we collect all the distributions of $Y_t(\underline{x})$ that can be reproduced by a distribution of U_t consistent with the observed data. A key simplification afforded by Assumption 1 is that, although we observe joint distributions $\mathcal{F}_{Y|X}$, the distribution of U_t is only required to match the marginals $\{\mathcal{F}_{Y_{t'}|X}\}_{t'=1}^{T}$, and we can combine these restrictions by taking intersection across t'. In this sense, a long panel plays an analogous role to that of an instrument with rich variation.

Theorem 1. *Suppose that Assumptions 1 and 2 hold. Then, the sharp identified set for* $\mathcal{F}_{Y_t(\underline{x})|X}$, *denoted by* $\mathsf{F}^*_{Y_t(\underline{x})|X}$*, is given by*

$$\mathsf{F}^*_{Y_t(\underline{x})|X} = \{\mathcal{F}_{Y_t(\underline{x})|X} : \exists \mathcal{F}_{U_t|X} \in \mathsf{F}^*_{U_t|X} \\ \quad s.t.\ \forall \mathcal{T} \in \mathsf{F}(\mathcal{Y}), F_{Y_t(\underline{x})|X=x}(\mathcal{T}) = F_{U_t|X=x}(\mathcal{U}(\mathcal{T}, \underline{x}; \theta_0))\ a.e.\ x \in \mathrm{Supp}(X)\}, \tag{2}$$

where $\mathsf{F}^*_{U_t|X}$ *collects the distributions of* U_t *consistent with the observed data in the sense that*

$$\mathsf{F}^*_{U_t|X} = \bigcap_{t'=1}^{T} \{\mathcal{F}_{U_t|X} : \forall \mathcal{T} \in \mathsf{F}(\mathcal{Y}), F_{Y_{t'}|X=x}(\mathcal{T}) = F_{U_t|X=x}(\mathcal{U}(\mathcal{T}, x_{t'}; \theta_0))\ a.e.\ x \in \mathrm{Supp}(X)\}.$$

Remark 2. *Point identification of* θ_0 *(Assumption 2) is imposed to fix ideas and is stronger than necessary. The identified set defined in (2) is sharp for a given value of* θ_0*. When point identification of* θ_0 *fails, we can still take the union of (2) over the sharp identified set for* θ_0 *to obtain the sharp identified set for* $\mathcal{F}_{Y_t(\underline{x})|X}$*.*

4 Implementation

By Theorem 1, the most straightforward way to implement $\mathsf{F}^*_{Y_t(\underline{x})|X}$ is to search over the space of distributions supported on

$$\mathbb{U}(x) = \left\{\mathcal{U}(y, \underline{x}; \theta_0) \cap \left(\bigcap_{t'=1}^{T} \mathcal{U}(y_{t'}, x_{t'}; \theta_0)\right) : (y, y_1, \ldots, y_T) \in \mathcal{Y}^{T+1}\right\}$$

for each $x \in \mathrm{Supp}(X)$. With discrete outcomes, $\mathbb{U}(x)$ is a finite partition of the space of U_t, and any point within each set in $\mathbb{U}(x)$ produces the same outcome under $\underline{x}, x_1, \ldots, x_T$. This extends the concept of the *minimal relevant partition* of Tebaldi et al. (2023) to general discrete choice models. Nonetheless, depending on T, the cardinality of $\mathcal{Y}$, and the structural function g, the cardinality of $\mathbb{U}(x)$ can be large, making the search computationally demanding. In this section, we provide tractable characterizations of $\mathsf{F}^*_{Y_t(\underline{x})|X}$ that avoid directly involving the distribution of U_t by exploiting the separable index restriction on g, with a focus on Examples 1 and 2. We start with a heuristic illustration in the binary choice model (1).

As shown in Figure 2, because of the separable index restriction, U-level sets are half intervals: $\mathcal{U}(1, x_t; \theta_0) = [-x_t^\top \beta_0, \infty)$. Hence, when we change the value of explanatory variables from observed to counterfactual ones, we see a set inclusion relationship between the corresponding U-

level sets, which can be translated into a comparison between the distributions of the observed and counterfactual outcomes:

$$-\underline{x}^\top\beta_0 \le (\ge) - x_t^\top\beta_0 \iff \mathcal{U}(1,\underline{x};\theta_0) \subseteq (\supseteq)\, \mathcal{U}(1,x_t;\theta_0) \iff F_{Y_t(\underline{x})|X=x}(\{1\}) \le (\ge) F_{Y_t|X=x}(\{1\}).$$

In this way, we generate identifying restrictions on $\mathcal{F}_{Y_t(\underline{x})|X}$ directly from $\mathcal{F}_{Y_t|X}$. Under Assumption 1, we can repeat this procedure using observed data from any period. The resulting identifying restrictions turn out to be sharp.

When we go beyond binary choice models, set inclusion relationships generally takes the form

$$\mathcal{U}(\mathcal{T},\underline{x};\theta_0) \subseteq (\supseteq)\, \mathcal{U}(\mathcal{T}',x_t;\theta_0)$$

for some $\mathcal{T},\mathcal{T}' \in \mathsf{F}(\mathcal{Y})$, implying that

$$F_{Y_t(\underline{x})|X=x}(\mathcal{T}) \le (\ge) F_{Y_t|X=x}(\mathcal{T}').$$

Given the separable index restriction on g, which is satisfied in Examples 1 and 2, the set inclusion relationships can be easily determined by examining the indices, and these set inclusion relationships can be shown to exhaust all the information on the distribution of $Y_t(\underline{x})$.

Example 1 (continued)**.** Define the generalized inverse of h as

$$h^-(y;\gamma) = \inf\{y^* : h(y^*;\gamma) \ge y\},\ \ y \in \mathcal{Y}.$$

Then, U-level sets satisfy

$$\mathcal{U}([y,\infty),x_t;\theta) = [-x_t^\top\beta + h^-(y;\gamma),\infty). \tag{3}$$

Also define

$$\begin{aligned}\overline{\mathbb{Y}}(x_t;\theta) &= \{([y,\infty),[y',\infty)) : (y,y') \in \mathcal{Y}, -\underline{x}^\top\beta + h^-(y;\gamma) \ge -x_t^\top\beta + h^-(y';\gamma)\},\\ \underline{\mathbb{Y}}(x_t;\theta) &= \{([y,\infty),[y',\infty)) : (y,y') \in \mathcal{Y}, -\underline{x}^\top\beta + h^-(y;\gamma) \le -x_t^\top\beta + h^-(y';\gamma)\}.\end{aligned}$$

We can predict the following set inclusion relationships:

$$\begin{aligned}(\mathcal{T},\mathcal{T}') \in \overline{\mathbb{Y}}(x_t;\theta) \quad &\Longleftrightarrow \quad \mathcal{U}(\mathcal{T},\underline{x};\theta) \subseteq \mathcal{U}(\mathcal{T}',x_t;\theta),\\ (\mathcal{T},\mathcal{T}') \in \underline{\mathbb{Y}}(x_t;\theta) \quad &\Longleftrightarrow \quad \mathcal{U}(\mathcal{T},\underline{x};\theta) \supseteq \mathcal{U}(\mathcal{T}',x_t;\theta).\end{aligned}$$

Example 2 (continued)**.** Note that for any $\mathcal{T} \subsetneq \{0,1,\ldots,J\}$ such that $\mathcal{T} \neq \emptyset$,

$$\mathcal{U}(\mathcal{T},x_t;\theta) = \Big\{U_t : \max_{j\in\mathcal{T}} x_{jt}^\top\beta + U_{jt} \geq \max_{k\notin\mathcal{T}} x_{kt}^\top\beta + U_{kt}\Big\}.$$

Define

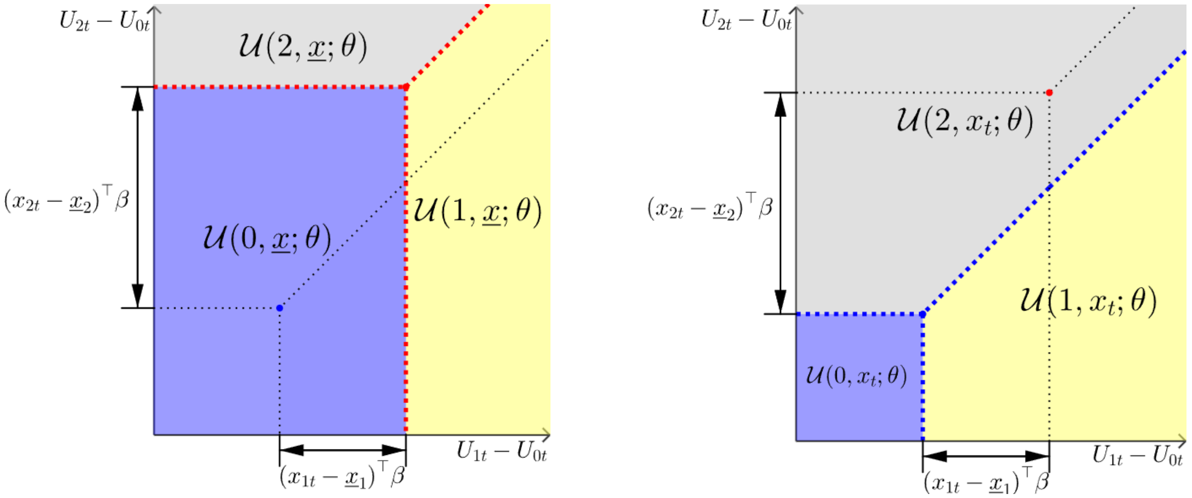
$$\mathbb{Y}(x_t;\theta) = \Big\{\mathcal{T} \subsetneq \{0,1,\ldots,J\} : \mathcal{T} \neq \emptyset, \min_{j\in\mathcal{T}} (x_{jt} - \underline{x}_j)^\top\beta \geq \max_{k\notin\mathcal{T}} (x_{kt} - \underline{x}_k)^\top\beta\Big\}. \tag{4}$$

We can predict the following set inclusion relationships:

$$\mathcal{T} \in \mathbb{Y}(x_t;\theta) \;\Rightarrow\; \mathcal{U}(\mathcal{T},\underline{x};\theta) \subseteq \mathcal{U}(\mathcal{T},x_t;\theta). \tag{5}$$

A proof of (5) is provided in Appendix A. It is helpful to understand (5) graphically. Consider the case of $J = 2$ and suppose that $(x_{2t} - \underline{x}_2)^\top\beta > (x_{1t} - \underline{x}_1)^\top\beta > 0$. Then, $\mathbb{Y}(x_t;\theta) = \{\{2\},\{2,1\}\}$. As shown in Figure 3, there are two set inclusion relationships:

$$\begin{aligned}\mathcal{U}(2,\underline{x};\theta) \quad &\subseteq \quad \mathcal{U}(2,x_t;\theta),\\ \mathcal{U}(2,\underline{x};\theta) \cup \mathcal{U}(1,\underline{x};\theta) \quad &\subseteq \quad \mathcal{U}(2,x_t;\theta) \cup \mathcal{U}(1,x_t;\theta).\end{aligned}$$


$U_{2t}-U_{0t}$
$\mathcal{U}(2,\underline{x};\theta)$
$(x_{2t}-\underline{x}_2)^\top\beta$
$\mathcal{U}(0,\underline{x};\theta)$
$\mathcal{U}(1,\underline{x};\theta)$
$U_{1t}-U_{0t}$
$(x_{1t}-\underline{x}_1)^\top\beta$
$U_{2t}-U_{0t}$
$\mathcal{U}(2,x_t;\theta)$
$(x_{2t}-\underline{x}_2)^\top\beta$
$\mathcal{U}(1,x_t;\theta)$
$\mathcal{U}(0,x_t;\theta)$
$U_{1t}-U_{0t}$
$(x_{1t}-\underline{x}_1)^\top\beta$

(a) U-Level Sets under the Counterfactual

(b) U-Level Sets for the Observed Data

Figure 3: Set Inclusion Relationships of U-Level Sets: Multinomial Choice Model

In general, to construct $\mathbb{Y}(x_t;\theta)$, one can simply rank the $J+1$ index function differences $\{(x_{jt}-\underline{x}_j)^\top\beta\}_{j=0}^J$ and collect the $\mathcal{T}$'s that contain the top j alternatives for $j=1,\dots J$.

With the set inclusion relationships of U-level sets discussed above, we are ready to present tractable characterizations of $\mathsf{F}^*_{Y_t(\underline{x})|X}$ for Examples 1 and 2 in Theorems 2 and 3, respectively. The proofs are in Appendix A.

Theorem 2. *Suppose that Assumptions 1 and 2 hold. Let g be specified as in Example 1. Then,*

$$\mathsf{F}^*_{Y_t(\underline{x})|X}=\bigcap_{t'=1}^{T}\Big\{\mathcal{F}_{Y_t(\underline{x})|X}:\forall(\mathcal{T},\mathcal{T}')\in\overline{\mathbb{Y}}(x_t;\theta_0),F_{Y_t(\underline{x})|X=x}(\mathcal{T})\le F_{Y_t|X=x}(\mathcal{T}'),$$
$$\forall(\mathcal{T},\mathcal{T}')\in\underline{\mathbb{Y}}(x_t;\theta_0),F_{Y_t(\underline{x})|X=x}(\mathcal{T})\ge F_{Y_t|X=x}(\mathcal{T}')\text{ a.e. }x\in\mathrm{Supp}(X)\Big\}\tag{6}$$

By Theorem 2, the sharp bounds on the counterfactual survival probability $F_{Y_t(\underline{x})|X=x}([y,\infty))$ are given by

$$\bigcap_{t'=1}^{T}\Bigg[\sup_{\substack{y':-\underline{x}^\top\beta_0+h^-(y;\gamma_0)\\ \ge-x_{t'}^\top\beta_0+h^-(y';\gamma_0)}}F_{Y_{t'}|X=x}([y',\infty)),\inf_{\substack{y':-\underline{x}^\top\beta_0+h^-(y;\gamma_0)\\ \le-x_{t'}^\top\beta_0+h^-(y';\gamma_0)}}F_{Y_{t'}|X=x}([y',\infty))\Bigg]$$

with the convention that $\sup\emptyset=0$ and $\inf\emptyset=1$. This result is similar to Theorem 2 of Botosaru and Muris (2024), where they allow the transformation function h to vary over time. Our framework can also accommodate time-varying h as long as it is point-identified. The key difference is that we establish the sharpness of their bounds.

Theorem 3. *Suppose that Assumptions 1 and 2 hold. Let g be specified as in Example 2. Then,*

$$\mathsf{F}^*_{Y_t(\underline{x})|X} = \bigcap_{t'=1}^{T} \{\mathcal{F}_{Y_t(\underline{x})|X} : \forall \mathcal{T} \in \mathbb{Y}(x_{t'};\theta_0), F_{Y_t(\underline{x})|X=x}(\mathcal{T}) \leq F_{Y_{t'}|X=x}(\mathcal{T}) \text{ a.e. } x \in \operatorname{Supp}(X)\}. \tag{7}$$

A collection of choice sets similar to (4) appears in Pakes and Porter (2024). They used the set inclusion relationship of U-level sets for the observed data between two time periods to derive identifying restrictions on the structural parameter θ_0. They also showed that when $T = 2$, these identifying restrictions are sharp and yield point identification under the additional conditions given in Shi et al. (2018). Our results further open up the possibility of counterfactual analysis built upon the knowledge of θ_0.

Remark 3. *The sharpness results in Theorems 2 and 7 are shown conditionally on each value of X. Therefore, the same conclusions hold if we allow the counterfactual evaluation point $\underline{x}$ to depend on X. For example, $\underline{x}$ can be the time average of X shifted by a small amount.*

5 Numerical Experiments

For Example 1, we consider the following data generating process:

$$Y_t = \sum_{j=0}^{J} 1\{\beta_0^{(1)} X_t^{(1)} + \beta_0^{(2)} X_t^{(2)} + U_t \geq \gamma_0^j\},\ t = 1, \ldots, T,$$

where $X_t^{(1)} \sim N(0, 0.5)$ and $U_t = A + V_t$ with $V_t \sim N(0, 0.5)$. We generate correlation between $X_t^{(2)}$ and A as follows. We define two equally sized latent populations of cross-sectional units. In the first population, $X_t^{(2)} \sim \text{Bernoulli}(0.5)$ and $A \sim N(1, 0.5)$. In the second population, $X_t^{(2)} = 0$ and $A \sim N(0, 0.5)$. We set $\beta_0^{(1)} = \beta_0^{(2)} = 1$. We consider three different numbers of categories: $J \in \{1, 2, 3\}$. We set $(\gamma_0^1, \gamma_0^2, \gamma_0^3) = (0, 1, 2)$. When $J = 1$, the model reduces to a binary response model.

Fixing a counterfactual value $\underline{x} = (-0.5, 1)$ for X_t, we are interested in the counterfactual survival probability $\Pr(Y_t(\underline{x}) \geq 1)$. We compute the sharp bounds on $\Pr(Y_t(\underline{x}) \geq 1)$ using Theorem

2 and noting that

$$\Pr(Y_t(\underline{x}) \geq 1) = \int F_{Y_t(\underline{x})|X=x}([1,\infty))dF_X(x),$$

where the integral is approximated by 5,000 random draws. Figure 4 shows the sharp bounds on $E[Y_t(\underline{x})]$ re-centered by the true value and divided by the scale of $Y_t(\underline{x})$ for $J \in \{1,2,3\}$ and $T \in \{1,2,\ldots,20\}$. We can see that the bounds tighten as T increases. There are substantial gains in identifying power when T increases from 1 to 10, but the incremental gains are less pronounced when T further increases from 10 to 20. The width of the (normalized) bounds do not differ much across J, especially when T is relatively large.

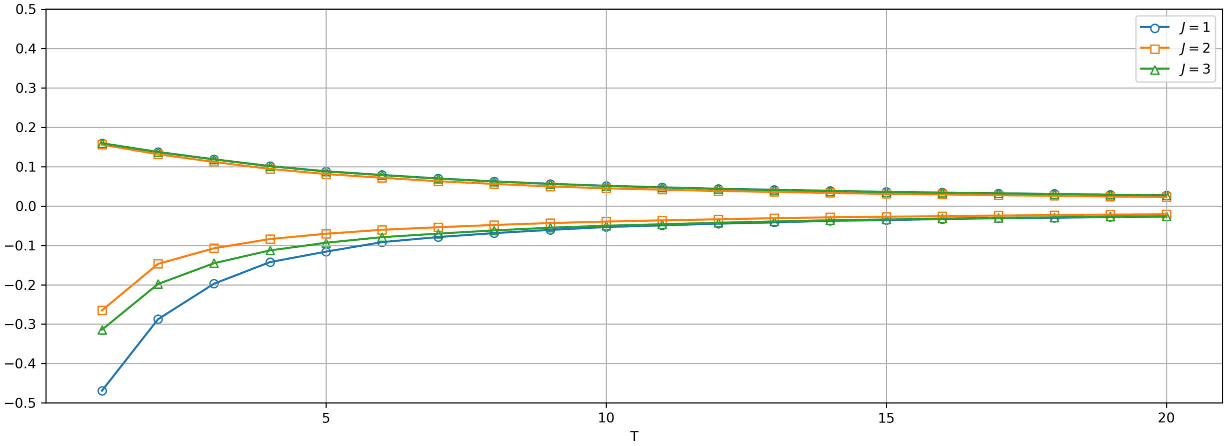


Figure 4: Sharp Bounds on $\Pr(Y_t(\underline{x}) \geq 1)$ in Ordered Choice Models

For Example 2, we consider the following data generating process:

$$Y_t = \max \arg \max_j Y^*_{jt}, \ t = 1, \ldots, T,$$

where the indirect utilities are given by

$$\begin{aligned} Y^*_{0t} &= 0, \\ Y^*_{jt} &= \beta^{(1)}_0 X^{(1)}_{jt} + \beta^{(2)}_0 X^{(2)}_{jt} + U_{jt}, \ j = 1, \ldots, J. \end{aligned}$$

Similar to Example 1, $X^{(1)}_{jt} \sim N(0, 0.5)$ $\forall j$ and $U_{jt} = A_j + V_{jt}$ $\forall j$, where $(V_{1t}, \ldots, V_{Jt})$ follows a zero mean multivariate normal distribution with a variance matrix that has 0.5 on the diagonal and 0.25 in all off-diagonal elements. To generate correlation between $X^{(2)}_{jt}$ and A_j, we again define two equally sized latent populations of cross-sectional units. In the first population, $X_{jt}^{(2)} \sim \text{Bernoulli}(0.5)\ \forall j$ and $A_j \sim N(1, 0.5)\ \forall j$. In the second population, $X_{jt}^{(2)} = 0\ \forall j$ and $A_j \sim N(0, 0.5)\ \forall j$. We set $\beta_0^{(1)} = \beta_0^{(2)} = 1$. We consider three different numbers of alternatives: $J \in \{1, 2, 3\}$. When $J = 1$, the model also reduces to a binary response model.

Fixing counterfactual values $\underline{x}_1 = (-0.5, 1)$ for X_{1t} and $\underline{x}_j = (0, 0)$ for $X_{jt}\ \forall j > 1$, we are interested in the probability of alternative 1 being chosen: $\Pr(Y_t(\underline{x}) = 1)$. We compute the sharp bounds on $\Pr(Y_t(\underline{x}) = 1)$ using Theorem 3 and noting that

$$\Pr(Y_t(\underline{x}) = 1) = \int F_{Y_t(\underline{x})|X=x}(\{1\}) dF_X(x),$$

where the integral is approximated by 5,000 random draws. Figure 5 shows the sharp bounds on $\Pr(Y_t(\underline{x}) = 1)$ re-centered by the true value for $J \in \{1, 2, 3\}$ and $T \in \{1, 2, \ldots, 20\}$. The trend in identifying power as T increases aligns with the pattern observed in Figure 4. Unlike in Figure 4, the bounds become wider when J increases.

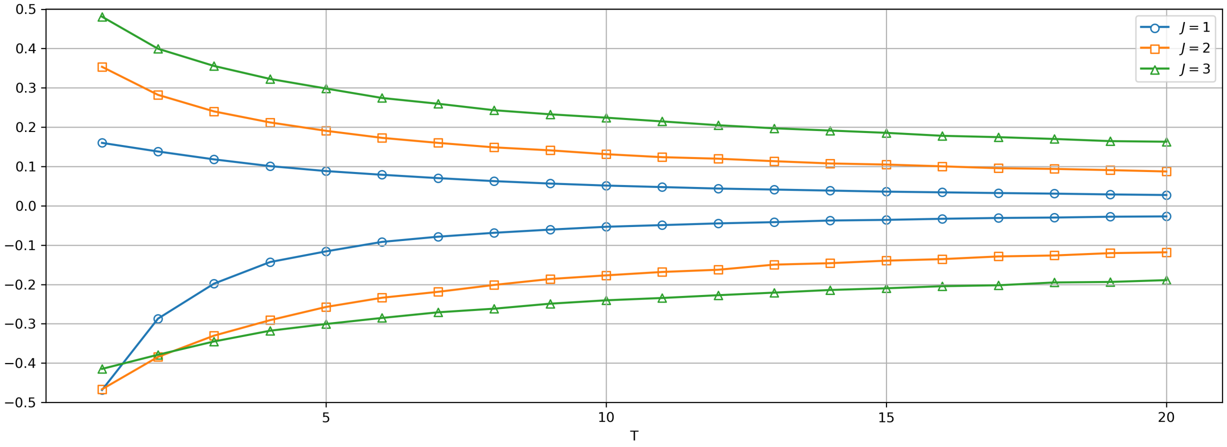


Figure 5: Sharp Bounds on $\Pr(Y_t(\underline{x}) = 1)$ in Multinomial Choice Models

6 Estimation and Inference

In this section, we focus on discrete outcomes. Let

$$\tau_0(x) = \{F_{Y_t|X=x}(\{y\}) : y \in \mathcal{Y}, t \in \{1, \ldots, T\}\}$$

denote the vector of observed conditional choice probabilities. We consider estimation and inference of aggregated intersection bounds that can be written as

$$\Psi(\theta_0) = E\Big[\min_{\lambda \in \Lambda(X;\theta_0)} \lambda^\top \tau_0(X)\Big], \tag{8}$$

where $\Lambda(x;\theta)$ is a known finite set.

Example 1 (continued)**.** We focus on ordered choice models. Fixing a counterfactual value $\underline{x}$ for X_t, the sharp bounds on counterfactual survival probabilities $\Pr(Y_t(\underline{x}) \geq j)$ take the form of (8). To see this, note that by Theorem 2, the bounds are given by $[E[\max_t \underline{\psi}_t(X;\theta_0)], E[\min_t \overline{\psi}_t(X;\theta_0)]]$, where

$$\begin{aligned} \underline{\psi}_t(x;\theta) &= F_{Y_t|X=x}(\{k : k \geq \min\{y \in \mathcal{Y} : -\underline{x}^\top\beta + h^-(j;\gamma) \leq -x_t^\top\beta + h^-(y;\gamma)\}\}), \\ \overline{\psi}_t(x;\theta) &= F_{Y_t|X=x}(\{k : k \geq \max\{y \in \mathcal{Y} : -\underline{x}^\top\beta + h^-(j;\gamma) \geq -x_t^\top\beta + h^-(y;\gamma)\}\}) \end{aligned}$$

with the convention that $\min\emptyset = \infty$. Since $\underline{\psi}_t(x;\theta)$ and $\overline{\psi}_t(x;\theta)$ are linear combinations of $\tau_0(x)$, we can write

$$\begin{aligned} \underline{\psi}_t(x;\theta) &= \underline{\lambda}_t(x;\theta)^\top \tau_0(x), \\ \overline{\psi}_t(x;\theta) &= \overline{\lambda}_t(x;\theta)^\top \tau_0(x), \end{aligned}$$

and define

$$\begin{aligned} \underline{\Lambda}(x;\theta) &= \{\underline{\lambda}_t(x;\theta) : t \in \{1,\ldots,T\}\}, \\ \overline{\Lambda}(x;\theta) &= \{\overline{\lambda}_t(x;\theta) : t \in \{1,\ldots,T\}\}. \end{aligned}$$

Then,

$$\begin{aligned} E[\max_t \underline{\psi}_t(X;\theta_0)] &= -E\Big[\min_{\lambda \in \underline{\Lambda}(x;\theta_0)} -\lambda^\top \tau_0(X)\Big], \\ E[\min_t \overline{\psi}_t(X;\theta_0)]] &= E\Big[\min_{\lambda \in \overline{\Lambda}(x;\theta_0)} \lambda^\top \tau_0(X)\Big]. \end{aligned}$$

Example 2 (continued)**.** Fixing a counterfactual value $\underline{x}$ for X_t, the sharp bounds on counterfactual choice probabilities $\Pr(Y_t(\underline{x}) = j)$ take the form of (8). To see this, note that by Theorem 3, the bounds are given by $[E[\max_t \underline{\psi}_t(X;\theta_0)], E[\min_t \overline{\psi}_t(X;\theta_0)]]$, where $\underline{\psi}_t(x;\theta)/\overline{\psi}_t(x;\theta)$ is the solution to the linear program

$$\begin{aligned} &\max / \min_{\vec{q}\in\Delta^{J+1}} \;\; q_j \\ &\text{s.t.} \sum_{j\in\mathcal{T}} q_j \leq F_{Y_t|X=x}(\mathcal{T}) \;\forall \mathcal{T} \in \mathbb{Y}(x_t;\theta). \end{aligned}$$

It turns out that $\underline{\psi}_t(x;\theta)$ and $\overline{\psi}_t(x;\theta)$ have closed forms:

$$\begin{aligned} \underline{\psi}_t(x;\theta) &= \begin{cases} F_{Y_t|X=x}(\{j\}) & \text{if } (x_{jt}-\underline{x}_j)^\top\beta \leq (x_{kt}-\underline{x}_k)^\top\beta,\ \forall k \\ 0 & \text{otherwise} \end{cases}, \\ \overline{\psi}_t(x;\theta) &= \begin{cases} F_{Y_t|X=x}(\{j\}) & \text{if } (x_{jt}-\underline{x}_j)^\top\beta \geq (x_{kt}-\underline{x}_k)^\top\beta,\ \forall k \\ F_{Y_t|X=x}(\{j\}\cup\{k:(x_{kt}-\underline{x}_k)^\top\beta > (x_{jt}-\underline{x}_j)^\top\beta\}) & \text{otherwise} \end{cases}. \end{aligned}$$

We can again see that $\underline{\psi}_t(x;\theta)$ and $\overline{\psi}_t(x;\theta)$ are linear combinations of $\tau_0(x)$.

To construct an estimator of $\Psi(\theta_0)$, we use cross-fitting to estimate τ_0.

Definition 1 (Cross-fitting)**.** Divide the data into K evenly-sized folds. For each fold $k = 1,\ldots,K$, use the other $K-1$ data folds to estimate τ_0; denote the resulting estimates by $\hat{\tau}^{(-k)}$. For each $i = 1,\ldots,N$, take $\hat{\tau}(X_i) = \hat{\tau}^{(-k_i)}(X_i)$, where k_i denotes the fold containing the ith observation.

We impose the following assumptions.

Assumption 3. *For all x and θ, $\max_{\lambda\in\Lambda(x;\theta)} \|\lambda\| \leq M$ for some $M > 0$.*

Assumption 4. *For all x, θ, and τ, $\arg\min_{\lambda\in\Lambda(x;\theta)} \lambda^\top\tau(x)$ is a singleton.*

Assumption 5. *The distribution of $\tau_0(X)$ is absolutely continuous with density bounded above.*

Assumption 6. *$\|\hat{\tau}-\tau_0\|_\infty = o_p(N^{-1/4})$, where $\|\tau\|_\infty = \sup_x \|\tau(x)\|$.*

Assumption 3 imposes boundedness on the objective function of the minimization problem and is satisfied in Examples 1 and 2. Assumption 4 requires the solution of the minimization problem to be unique. Assumption 5 is a sufficient condition for the margin condition (Lemma 1) that controls the concentration of the objective function in the neighborhood of the minimum. In other words, it ensures the minimum is separated from non-minimal values with high probability. The uniqueness of the optimal solution and the margin condition are also imposed in Semenova (2024) to derive inference for a general class of aggregated intersection bounds. We retain Assumption 5 because it is low-level and compatible with the sufficient conditions for Assumption 2. Assumption 6 requires the estimation error of $\hat{\tau}$ to vanish fast enough.

Let

$$I(Y) = \{1\{Y_t = y\} : y \in \mathcal{Y}, t \in \{1, \ldots, T\}\}$$

be a vector of binary indicators that is conformable with $\tau_0(x)$. Define

$$\lambda^*(x; \theta, \tau) = \underset{\lambda \in \Lambda(x;\theta)}{\arg\min}\, \lambda^\top \tau(x).$$

Given the first-step cross-fitted estimator $\hat{\tau}$ of τ_0, define

$$\hat{\Psi}(\theta) = \frac{1}{N} \sum_{i=1}^{n} \sum_{\lambda \in \Lambda(X_i;\theta)} 1\{\lambda^*(X_i; \theta, \hat{\tau}) = \lambda\} \lambda^\top I(Y_i).$$

Theorem 4. *Suppose that Assumptions 3-6 hold. Then, for a given θ,*

$$\sqrt{N}(\hat{\Psi}(\theta) - \Psi(\theta)) \xrightarrow{d} N(0, V(\theta)),$$

where $V(\theta) = E[\sum_{\lambda \in \Lambda(X;\theta)} 1\{\lambda^*(X; \theta, \tau_0) = \lambda\}(\lambda^\top I(Y))^2] - \Psi^2(\theta)$.

In view of Theorem 4, a natural idea is to plug in a first-step estimate $\hat{\theta}$ of θ_0 to obtain the final estimator $\hat{\Psi}(\hat{\theta})$. However, the asymptotic distribution $\hat{\Psi}(\hat{\theta})$ is complicated by the estimation error of $\hat{\theta}$. We give a heuristic explanation in the binary choice model (1). Note that θ enters $\Psi(\theta)$ only through Λ so that

$$|\Psi(\hat{\theta}) - \Psi(\theta_0)| = O(\Pr(\Lambda(X; \hat{\theta}) \neq \Lambda(X; \theta_0))).$$

For $\theta \neq \theta_0$, $\Lambda(x; \theta) \neq \Lambda(x; \theta_0)$ if for some t, $\text{sgn}((x_t - \underline{x})^\top \beta) \neq \text{sgn}((x_t - \underline{x})^\top \beta_0)$, which occurs with probability of order $O(\|\theta - \theta_0\|)$. Therefore, the estimation error of $\hat{\theta}$ becomes dominating in the expansion of $\hat{\Psi}(\hat{\theta})$ if $\hat{\theta}$ converges at a slower rate than $N^{-1/2}$, as is the case with the maximum estimator proposed by Manski (1987) and its smoothed version.

To utilize the asymptotic normality result in Theorem 4, we consider Bonferroni-type confidence intervals. To this end, define

$$\hat{V}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\sum_{\lambda\in\Lambda(X_i;\theta)} 1\{\lambda^*(X_i;\theta,\hat{\tau}) = \lambda\}(\lambda^\top I(Y_i))^2,$$

which is a consistent estimator of $V(\theta)$ for a given θ under Assumption 6. Also, suppose that we can construct a $(1-\alpha)$-confidence region for θ_0:

$$\lim_{N\to\infty} \Pr(\theta_0 \in \mathrm{CR}_N(\alpha)) = 1-\alpha. \tag{9}$$

For $0 \leq \delta \leq \alpha$, the Bonferroni confidence interval for $\Psi(\theta_0)$ is given by

$$\mathrm{CI}_N(\alpha,\delta) = \left[\inf_{\theta\in\mathrm{CR}_N(\delta)} \hat{\Psi}(\theta) - z_{1-(\alpha-\delta)/2}\sqrt{\hat{V}(\theta)/N}, \sup_{\theta\in\mathrm{CR}_N(\delta)} \hat{\Psi}(\theta) + z_{1-(\alpha-\delta)/2}\sqrt{\hat{V}(\theta)/N}\right]$$

Proposition 1. *Suppose that Assumptions 3-6 and (9) hold. Then, for any* $0 \leq \delta \leq \alpha$,

$$\lim_{N\to\infty} \Pr(\Psi(\theta_0) \in \mathrm{CI}_N(\alpha,\delta)) = 1-\alpha.$$

The literature on semiparametric inference for θ_0 has not yet converged on a single procedure. For panel data binary choice models, the asymptotic distribution of the maximum score estimator is that of the maximizer of a Gaussian process, which is hard to use for inference. One solution is to switch to the smoothed maximum score estimator proposed by Charlier, Melenberg, and van Soest (1995), but this requires selecting additional kernel functions and tuning parameters. An alternative is to use bootstrap-based methods. Abrevaya and Huang (2005) have shown that the classic bootstrap is inconsistent for the maximum score estimators. Valid inference may be conducted using subsampling (Delgado, Rodríguez-Poo, and Wolf, 2001), m-out-of-n bootstrap (Lee and Pun, 2006), the numerical bootstrap (Hong and Li, 2020), and a model-based bootstrap procedure that analytically modifies the criterion function (Cattaneo, Jansson, and Nagasawa, 2020). For panel data multinomial choice models, Khan et al. (2021) proposed a localized maximum score estimator, whose asymptotic distribution is also that of the maximizer of a Gaussian process. Khan et al. (2021) conjectured that both a smoothed maximum score approach and bootstrap-based procedures may be used for inference.

7 Empirical Illustration

7.1 Binary Choice Model: Female Labor Force Participation

In the first empirical illustration, we study women's labor force participation using data from the US Panel Study of Income Dynamics (PSID) and the British Household Panel Survey (BHPS). For the PSID, we use a sample from Fernández-Val (2009), which consists of $N = 1461$ women over $T = 9$ years between 1980-1988. Only married women aged 18-64 with husbands in the labor force in each sample period are included. For the BHPS, we construct a similar sample from all 1991-2008 waves, which consists of $N = 4602$ women. The sample is an unbalanced panel, in which any woman observed in at least two waves is included.

For illustrative purposes, we focus on the static binary choice model:

$$Y_{it} = 1\{X_{it}^{\top}\beta_0 + U_{it} \geq 0\},$$

where Y_t is the labor force participation indicator, and X_t includes the natural logarithm of the husband's income, the number of children in three age categories, and a quadratic function of age. Note that the husband's income and fertility may be jointly determined with the wife's labor force participation by unobserved factors such as the wife's household productivity. We assume that these factors are time-invariant so that Assumption 1 holds. Age categories in each sample differ slightly, with the PSID dividing children into 0-2, 3-5, and 6-17 years, and the BHPS into 0-2, 3-4, and 5-18 years. Descriptive statistics for both samples are given in Table 1.

The continuous variation in the husband's income enables the point identification of β_0. We estimate β_0 using the maximum-score-type objective function:

$$\sum_{i} \sum_{t>s} (Y_{it} - Y_{is}) \cdot \operatorname{sgn}((X_{it} - X_{is})^{\top}\beta).$$

Table 2 reports the point estimates of β_0. We see that the coefficients on the number of children in all three age categories are consistent across samples, exhibiting the same sign and similar magnitudes. While the coefficients on log husband's income also have the same sign in both samples, the magnitude is notably smaller in the BHPS sample. The coefficients on age and age squared indicate a concave relationship.

We are interested in counterfactual probabilities of labor force participation under various levels

Table 1: Descriptive Statistics

	Mean	Std. dev.
Panel A: PSID sample, 1980-1988		
Participation	0.72	0.45
Age	37.3	9.22
Kids 0-2	0.23	0.47
Kids 3-5	0.29	0.51
Kids 6-17	1.05	1.10
Husband's income (1995 \$1000)	42.29	40.01
No. observations	13149	
Panel B: BHPS sample, 1991-2008		
Participation	0.78	0.41
Age	41.9	10.02
Kids 0-2	0.12	0.34
Kids 3-4	0.12	0.34
Kids 5-18	0.74	0.98
Husband income (1995 £1000)	20.02	15.46
No. observations	35608	

of husband's income, which reflects the wife's reservation wage. We select counterfactual values $\underline{x}$ such that the log husband's income ranges from its 10th to 90th quantiles, and other explanatory variables are set to their medians. In the PSID sample, these choices correspond to hypothetical women who are 35 years old, have no children aged 0-2 or 3-5, and have one child between 6 and 17, and whose husband's income ranges from \$15K to \$68K. In the BHPS sample, these choices correspond to hypothetical women who are 41 years old, have no children aged 0-2 or 3-4, and have one child between 5 and 18, and whose husband's income ranges from £8K to £32K.

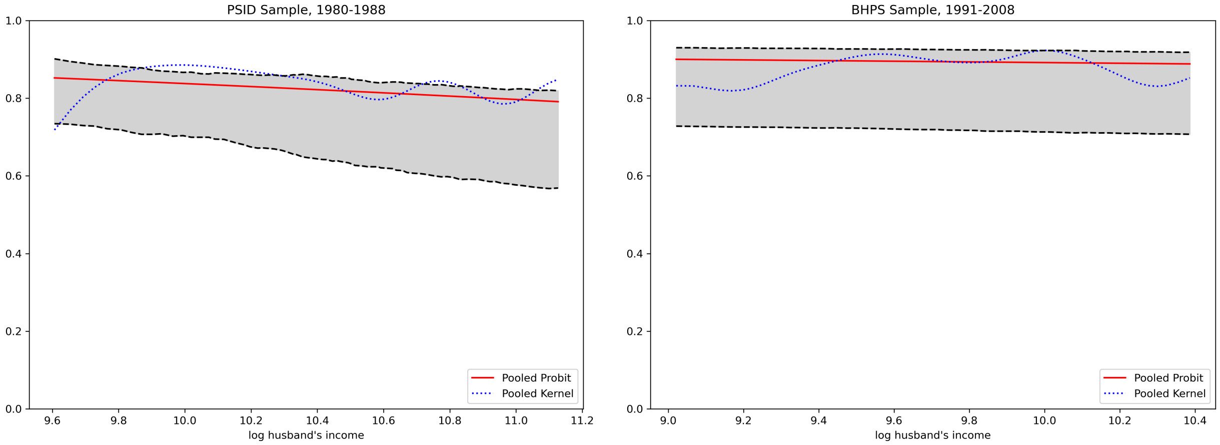
We calculate the sharp bounds on counterfactual probabilities of labor force participation using the estimator developed in Section 6 and plot them in Figure 6. To do this, we plug in the maximum-score estimates of β_0 in Table 2 and the estimates of observed conditional choice probabilities, $\tau_0(x)$, from the logistic regression of observed choices on X_{it} and $\frac{1}{T_i}\sum_{t=1}^{T_i} X_{it}$.[2] We see that for the PSID sample, the bounds are downward sloping with respect to the log husband's income with the lower bound steeper than the upper bound. In contrast, the bounds in the BHPS sample remain nearly flat. This pattern aligns with the coefficient estimates in Table 2 and suggests that the effect of the husband's income on women's labor force participation is more pronounced in the PSID sample.

[2] Note that each element of $\tau_0(x)$ can be written as $F_{Y_t|X=x}(\{y\}) = F_{U_t|X=x}(\mathcal{U}(y, x_t; \theta_0)) = G(x_t, x)$. Hence, the logistic regression of observed choices on some function of X_{it} and lower-dimensional statistics of X_i, such as $\frac{1}{T_i}\sum_{t=1}^{T_i} X_{it}$, can be viewed as a series logit approximation to $\tau_0(x)$.

Table 2: Estimated β_0

	PSID		BHPS	
	Max. Score	Pooled Probit	Max. Score	Pooled Probit
Kids 0-2	-1	-1	-1	-1
Kids 3-5	-0.565	-0.601		
Kids 3-4			-0.602	-0.686
Kids 6-17	-0.006	-0.166		
Kids 5-18			-0.014	-0.267
Log husband's income	-0.098	-0.351	-0.007	-0.050
Age/10	1.142	1.700	1.024	2.090
$(\text{Age}/10)^2$	-0.126	-0.266	-0.109	-0.275

Figure 6: Counterfactual Probabilities of Labor Force Participation



Notes: Black dashed lines, with the grey shaded area in between, represent the sharp bounds on counterfactual probabilities of labor force participation. Red solid/blue dotted lines represent probit/kernel predictions assuming exogeneity of X_{it}.

For comparison, we also plot the predictions assuming exogeneity of X_{it} with a probit specification (predictions based on a logit specification are quite similar and thus omitted). The associated coefficient estimates are reported in Table 2 under the column "Pooled Probit". We see that probit predictions tend to lie close to the upper bounds. In addition, we plot predictions under exogeneity of X_{it} based on the Nadaraya–Watson kernel regression.[3] The kernel predictions differ from probit predictions and exhibit a nonmonotonic relationship with the husband's income.

[3] We use Gaussian kernel and select the bandwidth using Silverman's rule of thumb.

7.2 Multinomial Choice Model: Saltine Cracker Purchases

In the second empirical illustration, we apply our approach to the optical-scanner panel data set on purchases of saltine crackers in the Rome (Georgia) market, collected by Information Resources Incorporated. The data set contains information on 3292 purchases of crackers by 136 households over a period of 2 years. There are three major national brands in the database: Nabisco, Sunshine, Keebler. Local brands are aggregated under the "Private" label. The data set also includes three explanatory variables, two of which are binary, and the other one is continuous. The first binary explanatory variable, "display", denotes whether or not a brand was on special display at the store at the time of purchase. The second binary explanatory variable, "feature", denotes whether or not a brand was featured in a newspaper advertisement at the time of purchase. The third explanatory variable is the "price", which corresponds to the actual price (in dollars) for the brand purchased and the shelf price for all other brands. Table 3 reports the descriptive statistics for each brand.

Table 3: Data Characteristics of Saltine Crackers

	Nabisco	Sunshine	Keebler	Private
Market Share	0.54	0.07	0.07	0.32
Display	0.34	0.13	0.11	0.10
Feature	0.09	0.04	0.04	0.05
Average Price	1.08	0.96	1.13	0.68

The dataset is an unbalanced panel data with the number of purchases varying across households i ($\equiv T_i, 14 \leq T_i \leq 77$). Write $\bar{\mathcal{J}} = \{1 = \text{Nabisco}, 2 = \text{Sunshine}, 3 = \text{Keebler}, 4 = \text{Private}\}$ for the choice set. For each household i, brand j, and purchase t, we use $X_{ijt}^{(1)}$, $X_{ijt}^{(2)}$, and $X_{ijt}^{(3)}$ to denote the three explanatory variables: the logarithm of "price", "display", and "feature", respectively. There are unobserved confounders, such as quality and intrinsic brand preferences, which are likely to remain invariant during the sample period. Hence, Assumption 1 is plausibly valid.

We follow Khan et al. (2021) to model the observed choice as

$$Y_{ijt} = 1\{Y_{ijt}^* > Y_{ikt}^*,\ \forall k \neq j\},$$

where the indirect utilities are given by

$$Y_{ijt}^{*} = -X_{ijt}^{(1)} + \beta_0^{(1)} X_{ijt}^{(2)} + \beta_0^{(2)} X_{ijt}^{(3)} + U_{ijt}, \quad j \in \bar{\mathcal{J}}, t = 1, \ldots, T_i,$$

where the coefficient on $X_{ijt}^{(1)}$ is normalized to be -1. $(\beta_0^{(1)}, \beta_0^{(2)})$ is point-identified because of rich variation in prices and can be estimated by minimizing a localized rank-based objective function

$$\sum_i \sum_{t>s} K_{h_n}(X_{i(-1)s}^{(1)} - X_{i(-1)t}^{(1)}) 1\{\tilde{X}_{i(-1)s} = \tilde{X}_{i(-1)t}\}(Y_{i1s} - Y_{i1t}) \cdot \operatorname{sgn}((X_{i1s} - X_{i1t})^{\top}\beta),$$

where $\beta = (-1, \beta^{(1)}, \beta^{(2)})^{\top}$, $\tilde{X}_{ijt} = (X_{ijt}^{(2)}, X_{ijt}^{(3)})'$, and $X_{i(-1)t}^{(1)}$ $(\tilde{X}_{i(-1)t})$ denotes the vector collecting $X_{ijt}^{(1)}$ $(\tilde{X}_{ijt})$ for all $j \in \bar{\mathcal{J}} \setminus \{1\}$. Following Khan et al. (2021), we choose the Gaussian kernel function and $h_n = 3\hat{\sigma} n^{-1/6} / \sqrt[3]{\log n}$, where $\hat{\sigma}$ is the standard deviation of the matching variable.

No other methods in the literature deliver counterfactual predictions for panel multinomial choice models. For comparison, we consider two parametric models, pooled multinomial logit and pooled multinomial probit, based on the indirect utility specification

$$Y_{ijt}^{*} = -\beta_0^{(0)} X_{ijt}^{(1)} + \beta_0^{(1)} X_{ijt}^{(2)} + \beta_0^{(2)} X_{ijt}^{(3)} + \alpha_j + V_{ijt}, \quad j \in \bar{\mathcal{J}}, t = 1, \ldots, T_i,$$

where V_{ijt} is independent of X_{ijt}, and $(\beta_0^{(0)}, \beta_0^{(1)}, \beta_0^{(2)})$ and alternative-specific intercepts α_j are parameters to be estimated.[4] Table 4 reports the point estimates of coefficients. For the pooled multinomial logit and probit models, we report the ratios of the coefficients on $X_{ijt}^{(2)}$ and $X_{ijt}^{(3)}$ to the absolute value of the coefficient on $X_{ijt}^{(1)}$.

Table 4: Parametric and Semiparametric Estimations of Coefficients

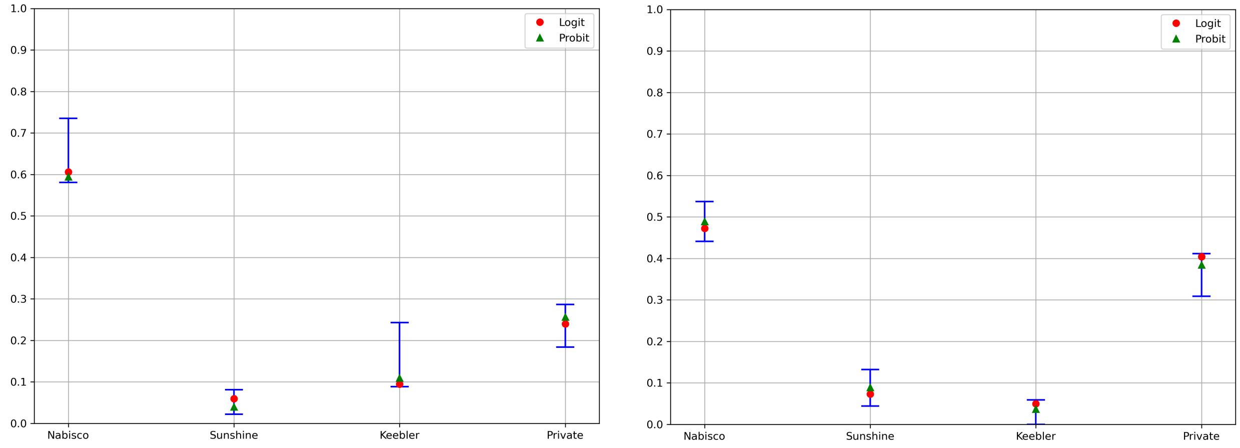
	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$
Semiparametric panel	0.0804	0.0859
Pooled multinomial logit	0.0330	0.1573
Pooled multinomial probit	0.0155	0.1108

We consider the counterfactual choice probabilities under two counterfactual values $\underline{x}$ and $\overline{x}$ for explanatory variables. The price vector for $\underline{x}$ is $\underline{p} = (1.09, 1.05, 1.05, 0.78)$ and the price vector for $\overline{x}$ is $\overline{p} = (1.09, 0.89, 1.21, 0.59)$. The display and feature statuses are fixed at zero for all brands for

[4] The parameter estimation of these models is conducted using Stata packages “cmclogit” and “cmcmmprobit”.

both $\underline{x}$ and $\overline{x}$. Moving from $\underline{x}$ to $\overline{x}$ corresponds to a simultaneous price change of multiple brands, which consists of a rise from the 25th percentile to the 75th percentile of the price for brand 3 (Keebler), and a fall from the 75th percentile to the 25th percentile of the price for brands 2 and 4 (Sunshine and Private), with the price of brand 1 (Nabisco) fixed at the median.

We calculate the sharp bounds on counterfactual choice probabilities using the estimator developed in Section 6. To do this, we plug in the semiparametric estimates of $(\beta_0^{(1)}, \beta_0^{(2)})$ in Table 4 and the estimates of observed conditional choice probabilities, $\tau_0(x)$, from multinomial logistic regression of observed choices on $\{(X_{ijt}, (X_{ijt}^{(1)})^2), \frac{1}{T_i}\sum_{t=1}^{T_i} X_{ijt}, \frac{1}{T_i}\sum_{t=1}^{T_i}(X_{ijt}^{(1)})^2)\}_{j\in\bar{\mathcal{J}}}$. Panels (a) and (b) of Figure 7 display the bounds under $\underline{x}$ and $\overline{x}$, respectively. For comparison, we also plot the predictions from pooled multinomial logit and probit models. We observe a market share decrease for brands 1 and 3 (Nabisco and Keebler) and a market share increase for brand 4 (Private), while the direction of the market share change for brand 2 (Sunshine) is ambiguous. Parametric predictions lie within semiparametric bounds, with some close to upper or lower limits. Consequently, parametric models might underestimate the market share change of brand 3 (Keebler).



(a) Under Price Vector $\underline{p} = (1.09, 1.05, 1.05, 0.78)$

(b) Under Price Vector $\overline{p} = (1.09, 0.89, 1.21, 0.59)$

Figure 7: Counterfactual Choice Probabilities

8 Extension: Dynamic Binary Choice Models

Although the main framework of this paper focuses on static models, the identification strategy based on the set inclusion relationship of U-level sets can be applied to dynamic models to derive (non-sharp) identifying restrictions on counterfactual distributions. To demonstrate this, we consider the dynamic panel data binary choice model:

$$Y_t = 1\{\rho_0 Y_{t-1} + X_t^\top \beta_0 + U_t \geq 0\}.$$

Let $\theta_0 = (\rho_0, \beta_0)$. We maintain Assumption 1, which is termed *partial stationarity* in Gao and Wang (2024) because the conditioning set only contains part of the explanatory variables. Identification of θ_0 under Assumption 1 is discussed in Khan et al. (2023) and Gao and Wang (2024). Fixing a counterfactual value $(\underline{y}, \underline{x})$ for (Y_{t-1}, X_t), we are interested in the distribution of the counterfactual outcome $Y_t(\underline{y}, \underline{x})$ that satisfies $Y_t(\underline{y}, \underline{x}) = 1\{\rho_0 \underline{y} + \underline{x}^\top \beta_0 + U_t \geq 0\}$. This is in line with the *dynamic potential outcome* model of Torgovitsky (2019).

We slightly modify the definition of U-level sets as

$$\mathcal{U}(y_t, y_{t-1}, x_t; \theta) = \{u_t : y_t = 1\{\rho y_{t-1} + x_t^\top \beta + u_t \geq 0\}\}.$$

The key observation is that for $y \in \{0, 1\}$,

$$U_t \in \mathcal{U}(y, Y_{t-1}, X_t; \theta_0) \text{ and } \mathcal{U}(y, Y_{t-1}, X_t; \theta_0) \subseteq \mathcal{U}(y, \underline{y}, \underline{x}; \theta_0) \;\Rightarrow\; U_t \in \mathcal{U}(y, \underline{y}, \underline{x}; \theta_0). \tag{10}$$

Note that

$$\begin{aligned} & 1\{\mathcal{U}(1, Y_{t-1}, X_t; \theta_0) \subseteq \mathcal{U}(1, \underline{y}, \underline{x}; \theta_0)\} \\ = \;& 1 - 1\{\mathcal{U}(0, Y_{t-1}, X_t; \theta_0) \subseteq \mathcal{U}(0, \underline{y}, \underline{x}; \theta_0)\} \\ = \;& 1\{Y_{t-1} = 1\} \cdot 1\{\rho_0 \underline{y} + \underline{x}^\top \beta_0 \geq \rho_0 + X_t^\top \beta_0\} + 1\{Y_{t-1} = 0\} \cdot 1\{\rho_0 \underline{y} + \underline{x}^\top \beta_0 \geq X_t^\top \beta_0\}. \end{aligned}$$

Taking the conditional expectation of (10) given $X = x$ yields

$$B_t^\ell(x; \theta_0) \leq \Pr(Y_t(\underline{y}, \underline{x}) = 1 | X = x) \leq B_t^u(x; \theta_0),$$

where

$$B_t^\ell(x;\theta) \quad = \quad \begin{cases} \Pr(Y_t = 1|X = x) & \text{if } \rho\underline{y} + \underline{x}^\top\beta \geq \max\{\rho + x_t^\top\beta, x_t^\top\beta\} \\ \Pr(Y_t = 1, Y_{t-1} = 0|X = x) & \text{if } x_t^\top\beta \leq \rho\underline{y} + \underline{x}^\top\beta < \rho + x_t^\top\beta \\ \Pr(Y_t = 1, Y_{t-1} = 1|X = x) & \text{if } \rho + x_t^\top\beta \leq \rho\underline{y} + \underline{x}^\top\beta < x_t^\top\beta \\ 0 & \text{otherwise} \end{cases},$$

$$B_t^u(x;\theta) \quad = \quad \begin{cases} 1 & \text{if } \rho\underline{y} + \underline{x}^\top\beta \geq \max\{\rho + x_t^\top\beta, x_t^\top\beta\} \\ 1 - \Pr(Y_t = 0, Y_{t-1} = 1|X = x) & \text{if } x_t^\top\beta \leq \rho\underline{y} + \underline{x}^\top\beta < \rho + x_t^\top\beta \\ 1 - \Pr(Y_t = 0, Y_{t-1} = 0|X = x) & \text{if } \rho + x_t^\top\beta \leq \rho\underline{y} + \underline{x}^\top\beta < x_t^\top\beta \\ \Pr(Y_t = 1|X = x) & \text{otherwise} \end{cases}.$$

The intuition is that when the counterfactual index is large or small enough to eliminate uncertainty in the set inclusion relationship of U-level sets, the bounds align with those in the static case. Otherwise, the bounds will depend on the distribution of the lagged outcome.

Assumption 1 allows us to use information across all periods to obtain tighter bounds. Eventually, the counterfactual probability $\Pr(Y_t(\underline{y}, \underline{x}) = 1)$ can be bounded as

$$E\Big[\max_t B_t^\ell(X;\theta_0)\Big] \leq \Pr(Y_t(\underline{y}, \underline{x}) = 1) \leq E\Big[\min_t B_t^u(X;\theta_0)\Big].$$

9 Conclusion

This paper establishes sharp identified sets of counterfactual distributions in semiparametric nonlinear panel data models, relying on mild assumptions such as time homogeneity on the distribution of unobserved heterogeneity and index separability on the structural function. We provide tractable implementation procedures for monotone transformation models and multinomial choice models. We examine factors affecting the informativeness of identified sets through numerical experiments. We also derive theoretical results for estimation and inference. Our approach is applied to empirical data on female labor force participation and purchases of saltine crackers. Finally, we discuss the potential extension of our identification strategy to dynamic settings.

References

Abrevaya, J. (2000): "Rank estimation of a generalized fixed-effects regression model," *Journal of Econometrics*, 95, 1–23.

Abrevaya, J. and J. Huang (2005): "On the Bootstrap of the Maximum Score Estimator," *Econometrica*, 73, 1175–1204.

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

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

Blundell, R. W. and J. L. Powell (2003): "Endogeneity in nonparametric and semiparametric regression models," in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Cambridge: Cambridge University Press, vol. 2, 312–357.

Botosaru, I. and C. Muris (2024): "Identification of time-varying counterfactual parameters in nonlinear panel models," *Journal of Econometrics*, 105639.

Botosaru, I., C. Muris, and K. Pendakur (2023): "Identification of time-varying transformation models with fixed effects, with an application to unobserved heterogeneity in resource shares," *Journal of Econometrics*, 232, 576–597.

Cattaneo, M. D., M. Jansson, and K. Nagasawa (2020): "Bootstrap-Based Inference for Cube Root Asymptotics," *Econometrica*, 88, 2203–2219.

Charlier, E., B. Melenberg, and A. H. O. van Soest (1995): "A smoothed maximum score estimator for the binary choice panel data model with an application to labour force participation," *Statistica Neerlandica*, 49, 324–342.

Chen, X., O. Linton, and I. Van Keilegom (2003): "Estimation of Semiparametric Models when the Criterion Function Is Not Smooth," *Econometrica*, 71, 1591–1608.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81, 535–580.

Chernozhukov, V., I. Fernández-Val, S. Hoderlein, H. Holzmann, and W. Newey (2015): "Nonparametric identification in panels using quantiles," *Journal of Econometrics*, 188, 378–392, heterogeneity in Panel Data and in Nonparametric Analysis in honor of Professor Cheng Hsiao.

Chernozhukov, V., I. Fernández-Val, and W. K. Newey (2019): "Nonseparable multinomial choice models in cross-section and panel data," *Journal of Econometrics*, 211, 104–116, annals Issue in Honor of Jerry A. Hausman.

Chesher, A. and A. M. Rosen (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85, 959–989.

Chesher, A., A. M. Rosen, and Y. Zhang (2024): "Robust Analysis of Short Panels," ArXiv: 2401.06611.

Chiong, K. X., Y.-W. Hsieh, and M. Shum (2021): "Bounds on counterfactuals in semiparametric discrete-choice models," in *Handbook of Research Methods and Applications in Empirical Microeconomics*, Edward Elgar Publishing, 223–237.

Delgado, M. A., J. M. Rodríguez-Poo, and M. Wolf (2001): "Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator," *Economics Letters*, 73, 241–250.

Fernández-Val, I. (2009): "Fixed effects estimation of structural parameters and marginal effects in panel probit models," *Journal of Econometrics*, 150, 71–85.

Gao, W. Y. and M. Li (2020): "Robust Semiparametric Estimation in Panel Multinomial Choice Models," ArXiv: 2009.00085.

Gao, W. Y. and R. Wang (2024): "Identification of Nonlinear Dynamic Panels under Partial Stationarity," ArXiv: 2401.00264.

Graham, B. S. and J. L. Powell (2012): "Identification and Estimation of Average Partial Effects in "Irregular" Correlated Random Coefficient Panel Data Models," *Econometrica*, 80, 2105–2152.

Gu, J., T. Russell, and T. Stringham (2024): "Counterfactual identification and latent space enumeration in discrete outcome models," Available at SSRN: https://ssrn.com/abstract=4188109 or http://dx.doi.org/10.2139/ssrn.4188109.

Hoderlein, S. and H. White (2012): "Nonparametric identification in nonseparable panel data models with generalized fixed effects," *Journal of Econometrics*, 168, 300–314.

Hong, H. and J. Li (2020): "The numerical bootstrap," *The Annals of Statistics*, 48, 397 – 412.

Khan, S., F. Ouyang, and E. Tamer (2021): "Inference on semiparametric multinomial response models," *Quantitative Economics*, 12, 743–777.

Khan, S., M. Ponomareva, and E. Tamer (2016): "Identification of panel data models with endogenous censoring," *Journal of Econometrics*, 194, 57–75.

——— (2023): "Identification of dynamic binary response models," *Journal of Econometrics*, 237, 105515.

Lee, S. M. S. and M. C. Pun (2006): "On m out of n Bootstrapping for Nonstandard M-Estimation With Nuisance Parameters," *Journal of the American Statistical Association*, 101, 1185–1197.

Liu, L., A. Poirier, and J.-L. Shiu (2021): "Identification and estimation of average partial effects in semiparametric binary response panel models," *arXiv preprint arXiv:2105.12891*.

Manski, C. F. (1987): "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica: Journal of the Econometric Society*, 357–362.

——— (2007): "Partial Identification of Counterfactual Choice Probabilities," *International Economic Review*, 48, 1393–1410.

Ouyang, F. and T. T. Yang (2023): "Semiparametric Discrete Choice Models for Bundles," ArXiv: 2306.04135.

——— (2024): "Semiparametric Estimation of Dynamic Binary Choice Panel Data Models," *Econometric Theory*, 1–40.

Pakes, A. and J. Porter (2024): "Moment inequalities for multinomial choice with fixed effects," *Quantitative Economics*, 15, 1–25.

Semenova, V. (2024): “Aggregated Intersection Bounds and Aggregated Minimax Values,” ArXiv: 2303.00982.

Shi, X., M. Shum, and W. Song (2018): “Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity,” *Econometrica*, 86, 737–761.

Tebaldi, P., A. Torgovitsky, and H. Yang (2023): “Nonparametric Estimates of Demand in the California Health Insurance Exchange,” *Econometrica*, 91, 107–146.

Torgovitsky, A. (2019): “Nonparametric Inference on State Dependence in Unemployment,” *Econometrica*, 87, 1475–1505.

Wang, R. (2023): “Testing and Identifying Substitution and Complementarity Patterns,” ArXiv: 2304.00678.

Appendix A Proofs

Proof of Theorem 1. Following Chesher and Rosen (2017), we adopt the notion of structures. In our case, a structure is a pair $m = (\theta, \mathcal{F}_{U|X})$. Each structure m delivers a conditional distribution $P_{Y|X}(\cdot|x; m)$ for each $x \in \text{Supp}(X)$. Let $\mathcal{P}_{Y|X}(m) = \{P_{Y|X}(\cdot|x; m) : x \in \text{Supp}(X)\}$. Let $\mathcal{M}$ be the set of structures that satisfy Assumption 1. Let $\mathcal{I}(\mathcal{M}, \mathcal{F}_{Y|X})$ denote the set of structures identified by $\mathcal{M}$ and $\mathcal{F}_{Y|X}$, that is, $m \in \mathcal{M}$ if m is admitted by $\mathcal{M}$ and $\mathcal{F}_{Y|X}$ and $\mathcal{P}_{Y|X}(m)$ agree. Then, the sharp identified set for $\mathcal{F}_{Y_t(\underline{x})|X}$ is defined as

$$\begin{aligned}\mathsf{F}^*_{Y_t(\underline{x})|X} = \{\mathcal{F}_{Y_t(\underline{x})|X} : \exists(\theta, \mathcal{F}_{U|X}) \in \mathcal{I}(\mathcal{M}, \mathcal{F}_{Y|X}) \\ \text{s.t. } \forall \mathcal{T} \in \mathsf{F}(\mathcal{Y}), F_{Y_t(\underline{x})|X=x}(\mathcal{T}) = F_{U_t|X=x}(\mathcal{U}(\mathcal{T}, \underline{x}; \theta)) \ a.e.\ x \in \text{Supp}(X)\}.\end{aligned}$$

Note that $\mathsf{F}^*_{Y_t(\underline{x})|X}$ depends on $\mathcal{I}(\mathcal{M}, \mathcal{F}_{Y|X})$ only through $(\theta, \{\mathcal{F}_{U_t|X}\}_{t=1}^T)$. Let $\mathcal{I}^*(\mathcal{M}, \mathcal{F}_{Y|X})$ denote the projection of $\mathcal{I}(\mathcal{M}, \mathcal{F}_{Y|X})$ onto $(\theta, \{\mathcal{F}_{U_t|X}\}_{t=1}^T)$. Then,

$$\begin{aligned}\mathsf{F}^*_{Y_t(\underline{x})|X} = \{\mathcal{F}_{Y_t(\underline{x})|X} : \exists(\theta, \{\mathcal{F}_{U_t|X}\}_{t=1}^T) \in \mathcal{I}^*(\mathcal{M}, \mathcal{F}_{Y|X}) \\ \text{s.t. } \forall \mathcal{T} \in \mathsf{F}(\mathcal{Y}), F_{Y_t(\underline{x})|X=x}(\mathcal{T}) = F_{U_t|X=x}(\mathcal{U}(\mathcal{T}, \underline{x}; \theta)) \ a.e.\ x \in \text{Supp}(X)\}.\end{aligned} \tag{11}$$

In static models, $(\theta, \{\mathcal{F}_{U_t|X}\}_{t=1}^T)$ only deliver the marginals of $P_{Y|X}(\cdot|x; m)$. By Sklar's theorem, there exists a collection of T-variate copula $\mathcal{C}_X = \{C_X(\cdot|x) : x \in \text{Supp}(X)\}$ such that $C_X(\cdot|x)$ reproduces the dependence structure of $P_{Y|X}(\cdot|x; m)$. In this sense, $(\theta, \{\mathcal{F}_{U_t|X}\}_{t=1}^T, \mathcal{C}_X)$ is observational equivalent to m. Since Assumption 1 only restricts $\{\mathcal{F}_{U_t|X}\}_{t=1}^T$, we can set $\mathcal{C}_X$ to be the collection of copulas associated with $\mathcal{F}_{Y|X}$ and require $(\theta, \{\mathcal{F}_{U_t|X}\}_{t=1}^T)$ to satisfy Assumption 1 and be consistent with the marginals of $\mathcal{F}_{Y|X}$. Hence,

$$\begin{aligned}\mathcal{I}^*(\mathcal{M}, \mathcal{F}_{Y|X}) = \{(\theta, \{\mathcal{F}_{U_t|X}\}_{t=1}^T) : \text{ Assumption 1 holds and } \forall t \in \{1, \ldots, T\}, \forall \mathcal{T} \in \mathsf{F}(\mathcal{Y}), \\ F_{Y_t|X=x}(\mathcal{T}) = F_{U_t|X=x}(\mathcal{U}(\mathcal{T}, x_t; \theta)) \ a.e.\ x \in \text{Supp}(X)\}.\end{aligned}$$

Finally, by Assumption 2, we can further write

$$\begin{aligned}\mathcal{I}^*(\mathcal{M}, \mathcal{F}_{Y|X}) = \{\theta_0\} \times \{\{\mathcal{F}_{U_t|X}\}_{t=1}^T : \text{ Assumption 1 holds and } \forall t \in \{1, \ldots, T\}, \forall \mathcal{T} \in \mathsf{F}(\mathcal{Y}), \\ F_{Y_t|X=x}(\mathcal{T}) = F_{U_t|X=x}(\mathcal{U}(\mathcal{T}, x_t; \theta_0)) \ a.e.\ x \in \text{Supp}(X)\}\end{aligned}$$

$$= \{\theta_0\} \times \bigcap_{t'=1}^{T} \{\mathcal{F}_{U_t|X} : \forall \mathcal{T} \in \mathsf{F}(\mathcal{Y}),$$

$$F_{Y_{t'}|X=x}(\mathcal{T}) = F_{U_t|X=x}(\mathcal{U}(\mathcal{T}, x_{t'}; \theta_0)) \; a.e. \; x \in \text{Supp}(X)\}. \tag{12}$$

The result follows by plugging (12) into (11). □

Proof of (5). Fix $\mathcal{T} \in \mathbb{Y}(x_t;\theta)$. For any $j \in \mathcal{T}$ and $k \notin \mathcal{T}$, $(x_{jt} - \underline{x}_j)^\top \beta \geq (x_{kt} - \underline{x}_k)^\top \beta$. Rearranging, $(x_{jt} - x_{kt})^\top \beta \geq (\underline{x}_j - \underline{x}_k)^\top \beta$. Take any $U_t \in \mathcal{U}(\mathcal{T}, \underline{x}; \theta)$. Then, there exists $j \in \mathcal{T}$ such that for any $k \notin \mathcal{T}$,

$$U_{kt} - U_{jt} \leq (\underline{x}_j - \underline{x}_k)^\top \beta \leq (x_{jt} - x_{kt})^\top \beta.$$

Hence, $U_t \in \mathcal{U}(\mathcal{T}, x_t; \theta)$. □

Proof of Theorem 2. By definition, $\mathcal{F}_{U_t|X} \in \mathsf{F}^*_{U_t|X}$ if and only if $\forall y' \in \mathcal{Y}$, $\forall t' \in \{1, \ldots, T\}$,

$$F_{Y_{t'}|X=x}([y', \infty)) = F_{U_t|X=x}(\mathcal{U}([y', \infty), x_{t'}; \theta_0)) \; a.e. \; x \in \text{Supp}(X).$$

It follows that

$$\begin{aligned}
\mathsf{F}^*_{Y_t(\underline{x})|X} = \{&\mathcal{F}_{Y_t(\underline{x})|X} : \exists \mathcal{F}_{U_t|X} \text{ s.t. } \forall y \in \mathcal{Y}, \forall t' \in \{1, \ldots, T\}, \\
&F_{Y_t(\underline{x})|X=x}([y, \infty)) = F_{U_t|X=x}(\mathcal{U}([y, \infty), \underline{x}; \theta_0)), \\
&F_{Y_{t'}|X=x}([y, \infty)) = F_{U_t|X=x}(\mathcal{U}([y, \infty), x_{t'}; \theta_0)) \; a.e. \; x \in \text{Supp}(X)\} \\
= \{&\mathcal{F}_{Y_t(\underline{x})|X} : \exists \mathcal{F}_{U_t|X} \text{ s.t. } \forall y \in \mathcal{Y}, \forall t' \in \{1, \ldots, T\}, \\
&F_{Y_t(\underline{x})|X=x}([y, \infty)) = F_{U_t|X=x}([-\underline{x}^\top \beta_0 + h^-(y, \gamma_0), \infty)), \\
&F_{Y_{t'}|X=x}([y, \infty)) = F_{U_t|X=x}([-x_{t'}^\top \beta_0 + h^-(y, \gamma_0), \infty)) \; a.e. \; x \in \text{Supp}(X)\},
\end{aligned}$$

where the second equality follows from (3). Taking $\mathcal{F}_{Y_t(\underline{x})|X}$ from the right-hand side of (6), we want to show that $\mathcal{F}_{Y_t(\underline{x})|X} \in \mathsf{F}^*_{Y_t(\underline{x})|X}$, which amounts to for all $x \in \text{Supp}(X)$ exhibiting $F_{U_t|X=x}$ satisfying $\forall y \in \mathcal{Y}$,

$$\begin{aligned}
F_{Y_t(\underline{x})|X=x}([y, \infty)) &= F_{U_t|X=x}([-\underline{x}^\top \beta_0 + h^-(y, \gamma_0), \infty)), \\
F_{Y_{t'}|X=x}([y, \infty)) &= F_{U_t|X=x}([-x_{t'}^\top \beta_0 + h^-(y, \gamma_0), \infty)), \; t' = 1, \ldots, T.
\end{aligned}$$

Fix $x \in \text{Supp}(X)$. The desired $F_{U_t|X=x}$ can be constructed as follows. Define

$$\begin{aligned} p_{t'}(y) &= F_{Y_{t'}|X=x}([y,\infty)),\ t'=1,\ldots,T, \\ p_{T+1}(y) &= F_{Y_t(\underline{x})|X=x}([y,\infty)), \\ \underline{u}_{t'}(y) &= -x_{t'}^\top\beta_0 + h^-(y,\gamma_0),\ t'=1,\ldots,T, \\ \underline{u}_{T+1}(y) &= -\underline{x}^\top\beta_0 + h^-(y,\gamma_0). \end{aligned}$$

Then, (6) ensures that for any $t' \in \{1,\ldots,T\}$ and $y, y' \in \mathcal{Y}$,

$$\begin{aligned} \underline{u}_{T+1}(y) \geq \underline{u}_{t'}(y') &\iff p_{T+1}(y) \leq p_{t'}(y'), \\ \underline{u}_{T+1}(y) \leq \underline{u}_{t'}(y') &\iff p_{T+1}(y) \geq p_{t'}(y'), \end{aligned}$$

Also, by Lemma 1 of Botosaru et al. (2023), Assumption 2 ensures that for any $t', t'' \in \{1,\ldots,T\}$ and $y, y' \in \mathcal{Y}$,

$$\underline{u}_{t'}(y) \leq \underline{u}_{t''}(y') \iff p_{t'}(y) \geq p_{t''}(y').$$

Put together, we have for any $t', t'' \in \{1,\ldots,T+1\}$ and $y, y' \in \mathcal{Y}$,

$$\underline{u}_{t'}(y) \leq \underline{u}_{t''}(y') \iff p_{t'}(y) \geq p_{t''}(y'). \tag{13}$$

For $u \in \mathbb{R}$, define

$$(t^*(u), y^*(u)) = \underset{(t',y)\in\{1,\ldots,T+1\}\times\mathcal{Y}:\underline{u}_{t'}(y)\leq u}{\arg\max} \underline{u}_{t'}(y).$$

We can set

$$F_{U_t|X=x}([u,\infty)) = p_{t^*(u)}(y^*(u)),\ u \in \mathbb{R}.$$

We now show that $F_{U_t|X=x}$ satisfies the monotonicity requirement of a CDF, i.e.,

$$F_{U_t|X=x}([u,\infty)) \geq F_{U_t|X=x}([u',\infty)),\ \forall u \leq u'.$$

To see this, note that by definition,

$$\underline{u}_{t^*(u)}(y^*(u)) \leq \underline{u}_{t^*(u')}(y^*(u')).$$

which implies that

$$F_{U_t|X=x}([u,\infty)) = p_{t^*(u)}(y^*(u)) \geq p_{t^*(u')}(y^*(u')) = F_{U_t|X=x}([u',\infty)),$$

where the inequality follows from (13). □

Proof of Theorem 3. Taking $\mathcal{F}_{Y_t(\underline{x})|X}$ from the right-hand side of (7), we want to show that $\mathcal{F}_{Y_t(\underline{x})|X} \in \mathsf{F}^*_{Y_t(\underline{x})|X}$, which amounts to for all $x \in \operatorname{Supp}(X)$ exhibiting $F_{U_t|X=x}$ satisfying

$$\begin{aligned} F_{Y_t(\underline{x})|X=x}(\{j\}) &= F_{U_t|X=x}(\mathcal{U}(j,\underline{x};\theta_0)), \\ F_{Y_{t'}|X=x}(\{j\}) &= F_{U_t|X=x}(\mathcal{U}(j,x_{t'};\theta_0)), \end{aligned}$$

for all $j \in \{0,1,\ldots,J\}$ and $t' \in \{1,\ldots,T\}$. Fix $x \in \operatorname{Supp}(X)$. Define $\mathcal{U}_{j_1,\ldots,j_T,j'} = \mathcal{U}(j_1,x_1;\theta_0) \cap \cdots \cap \mathcal{U}(j_T,x_T;\theta_0) \cap \mathcal{U}(j',\underline{x};\theta_0)$ and $q_{j_1,\ldots,j_T,j'} = F_{U_t|X=x}(\mathcal{U}_{j_1,\ldots,j_T,j'})$. Note that $q_{j_1,\ldots,j_T,j'} = 0$ if $\mathcal{U}_{j_1,\ldots,j_T,j'} = \emptyset$. The probabilities $q = \{q_{j_1,\ldots,j_T,j'} : \mathcal{U}_{j_1,\ldots,j_T,j'} \neq \emptyset\}$ are the building blocks for constructing $F_{U_t|X=x}$. We can rephrase our task as exhibiting $q_{j_1,\ldots,j_T,j'} \geq 0$ satisfying

$$\sum_{(j_1,\ldots,j_T,j'):\ \mathcal{U}_{j_1,\ldots,j_T,j'}\neq\emptyset,\ j'=j} q_{j_1,\ldots,j_T,j'} = F_{Y_t(\underline{x})|X=x}(\{j\}), \tag{14}$$

$$\sum_{(j_1,\ldots,j_T,j'):\ \mathcal{U}_{j_1,\ldots,j_T,j'}\neq\emptyset,\ j_{t'}=j} q_{j_1,\ldots,j_T,j'} = F_{Y_{t'}|X=x}(\{j\}), \tag{15}$$

for all $j \in \{0,1,\ldots,J\}$ and $t' \in \{1,\ldots,T\}$. Let

$$p^{\text{ct}} = \begin{bmatrix} F_{Y_t(\underline{x})|X=x}(\{0\}) \\ F_{Y_t(\underline{x})|X=x}(\{1\}) \\ \vdots \\ F_{Y_t(\underline{x})|X=x}(\{J\}) \end{bmatrix} \text{ and } p^{\text{ob}}_{t'} = \begin{bmatrix} F_{Y_{t'}|X=x}(\{0\}) \\ F_{Y_{t'}|X=x}(\{1\}) \\ \vdots \\ F_{Y_{t'}|X=x}(\{J\}) \end{bmatrix},\ t' = 1,\ldots,T.$$

Let Q^{ct} be the matrix with elements in $\{0,1\}$ such that (14) can be restated as $Q^{\text{ct}}q = p^{\text{ct}}$ and let $Q^{\text{ob}}_{t'}$ be the matrix with elements in $\{0,1\}$ such that (15) can be restated as $Q^{\text{ob}}_{t'}q = p^{\text{ob}}_{t'}$. Our task can be summarized as showing that $\exists q \geq 0$ such that: (A) $Q^{\text{ct}}q = p^{\text{ct}}$ and (B) $Q^{\text{ob}}_{t'}q = p^{\text{ob}}_{t'}$, $\forall t'$. Let $\{z^{t'} = (z^{t'}_0, z^{t'}_1, \ldots, z^{t'}_J)^\mathsf{T}\}_{t'=1}^T$ and $w = (w_0, w_1, \ldots, w_J)^\mathsf{T}$ be $(J+1)$-dimensional constant vectors.

Farkas's Lemma states that if

$$w^{\intercal} Q^{\mathrm{ct}} + \sum_{t'=1}^{T} (z^{t'})^{\intercal} Q_{t'}^{\mathrm{ob}} \geq 0 \text{ implies } w^{\intercal} p^{\mathrm{ct}} + \sum_{t'=1}^{T} (z^{t'})^{\intercal} p_{t'}^{\mathrm{ob}} \geq 0,$$

then $\exists q \geq 0$ satisfying constraints (A) and (B). For each $t' \in \{1, \ldots, T\}$, there exists a weak ordering for $\{(x_{jt'} - \underline{x}_j)^{\top} \beta_0\}_{j=0}^{J}$. Let $M_{t'}(j)$ denote the rank of alternative j in this ordering and $M_{t'}^{-1}$ denote the inverse mapping. Then, $\{M_{t'}^{-1}(J), \ldots, M_{t'}^{-1}(j)\} \in \mathbb{Y}(x_{t'}; \theta_0)$ for $j > 0$. For any $\{a_j^{t'}\}_{j=0,1,\ldots,J,t'=1,\ldots,T} \in \mathbb{R}$,

$$\begin{aligned}
& w^{\intercal} p^{\mathrm{ct}} + \sum_{t'=1}^{T} (z^{t'})^{\intercal} p_{t'}^{\mathrm{ob}} \\
= \; & \sum_{j=0}^{J} w_j F_{Y_t(\underline{x})|X=x}(\{j\}) + \sum_{t'=1}^{T} \sum_{j=0}^{J} z_j^{t'} F_{Y_{t'}|X=x}(\{j\}) \\
= \; & \sum_{t'=1}^{T} \sum_{j=0}^{J} a^{t'}_{M_{t'}^{-1}(j)} \underbrace{(F_{Y_{t'}|X=x}(\{M_{t'}^{-1}(J), \ldots, M_{t'}^{-1}(j)\}) - F_{Y_t(\underline{x})|X=x}(\{M_{t'}^{-1}(J), \ldots, M_{t'}^{-1}(j)\}))}_{\geq 0 \text{ by } (7)} \\
& + \sum_{j=0}^{J} \Big(w_j + \sum_{t'=1}^{T} \sum_{\ell: M_{t'}(\ell) \leq M_{t'}(j)} a_\ell^{t'} \Big) F_{Y_t(\underline{x})|X=x}(\{j\}) + \sum_{t'=1}^{T} \sum_{j=0}^{J} \Big(z_j^{t'} - \sum_{\ell: M_{t'}(\ell) \leq M_{t'}(j)} a_\ell^{t'} \Big) F_{Y_{t'}|X=x}(\{j\}).
\end{aligned}$$

Therefore, given $w^{\intercal} Q^{\mathrm{ct}} + \sum_{t'=1}^{T} (z^{t'})^{\intercal} Q_{t'}^{\mathrm{ob}} \geq 0$, we have $w^{\intercal} p^{\mathrm{ct}} + \sum_{t'=1}^{T} (z^{t'})^{\intercal} p_{t'}^{\mathrm{ob}} \geq 0$ if we can find $\{a_j^{t'}\}_{j=0,1,\ldots,J,t'=1,\ldots,T} \in \mathbb{R}$ satisfying

$$\begin{aligned}
& w_j + \sum_{t'=1}^{T} \sum_{\ell: M_{t'}(\ell) \leq M_{t'}(j)} a_\ell^{t'} \geq 0, \ \forall j, \\
& z_j^{t'} - \sum_{\ell: M_{t'}(\ell) \leq M_{t'}(j)} a_\ell^{t'} \geq 0, \ \forall j, t', \\
& a_j^{t'} \geq 0 \text{ if } M_{t'}(j) > 0, \ \forall t'.
\end{aligned}$$

From the examination of matrices Q^{ct} and $Q_1^{\mathrm{ob}}, \ldots, Q_T^{\mathrm{ob}}$, $w^{\intercal} Q^{\mathrm{ct}} + \sum_{t'=1}^{T} (z^{t'})^{\intercal} Q_{t'}^{\mathrm{ob}} \geq 0$ yields

$$w_{j'} + \sum_{t'=1}^{T} z_{j_{t'}}^{t'} \geq 0 \text{ if } \mathcal{U}_{j_1, \ldots, j_T, j'} \neq \emptyset.$$

For $j = 0, 1, \dots, J$, let

$$\begin{aligned}
\underline{a}_j^1 &= \min_{\ell\colon \mathcal{U}_{\ell, j_2, \dots, j_T, j} \neq \emptyset} z_\ell^1, \\
\underline{a}_j^{t'} &= \min_{\ell\colon \mathcal{U}_{\dots, j_{t'-1}, \ell, j_{t'+1}, \dots, j} \neq \emptyset} z_\ell^{t'},\ 1 < t' < T, \\
\underline{a}_j^T &= \min_{\ell\colon \mathcal{U}_{j_1, \dots, j_{T-1}, \ell, j} \neq \emptyset} z_\ell^T.
\end{aligned}$$

Then, $w_j + \sum_{t'=1}^T \underline{a}_j^{t'} \geq 0,\ \forall j$. Since $\mathcal{U}_{j_1, \dots, j_T, j'} \neq \emptyset$ when $j_1 = \cdots = j_T = j'$, we also have $\underline{a}_j^{t'} \leq z_j^{t'},\ \forall j, t'$. Moreover, note that $\mathcal{U}_{j_1, \dots, j_T, j'} \neq \emptyset$ implies that $M_{t'}(j_{t'}) \geq M_{t'}(j'),\ \forall t'$. Hence, $\underline{a}_{M_{t'}^{-1}(j)}^{t'}$ is increasing in j. The desired $\{a_j^{t'}\}_{j=0,1,\dots,J, t'=1,\dots,T}$ can be constructed as follows:

$$\begin{aligned}
a_{M_{t'}^{-1}(0)}^{t'} &= \underline{a}_{M_{t'}^{-1}(0)}^{t'}, \\
a_{M_{t'}^{-1}(j)}^{t'} &= \underline{a}_{M_{t'}^{-1}(j)}^{t'} - \underline{a}_{M_{t'}^{-1}(j-1)}^{t'},\ j = 1, \dots, J.
\end{aligned}$$

It remains to construct $F_{U_t|X=x}$. For each $\mathcal{U}_{j_1, \dots, j_T, j'} \neq \emptyset$, choose a point $r_{j_1, \dots, j_T, j'} \in \mathcal{U}_{j_1, \dots, j_T, j'}$. Then, define $F_{U_t|X=x}$ to be the discrete distribution on support points $r_{j_1, \dots, j_T, j'}$ with $F_{U_t|X=x}(\{r_{j_1, \dots, j_T, j'}\}) = q_{j_1, \dots, j_T, j'}$. Now we can conclude that (7) holds. □

Proof of Theorem 4. For each function $f : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}$, let $\mathbb{G}_N(f(Y, X)) = N^{-1/2} \sum_{i=1}^N (f(Y_i, X_i) - E[f(Y_i, X_i)])$. By the standard decomposition, we have

$$\begin{aligned}
\sqrt{N}(\hat{\Psi}(\theta) - \Psi(\theta)) = \mathbb{G}_n\Big(&\sum_{\lambda \in \Lambda(X;\theta)} 1\{\lambda^*(X; \theta, \tau_0) = \lambda\} \lambda^\top I(Y) \Big) && (16) \\
&+ \mathbb{G}_n\Big(\sum_{\lambda \in \Lambda(X;\theta)} (1\{\lambda^*(X; \theta, \hat{\tau}) = \lambda\} - 1\{\lambda^*(X; \theta, \tau_0) = \lambda\}) \lambda^\top I(Y) \Big) && (17) \\
&+ \sqrt{N} E\Big[\sum_{\lambda \in \Lambda(X;\theta)} (1\{\lambda^*(X; \theta, \hat{\tau}) = \lambda\} - 1\{\lambda^*(X; \theta, \tau_0) = \lambda\}) \lambda^\top I(Y) \Big]. && (18)
\end{aligned}$$

To show (17) and (18) are $o_p(1)$, we will use the following lemma:

Lemma 1. *Suppose that Assumptions 3 and 5 hold. Then, for all θ, there exists $C > 0$ such that for any $\delta \geq 0$,*

$$\Pr\left(0 < \min_{\lambda \in \Lambda(X;\theta) : \lambda \neq \lambda^*(X;\theta,\tau_0)} (\lambda - \lambda^*(X; \theta, \tau_0))^\top \tau_0(X) \leq \delta \right) \leq C\delta.$$

First, by Assumption 6, (17) is $o_p(1)$ if the stochastic equicontinuity property holds: for all positive values $\delta_N = o(1)$,

$$\sup_{\|\tau-\tau_0\|_\infty \leq \delta_N} \Big| \mathbb{G}_n\Big(\sum_{\lambda \in \Lambda(X;\theta)} (1\{\lambda^*(X;\theta,\tau)=\lambda\} - 1\{\lambda^*(X;\theta,\tau_0)=\lambda\})\lambda^\top I(Y) \Big) \Big| = o_p(1).$$

To this end, note that by Assumption 3,

$$\Big| \sum_{\lambda \in \Lambda(X;\theta)} (1\{\lambda^*(X;\theta,\tau)=\lambda\} - 1\{\lambda^*(X;\theta,\tau_0)=\lambda\})\lambda^\top I(Y) \Big| \leq M \cdot 1\{\lambda^*(X;\theta,\tau) \neq \lambda^*(X;\theta,\tau_0)\},$$

where

$$\begin{aligned} & 1\{\lambda^*(X;\theta,\tau) \neq \lambda^*(X;\theta,\tau_0)\} \\ = \; & 1\{0 < (\lambda^*(X;\theta,\tau) - \lambda^*(X;\theta,\tau_0))^\top \tau_0(X) < (\lambda^*(X;\theta,\tau) - \lambda^*(X;\theta,\tau_0))^\top(\tau_0(X) - \tau(X))\} \\ \leq \; & 1\Big\{0 < \min_{\lambda \in \Lambda(X;\theta): \lambda \neq \lambda^*(X;\theta,\tau_0)} (\lambda - \lambda^*(X;\theta,\tau_0))^\top \tau_0(X) \leq M\|\tau - \tau_0\|_\infty \Big\} \end{aligned}$$

It follows that

$$\begin{aligned} & E\Big[\sup_{\|\tau-\tau_0\|_\infty \leq \delta_N} \Big| \sum_{\lambda \in \Lambda(X;\theta)} (1\{\lambda^*(X;\theta,\tau)=\lambda\} - 1\{\lambda^*(X;\theta,\tau_0)=\lambda\})\lambda^\top I(Y) \Big| \Big] \\ \leq \; & \Pr\Big(0 < \min_{\lambda \in \Lambda(X;\theta): \lambda \neq \lambda^*(X;\theta,\tau_0)} (\lambda - \lambda^*(X;\theta,\tau_0))^\top \tau_0(X) \leq \delta_N \Big). \end{aligned}$$

By Lemma 1 and Theorem 3 of Chen, Linton, and Van Keilegom (2003), (17) is $o_p(1)$. Second, for (18), we observe that

$$\begin{aligned} & E\Big[\sum_{\lambda \in \Lambda(X;\theta)} (1\{\lambda^*(X;\theta,\hat\tau)=\lambda\} - 1\{\lambda^*(X;\theta,\tau_0)=\lambda\})\lambda^\top I(Y) \Big| \hat\tau \Big] \\ = \; & E\Big[\sum_{\lambda \in \Lambda(X;\theta)} (1\{\lambda^*(X;\theta,\hat\tau)=\lambda\} - 1\{\lambda^*(X;\theta,\tau_0)=\lambda\})\lambda^\top \tau_0(X) \Big| \hat\tau \Big] \\ = \; & E[(\lambda^*(X;\theta,\hat\tau) - \lambda^*(X;\theta,\tau_0))^\top \tau_0(X) 1\{(\lambda^*(X;\theta,\hat\tau) - \lambda^*(X;\theta,\tau_0))^\top \tau_0(X) > 0\} | \hat\tau] \\ \leq \; & E\Big[(\lambda^*(X;\theta,\hat\tau) - \lambda^*(X;\theta,\tau_0))^\top (\tau_0(X) - \hat\tau(X)) \\ & \cdot 1\{0 < (\lambda^*(X;\theta,\hat\tau) - \lambda^*(X;\theta,\tau_0))^\top \tau_0(X) < (\lambda^*(X;\theta,\hat\tau) - \lambda^*(X;\theta,\tau_0))^\top (\tau_0(X) - \hat\tau(X))\} \Big| \hat\tau \Big] \\ \leq \; & M\|\hat\tau - \tau_0\|_\infty \Pr\Big(0 < \min_{\lambda \in \Lambda(X;\theta): \lambda \neq \lambda^*(X;\theta,\tau_0)} (\lambda - \lambda^*(X;\theta,\tau_0))^\top \tau_0(X) \leq M\|\hat\tau - \tau_0\|_\infty \Big| \hat\tau \Big) \\ \leq \; & CM^2 \|\hat\tau - \tau_0\|_\infty^2, \end{aligned}$$

where the last inequality follows from Lemma 1. Then, by Assumption 6, (18) is $o_p(1)$. Now we can apply the central limit theorem to (16) to obtain the desired result. □