

Identification of Dynamic Nonlinear Panel Models under Partial Stationarity*

Wayne Yuan Gao[†] and Rui Wang[‡]

January 10, 2024

Abstract

This paper studies identification for a wide range of nonlinear panel data models, including binary choice, ordered response, and other types of limited dependent variable models. Our approach accommodates dynamic models with any number of lagged dependent variables as well as other types of (potentially contemporary) endogeneity. Our identification strategy relies on a partial stationarity condition, which not only allows for an unknown distribution of errors but also for temporal dependencies in errors. We derive partial identification results under flexible model specifications and provide additional support conditions for point identification. We demonstrate the robust finite-sample performance of our approach using Monte Carlo simulations, and apply the approach to analyze the empirical application of income categories using various ordered choice models.

Keywords: Panel Discrete Choice Models; Stationarity; Dynamic Models; Partial Identification; Endogeneity

*We are grateful to Jason Blevins, Robert de Jong, Kyoo il Kim, Louise Laage, Eric Mbakop, Liyang Sun, and Jeffrey Wooldridge for helpful comments and suggestions.

[†]Department of Economics, University of Pennsylvania, 133 S 36th St, Philadelphia, PA 19104, USA. Email: waynegao@upenn.edu

[‡]Department of Economics, Ohio State University, 1945 N High St, Columbus, OH 43210, USA. Email: wang.16498@osu.edu

1 Introduction

This paper studies semiparametric partial identification of a wide range of panel data models with limited dependent variables, including various discrete choice models, censored dependent variable models, and nonseparable models. In particular, our approach accommodates dynamic models with any number of lagged dependent variables as well as other types of endogenous covariates.

To fix ideas, we start with the following dynamic binary choice model, which is on its own of considerable theoretical and applied interest. Sections 3-6 generalize the approach to other limited dependent variable models. Specifically, consider

$$Y_{it} = \mathbb{1}\{Z'_{it}\beta_0 + X'_{it}\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\}, \quad (1)$$

where $Y_{it} \in \{0, 1\}$ denotes a binary outcome variable for individual $i = 1, 2, \dots$ and time $t = 1, \dots, T$, while $Z_{it} \in \mathcal{R}^{d_z}$ denotes exogenous covariates, $X_{it} \in \mathcal{R}^{d_x}$ denotes potentially endogenous covariates, $\alpha_i \in \mathcal{R}$ denotes the unobserved fixed effect for individual i , and ϵ_{it} denotes the unobserved time-varying error term for individual i at time t . The objective is to identify the parameter $\theta_0 := (\beta'_0, \gamma'_0)'$ using a panel of observed variables $(Z_{it}, X_{it}, Y_{it})_{it}$. We focus on short panels, where the number of time periods $T \geq 2$ is fixed and finite.

The identification of model (1) has been explored in the literature under various assumptions. For example, Chamberlain (1980) examines identification under the logistic distribution of ϵ_{it} and the independence of ϵ_{it} with respect to $(\alpha_i, \{(Z_{it}, X_{it})\}_{t=1}^T)$. Subsequently, Manski (1987) relaxes the distributional assumption and employs the following conditional stationarity of ϵ_{it} to achieve identification:

$$\epsilon_{is} \mid Z_{is}, Z_{it}, X_{is}, X_{it}, \alpha_i \stackrel{d}{\sim} \epsilon_{it} \mid Z_{is}, Z_{it}, X_{is}, X_{it}, \alpha_i \quad \text{for any } s, t \leq T. \quad (2)$$

This condition is also referred to as “group stationarity” or “group homogeneity” and has also been exploited in studies such as Chernozhukov et al. (2013), Shi, Shum, and Song (2018) and Pakes and Porter (2022). Condition (2) does not impose parametric restrictions on the distributions of ϵ_{it} and allows dependence between the fixed effect α_i and the covariates (Z_{it}, X_{it}) . However, condition (2) does impose substantial restriction on the dependence between (Z_{it}, X_{it}) and the time-varying error term ϵ_{it} : it effectively requires that all covariates in (Z_{it}, X_{it}) are exogeneous with respect to the time varying error ϵ_{it} .¹

In many economic applications, certain components of the observable covariates, namely X_{it} , may exhibit endogeneity. For example, in a *dynamic* setting where X_{it} includes the lagged outcome variable $Y_{i,t-1}$, then endogeneity of $Y_{i,t-1}$ with respect to $\epsilon_{i,t-1}$ arises im-

¹For instance, suppose $\mathbb{E}[\epsilon_{it} \mid Z_{is}, Z_{it}, X_{is}, X_{it}] = X'_{it}\eta$, then the conditional distributions of ϵ_{it} and ϵ_{is} cannot be the same as long as $X'_{it}\eta \neq X'_{is}\eta$, so condition (2) fails in general.

diately. For another example, if X_{it} includes “price” or other variables that may be endogenously chosen by economic agents, then the exogeneity restriction imposed by condition (2) will again fail to hold.

We propose in this paper a general identification approach for various nonlinear panel models in the presence of endogenous regressors, using the following “partial stationarity” condition, which can be viewed as a weaker version of the stationarity condition (2) above:

$$\epsilon_{is} \mid Z_{is}, Z_{it}, \alpha_i \stackrel{d}{\sim} \epsilon_{it} \mid Z_{is}, Z_{it}, \alpha_i \quad \text{for any } s, t \leq T. \quad (3)$$

Our partial stationarity condition (3), as its name suggests, only requires that the errors are stationary across time periods conditional on the realizations of a subvector of the covariates (i.e., the exogenous covariates denoted by Z) while allowing the remaining covariates (denoted by X) to be endogenous in arbitrary manners. In short, condition (3) imposes exogeneity conditions only on exogenous covariates. Alternatively, we can interpret condition (3) as an assumption of the existence of *some* covariates being exogenous.²

We describe how to exploit the partial stationarity condition (3) to derive the identified set on the model parameters θ_0 through conditional moment inequalities, which take the form of lower and upper bounds for the conditional distribution $\epsilon_{it} + \alpha_i \mid Z_{is}, Z_{it}$, solely as functions of observed variables and the model parameters θ_0 . We show that these bounds must have nonzero intersections over time under the partial stationarity assumption, thereby forming identifying restrictions for the parameter θ_0 .

While extensive work exists on nonlinear panel models under stationarity, previous studies typically focus on individual models. The identification strategies are often context specific and the identification results may have various complicated representations, as seen in studies such as [Khan, Ponomareva, and Tamer \(2023\)](#) and [Pakes and Porter \(2022\)](#). Our approach offers the advantage of providing a simple and unified characterization of the identified set for a broad range of static and dynamic panel models, irrespective of the specific types of variables (discrete/continuous outcome and covariates) and forms of endogeneity (lagged/contemporary endogenous regressors). Furthermore, our strategy does not rely on scalar-additive structures that are often imposed in various parametric and semiparametric models, and can be further extended to accommodate nonseparable panel data models.

We demonstrate the sharpness of the identified set for binary choice models when the support of X_{it} is finite. More precisely, we show that, for any θ that satisfies all the conditional moment inequalities we derived, we can construct an observationally equivalent joint distribution of the observed and unobserved variables in our model. While our main result is about set identification, we also provide sufficient conditions for the point identification of

²Condition (3) also accommodates the standard stationarity assumption conditional on all covariates.

the coefficients β_0 on exogenous covariates (under scale normalization) as well as the signs of the coefficients on endogenous covariates γ_0 .

Our identification strategy based on partial stationarity can be applied more broadly beyond the context of dynamic binary choice models. We demonstrate its applicability to other limited dependent variable models, such as ordered response models, multinomial choice models, and censored outcome models. The results of our approach accommodates both static and dynamic settings across all these models. Furthermore, we illustrate the adaptability of our key strategy to nonseparable semiparametric models with endogeneity.

We characterize the identified set using a collection of conditional moment inequalities, based on which estimation and inference can be conducted using established econometric methods in the literature, such as Chernozhukov, Hong, and Tamer (2007), Andrews and Shi (2013), and Chernozhukov, Lee, and Rosen (2013). Through Monte Carlo simulations, we demonstrate that our identification method yields informative and robust finite-sample confidence intervals for coefficients in both static and dynamic models.

Literature Review

Our paper contributes directly to the line of econometric literature on semiparametric panel discrete choice models. Dating back to Manski (1987), a series of work exploits “full” stationarity conditions for identification, such as Chernozhukov et al. (2013), Khan, Ponomareva, and Tamer (2016), Shi, Shum, and Song (2018), Pakes and Porter (2022), Khan, Ouyang, and Tamer (2021), Khan, Ponomareva, and Tamer (2023), Gao and Li (2020), and Wang (2022). As discussed above, full stationarity conditions given all observable covariates effectively require that all covariates are exogenous with no dynamic effects (i.e., lagged dependent variables). In contrast, we exploit the “partial” stationarity condition, thereby allowing for lagged dependent variables as well as other endogenous covariates.

In the literature on dynamic discrete choice models, our paper is most closely related to Khan, Ponomareva, and Tamer (2023, KPT thereafter), who studies the following dynamic panel binary choice model

$$Y_{it} = \mathbb{1}\{Z'_{it}\beta_0 + Y_{i,t-1}\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\}, \quad (4)$$

where the one-period lagged dependent variable $Y_{i,t-1} \in \{0, 1\}$ serves as the endogenous covariate, and Z_{it} are exogenous covariates. KPT exploits the following assumption:

$$\epsilon_{is} \mid Z_{it}, Z_{is}, \alpha_i \stackrel{d}{\sim} \epsilon_{it} \mid Z_{it}, Z_{is}, \alpha_i \quad \text{for any } s, t \leq T.$$

which is exactly a “partial stationarity” condition in the specific context of (4), and derives

the sharp identified set for θ_0 by explicitly enumerating the realizations of the one-period lagged outcome variable $Y_{i,t-1}$ (across two periods t, s). In contrast, our model (1), along with the “partial stationarity” condition, is stated with more general specifications of the endogenous covariates X_{it} . The covariates X_{it} can include more than one lagged dependent variables (e.g. $Y_{i,t-1}, Y_{i,t-2}, \dots$) and other endogenous variables (such as “price” if Y_{it} represents the purchase of a particular product), which may be continuously valued. Consequently, our identification strategy is substantially different from that of KPT, and can be applied more broadly to various other dynamic nonlinear panel models. In the specific model (4), we show that the identifying restrictions we derived are equivalent to those derived in KPT and thus both approaches lead to sharp identification. Relatedly, Mbakop (2023) proposes a computational algorithm to derive conditional moment inequalities in a general class of dynamic discrete choice models (potentially with multiple lags). The focus of Mbakop (2023) is on scenarios where the lagged discrete outcome variables are the only endogenous covariates in the model, and the proposed algorithm relies on the discreteness of these variables. Relative to these works, our paper introduces an analytic approach that directly applies to a more general class of dynamic binary choice models, as well as other types of models with continuous limited dependent variables and any number of endogenous covariates, regardless of whether they are discrete or continuous and whether they take the form of lagged outcome variables or not.

Our identification strategy relies on a type of stationarity condition, while alternative approaches utilize other notions of exogeneity. For example, Aristodemou (2021) exploits a type of “full independence” assumption to identify dynamic binary choice models. The “full independence” assumption essentially requires that the time-varying errors from all time periods and the exogenous variables from all time periods are independent (conditional on initial conditions), but does not make intertemporal restrictions on the errors (such as stationarity). Hence, such “full independence” assumption and the partial stationarity assumption in our paper do not nest each other as special cases. Chesher, Rosen, and Zhang (2023) applies the framework of generalized instrumental variables (Chesher and Rosen, 2017) to the context of various dynamic discrete choice models with fixed effects, and utilizes a similar “full independence” assumption (Aristodemou, 2021) for identification.³ More differently, some other papers work with sequential exogeneity in various dynamic panel mod-

³Our identification strategy shares some conceptual similarity with the idea of generalized instrumental variable (GIV) in Chesher and Rosen (2017), who proposes a general approach for representing the identified set of structural models with endogeneity. Chesher and Rosen (2017), Chesher and Rosen (2020), and Chesher, Rosen, and Zhang (2023) demonstrate how the GIV framework can be applied to various settings, but focus mostly on the use of exclusion restrictions and/or full independence assumptions. In this paper, we neither impose exclusion restrictions nor independence assumptions but instead explore identification under a partial stationarity condition.

els and provide (non-)identification results under different model restrictions. For example, [Shiu and Hu \(2013\)](#) imposes a high-level invertibility condition along with a restriction that rules out the dependence of covariates on past dependent variables. More recently, [Bonhomme, Dano, and Graham \(2023\)](#) investigates panel binary choice models with a single binary predetermined covariate under *sequential exogeneity*, whose evolution may depend on the past history of outcome and covariate values. The sequential exogeneity condition considered in these papers and the partial stationarity condition in ours again do not nest each other as special cases: in particular, our current paper accommodates contemporarily endogenous covariates that violate sequential exogeneity. In summary, the key assumptions, identification strategy, and identification results of these studies are substantially different from and thus not directly comparable to those in our current paper.

Our paper is also complementary to the line of literature that studies dynamic logit models with fixed effects for binary, ordered responses, or multinomial choice models. This literature typically assumes that time-varying errors are conditionally independent across time, independent from all other variables, and follow the logistic distribution. The logit assumption in panel data models has long been studied, such as in [Chamberlain \(1984\)](#) and [Chamberlain \(2010\)](#). In the context of dynamic discrete choice models, [Honoré and Kyriazidou \(2000\)](#) first shows how to conduct differencing of fixed effects under the logit assumption, while recent papers by [Honoré and Weidner \(2020\)](#) and [Dano \(2023\)](#) illustrate how to systematically obtain moment conditions free of fixed effects and time-varying errors. [Honoré, Muris, and Weidner \(2021\)](#) extends the approach in [Honoré and Weidner \(2020\)](#) to dynamic ordered logit models. Meanwhile, [Dobronyi, Gu, and Kim \(2021\)](#) derives sharp identification for dynamic logit models using a different approach based on truncated moments. Alternatively, [Honoré and Tamer \(2006\)](#) proposes a linear programming method to obtain bounds on model parameters and average marginal effects under logit and other parametric error distributions. In addition, [Davezies, D’Haultfoeuille, and Laage \(2021\)](#) provides analytic bounds on average marginal effects in static logit models. Relative to this line of literature, our paper does not require parametric (logistic) or conditional independence assumptions, and provides general semiparametric identification results for various dynamic panel models.

The rest of the paper is organized as follows. [Section 2](#) studies the sharp identification of binary choice models with endogenous covariates. We provide sufficient conditions to achieve point identification and explore two classic models: static and dynamic models. [Sections 3-5](#) apply the identification approach to ordered response models, multinomial choice models, and models with censored dependent variables. [Section 6](#) broadens the scope to general nonseparable models. [Section 7](#) presents simulation results about the finite-sample

performances of our approach and Section 8 explores the empirical application of income categories using various ordered response models. We conclude with Section 9.

2 Binary Choice Model

2.1 Model

In this section, we focus on the identification of dynamic binary choice models with endogenous covariates. Specifically, consider the following binary choice model:

$$Y_{it} = \mathbb{1}\{Z'_{it}\beta_0 + X'_{it}\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\}, \quad (5)$$

where $Y_{it} \in \{0, 1\}$ denotes the binary dependent variable for individual $i \in \{1, 2, \dots\}$ and time $t \leq T$, $Z_{it} \in \mathcal{R}^{d_z}$ denotes exogenous covariates, $X_{it} \in \mathcal{R}^{d_x}$ denotes potentially endogenous covariates, $\alpha_i \in \mathcal{R}$ denotes the unobserved fixed effect for individual i , and ϵ_{it} denotes the unobserved time-varying error term for individual i at time t . The objective is to identify the parameter $\theta_0 := (\beta'_0, \gamma'_0)'$ using a short panel of observed variables $(Z_{it}, X_{it}, Y_{it})_{it}$, where the number of time periods $T \geq 2$ is fixed and finite.

The identification of model (5) has been explored in the literature under various assumptions.⁴ Specifically, Manski (1987) exploits the following conditional stationarity (also referred to as “group homogeneity”) of ϵ_{it} to achieve identification:

$$\epsilon_{is} \mid Z_{is}, Z_{it}, X_{is}, X_{it}, \alpha_i \stackrel{d}{\sim} \epsilon_{it} \mid Z_{is}, Z_{it}, X_{is}, X_{it}, \alpha_i \quad \text{for any } s, t \leq T. \quad (6)$$

The above condition does not impose parametric restrictions on the distributions of ϵ_{it} and allows for dependence of ϵ_{it} over time. However, condition (2) does impose substantial restriction on the dependence between (Z_{it}, X_{it}) and the time-varying error ϵ_{it} . For instance, suppose $\mathbb{E}[\epsilon_{it} \mid Z_{is}, Z_{it}, X_{is}, X_{it}] = Z'_{it}\eta_1 + X'_{it}\eta_2$, then condition (2) fails in general. Hence, condition (2) can also be interpreted as a form of exogeneity condition on ϵ_{it} and all covariates (Z_{it}, X_{it}) . Importantly, in a *dynamic* setting where X_{it} includes the lagged dependent variable $Y_{i,t-1}$, then condition (2) is no longer a justifiable assumption to work with.

Our paper aims to study identification of θ_0 robust to endogeneity of X_{it} , where X_{it} can be arbitrarily dependent with $(\alpha_i, \epsilon_{it})$. Before introducing our assumptions, we first provide several applications to illustrate this endogeneity.

Example 1 (Dynamic Models). Khan, Ponomareva, and Tamer (2023) studies the following dynamic model with one lagged dependent variable:

$$Y_{it} = \mathbb{1}\{Z'_{it}\beta_0 + Y_{it-1}\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\},$$

⁴See Section 1 for more detailed discussions on related literature.

under the stationarity assumption only given exogenous covariates Z_{it} : $\epsilon_{is} \mid Z_{is}, Z_{it}, \alpha_i \stackrel{d}{\sim} \epsilon_{it} \mid Z_{is}, Z_{it}, \alpha_i$. In this model, the full stationarity condition may naturally fail due to its dynamic nature. For example, suppose ϵ_{it} is i.i.d. across time and independent of $(\{Z_{it}\}_{t=1}^T, \alpha_i)$. Given $Z_{is} = Z_{it} = z, \alpha_i = a$, it is clear that the conditional distribution of $\epsilon_{it} \mid Y_{it-1}, Y_{it}, z, a$ will be different from $\epsilon_{it+1} \mid Y_{it-1}, Y_{it}, z, a \stackrel{d}{\sim} \epsilon_{it+1}$, since Y_{it} is informative for ϵ_{it} .

Example 2 (Omitted Variable). Consider the true model is given as $Y_{it} = \mathbb{1}\{Z'_{it}\beta_0 + X'_{it}\gamma_0 + S_{it}\eta_0 + \alpha_i + v_{it} \geq 0\}$, where v_{it} is independent of $(\alpha_i, \{Z_{it}, X_{it}, S_{it}\}_{t=1}^T)$. However, the variable S_{it} is omitted and we instead estimate the following model:

$$Y_{it} = \mathbb{1}\{Z'_{it}\beta_0 + X'_{it}\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\},$$

where $\epsilon_{it} = v_{it} + S_{it}\eta_0$. The covariate X_{it} can be correlated with ϵ_{it} through the omitted variable S_{it} , provided that the omitted variable S_{it} is correlated with X_{it} . Then the conditional distribution of $\epsilon_{it} \mid (Z_{is}, Z_{it}, X_{is}, X_{it}, \alpha_i)$ may differ from $\epsilon_{is} \mid (Z_{is}, Z_{it}, X_{is}, X_{it}, \alpha_i)$ since the distribution of $S_{it} \mid X_{it}$ could be different from the distribution $S_{is} \mid X_{is}$ when $X_{it} \neq X_{is}$.

Example 3 (Measurement Error). Consider the true model is given as $Y_{it} = \mathbb{1}\{Z'_{it}\beta_0 + X_{it}^*\gamma_0 + \alpha_i + v_{it} \geq 0\}$, where v_{it} is independent of $(\alpha_i, \{Z_{it}, X_{it}^*\}_{t=1}^T)$. However, the covariate X_{it}^* may not be observed and we only observe its measurement X_{it} which may be subject to a measurement error $X_{it} = X_{it}^* + e_{it}$. Suppose we estimate the following model using the measurement X_{it} as a regressor:

$$Y_{it} = \mathbb{1}\{Z'_{it}\beta_0 + X'_{it}\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\},$$

where $\epsilon_{it} = v_{it} - e'_{it}\gamma_0$. The error term ϵ_{it} would be correlated with the measurement X_{it} through the measurement error e_{it} . Similarly, the conditional stationarity in (2) could fail since X_{it} is informative for e_{it} such that the conditional distribution of $\epsilon_{it} \mid z, X_{is}, X_{it}, \alpha_i$ would vary over time given different covariates $X_{is} \neq X_{it}$ while holding $Z_{is} = Z_{it} = z$.

Motivated by the above examples, this paper aims to establish identification of θ_0 , while being robust to endogeneity of X_{it} with respect to unobserved factors $(\alpha_i, \epsilon_{it})$. Below we propose an approach that accommodates various types of endogeneity, regardless of whether the endogenous covariates X_{it} are discrete lagged outcome variables (as in Example 1) or continuously-valued covariates with contemporary endogeneity as in other examples. To achieve this, we introduce the following partial conditional stationarity assumption:

Assumption 1 (Partial Stationarity). *The conditional distribution of $\epsilon_{it} \mid Z_{is}, Z_{it}, \alpha_i$ is stationary over time:*

$$\epsilon_{is} \mid Z_{is}, Z_{it}, \alpha_i \stackrel{d}{\sim} \epsilon_{it} \mid Z_{is}, Z_{it}, \alpha_i \quad \text{for any } s, t \leq T.$$

Assumption 1 only assumes the stationarity given the covariate (Z_{is}, Z_{it}) , but does not impose any restrictions on the endogenous covariate X_{it} . The standard full stationarity condition (2) is nested in Assumption 1, and it is in general stronger than Assumption 1. The main distinction is that Assumption 1 allows X_{it} to be arbitrarily dependent with both the fixed effect α_i and the time-changing error ϵ_{it} . In addition, their correlation can be different across time. Assumption 1 also shares similar features with condition (2), in that it does not impose any distributional assumption on ϵ_{it} and allows ϵ_{it} to be correlated across time. The standard i.i.d. assumption of ϵ_{it} across time or independence of ϵ_{it} with $(\alpha_i, Z_{it}, X_{it})$ are also nested in Assumption 1 as special cases.

2.2 Identification Strategy

We now explain our key identification strategy based on partial stationarity. While it applies more generally, we first illustrate the key idea in the specific context of the dynamic binary choice model. Sections 3-6 provide generalizations of the idea to various models.

Let $v_{it} := -(\epsilon_{it} + \alpha_i)$, $W_{it} := (Z_{it}, X_{it})$, and $W_{ist} := (W_{is}, W_{it})$. The conditional choice probability is given as follows:

$$\Pr(Y_{it} = 1 \mid W_{ist} = w_{st}) = \Pr(v_{it} \leq w'_t \theta_0 \mid W_{ist} = w_{st}).$$

When assuming all covariates are exogenous under condition (2), variation in the choice probability is solely due to variation in the covariate index $w'_t \theta_0$.⁵ Therefore, as shown in Manski (1987), the identified set for θ_0 is characterized by the condition: an increase in the value of $w'_t \theta_0$ over time implies an increase in the choice probability over time.

However, Manski (1987)'s method is not applicable to our setting with Assumption 1, given that the endogenous regressor X_{it} can be correlated with v_{it} . When the covariate X_{it} changes, the conditional distribution of v_{it} might potentially change as well. Variation in the choice probability contains a mixture of changes in the covariate index and the conditional distribution of v_{it} . Hence, it is not feasible to derive identifying restrictions on θ_0 merely by examining the intertemporal variation in the choice probability.

Our analysis relies on the partial stationarity condition in Assumption 1, which implies the stationarity of v_{it} given $Z_{ist} := (Z_{is}, Z_{it})$:

$$v_{is} \mid Z_{ist} \stackrel{d}{\sim} v_{it} \mid Z_{ist}.$$

The identification strategy proceeds in two steps. We first derive lower and upper bounds for the distribution of $v_{it} \mid Z_{ist}$ at each period t . The derived bounds at each period t are

⁵Condition 2 implies the stationarity of v_{it} : $v_{is} \mid W_{ist} \stackrel{d}{\sim} v_{it} \mid W_{ist}$.

functions of observed variables and the parameter θ_0 . Since the distribution of $v_{it} | Z_{ist}$ is the same over time, then the bounds of the distribution across time should have intersections. This restriction serves as the identifying restriction for θ_0 . Next, we show how to derive bounds on the distribution of $v_{it} | Z_{ist}$.

Let $F_{v_t|W_{st}}(\cdot | w_{st})$ denote the conditional CDF of v_{it} given $W_{ist} = w_{st}$. For any fixed point $a \in \mathcal{R}$, we first derive bounds for the conditional distribution $F_{v_t|W_{st}}(a | w_{st})$, and then establish bounds on $F_{v_t|Z_{st}}(a | z_{st})$ by taking expectation over the covariate X_{ist} given $Z_{ist} = z_{st}$.

Lower bound. The conditional choice probability given $W_{ist} = w_{st}$ is given as

$$\Pr(Y_{it} = 1 | W_{ist} = w_{st}) = \Pr(v_{it} \leq w'_t \theta_0 | W_{ist} = w_{st}) = F_{v_t|W_{st}}(w'_t \theta_0 | w_{st}).$$

For any fixed point a , by the monotonicity of a distribution, $F_{v_t|W_{st}}(a | w_{st})$ can be bounded below using the observed choice probability:

$$F_{v_t|W_{st}}(a | w_{st}) \geq \Pr(Y_{it} = 1 | W_{ist} = w_{st}) \mathbb{1}\{w'_t \theta_0 \leq a\}.$$

The above inequality holds since either a is larger than the covariate index $w'_t \theta_0$ so we know $F_{v_t|W_{st}}(a | w_{st}) \geq F_{v_t|W_{st}}(w'_t \theta_0 | w_{st})$ or the right hand side is equal to zero when a is smaller than $w'_t \theta_0$.

As Assumption 1 is on the distribution of v_{it} given Z_{ist} , we can establish the lower bound for $F_{v_t|Z_{st}}(a | z_{st})$ by taking expectation over covariates X_{ist} given $Z_{ist} = z_{st}$:

$$\begin{aligned} F_{v_t|Z_{st}}(a | z_{st}) &\geq \int_{x_{st}} \Pr(Y_{it} = 1 | W_{ist} = w_{st}) \mathbb{1}\{w'_t \theta_0 \leq a\} dF_{X_{st}|Z_{st}}(x_{st} | z_{st}) \\ &= \Pr(Y_{it} = 1, z'_t \beta_0 + X'_{it} \gamma_0 \leq a | z_{st}). \end{aligned}$$

The above lower bound only depends on observed variables (W_{ist}, Y_{it}) and θ_0 , so it is identified up to the parameter θ_0 . This bound will help construct identifying restrictions on θ_0 .

Upper bound. The idea of constructing the upper bound is similar. For any $a \in \mathcal{R}$, the conditional distribution $F_{v_t|W_{st}}$ can be bounded above:

$$F_{v_t|W_{st}}(a | w_{st}) \leq \Pr(Y_{it} = 1 | W_{ist} = w_{st}) \mathbb{1}\{w'_t \theta_0 \geq a\} + \mathbb{1}\{w'_t \theta_0 < a\}.$$

Then, taking expectation over covariates X_{ist} given $Z_{ist} = z_{st}$ leads to the upper bound for $F_{v_t|Z_{st}}(a | z_{st})$:

$$\begin{aligned} F_{v_t|Z_{st}}(a | z_{st}) &\leq \int_{x_{st}} \left\{ \Pr(Y_{it} = 1 | W_{ist} = w_{st}) \mathbb{1}\{w'_t \theta_0 \geq a\} + \mathbb{1}\{w'_t \theta_0 < a\} \right\} dF_{X_{st}|Z_{st}}(x_{st} | z_{st}) \\ &= \Pr(Y_{it} = 1, z'_t \beta_0 + X'_{it} \gamma_0 \geq a | z_{st}) + \Pr(z'_t \beta_0 + X'_{it} \gamma_0 < a | z_{st}) \\ &= 1 - \Pr(Y_{it} = 0, z'_t \beta_0 + X'_{it} \gamma_0 \geq a | z_{st}). \end{aligned}$$

Given the lower and upper bound for the distribution $F_{v_t|Z_{st}}(a | z_{st})$, we are ready to establish identifying conditions for θ_0 . Since Assumption 1 requires that $v_t | Z_{ist} \stackrel{d}{\sim} v_s | Z_{ist}$, it implies the upper bound for the distribution of $v_s | Z_{ist}$ must be larger than the lower bound for the distribution of $v_t | Z_{ist}$ otherwise Assumption 1 cannot hold. This restriction characterizes an identified set for θ_0 , presented in the following proposition.

Proposition 1. *Under Assumption 1, an identified set $\Theta_{I,1}$ for θ_0 is the set of parameters that satisfy the following conditions:*

$$1 - \Pr(Y_{is} = 0, z'_s\beta + X'_{is}\gamma \geq a | z_{st}) \geq \Pr(Y_{it} = 1, z'_t\beta + X'_{it}\gamma \leq a | z_{st}), \quad (7)$$

for any $a \in \mathcal{R}$, any $s, t \leq T$, and any $z_{st} \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_z}$.

Proposition 1 characterizes an identified set for θ_0 , using the conditional joint distribution of (Y_{it}, X_{it}) given Z_{ist} across any pair of periods (s, t) . The identification result allows for endogeneity in the covariate X_{ist} and does not rely on any excluded instruments.⁶

The identifying condition in (7) can be reformulated as conditional moment inequalities, by substituting the probability $\Pr(Y_{it} = 0, z'_t\beta_0 + X'_{it}\gamma_0 \geq a | z_{st})$ with the conditional expectation $\mathbb{E}[\mathbb{1}\{Y_{it} = 0, z'_t\beta_0 + X'_{it}\gamma_0 \geq a\} | z_{st}]$. Consequently, the estimation and inference for θ_0 can be conducted using established methods such as Chernozhukov, Hong, and Tamer (2007), Andrews and Shi (2013), and Chernozhukov, Lee, and Rosen (2013).

Remark 1. *Our “partial stationarity” condition accommodates the “full stationarity” condition (2) as a special case. Under full stationarity, or in other words, when all covariates are exogenous, Proposition 1 specializes to the following “maximum-score-type” identifying restrictions in Manski (1987):*

$$w'_s\theta_0 \geq w'_t\theta_0 \Leftrightarrow 1 - \Pr(Y_{is} = 0 | w_{st}) = \Pr(Y_{is} = 1 | w_{st}) \geq \Pr(Y_{it} = 1 | w_{st}).$$

Remark 2. *In the special case where the only endogenous covariate is one lagged outcome variable, as studied in Khan, Ponomareva, and Tamer (2023), the identifying restrictions yielded by Proposition 1 become*

$$1 - \Pr(Y_{is} = 0, z'_s\beta_0 + Y_{i,s-1}\gamma_0 \geq a | z_{st}) \geq \Pr(Y_{it} = 1, z'_t\beta_0 + Y_{i,t-1}\gamma_0 \leq a | z_{st}),$$

for any $a \in \{z'_t\beta_0, z'_t\beta_0 + \gamma_0, z'_s\beta_0, z'_s\beta_0 + \gamma_0\}$. As shown in Appendix A.4, the above identifying condition on θ_0 implies the identifying conditions in Khan, Ponomareva, and Tamer (2023, Theorem 1, p. 8-9), under different values of $(z'_s\beta_0, z'_t\beta_0, \gamma_0)$.

⁶When a sub-vector Z_{it}^{exc} of Z_{it} is assumed to be excluded from the binary choice model, then the identifying restriction in (7) is adjusted by setting the coefficient of Z_{it}^{exc} as zero: $\beta_0^{exc} = 0$.

Proposition 1 does not impose restrictions on the support of the endogenous covariate X_{it} , allowing it to be either discrete or continuous. When X_{it} is continuous, then Proposition 1 requires condition (7) to hold for any $a \in \mathcal{R}$ so it exploits the whole distribution information of (X_{it}, Y_{it}) at any value a . For the estimation and inference of θ_0 , we can follow the conventional inference method to discretize the space of \mathcal{R} and pick K_n points for a . In the case of discrete X_{it} , the dimension of a can be further reduced, and the identified set $\Theta_{I,1}$ can be characterized by finite number of restrictions, shown in the following corollary.

Corollary 1. *When the endogenous covariate $X_{it} \in \{a_1, \dots, a_K\}$ only takes finite number of values, then the identified set $\Theta_{I,1}$ is the set of parameters θ that satisfy condition (7) for any $a \in \{z'_s\beta + a'_k\gamma, z'_t\beta + a'_k\gamma\}_{k=1}^K$, any $s, t \leq T$, and any $z_{st} \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_z}$.*

Corollary 1 reduces the number of restrictions in Proposition 1 to $2K$ number of moment restrictions for any pair of two periods s, t . For the dynamic model where X_{it} includes one lagged dependent variable $X_{it} = Y_{i,t-1} \in \{0, 1\}$, there are only four conditional restrictions for θ_0 . When X_{it} includes j lagged dependent variable $X_{it} = (Y_{i,t-1}, Y_{i,t-2}, \dots, Y_{i,t-j})$, then there will be 2^{j+1} restrictions.

We show that $\Theta_{I,1}$ is an identified set for θ_0 , however, it is still unclear whether condition (7) exploits all the available information for θ_0 . The following theorem establishes the sharpness of the identified set $\Theta_{I,1}$ when X_{it} takes finite number of values.

Theorem 1 (Sharpness). *Under Assumption 1, the identified set $\Theta_{I,1}$ is sharp when X_{it} takes finite number of values for any $t \leq T$.*

Proposition 1 provides identification results without restrictions on the support of X_{it} , and Theorem 1 shows that the results already exhausted all available information for θ_0 when X_{it} only takes finite number of values. This results implies that for dynamic models with any number of lagged dependent variables, the identified set $\Theta_{I,1}$ is the smallest set we can obtain.

2.3 Point Identification

Proposition 1 characterizes the sharp identified set for θ_0 by only imposing Assumption 1. This section provides sufficient conditions to achieve point identification for β_0 (up to scale) and the sign of γ_0 under support conditions on the exogenous covariate Z_{it} . We focus on the scenario where the endogenous covariate X_{it} is discrete $X_{it} \in \mathcal{X} = \{a_1, a_2, \dots, a_K\}$ and there are only two periods $T = 2$.

To point identify β_0 , the first step is to determine the sign of the covariate index $(Z_{i2} - Z_{i1})'\beta_0$ under certain variation of observed choice probability. To identify the sign of $(Z_{i2} -$

$Z_{i1})'\beta_0$, we define the following two sets:

$$\begin{aligned}\mathcal{Z}_1 &:= \left\{ (z_1, z_2) \mid \exists x \in \mathcal{X} \text{ s.t. } 1 - \Pr(Y_{i1} = 0, X_{i1} = x \mid z_{12}) < \Pr(Y_{i2} = 1, X_{i2} = x \mid z_{12}) \right\}, \\ \mathcal{Z}_2 &:= \left\{ (z_1, z_2) \mid \exists x \in \mathcal{X} \text{ s.t. } 1 - \Pr(Y_{i1} = 1, X_{i1} = x \mid z_{12}) < \Pr(Y_{i2} = 0, X_{i2} = x \mid z_{12}) \right\}.\end{aligned}$$

Let $\mathcal{Z} := \mathcal{Z}_1 \cup \mathcal{Z}_2$. Let $\Delta Z_i = Z_{i2} - Z_{i1}$ and $\Delta \mathcal{Z}$ be defined as

$$\Delta \mathcal{Z} := \left\{ \Delta z := z_2 - z_1 \mid (z_1, z_2) \in \mathcal{Z} \right\}.$$

As shown in Appendix A.3, when Δz satisfies $\Delta z \in \Delta \mathcal{Z}$, the sign of $\Delta z'\beta_0$ is identified. In the definition of the two sets $\mathcal{Z}_1, \mathcal{Z}_2$, we only need the existence of one value in the support of \mathcal{X} such that the choice probability in the two sets are observed. When observing such choice probability, the sign of $\Delta z'\beta_0$ is identified. Then β_0 can be identified up to scale under the standard large support condition on Δz .

Let Δz^j denote the j th element of Δz . The following is the support condition on the covariate.

Assumption 2 (Support Condition). (1) $\Delta \mathcal{Z}$ is not contained in any proper linear subspace of \mathcal{R}^{d_z} ; (2) for any $\Delta z \in \Delta \mathcal{Z}$, there exists one element Δz^{j^*} such that $\beta_0^{j^*} \neq 0$, and the conditional support of Δz^{j^*} is \mathcal{R} given $\Delta z \setminus \Delta z^{j^*}$, where $\Delta z \setminus \Delta z^{j^*}$ denote the remaining components of Δz besides Δz^{j^*} .

Proposition 2. Under Assumptions 1-2, the parameter β_0 is point identified up to scale.

We provide point identification for β_0 with two periods $T = 2$. When there are more than two periods, then we only require the existence of two periods, satisfying Assumption 2. As shown in Manski (1987), the large support assumption is necessary to point identify β_0 , as without it, there exists some $b \neq k\beta_0$ such that $\Delta z'b$ has the same sign with $\Delta z'\beta_0$ when Δz has bounded support.

The parameter γ_0 in general can be only partially identified given potential endogeneity of X_{it} and flexible structures on $(\alpha_i, \epsilon_{it})$. Nevertheless, we can still bound the value $(x_1 - x_2)'\gamma_0$ and identify the sign of γ_0 under certain choice probabilities. We derive the sufficient conditions to identify the sign of γ_0 .

Let x^j denote the j -th element of x and γ_0^j denote the j -th coefficient of γ_0 . We define the following two sets:

$$\begin{aligned}\mathcal{Z}_3^j &:= \left\{ (z_1, z_2) \mid \exists x_1, x_2 \in \mathcal{X} \text{ with } x_1^j \neq x_2^j, x_1^m = x_2^m \forall m \neq j \text{ s.t.} \right. \\ &\quad \left. 1 - \Pr(Y_{i1} = 0, X_{i1} = x_1 \mid z_{12}) < \Pr(Y_{i2} = 1, X_{i2} = x_2 \mid z_{12}) \right\}; \\ \mathcal{Z}_4^j &:= \left\{ (z_1, z_2) \mid \exists x_1, x_2 \in \mathcal{X} \text{ with } x_1^j \neq x_2^j, x_1^m = x_2^m, \forall m \neq j \text{ s.t.} \right. \\ &\quad \left. 1 - \Pr(Y_{i1} = 1, X_{i1} = x_1 \mid z_{12}) < \Pr(Y_{i2} = 0, X_{i2} = x_2 \mid z_{12}) \right\}.\end{aligned}$$

From the identifying results in Proposition 1, the value of $(x_1^j - x_2^j)\gamma_0^j$ can be bounded when (z_1, z_2) belong to the two sets:

$$\begin{aligned}(z_1, z_2) \in \mathcal{Z}_3^j &\implies (x_1^j - x_2^j)\gamma_0^j < \Delta z' \beta_0, \\ (z_1, z_2) \in \mathcal{Z}_4^j &\implies (x_1^j - x_2^j)\gamma_0^j > \Delta z' \beta_0.\end{aligned}$$

Then the sign of γ_0^j is identified if either the sign of $\Delta z' \beta_0$ is identified as negative when $(z_1, z_2) \in \mathcal{Z}_2$ or as positive when $(z_1, z_2) \in \mathcal{Z}_1$.

Proposition 3. *Under Assumptions 1, and for any $1 \leq j \leq d_x$, either $\mathcal{Z}_3^j \cap \mathcal{Z}_2 \neq \emptyset$ or $\mathcal{Z}_4^j \cap \mathcal{Z}_1 \neq \emptyset$, then the sign of γ_0 is identified.*

When the endogenous variable X_{it} is a scalar, e.g., the lagged dependent variable $X_{it} = Y_{i,t-1}$, then the definition of the two sets $\mathcal{Z}_3^j, \mathcal{Z}_4^j$ can be simplified as there existing $x_1 \neq x_2$ such that the corresponding choice probability is observed. Besides the sign of γ_0 , the identification results can also bound the value of γ_0 from variation in the exogenous covariates.

When X_{it} is multi-dimensional such as including two lagged dependent variable $X_{it} = (Y_{i,t-1}, Y_{i,t-2})$ with $\gamma_0 = (\gamma_0^1, \gamma_0^2)$, then γ_0^1 is identified when the required choice probability in the two sets $\mathcal{Z}_3^1, \mathcal{Z}_4^1$ are observed for $(Y_{i,1}, Y_{i,0}) = (1, 1), (Y_{i,2}, Y_{i,1}) = (0, 1)$ or $(Y_{i,1}, Y_{i,0}) = (0, 0), (Y_{i,2}, Y_{i,1}) = (1, 0)$. We provide general sufficient conditions to identify the sign of γ_0 , which may be stronger than necessary and can be relaxed in certain scenarios. For example, when we know that $\gamma_0^1 + \gamma_0^2 > 0$ while $\gamma_0^1 < 0$, we can infer that $\gamma_0^2 > 0$ without requiring additional assumptions on the two sets $\mathcal{Z}_3^2, \mathcal{Z}_4^2$.

3 Ordered Response Model

3.1 Model and Identification

Consider a setting where the outcome variable Y_{it} is discrete and takes J values: $Y_{it} \in \{y_1, \dots, y_J\}$. For example, Y_{it} can be different categories of income, health outcomes, or educational attainment. We study the following panel ordered response model:

$$\begin{aligned}Y_{it}^* &= Z_{it}'\beta_0 + X_{it}'\gamma_0 + \alpha_i + \epsilon_{it}, \\ Y_{it} &= \sum_{j=1}^J y_j \mathbb{1}\{b_j < Y_{it}^* \leq b_{j+1}\},\end{aligned}\tag{8}$$

where Y_{it}^* denotes the latent dependent variable, $b_1 = -\infty, b_{J+1} = +\infty$, and the remaining threshold parameters b_j (where $b_j \leq b_{j+1}$) can be either known or unknown for $2 \leq j \leq J-1$. The binary choice model in (1) is nested with $J = 2$ and $b_2 = 0$. Similar to model (1), X_{it}

is the potentially endogenous covariate that can be arbitrarily dependent on $(\alpha_i, \epsilon_{it})$, such as the lagged dependent variable $Y_{i,t-1}$, while Z_{it} is the exogenous covariate that satisfies Assumption 1:

$$\epsilon_{is} \mid Z_{is}, Z_{it}, \alpha_i \stackrel{d}{\sim} \epsilon_{it} \mid Z_{is}, Z_{it}, \alpha_i \quad \text{for any } s, t \leq T.$$

We now show how our key identification strategy based on partial stationarity can also be exploit to partially identify θ_0 . Similar to Section 2, we still seek to establish both lower and upper bounds for the distribution $F_{v_t|Z_{st}}(a \mid z_{st})$, where $v_{it} := \epsilon_{it} + \alpha_i$. The distinction is that in the ordered choice setting we can obtain J different (upper and lower) bounds, which can then be aggregated over to form tighter bounds as shown below.

1. Lower bound. By the monotonicity of a distribution, we first derive a lower bound for the distribution $v_t \mid w_{st}$ conditional on all covariates $W_{ist} = w_{st}$, given as

$$\begin{aligned} F_{v_t|W_{st}}(a \mid w_{st}) &= \Pr(v_t \leq a \mid w_{st}) \\ &\geq \Pr(v_t \leq b_{j+1} - w'_t \theta_0 \mid w_{st}) \mathbb{1}\{b_{j+1} - w'_t \theta_0 \leq a\}, \end{aligned}$$

where the above lower bound holds for any choice $j \leq J$. Since the interval $(-\infty, b_{j+1} - w'_t \theta_0]$ can be expressed as the union of multiple intervals: $(-\infty, b_{j+1} - w'_t \theta_0] = \bigcup_{k=1}^j (b_k - w'_t \theta_0, b_{k+1} - w'_t \theta_0]$, the above lower bound can be also written as

$$\begin{aligned} F_{v_t|W_{st}}(a \mid w_{st}) &\geq \sum_{k=1}^j \Pr(b_k - w'_t \theta_0 < v_t \leq b_{k+1} - w'_t \theta_0 \mid w_{st}) \mathbb{1}\{b_{j+1} - w'_t \theta_0 \leq a\} \\ &= \sum_{k=1}^j \Pr(Y_{it} = y_k \mid w_{st}) \mathbb{1}\{b_{j+1} - w'_t \theta_0 \leq a\}, \end{aligned}$$

for any $j \leq J$. Fixing the constant a , there exists a maximum choice j_a^* , defined as $j_a^* := \max\{j : b_{j+1} - w'_t \theta_0 \leq a\}$. Then, the largest lower bound is the sum of choice probabilities up to the maximum choice j_a^* :

$$F_{v_t|W_{st}}(a \mid w_{st}) \geq \sum_{k=1}^{j_a^*} \Pr(Y_{it} = y_k \mid w_{st}) \mathbb{1}\{b_{j_a^*+1} - w'_t \theta_0 \leq a\}.$$

An equivalent expression for this lower bound is given as:

$$F_{v_t|W_{st}}(a \mid w_{st}) \geq \sum_{j=1}^J \Pr(Y_{it} = y_j \mid w_{st}) \mathbb{1}\{b_{j+1} - w'_t \theta_0 \leq a\},$$

as the indicator function is zero $\mathbb{1}\{b_{j+1} - w'_t \theta_0 \leq a\} = 0$ for any choice $j > j_a^*$.

2. Upper bound. Similarly, we can derive the upper bound for $F_{v_t|w_{st}}(a \mid w_{st})$ using

the sum of conditional choice probabilities, given as follows:

$$\begin{aligned}
F_{v_t|W_{st}}(a | w_{st}) &= 1 - \Pr(v_t > a | w_{st}) \\
&\leq 1 - \sum_{j=1}^J \Pr(b_j - w'_t \theta_0 < v_t \leq b_{j+1} - w'_t \theta_0 | w_{st}) \mathbb{1}\{b_j - w'_t \theta_0 \geq a\} \\
&\leq 1 - \sum_{j=1}^J \Pr(Y_{it} = y_j | w_{st}) \mathbb{1}\{b_j - w'_t \theta_0 \geq a\}.
\end{aligned}$$

Then, the upper bound for the distribution $F_{v_t|z_{st}}(a | z_{st})$ is derived as follows by taking expectation over covariate X_{ist} given z_{st} :

$$F_{v_t|z_{st}}(a | z_{st}) \leq 1 - \sum_{j=1}^J \Pr(Y_{it} = y_j, b_j - z'_t \beta_0 - X'_{it} \gamma_0 \geq a | z_{st}).$$

Given the established lower and bounds on the distribution of $v_t | z_{st}$, an identified set for θ_0 is characterized in the following proposition.

Proposition 4. *Under Assumption 1, the identified set $\Theta_{I,2}$ for θ_0 is the set of parameters θ that satisfy the following conditions:*

$$1 - \sum_{j=1}^J \Pr(Y_{is} = y_j, b_j - z'_s \beta - X'_{is} \gamma \geq a | z_{st}) \geq \sum_{j=1}^J \Pr(Y_{it} = y_j, b_{j+1} - z'_t \beta - X'_{it} \gamma \leq a | z_{st}), \quad (9)$$

for any $a \in \mathcal{R}$, any $s, t \leq T$, and any $z_{st} \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_z}$.

Proposition 4 characterizes the identified set for θ_0 for the general ordered response model with potential endogeneity. This result allows for both static and dynamic models, as discussed in Section 3.2. In comparison to the binary choice model, Proposition 4 aggregates information from all J ordered responses across different time periods to identify θ_0 . When there are only two choices ($J = 2$), this result simplifies to the identifying condition (7) in Proposition 1, with $y_1 = 0, y_2 = 1$ and changing a to $-a$. Again, the result in Proposition 4 accommodates both discrete and continuous endogenous covariates. When the endogenous X_{it} only takes a finite number of values, such as lagged dependent variable, we can reduce the number of restrictions in Proposition 4 to a finite number, as shown in the next corollary.

Corollary 2. *Under Assumption 1 and X_{it} only takes finite number of values $X_{it} \in \{a_1, \dots, a_K\}$, the identified set $\Theta_{I,2}$ is the set of parameters θ that satisfy condition (9) for any $a \in \mathcal{Q}_2 := \{b_j - z'_s \beta - a'_1 \gamma, \dots, b_j - z'_s \beta - a'_K \gamma, b_j - z'_t \beta - a'_1 \gamma, \dots, b_j - z'_t \beta - a'_K \gamma\}_{j=2}^J$, any $s, t \leq T$, and any $z_{st} \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_z}$.*

Similar to Corollary 1, Corollary 2 reduces the number of conditional moment inequalities to $2K \times (J - 1)$ numbers for any pair of two periods. For the dynamic model with one lagged

dependent variable $X_{it} = Y_{i,t-1} \in \{y_1, \dots, y_J\}$, then there will be $2J(J - 1)$ number of restrictions.

3.2 Applications

In this section, we apply the results in Proposition 4 to explore two panel ordered choice model. The first one examines a static model where all covariates are exogenous, while the other studies a dynamic model with one lagged dependent variable. To the best of our knowledge, these two scenarios have not been explored in the literature within the framework of the stationarity assumption.

Static model: we consider a static model where Assumption 1 holds conditional on all covariates W_{ist} . The identifying restriction in Proposition 4 is given as

$$1 - \sum_{j=1}^J \Pr(Y_{is} = y_j \mid w_{st}) \mathbb{1}\{b_j - w'_s \theta_0 \geq a\} \geq \sum_{j=1}^J \Pr(Y_{it} = y_j \mid w_{st}) \mathbb{1}\{b_{j+1} - w'_t \theta_0 \leq a\}.$$

We can further simplify the above condition by getting rid of the parameter a and transform it into finite number of conditional moment inequalities. The following lemma presents the identified set for $\Theta_{I,2}$ for this static model.

Lemma 1. *Assume that $\epsilon_{is} \mid (W_{ist}, \alpha_i) \stackrel{d}{\sim} \epsilon_{it} \mid (W_{ist}, \alpha_i)$, the identified set $\Theta_{I,2}$ for θ_0 is the set of parameters θ that satisfy the following conditions:*

$$b_{j_1+1} - w'_s \theta \geq b_{j_2+1} - w'_t \theta \implies \sum_{j=1}^{j_1} \Pr(Y_{is} = y_j \mid w_{st}) \geq \sum_{j=1}^{j_2} \Pr(Y_{it} = y_j \mid w_{st}),$$

for any $1 \leq j_1, j_2 \leq J - 1$, $s, t \leq T$, and $w_{st} \in \mathcal{R}^d \times \mathcal{R}^d$.

In the static model, variation in conditional choice probability only comes from changes in the covariate W_{it} . Thus, we can directly establish the relationship between the variation in choice probability and the changes in covariates over time, eliminating the parameter a . Different from the binary choice model, the results in Lemma 1 also exploit intertemporal variation in the sum of multiple choices rather than investigating a single choice. Moreover, we can utilize variations in the sum of different choices across various periods to identify θ_0 .

Besides the static model, the results in Proposition 4 can also be applied to study a dynamic ordered response model, where people's choice at the current period (t) can depend on their choice in the last period ($t - 1$).

Dynamic model: consider the following dynamic ordered response model with one

lagged dependent variable:

$$Y_{it}^* = Z'_{it}\beta_0 + Y_{it-1}\gamma_0 + \alpha_i + \epsilon_{it},$$

$$Y_{it} = \sum_{j=1}^J y_j \mathbb{1}\{b_j < Y_{it}^* \leq b_{j+1}\}.$$

In this example, the endogenous covariate is the lagged dependent variable $X_{it} = Y_{it-1} \in \{y_1, \dots, y_J\}$. The identifying restriction in Proposition 4 is presented as follows:

$$1 - \sum_{j=1}^J \Pr(Y_{is} = y_j, b_j - z'_s\beta - Y_{is-1}\gamma \geq a \mid z_{st})$$

$$\geq \sum_{j=1}^J \Pr(Y_{it} = y_j, b_{j+1} - z'_t\beta - Y_{it-1}\gamma \leq a \mid z_{st}),$$

for any $a \in \{b_j - z'_s\beta - y_1\gamma, \dots, b_j - z'_s\beta - y_J\gamma, b_j - z'_t\beta - y_1\gamma, \dots, b_j - z'_t\beta - y_J\gamma\}_{j=2}^J$.

For the dynamic model, the identified set $\Theta_{I,2}$ is characterized by $2J(J-1)$ number of conditional inequalities for any pair of two periods (s, t) . This approach can also allow for dynamic models with more than one lagged dependent variable, such as $X_{it} = (Y_{i,t-1}, Y_{i,t-2})$. Furthermore, it can accommodate dynamic models that allow for heterogeneous effects from different choices in the last period, as given below:

$$Y_{it}^* = Z'_{it}\beta_0 + \sum_{j=2}^J \mathbb{1}\{Y_{it-1} = j\}\gamma_{0,j} + \alpha_i + \epsilon_{it}.$$

3.3 Point Identification

This section provides sufficient conditions to achieve point identification for β_0 (up to scale) and the sign of γ_0 . Similar to Section 2.3, we still focus on the scenario where the endogenous covariate X_{it} only takes finite number of values $X_{it} \in \mathcal{X} = \{a_1, a_2, \dots, a_K\}$, and we illustrate the concept using two periods ($T = 2$).

To identify β_0 , we define the following two sets:

$$\mathcal{Z}_{1,order} := \left\{ (z_1, z_2) \mid \exists x \in \mathcal{X}, 2 \leq k \leq J \text{ s.t.} \right.$$

$$\left. 1 - \sum_{j=k}^J \Pr(Y_{i1} = y_j, X_{i1} = x \mid z_{12}) < \sum_{j=1}^{k-1} \Pr(Y_{i2} = y_j, X_{i2} = x \mid z_{12}) \right\};$$

$$\mathcal{Z}_{2,order} := \left\{ (z_1, z_2) \mid \exists x \in \mathcal{X}, 2 \leq k \leq J \text{ s.t.} \right.$$

$$\left. 1 - \sum_{j=1}^{k-1} \Pr(Y_{i1} = y_j, X_{i1} = x \mid z_{12}) < \sum_{j=k}^J \Pr(Y_{i2} = y_j, X_{i2} = x \mid z_{12}) \right\}.$$

Let $\mathcal{Z}_{order} := \mathcal{Z}_{1,order} \cup \mathcal{Z}_{2,order}$. Let $\Delta Z_i = Z_{i2} - Z_{i1}$ and $\Delta \mathcal{Z}_{order}$ be defined as

$$\Delta \mathcal{Z}_{order} := \left\{ \Delta z := z_2 - z_1 \mid (z_1, z_2) \in \mathcal{Z}_{order} \right\}.$$

Assumption 3 (Support Condition). (1) $\Delta \mathcal{Z}_{order}$ is not contained in any proper linear subspace of \mathcal{R}^{d_z} ; (2) for any $\Delta z \in \Delta \mathcal{Z}_{order}$, there exists at least one element Δz^{j*} with $\beta_0^{j*} \neq 0$, and the conditional support of Δz^{j*} is \mathcal{R} given $\Delta z \setminus \Delta z^{j*}$, where $\Delta z \setminus \Delta z^{j*}$ denote the remaining components of Δz besides Δz^{j*} .

Proposition 5. Under Assumptions 1 & 3, the parameter β_0 is point identified up to scale.

Similar to Proposition 2, the strategy is to first identify the sign of $\Delta z' \beta_0$ from certain variation in the choice probability, as defined in the two sets $(\mathcal{Z}_{1,order}, \mathcal{Z}_{2,order})$. Given the sign of $\Delta z' \beta_0$ is identified, then the parameter β_0 is point identified up to scale under the conventional large support condition on Δz . For the static model (with $\gamma_0 = 0$), all coefficients β_0 are point identified under the support condition.

To identify the sign of γ_0 , we define the following two sets: for $j \leq d_x$,

$$\mathcal{Z}_{3,order}^j := \left\{ (z_1, z_2) \mid \exists x_1, x_2 \in \mathcal{X} \text{ with } x_1^j \neq x_2^j, x_1^m = x_2^m \forall m \neq j, \exists 2 \leq k \leq J \text{ s.t.} \right. \\ \left. 1 - \sum_{j=k}^J \Pr(Y_{i1} = y_j, X_{i1} = x_1 \mid z_{12}) < \sum_{j=1}^{k-1} \Pr(Y_{i2} = y_j, X_{i2} = x_2 \mid z_{12}) \right\};$$

$$\mathcal{Z}_{4,order}^j := \left\{ (z_1, z_2) \mid \exists x_1, x_2 \in \mathcal{X} \text{ with } x_1^j \neq x_2^j, x_1^m = x_2^m \forall m \neq j, \exists 2 \leq k \leq J \text{ s.t.} \right. \\ \left. 1 - \sum_{j=1}^{k-1} \Pr(Y_{i1} = y_j, X_{i1} = x_1 \mid z_{12}) < \sum_{j=k}^J \Pr(Y_{i2} = y_j, X_{i2} = x_2 \mid z_{12}) \right\}.$$

Proposition 6. Under Assumptions 1, and for any $1 \leq j \leq d_x$, either $\mathcal{Z}_{3,order}^j \cap \mathcal{Z}_{2,order} \neq \emptyset$ or $\mathcal{Z}_{4,order}^j \cap \mathcal{Z}_{1,order} \neq \emptyset$, then the sign of γ_0 is identified.

For the dynamic model with one lagged dependent variable $X_{it} = Y_{i,t-1}$, then the joint choice probabilities defined in the two sets $\mathcal{Z}_{3,order}^j, \mathcal{Z}_{4,order}^j$ can be simplified as taking different values for the lagged dependent variable: $Y_{i0} = y_{j_1}, Y_{i1} = y_{j_2}$ with $y_{j_1} \neq y_{j_2}$. The identifying restriction in Proposition 4, in general, can bound the value of $(y_j - y_k)\gamma_0$ using the variation in Z_{it} . Proposition 6 ensures that its bound excludes zero for some (y_j, y_k) , thereby identifying the sign of γ_0 . When there are more than two periods ($T > 2$), then point identification of β_0 and the sign of γ_0 is achieved when there exists two periods such that the assumptions are satisfied.

4 Multinomial Choice Model

4.1 Model and Identification

This section applies our key identification strategy to panel multinomial choice model with endogeneity. Specifically, consider a set of choice alternatives $\mathcal{J} = \{0, 1, \dots, J\}$. Let u_{ijt} denote the utility for individual i for choice j at time t , which depends on three components: observed covariate $W_{ijt} = (Z'_{ijt}, X'_{ijt})'$, unobserved fixed effects α_{ij} , and unobserved time-varying preference shock ϵ_{ijt} . The utility of choice 0 (outside option) is normalized to zero: $u_{i0t} = 0$. Let $Y_{it} \in \mathcal{J}$ denote individual i 's choice at time t . We study the following panel multinomial choice model:

$$\begin{aligned} u_{ijt} &= Z'_{ijt}\beta_0 + X'_{ijt}\gamma_0 + \alpha_{ij} + \epsilon_{ijt}, \\ Y_{it} &= \arg \max_{j \in \mathcal{J}} u_{ijt}, \end{aligned}$$

where X_{ijt} denotes the potentially endogenous covariate, such as the lagged dependent variable or endogenously determined prices, and Z_{ijt} denotes the exogenous covariates that satisfies the partial stationarity condition (Assumption 1)

$$\epsilon_{is} \mid Z_{is}, Z_{it}, \alpha_i \stackrel{d}{\sim} \epsilon_{it} \mid Z_{is}, Z_{it}, \alpha_i \quad \text{for any } s, t \leq T.$$

with $Z_{it} = \{Z_{ijt}\}_{j \in \mathcal{J}}$, $\alpha_i = \{\alpha_{ij}\}_{j \in \mathcal{J}}$ and $\epsilon_{it} = \{\epsilon_{ijt}\}_{j \in \mathcal{J}}$ defined to collect terms across all J choice alternatives.

The identification of θ_0 has been studied under the standard full stationarity condition given all covariates W_{ist} in different models, including Pakes and Porter (2022), Shi, Shum, and Song (2018), Gao and Li (2020), and Wang (2022). The main focus of this paper is to derive identification for θ_0 , while allowing for the endogenous covariate X_{it} . As discussed in Section 4.2, our approach also accommodates the standard stationarity assumption, and the results are consistent with those in Pakes and Porter (2022).

Let $v_{ijt} = \alpha_{ij} + \epsilon_{ijt}$ and $v_{it} = \{v_{ijt}\}_{j \in \mathcal{J}}$. The identification strategy is still to establish bounds on the distribution of $v_{it} \mid Z_{ist}$ and derive identifying restrictions on θ_0 based on partial stationarity. Here, the distinction is that the error term v_{it} is multi-dimensional instead of one-dimensional in binary or ordered choice models. Moreover, consumers' choice involves the comparison between different choices.

The observed choice probability of selecting j at period t is given as

$$\Pr(Y_{it} = j \mid w_{st}) = \Pr(w'_{jt}\theta_0 + v_{ijt} \geq w'_{kt}\theta_0 + v_{ikt} \quad \forall k \in \mathcal{J} \mid w_{st}).$$

The above probability restricts the distribution $\Pr(v_{ikt} - v_{ijt} \leq (w_{jt} - w_{kt})'\theta_0 \quad \forall k \neq j \mid w_{st})$. From observed choice probability, we can derive bounds for the distribution $\Pr(v_{ikt} - v_{ijt} \leq$

$a_{jk} \forall k \neq j \mid z_{st}$) as a function of θ_0 and observed variable (X_{it}, Z_{it}, Y_{it}) . Therefore, the identifying condition for θ_0 is characterized by the restriction that the bounds over different periods must intersect. More generally, we observe the conditional probability $\Pr(Y_{it} \in K \mid w_{st})$ for any set $K \subset \mathcal{J}$ and can bound the probability $\Pr(v_{ikt} - v_{ijt} \leq a_{jk} \exists j \in K, \forall k \in K^c \mid z_{st})$, where $K^c := \mathcal{J} \setminus K$. In the case where K is a singleton ($K = \{j\}$), it then becomes the conditional probability of selecting a single choice j . The following proposition characterizes the identification results for θ_0 .

Proposition 7. *Under Assumption 1, an identified set $\Theta_{I,3}$ for θ_0 is the set of parameters $\theta = (\beta, \gamma)$ that satisfy the following condition:*

$$\begin{aligned} 1 - \Pr(Y_{is} \in K^c, (z_{js} - z_{ks})'\beta + (X_{ijs} - X_{iks})'\gamma \geq a_{jk} \forall j \in K, k \in K^c \mid z_{st}) \\ \geq \Pr(Y_{it} \in K, (z_{jt} - z_{kt})'\beta + (X_{ijt} - X_{ikt})'\gamma \leq a_{jk} \forall j \in K, k \in K^c \mid z_{st}), \end{aligned} \quad (10)$$

for any subset $K \subset \mathcal{J}$, any $a_{jk} \in \mathcal{R}$ with $j \in K$ and $k \in K^c$, any $s, t \leq T$, and any $z_{st} \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_z}$.

Corollary 3. *Under Assumption 1, and when $X_{ijt} - X_{ikt}$ only takes finite number of values $X_{ijt} - X_{ikt} \in \mathcal{X}_{jk} := \{x_{jk,1}, \dots, x_{jk,M}\}$, then the identified set $\Theta_{I,3}$ is the set of parameters θ that satisfy condition (10) for any subset $K \subset \mathcal{J}$, any $a_{jk} \in \{(z_{js} - z_{ks})'\beta + x'_{jk}\gamma, (z_{jt} - z_{kt})'\beta + x'_{jk}\gamma\}_{x_{jk} \in \mathcal{X}_{jk}}$ with $j \in K$ and $k \in K^c$, any $s, t \leq T$, and any $z_{st} \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_z}$.*

Proposition 7 provides general identification results for multinomial choice models with discrete or continuous endogenous covariates. Corollary 3 further reduces the number of restrictions in Proposition 7 to a finite number when the difference in the endogenous covariate between choices, $X_{ijt} - X_{ikt}$, only takes finite number of values. The results accommodate both static models and dynamic models with any number of lagged dependent variables. Point identification can be achieved under conditions similar to those presented in Section 2.3 and 3.3, so further analysis is omitted here.

4.2 Applications

We examine the two applications of Proposition 7 for the panel multinomial choice model. The first one studies the static model, where all covariates W_{it} are exogenous ($\epsilon_{is} \mid W_{ist}, \alpha_i \stackrel{d}{\sim} \epsilon_{it} \mid W_{ist}, \alpha_i$), which is the same setting as Pakes and Porter (2022). In this scenario, we show that Proposition 7 boils down to the identifying results in Pakes and Porter (2022). Another one explores the dynamic multinomial choice model, incorporating one lagged dependent variable $Y_{i,t-1}$.

Static model: Pakes and Porter (2022) characterizes the sharp identified set for θ_0 under the stationarity assumption given all covariates: $\epsilon_{is} \mid W_{is}, W_{it}, \alpha_i \stackrel{d}{\sim} \epsilon_{it} \mid W_{is}, W_{it}, \alpha_i$. Now, we show that the identifying restriction in (10) is consistent with the results in Pakes and Porter (2022).

Under full stationarity (given all covariates W_{it}), condition (10) becomes

$$\begin{aligned} 1 - \Pr(Y_{is} \in K^c, (w_{js} - w_{ks})' \theta_0 \geq a_{jk} \ \forall j \in K, k \in K^c \mid w_{st}) \\ \geq \Pr(Y_{it} \in K, (w_{jt} - w_{kt})' \theta_0 \leq a_{jk} \ \forall j \in K, k \in K^c \mid w_{st}). \end{aligned}$$

The above equation is only informative when $a_{jk} \in [(w_{jt} - w_{kt})' \theta_0, (w_{js} - w_{ks})' \theta_0]$ for any $j \in K, k \in K^c$; otherwise either the upper bound becomes one or the lower bound becomes zero in the above restriction. When $a \in [(w_{jt} - w_{kt})' \theta_0, (w_{js} - w_{ks})' \theta_0]$, the above restriction is equivalent to the following condition: for any $K \subset \mathcal{J}$,

$$\begin{aligned} \text{If } (w_{js} - w_{ks})' \theta_0 \geq (w_{jt} - w_{kt})' \theta_0 \ \forall j \in K, k \in K^c \\ \implies 1 - \Pr(Y_{is} \in K^c \mid w_{st}) = \Pr(Y_{is} \in K \mid w_{st}) \geq \Pr(Y_{it} \in K \mid w_{st}), \end{aligned}$$

which aligns with the identification result in Pakes and Porter (2022) (Proposition 1, P. 12).

Dynamic model: Proposition 7 is also applicable to the following dynamic panel multinomial choice model:

$$u_{ijt} = Z'_{ijt} \beta_0 + \mathbb{1}\{Y_{i,t-1} = j\} \gamma_0 + \alpha_{ij} + \epsilon_{ijt}.$$

The above model allows individuals' current utility at time t to depend on their choices at the last period $t-1$. In this model, the endogenous variable X_{ijt} is the lagged dependent variable $Y_{i,t-1}$, and the difference of the endogenous covariate between choices only takes three values: $X_{ijt} - X_{ikt} \in \mathcal{X}_{jk} := \{1, -1, 0\}$. Then the identified set for $\Theta_{I,3}$ is characterized by the condition in Corollary 2 with $a_{jk} \in \{(z_{js} - z_{ks})' \beta + \gamma, (z_{js} - z_{ks})' \beta - \gamma, (z_{js} - z_{ks})' \beta, (z_{jt} - z_{kt})' \beta + \gamma, (z_{jt} - z_{kt})' \beta - \gamma, (z_{jt} - z_{kt})' \beta\}$.

5 Censored Dependent Variable Model

The previous sections primarily investigate discrete choice models, while our approach also applies to models with continuous dependent variables, including those with censored or interval outcomes. To illustrate, we focus on censored outcomes models below.

Consider the following panel models with censored outcomes:

$$\begin{aligned} Y_{it}^* &= Z'_{it} \beta_0 + X'_{it} \gamma_0 + \alpha_i + \epsilon_{it}, \\ Y_{it} &= \max\{Y_{it}^*, 0\}, \end{aligned}$$

where X_{it} represents the endogenous covariate, Z_{it} stands for the exogenous covariate that satisfies Assumption 1, Y_{it}^* denotes the latent outcome not observed in the data, and Y_{it} represents the observed outcome, which is censored at zero. The threshold for censoring can be replaced with other nonzero constants. The identification strategy is still to derive conditions for θ_0 based on the stationarity assumption. Here the conditional distribution of $\alpha_i + \epsilon_{it} \mid Z_{st}$ is not identified since we only observe the censoring outcome Y_{it} . While we can still bound the conditional distribution $\alpha_i + \epsilon_{it} \mid Z_{st}$ using variation in the joint distribution (Y_{it}, Z_{it}, X_{it}) . The following proposition presents the identification results of θ_0 with censoring outcomes.

Proposition 8. *Under Assumption 1, an identified set for θ_0 is the set of parameters θ that satisfy the following condition:*

$$\Pr(0 < Y_{is} \leq z'_s \beta + X'_{is} \gamma - a \mid z_{st}) + \Pr(Y_{is} = 0 \mid z_{st}) \geq \Pr(Y_{it} \leq z'_t \beta + X'_{it} \gamma - a \mid z_{st}),$$

for any $a \in \mathcal{R}$, any $(s, t) \leq T$, and any $z_{st} \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_z}$.

Similar to discrete choice models studied in previous sections, Proposition 8 characterizes an identified set for θ_0 by exploiting the variation in the joint distribution $(Y_{it}, X_{it}) \mid Z_{ist}$ over time and the variation in the exogenous covariates Z_{ist} . The bounds on the distribution $\alpha_i + \epsilon_{it} \mid Z_{st}$ can be derived either from the probability $\Pr(0 < Y_{it} \leq y \mid z_{st})$ or $\Pr(Y_{it} = 0 \mid z_{st})$, depending on the value of the covariate index $z'_t \beta + X'_{it} \gamma$. This result still accommodates both static and dynamic models with censored outcomes.

Static model: consider that the standard stationarity assumption holds, where $\epsilon_{it} \mid \alpha_i, W_{ist} \stackrel{d}{\sim} \epsilon_{is} \mid \alpha_i, W_{ist}$. Then, the identifying condition in Proposition 8 is given as

$$\Pr(0 < Y_{is} \leq w'_s \theta - a) + \Pr(Y_{is} = 0 \mid w_{st}) \geq \Pr(Y_{it} \leq w'_t \theta - a \mid w_{st}). \quad (11)$$

The above restriction is informative only when $w'_t \theta \geq a$, otherwise the lower bound becomes zero. We discuss two cases for the constant a . When $w'_s \theta \leq a \leq w'_t \theta$, then condition (11) becomes

$$\Pr(Y_{is} = 0 \mid w_{st}) \geq \Pr(Y_{it} \leq w'_t \theta - a \mid w_{st}).$$

When a satisfies $a \leq \min\{w'_s \theta, w'_t \theta\}$, condition (11) transforms into

$$\begin{aligned} \Pr(0 < Y_{is} \leq w'_s \theta - a \mid w_{st}) + \Pr(Y_{is} = 0 \mid w_{st}) &= \Pr(Y_{is} \leq w'_s \theta - a \mid w_{st}) \\ &\geq \Pr(Y_{it} \leq w'_t \theta - a \mid w_{st}). \end{aligned}$$

Since the above condition needs to hold for any (s, t) and is symmetric in (s, t) , it becomes equalities after exchanging s and t .

The following lemma derives the identified set for θ_0 using both conditional moment

inequalities and equalities.

Lemma 2. *Assuming that $\epsilon_{it} \mid \alpha_i, W_{ist} \stackrel{d}{\sim} \epsilon_{is} \mid \alpha_i, W_{ist}$, an identified set for θ_0 is characterized by the set of parameters θ that satisfy the following conditions:*

$$\begin{cases} \text{If } w'_s\theta \leq a \leq w'_t\theta \implies \Pr(Y_{is} = 0 \mid w_{st}) \geq \Pr(Y_{it} \leq w'_t\theta - a \mid w_{st}), \\ \text{If } a \leq \min\{w'_s\theta, w'_t\theta\} \implies \Pr(Y_{is} \leq w'_s\theta - a \mid w_{st}) = \Pr(Y_{it} \leq w'_t\theta - a \mid w_{st}), \end{cases}$$

for any $s, t \leq T$ and any $w_{st} \in \mathcal{R}^d \times \mathcal{R}^d$.

Dynamic model: Proposition 8 also accommodates the following dynamic model with the lagged outcome $Y_{i,t-1}$:

$$\begin{aligned} Y_{it}^* &= Z'_{it}\beta_0 + Y_{i,t-1}\gamma_0 + \alpha_i + \epsilon_{it}, \\ Y_{it} &= \max\{Y_{it}^*, 0\}. \end{aligned}$$

In this model, since the endogenous variable $X_{it} = Y_{i,t-1} \in [0, \infty)$ can be continuous, we are not able to further simplify the identifying condition in Proposition 8. Appendix A.10 also studies dynamic models with the latent lagged outcome $Y_{i,t-1}^*$. Consequently, the results in Proposition 8 need to be adjusted as the endogenous variable $X_{it} = Y_{i,t-1}^*$ is not observed.

6 Generalization to Nonseparable Models

The key idea underlying our identification strategy generalizes further beyond the models considered before. In this section, we show that our approach relies only on two conditions: partial stationarity and “index monotonicity.” Additional model specifications and restrictions, such as the scalar-additivity of fixed effects and time-varying errors, are not needed for our identification strategy.

To illustrate, consider the following nonseparable semiparametric model:

$$Y_{it} = G\left(W'_{it}\theta_0, \alpha_i, \epsilon_{it}\right),$$

where α_i is the individual fixed effect of arbitrary dimension, ϵ_{it} is the time-varying error of arbitrary dimension, W_{it} is a vector of observable covariates, $\theta_0 \in \mathcal{R}^d$ is a conformable vector of parameter, and the function G is allowed to be unknown, nonseparable but assumed to be weakly monotone in the the parametric index:

Assumption 4 (Monotonicity). *The mapping $\delta \mapsto G(\delta, \alpha, \epsilon)$ is weakly increasing in $\delta \in \mathcal{R}$ for each realization of (α, ϵ) .*

Note that, by setting α_i, ϵ_{it} to be scalar-valued, and $G(W'_{it}\theta_0, \alpha_i, \epsilon_{it}) = \mathbb{1}\{W'_{it}\theta_0 + \alpha_i + \epsilon_{it} \geq 0\}$, we obtain the dynamic binary choice model in Section 2, where G is by construction weakly increasing in $W'_{it}\theta_0$.

As before, we decompose W_{it} , and correspondingly θ_0 , into two components, $W_{it} = (Z'_{it}, X'_{it})'$ and $\theta_0 = (\beta'_0, \gamma'_0)'$ so that

$$W'_{it}\theta_0 = Z'_{it}\beta_0 + X'_{it}\gamma_0,$$

where Z_{it} denotes exogenous covariates while X_{it} denotes endogenous covariates, with the precise definition of exogeneity encoded by the partial stationarity in Assumption 1.

Now we show how partial stationarity can be exploited in conjunction with weak monotonicity (Assumption 4) to obtain identifying restrictions in the presence of endogeneity.

Let \mathcal{Y} denote the support of Y_{it} . For any $a \in \mathcal{R}$ and $y \in \mathcal{Y}$, conditional on $W_{ist} = w_{st}$, we can bound $\mathbb{1}\{G(a, \alpha_i, \epsilon_{it}) \geq y\}$ below as

$$\begin{aligned} \mathbb{1}\{G(a, \alpha_i, \epsilon_{it}) \geq y\} &\geq \mathbb{1}\{w'_t\theta_0 \leq a\} \mathbb{1}\{G(a, \alpha_i, \epsilon_{it}) \geq y\} \\ &\geq \mathbb{1}\{w'_t\theta_0 \leq a\} \mathbb{1}\{G(w'_t\theta_0, \alpha_i, \epsilon_{it}) \geq y\}, \end{aligned}$$

and above as

$$\begin{aligned} &\mathbb{1}\{G(a, \alpha_i, \epsilon_{it}) \geq y\} \\ &\leq \mathbb{1}\{w'_t\theta_0 > a\} \mathbb{1}\{G(a, \alpha_i, \epsilon_{it}) \geq y\} + \mathbb{1}\{w'_t\theta_0 \leq a\} \\ &\leq \mathbb{1}\{w'_t\theta_0 > a\} \mathbb{1}\{G(w'_t\theta_0, \alpha_i, \epsilon_{it}) \geq y\} + \mathbb{1}\{w'_t\theta_0 \leq a\} \\ &= 1 - \mathbb{1}\{w'_t\theta_0 > a\} \mathbb{1}\{G(w'_t\theta_0, \alpha_i, \epsilon_{it}) < y\}. \end{aligned}$$

Then, the conditional probability $\mathbb{P}(G(a, \alpha_i, \epsilon_{it}) \geq y \mid w_{st})$ given $W_{ist} = w_{st}$ can be bounded as follows:

$$\begin{aligned} \mathbb{P}(Y_{it} \geq y \mid w_{st}) \mathbb{1}\{w'_t\theta_0 \leq a\} &\leq \mathbb{P}(G(a, \alpha_i, \epsilon_{it}) \geq y \mid w_{st}) \\ &\leq 1 - \mathbb{P}(Y_{it} < y \mid w_{st}) \mathbb{1}\{w'_t\theta_0 > a\}. \end{aligned}$$

Taking conditional expectation of the above with respect to X_{ist} given $Z_{ist} = z_{st}$ yields

$$\begin{aligned} &\mathbb{P}(Y_{is} \geq y, W'_{is}\theta_0 \leq a \mid z_{st}) \\ &\leq \mathbb{P}(G(a, \alpha_i, \epsilon_{is}) \geq y \mid z_{st}) = \mathbb{P}(G(a, \alpha_i, \epsilon_{it}) \geq y \mid z_{st}) \text{ by Assumption 1} \\ &\leq 1 - \mathbb{P}(Y_{it} < y, W'_{it}\theta_0 > a \mid z_{st}). \end{aligned}$$

The key difference of (6) relative to the corresponding identifying restrictions in previous sections lies in that the “middle term” in (6) is no longer the conditional CDF of $\alpha_i + \epsilon_{it}$, but a conditional probability about the event $G(a, \alpha_i, \epsilon_{is}) \geq y$, with the latter representation not dependent on scalar-additivity of fixed effects and time-varying errors.

We summarize the identifying restriction above by the following proposition:

Proposition 9. *Under Assumptions 1 and 4, the following condition holds:*

$$1 - \mathbb{P}(Y_{it} < y, z'_t \beta_0 + X'_{it} \gamma_0 > a \mid z_{st}) \geq \mathbb{P}(Y_{is} \geq y, z'_s \beta_0 + X'_{it} \gamma_0 \leq a \mid z_{st}), \quad (12)$$

for any $a \in \mathcal{R}$, $y \in \mathcal{Y}$, $s \neq t \leq T$, and $z_{st} \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_z}$.

Note that in the binary choice setting of Section 2, it suffices to set $y = 1$ in (12), which then coincides with the identifying condition (7) in Proposition 1. It is also straightforward to show that our identification results in Lemma 9 can be adapted to the ordered response model in Section 3, the multinomial choice model in Section 4, and the censored outcome model in Section 5 without the scalar additive specifications.

7 Simulation

This section examines the finite sample performance of our identification approaches using Monte Carlo simulations. Since the literature has extensively studied binary choice models, we focus on the static and dynamic ordered choice models explored in Section 3 as examples to illustrate the approach. We implement the kernel-based CLR inference approach proposed in the papers by Chernozhukov, Lee, and Rosen (2013) and Chen and Lee (2019), developed to construct confidence interval based on general conditional moment inequalities.

7.1 Static Ordered Choice Model

This section explores a static ordered choice model with three choices $Y_{it} \in \{1, 2, 3\}$. We consider the following two-period model with $T = 2$, and the latent dependent variable Y_{it}^* is generated as:

$$Y_{it}^* = Z_{it}^1 \beta_{01} + Z_{it}^2 \beta_{02} + \alpha_i + \epsilon_{it},$$

where the covariate Z_{it}^k satisfies $Z_{it}^k \sim \mathcal{N}(0, \sigma_z)$ for $k \in \{1, 2\}$; the fixed effects α_i are given as $\alpha_i = \sum_{t=1}^T (Z_{it}^1 + Z_{it}^2) / (4 * \sigma_z * T)$, so they are correlated with the covariates; the error term $(\epsilon_{i1}, \epsilon_{i2})$ follows the normal distribution $\mathcal{N}(\mu, \Sigma)$ with $\mu = (0, 0)$ and $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The true parameter is $\beta_0 := (\beta_{0,1}, \beta_{0,2})' = (1, 1)'$, the repetition number is $B = 200$, and the sample size is $n = \{2000, 8000\}$. We consider three specifications for $\sigma_z \in \{1, 1.5, 2\}$ and $\rho \in \{0, 0.25, 0.5\}$.

The observed dependent variable Y_{it} is given as

$$Y_{it} = 1 * (Y_{it}^* \leq b_2) + 2 * (0 \leq Y_{it}^* \leq b_2) + 3 * (Y_{it}^* > b_3),$$

where $b_2 = -1$ and $b_3 = 1$.

We consider the covariates Z_{it} to be exogenous and Lemma 1 characterizes the identified set for β_0 using the following conditional moment inequalities:

$$E[g(Z_{ist}, Y_{ist}; \beta_0) \mid z_{st}] \geq 0,$$

where

$$g(Z_{ist}, Y_{ist}; \beta_0) = \begin{cases} \mathbb{1}\{b_2 - Z'_{is}\beta \geq b_2 - Z'_{it}\beta_0\}(\mathbb{1}\{Y_{is} = 1\} - \mathbb{1}\{Y_{it} = 1\}); \\ \mathbb{1}\{b_2 - Z'_{is}\beta \geq b_3 - Z'_{it}\beta_0\}(\mathbb{1}\{Y_{is} = 1\} - \mathbb{1}\{Y_{it} \in \{1, 2\}\}); \\ \mathbb{1}\{b_3 - Z'_{is}\beta \geq b_2 - Z'_{it}\beta_0\}(\mathbb{1}\{Y_{is} \in \{1, 2\}\} - \mathbb{1}\{Y_{it} = 1\}); \\ \mathbb{1}\{b_3 - Z'_{is}\beta \geq b_3 - Z'_{it}\beta_0\}(\mathbb{1}\{Y_{is} \in \{1, 2\}\} - \mathbb{1}\{Y_{it} \in \{1, 2\}\}). \end{cases}$$

The first element β_{01} of the parameter β_0 is normalized to one, and we are interested in conducting inference for the parameter β_{02} using the CLR approach. Tables 1 and 2 report the average confidence interval (CI) for β_{02} , the coverage probability (CP), the average length of the CI (length), the power of the test at zero (power), and the mean absolute deviation of the lower bound (l_{MAD}) and upper bound (u_{MAD}) of the CI.

Table 1: Performance of β_{02} under different values of σ_z ($\rho = 0.25$)

σ_z	CI	CP	length	Power	l_{MAD}	u_{MAD}
$N = 2000$						
$\sigma_z = 1$	[0.537, 1.760]	0.876	1.222	1.000	0.476	0.784
$\sigma_z = 1.5$	[0.556, 1.768]	0.934	1.212	1.000	0.454	0.773
$\sigma_z = 2$	[0.567, 1.791]	0.950	1.224	1.000	0.440	0.796
$N = 8000$						
$\sigma_z = 1$	[0.570, 1.532]	0.939	0.962	1.000	0.439	0.548
$\sigma_z = 1.5$	[0.607, 1.561]	0.975	0.954	1.000	0.398	0.563
$\sigma_z = 2$	[0.618, 1.571]	0.985	0.953	1.000	0.383	0.573

As shown in Tables 1 and 2, our approach exhibits robust performance across various specifications of standard deviation σ and correlation coefficients ρ . The coverage probabilities of the 95% confidence interval (CI) for β_{02} are close to the nominal level, the length of the CI is reasonably small, and the CI consistently excludes zero. When the sample size increases, there is a significant decrease in CI length, an improvement in coverage probability, and a reduction of the mean absolute deviation (MAD) for the lower and upper bounds of the CI. Overall, these results demonstrate the good performance of our approach in different DGP designs.

Table 2: Performance of β_{02} under different values of ρ ($\sigma_z = 1$)

ρ	CI	CP	length	Power	l_{MAD}	u_{MAD}
$N = 2000$						
$\rho = 0$	[0.537, 1.755]	0.895	1.218	1.000	0.476	0.773
$\rho = 0.25$	[0.537, 1.760]	0.876	1.222	1.000	0.476	0.784
$\rho = 0.5$	[0.511, 1.765]	0.909	1.254	1.000	0.497	0.785
$N = 8000$						
$\rho = 0$	[0.584, 1.553]	0.933	0.969	1.000	0.436	0.568
$\rho = 0.25$	[0.570, 1.532]	0.939	0.962	1.000	0.439	0.548
$\rho = 0.5$	[0.573, 1.526]	0.934	0.954	1.000	0.442	0.541

7.2 Dynamic Ordered Choice Model

In this section, we investigate a dynamic ordered choice model with one lagged dependent variable $Y_{i,t-1}$. The latent dependent variable Y_{it}^* is generated as follows:

$$Y_{it}^* = Z_{it}\beta_0 + Y_{i,t-1}\gamma_0 + \alpha_i + \epsilon_{it}.$$

where the endogenous variable is the lagged dependent variable $Y_{i,t-1}$. We study three periods $T = 3$ to illustrate our approach with multiple periods. The DGP is similar: the exogenous covariate Z_{it} satisfies $Z_{it} \sim \mathcal{N}(0, \sigma_z)$; the fixed effects α_i are given as $\alpha_i = \sum_{t=1}^T Z_{it}/(4*\sigma_z*T)$; the error term $(\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3})$ follows the normal distribution $\mathcal{N}(\mu, \Sigma)$ with $\mu = (0, 0, 0)$ and $\Sigma = (0.5 \ c \ c; c \ 0.5 \ c; c \ c \ 0.5)$, where $c = 0.5*\rho$. The true parameter is $\theta_0 := (\beta_0, \gamma_0)' = (1, 1)'$, the repetition number is $B = 200$, and the sample size is $n \in \{2000, 8000\}$. We consider three specifications for $\sigma_z \in \{1, 1.5, 2\}$ and $\rho \in \{0, 0.25, 0.5\}$.

The observed dependent variable Y_{it} is given as

$$Y_{it} = 1 * (Y_{it}^* \leq b_2) + 2 * (0 \leq Y_{it}^* \leq b_2) + 3 * (Y_{it}^* > b_3),$$

for $1 \leq t \leq T$. The initial value $Y_{i0} \in \{1, 2, 3\}$ is generated independently of all variables, and follow the distribution $\Pr(Y_{i0} = 1) = 0.6, \Pr(Y_{i0} = 2) = \Pr(Y_{i0} = 3) = 0.2$.

In this dynamic model, the covariates $Z_i := (Z_{it})_{t=1}^T$ and the initial value Y_{i0} are exogenous, while the lagged variable $Y_{i,t-1}$ is endogenous. Proposition 4 characterizes the identified set for θ_0 with the following conditional moment inequalities:

(1) When $s \in \{2, 3\}$,

$$\begin{aligned}
1 - \sum_{j=2}^3 \Pr(Y_{i1} = y_j \mid z, y_0) * \mathbb{1}\{b_j - z'_1\beta - y_0\gamma \geq a\} \\
\geq \sum_{j=1}^2 \Pr(Y_{is} = y_j, b_{j+1} - z'_s\beta - Y_{is-1}\gamma \leq a \mid z, y_0), \\
1 - \sum_{j=2}^3 \Pr(Y_{is} = y_j, b_j - z'_s\beta - Y_{is-1}\gamma \geq a \mid z, y_0) \\
\geq \sum_{j=1}^2 \Pr(Y_{i1} = y_j \mid z, y_0) * \mathbb{1}\{b_{j+1} - z'_1\beta - y_0\gamma \leq a\},
\end{aligned}$$

for any $a \in \{b_j - z'_1\beta - y_0\gamma, b_j - z'_s\beta - \gamma, b_j - z'_s\beta - 2\gamma, b_j - z'_s\beta - 3\gamma\}_{j=2}^T$;

(2) When $s, t \in \{2, 3\}$,

$$\begin{aligned}
1 - \sum_{j=2}^3 \Pr(Y_{is} = y_j, b_j - z'_s\beta - Y_{is-1}\gamma \geq a \mid z, y_0) \\
\geq \sum_{j=1}^2 \Pr(Y_{it} = y_j, b_{j+1} - z'_t\beta - Y_{it-1}\gamma \leq a \mid z, y_0),
\end{aligned}$$

for any $a \in \{b_j - z'_s\beta - \gamma, b_j - z'_s\beta - 2\gamma, b_j - z'_s\beta - 3\gamma, b_j - z'_t\beta - \gamma, b_j - z'_t\beta - 2\gamma, b_j - z'_t\beta - 3\gamma\}_{j=2}^3$.

We normalize the first parameter β_0 to one, and report the performance of the coefficient γ_0 for the lagged dependent variable. Tables 3 and 4 illustrate that our approach yields robust and informative results for the dynamic ordered choice model across various DGP specifications. The coverage probability of the CI nearly reaches 95%, and the CI consistently excludes zero, producing significant coefficients. These results remain similar across different values of correlation coefficients. When the standard deviation σ_z increases, the length of the CI also experiences a slight increase. This phenomenon occurs because, in the dynamic model, only partial identification is achieved, and the bound for γ_0 depends on the variation in $\Delta z'\beta_0$. A larger variation in $\Delta z'\beta_0$ may result in a wider identified set in this specification, but it still provides informative results. As the sample size increases, the confidence interval shrinks, and concurrently, the coverage probability improves in all specifications.

Table 3: Performance of γ_0 under different values of σ_z ($\rho = 0.25$)

σ_z	CI	CP	length	Power	l_{MAD}	u_{MAD}
$N = 2000$						
$\sigma_z = 1$	[0.446, 1.606]	0.935	1.160	1.000	0.565	0.625
$\sigma_z = 1.5$	[0.375, 1.673]	0.959	1.298	1.000	0.629	0.693
$\sigma_z = 2$	[0.311, 1.730]	0.960	1.418	1.000	0.700	0.739
$N = 8000$						
$\sigma_z = 1$	[0.529, 1.495]	0.969	0.966	1.000	0.473	0.504
$\sigma_z = 1.5$	[0.460, 1.559]	0.965	1.100	1.000	0.548	0.564
$\sigma_z = 2$	[0.427, 1.585]	0.985	1.158	1.000	0.573	0.589

Table 4: Performance of γ_0 under different values of ρ ($\sigma_z = 1$)

ρ	CI	CP	length	Power	l_{MAD}	u_{MAD}
$N = 2000$						
$\rho = 0$	[0.472, 1.593]	0.932	1.121	1.000	0.550	0.607
$\rho = 0.25$	[0.446, 1.606]	0.935	1.160	1.000	0.565	0.625
$\rho = 0.5$	[0.457, 1.631]	0.943	1.173	1.000	0.548	0.648
$N = 8000$						
$\rho = 0$	[0.528, 1.472]	0.958	0.945	1.000	0.475	0.487
$\rho = 0.25$	[0.529, 1.495]	0.969	0.966	1.000	0.473	0.504
$\rho = 0.5$	[0.535, 1.515]	0.975	0.980	1.000	0.467	0.519

8 Empirical Application

In this section, we apply our proposed approach to explore the empirical analysis of income categories using the NLSY79 dataset. The dependent variable is three categories of (log) income, denoted by the three values $\{1, 2, 3\}$, indicating whether an individual falls within the top 33.3% highest income bracket, the 33.3%-66.6% highest income range, and the lowest 33.3% income tier, respectively. We include two covariates in this analysis: one is tenure, defined as the total duration (in weeks) with the current employer, and the other is the residence indicator for whether one lives in an urban or rural area.⁷ We use two periods of panel data from the years 1982 and 1983 as well as the income data from 1981 as initial values, and there are $n = 5259$ individuals in each period. The following table presents the summary statistics of these variables.

Table 5: Application: Summary Statistics

	income category	residence	tenure /100
mean	1.990	0.799	0.825
s.d.	0.810	0.401	0.738
25% quantile	1.000	1.000	0.220
median	2.000	1.000	0.605
75% quantile	3.000	1.000	1.280
minimum	1.000	1.000	0.010
maximum	3.000	1.000	4.850

We adopt various ordered response models introduced in Section 3 to analyze the income category. The first model is the standard static model without any endogeneity. The second is the static model, while treating residence as an endogenous covariate. Residence is potentially endogenous since the choice of living area is typically endogenously determined and may be correlated with individuals' unobserved ability or preference. The last model considers the dynamic model with one lagged dependent variable, allowing people's income in current periods to depend on their income in the last period. All three models allow for individual fixed effects and do not impose any parametric distributions on time-changing shocks. Proposition 4 characterizes the identified set of the model coefficients for these three models using conditional moment inequalities. Similar to Section 7, we exploit the kernel-

⁷This dataset also contains other crucial factors for income such as gender and race. However, these variables are time-invariant and cannot be included for panel models with fixed effects.

based CLR inference method to construct confidence intervals. The coefficient of the variable ‘residence’ is normalized to one. Table 6 reports the confidence intervals for the coefficients of the covariate ‘tenure’ and the lagged dependent variable (when applicable).

Table 6: Application: Income Categories

	$\beta_{0,1}$ (residence)	$\beta_{0,2}$ (tenure)	γ_0 (lag)
exogenous static model	1	[0.612, 0.939]	-
endogenous static model	1	[0.041, 0.939]	-
dynamic model	1	[0.531, 0.694]	[0.286, 0.612]

As shown in Table 6, tenure exhibits a significantly positive effect on the income category across all specifications. When allowing for the endogeneity of residence, the confidence interval for tenure becomes wider, as we need to account for all possible correlations between residence and unobserved heterogeneity. The results from the dynamic model show that the income category in the current period is also positively affected by last period’s income, and this effect is significant. Furthermore, this analysis demonstrates the flexibility of our approach, which can not only allow for endogeneity introduced by dynamics but also account for contemporary endogeneity.

9 Conclusion

We introduce a general method to identify nonlinear panel data models based on a partial stationarity condition. This approach accommodates dynamic models with an arbitrary finite number of lagged outcome variables and other types of endogenous covariates. We demonstrate how our key identification strategy can be applied to obtain informative identifying restrictions in various limited dependent variable models, including binary choice, ordered response, multinomial choice, as well as censored dependent variable models. Finally, we further extend this approach to study general nonseparable models.

There are some natural directions for follow-up research. In this paper we focus on the identification of model parameters, but it would also be interesting to investigate how our identification strategy can be exploited to obtain informative bounds on average marginal effects and other counterfactual parameters, say, following the approach proposed in [Botosaru and Muris \(2022\)](#).⁸ Additionally, the idea of bounding an endogenous object

⁸[Botosaru and Muris \(2022\)](#) proposes an approach to obtain bounds on counterfactual CCPs in semi-parametric dynamic panel data models, assuming that the index parameters are (partially) identified.

(parametric index in our case) by an arbitrary constant so as to obtain an object free of endogeneity issues may have broader applicability beyond the models studied in this work, and it remains to see whether our key identification strategy can be further adapted to other structures.

A Appendix

A.1 Proof of Lemma 1

Proof. The identifying condition (7) for the parameter θ in Proposition 1 is given as

$$1 - \Pr(Y_{is} = 0, z'_s\beta + X'_{is}\gamma \geq a \mid z_{st}) \geq \Pr(Y_{it} = 1, z'_t\beta + X'_{it}\gamma \leq a \mid z_{st}). \quad (13)$$

Now we show that when $X_{it} \in \{a_1, \dots, a_K\}$, the above condition is informative only at the $2K$ points $a \in \mathcal{Q}_1 = \{z'_t\beta + a'_1\gamma, \dots, z'_t\beta + a'_K\gamma, z'_s\beta + a'_1\gamma, \dots, z'_s\beta + a'_K\gamma\}$. Let $q_k \in \mathcal{Q}_1$ denote an element in \mathcal{Q} . Without loss of generalization, we can rank q_k as $q_1 \leq q_2 \leq \dots \leq q_{2K}$.

When $a > q_{2K}$, then $a > z'_s\beta + \max_k a'_k\gamma$. Therefore, the upper bound becomes $1 - \Pr(Y_{is} = 0, z'_s\beta + X'_{is}\gamma \geq a \mid z_{st}) = 1$ and the identifying condition holds for any θ . Therefore, there is no information for θ when $a > q_{2K}$.

When $a < q_1$, similarly the lower bound is zero $\Pr(Y_{it} = 1, z'_t\beta + X'_{it}\gamma \leq a \mid z_{st}) = 0$ and the inequality has no identifying power for θ .

When $a \in (q_k, q_{k+1})$ for some k , the identifying restriction at a can be implied by the restriction at $a = q_{k+1}$. It is because the upper bound at q_{k+1} is the same as a since there does not exist any point $z'_s\beta + a'_k\gamma$ that lies in (q_k, q_{k+1}) . The lower bound $\Pr(Y_{it} = 1, z'_t\beta + X'_{it}\gamma \leq a \mid z_{st})$ is weakly larger at q_{k+1} than a since it is weakly increasing. Then, any parameter that satisfies condition (13) evaluated at q_{k+1} also satisfies the one at a , so condition (13) evaluated at a does not contain additional information for θ .

□

A.2 Proof of Theorem 1

To prove sharpness of the identified set $\Theta_{I,1}$, we seek to show that, for any $\theta \in \Theta_{I,1}$, there exists an underlying DGP characterized by the conditional distribution of $(v_{is}, v_{it}) \mid W_{ist}$ that produces the same observed choice probability and satisfies Assumption 1. The following construction is for one candidate parameter $\theta \in \Theta_{I,1}$. We set the fixed effect to be zero $\alpha_i = 0$, so that $v_{it} := -(\epsilon_{it} + \alpha_i) = -\epsilon_{it}$.

Let $P_{st}(j, k \mid w_{st}) := \Pr(Y_{is} = j, Y_{it} = k \mid W_{ist} = w_{st})$ denote the observed joint probability of selecting j at period s and selecting k at period t . The first requirement for sharpness

is that the constructed distribution of $(v_{is}, v_{it}) \mid W_{ist}$ matches the observed joint choice probability according to the model: given $W_{ist} = w_{st}$,

$$\begin{aligned} P_{st}(1, 1 \mid w_{st}) &= \Pr(v_{is} \leq w'_s \theta, v_{it} \leq w'_t \theta \mid w_{st}), \\ P_{st}(1, 0 \mid w_{st}) &= \Pr(v_{is} \leq w'_s \theta \mid w_{st}) - \Pr(v_{is} \leq w'_s \theta, v_{it} \leq w'_t \theta \mid w_{st}), \\ P_{st}(0, 1 \mid w_{st}) &= \Pr(v_{it} \leq w'_t \theta \mid w_{st}) - \Pr(v_{is} \leq w'_s \theta, v_{it} \leq w'_t \theta \mid w_{st}), \end{aligned} \quad (14)$$

and $P_{st}(0, 0 \mid w_{st})$ can be matched automatically when the above conditions hold.

Another condition for sharpness requires the marginal distribution of $v_{it} \mid W_{ist}$ to satisfy the following stationarity condition in Assumption 1:

$$F_{v_{it} \mid Z_{st}}(a \mid z_{st}) = F_{v_{is} \mid Z_{st}}(a \mid z_{st}) \quad \forall a \in \mathcal{R}. \quad (15)$$

We can focus on a reduced problem, which is to find a marginal distribution $v_{it} \mid W_{ist}$ that matches the marginal choice probability and satisfies the stationarity assumption. Let $P_t(w_{st}) := \Pr(Y_{it} = 1 \mid W_{ist} = w_{st})$ and $P_s(w_{st}) := \Pr(Y_{is} = 1 \mid W_{ist} = w_{st})$ denote the observed marginal choice probability at t and s , given as follows:

$$\begin{aligned} P_t(w_{st}) &= \Pr(v_{it} \leq w'_t \theta \mid W_{ist} = w_{st}), \\ P_s(w_{st}) &= \Pr(v_{is} \leq w'_s \theta \mid W_{ist} = w_{st}). \end{aligned} \quad (16)$$

If we can construct a marginal distribution satisfying conditions (15) and (16), then the joint choice probability in (14) is matched by setting the joint distribution of $(v_{is}, v_{it}) \mid W_{ist}$ at $(w'_s \theta, w'_t \theta)$ as $P_{st}(1, 1 \mid w_{st}) = \Pr(v_{is} \leq w'_s \theta, v_{it} \leq w'_t \theta \mid w_{st})$. This construction is feasible since there is no other assumptions on the joint distribution $(v_{is}, v_{it}) \mid W_{ist}$. Moreover, other choice probabilities such as $P_{st}(1, 0 \mid w_{st})$ can be computed using the marginal probability $P_s(w_{st})$ subtracting $P_{st}(1, 1 \mid w_{st})$.

Now, for any $\theta \in \Theta_{I,1}$, our objective is to construct the distribution of $v_{it} \mid W_{st}$ and $v_{is} \mid W_{st}$ such that conditions (15) and (16) hold. Consider that the endogenous covariate X_{it} only takes K values: $X_{it} \in \mathcal{X} = \{a_1, \dots, a_K\}$ for any $t \leq T$.⁹ Given $Z_{it} = z_t$, the covariate index $z'_t \beta + x'_t \gamma$ belongs to the K points $\{z'_t \beta + a'_1 \gamma, \dots, z'_t \beta + a'_K \gamma\}$. Without loss of generalization, we can rank the K points as $a'_1 \gamma \leq a'_2 \gamma \leq \dots \leq a'_K \gamma$ since we can reindex the K values of $\{a_1, \dots, a_K\}$ such that this ranking is satisfied. Let $m_k := z'_t \beta + a'_k \gamma$ denote the K points for which $w'_t \theta$ can take, and $n_k := z'_s \beta + a'_k \gamma$ for $1 \leq k \leq K$ denote the K points $w'_s \theta$ can take. Let $\bar{m} = \max\{z'_t \beta, z'_s \beta\} + a'_K \gamma + \epsilon$ for some $\epsilon > 0$.

We first look at the stationarity condition in (15). Section 2.2 derives bounds for the

⁹The covariate X_{it} is allowed to have different support across time t , then \mathcal{X} is the union of the support of X_{it} over all periods.

distribution $F_{v_t|Z_{st}}(a | z_{st})$, given as follows:

$$\begin{aligned} F_{v_t|Z_{st}}(a | z_{st}) &\leq \sum_{x_{st} \in \mathcal{X}} \left\{ F_{v_t|W_{st}}(w'_t \theta | w_{st}) \mathbb{1}\{w'_t \theta \geq a\} + \mathbb{1}\{w'_t \theta < a\} \right\} \Pr(X_{st} = x_{st} | z_{st}) \\ &:= U_t(a | z_{st}), \\ F_{v_t|Z_{st}}(a | z_{st}) &\geq \sum_{x_{st} \in \mathcal{X}} F_{v_t|W_{st}}(w'_t \theta | w_{st}) \mathbb{1}\{w'_t \theta \leq a\} \Pr(X_{st} = x_{st} | z_{st}) := L_t(a | z_{st}). \end{aligned}$$

Given the definition of the identified set $\Theta_{I,1}$, for $\theta \in \Theta_{I,1}$, the following condition holds for any $a \in \mathcal{R}$,

$$\max_{s,t} \{L_s(a | z_{st}), L_t(a | z_{st})\} \leq \min_{s,t} \{U_t(a | z_{st}), U_s(a | z_{st})\},$$

Let $\tilde{L}(a | z_{st}) = \max_{s,t} \{L_t(a | z_{st}), L_s(a | z_{st})\}$, the stationarity condition (15) is satisfied by constructing the conditional distribution of $F_{v_s|Z_{st}}^*(a | z_{st}) = F_{v_t|Z_{st}}^*(a | z_{st})$ as follows:

$$F_{v_s|Z_{st}}^*(a | z_{st}) = F_{v_t|Z_{st}}^*(a | z_{st}) = \begin{cases} 0 & \text{if } a < m_1, \\ \tilde{L}(a | z_{st}) & \text{if } m_1 \leq a < \bar{m}, \\ 1 & \text{if } \bar{m} \leq a. \end{cases} \quad (17)$$

The above construction guarantees that the marginal distribution $F_{v_s|Z_{st}}^*(a | z_{st})$ lies between the bounds $[L_s(a | z_{st}), U_s(a | z_{st})]$. From the definition of $L_t(a)$, we know that $L_t(a | z_{st}) \geq L_t(a' | z_{st})$ for any $a \geq a'$. Therefore, the monotonicity of the above distribution is satisfied as $\max_{s,t} \{L_t(a | z_{st}), L_s(a | z_{st})\} \geq \max_{s,t} \{L_t(a' | z_{st}), L_s(a' | z_{st})\}$ for any $a \geq a'$. Additionally, the above construction satisfies the stationarity assumption given $F_{v_s|Z_{st}}^*(a | z_{st}) = F_{v_t|Z_{st}}^*(a | z_{st})$ for any a .

We transform the two requirements of sharpness into finding a distribution $F_{v_t|W_{st}}^*(a | w_{st})$ to generate the marginal distribution $F_{v_t|Z_{st}}^*(a | z_{st})$ in (17) and also match the marginal choice probability given in condition (16).

Construction

We focus on the construction for $v_{it} | W_{ist}$, and the construction for $v_{is} | W_{ist}$ at period s is similar so it is omitted here. The construction proceeds in four steps. We first constructs a distribution to match the lower and upper bounds $L_t(a | z_{st})$ and $U_t(a | z_{st})$. Then, we construct the distribution $F_{v_t|W_{st}}^*(a | w_{st})$ at the points $\{m_1, \dots, m_K\}$ with $m_k = z'_t \beta + a'_k \gamma$. The third step look at the construction at the points $\{n_1, \dots, n_K\}$ with $n_k = z'_s \beta + a'_k \gamma$. The final step examines the points a that are not in $\{m_1, \dots, m_K, n_1, \dots, n_K\}$.

Step 1:

We first show that the bounds $U_t(a | z_{st})$ and $L_t(a | z_{st})$ can be achieved under a

distribution of $v_{it} \mid W_{ist}$ that matches the marginal choice in (16). From its definition, the upper bound $U_t(a \mid z_{st})$ can be achieved under the following distribution $F_{v_t|W_{st}}^U(a \mid w_{st})$:

$$F_{v_t|W_{st}}^U(a \mid w_{st}) = \begin{cases} 0 & \text{if } a < m_1, \\ P_t(w_{st}) & \text{if } m_1 \leq a < w'_t\theta + \epsilon, \\ 1 & \text{if } w'_t\theta + \epsilon \leq a, \end{cases}$$

where $\epsilon > 0$ is used to satisfy the right-continuity property of the distribution.

Similarly, the lower bound function $L_t(a \mid z_{st})$ can be achieved under the following distribution $F_{v_t|W_{st}}^L(a \mid w_{st})$:

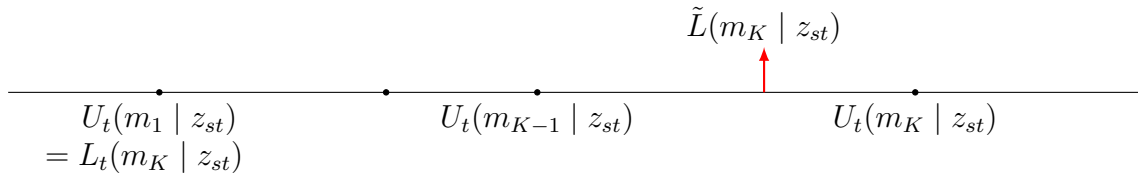
$$F_{v_t|W_{st}}^L(a \mid w_{st}) = \begin{cases} 0 & \text{if } a < w'_t\theta, \\ P_t(w_{st}) & \text{if } w'_t\theta \leq a < \bar{m}, \\ 1 & \text{if } \bar{m} \leq a \end{cases}$$

From the definition of the distributions $F_{v_t|W_{st}}^U(a \mid w_{st})$, $F_{v_t|W_{st}}^L(a \mid w_{st})$, they both satisfy condition (16) which requires $F_{v_t|W_{st}}^U(w'_t\theta \mid w_{st}) = F_{v_t|W_{st}}^L(w'_t\theta \mid w_{st}) = P_t(w_{st})$.

Step 2: this step establishes the construction of the distribution $F_{v_t|W_{st}}^*(a \mid w_{st})$ at the points m_k . The observed marginal distribution given in condition (16) only restricts the distribution of $v_{it} \mid W_{ist}$ at each point m_k , given $w_{st} = (z_{st}, x_s, a_k)$,

$$P_t(w_{st}) = \Pr(v_{it} \leq m_k \mid z_{st}, x_s, a_k).$$

(i) We start with the largest value $m_K = z'_t\beta + a'_K\gamma$. From the definition of $U_t(a \mid z_{st})$, $L_t(a \mid z_{st})$, we know that $U_t(m_1 \mid z_{st}) = L_t(m_K \mid z_{st}) \leq \tilde{L}(m_K \mid z_{st}) \leq U_t(m_K \mid z_{st})$, which is described in the following graph:



Given the K points $U_t(m_1 \mid z_{st}) \leq U_t(m_2 \mid z_{st}) \leq \dots \leq U_t(m_K \mid z_{st})$, we can find $2 \leq j_1 \leq K$ such that $U_t(m_{j_1-1} \mid z_{st}) \leq \tilde{L}(m_K \mid z_{st}) \leq U_t(m_{j_1} \mid z_{st})$. Then we can express $\tilde{L}(m_K \mid z_{st})$ as follows: for some $\alpha_1 \in [0, 1]$,

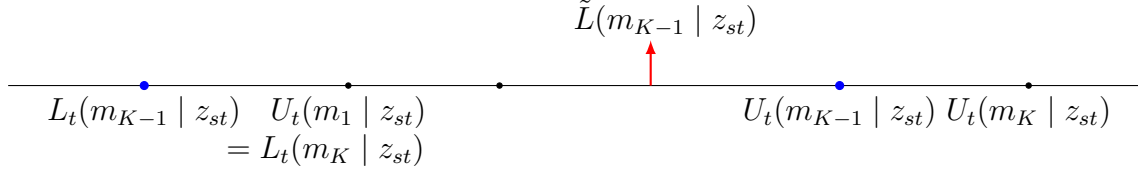
$$\tilde{L}(m_K \mid z_{st}) = \alpha_1 U_t(m_{j_1-1} \mid z_{st}) + (1 - \alpha_1) U_t(m_{j_1} \mid z_{st}).$$

Accordingly, we construct the distribution $F_{v_t|W_{st}}^*(m_K \mid w_{st})$ as follows:

$$F_{v_t|W_{st}}^*(m_K \mid w_{st}) = \alpha_1 F_t^U(m_{j_1-1} \mid w_{st}) + (1 - \alpha_1) F_t^U(m_{j_1} \mid w_{st}).$$

The above construction generates the marginal distribution $\tilde{L}(m_K | z_{st})$. Moreover, it matches the marginal choice probability when $x_t = a_K$ in condition (16) since $F_t^U(m_j | w_{st}) = P_t(w_{st})$ if $x_t = a_K$ for any $1 \leq j \leq K$ from the definition of $F_t^U(m_j | w_{st})$.

(ii) Now consider that $a = m_{K-1}$. Given that $L_t(m_{K-1} | z_{st}) \leq \tilde{L}(m_{K-1} | z_{st}) \leq U_t(m_{K-1} | z_{st})$, there either exists $1 \leq j_2 \leq K-1$ such that $\tilde{L}(m_{K-1} | z_{st}) \in [U_t(m_{j_2-1} | z_{st}), U_t(m_{j_2} | z_{st})]$ or $\tilde{L}(m_{K-1} | z_{st}) \in [L_t(m_{K-1} | z_{st}), U_t(m_1 | z_{st})]$, given in the following graph:



Then $\tilde{L}(m_{K-1} | z_{st})$ can be expressed as follows, for $\alpha_2 \in [0, 1]$ and $\tilde{\alpha}_2 \in [0, 1]$,

$$\begin{aligned} & \tilde{L}(m_{K-1} | z_{st}) \\ = & \begin{cases} \alpha_2 U_t(m_{j_2-1} | z_{st}) + (1 - \alpha_2) U_t(m_{j_2} | z_{st}) & \text{if } \tilde{L}(m_{K-1} | z_{st}) \in [U_t(m_{j_2-1} | z_{st}), U_t(m_{j_2} | z_{st})], \\ \tilde{\alpha}_2 L_t(m_{K-1} | z_{st}) + (1 - \tilde{\alpha}_2) U_t(m_1 | z_{st}) & \text{if } \tilde{L}(m_{K-1} | z_{st}) \in [L_t(m_{K-1} | z_{st}), U_t(m_1 | z_{st})]. \end{cases} \end{aligned}$$

Accordingly, the distribution $F_{v_t|W_{st}}^*(m_{K-1} | w_{st})$ at m_{K-1} is constructed as

$$\begin{aligned} & F_{v_t|W_{st}}^*(m_{K-1} | w_{st}) \\ = & \begin{cases} \alpha_2 F_t^U(m_{j_2-1} | w_{st}) + (1 - \alpha_2) F_t^U(m_{j_2} | w_{st}) & \text{if } \tilde{L}(m_{K-1} | z_{st}) \in [U_t(m_{j_2-1} | z_{st}), U_t(m_{j_2} | z_{st})], \\ \tilde{\alpha}_2 F_t^L(m_{K-1} | z_{st}) + (1 - \tilde{\alpha}_2) F_t^U(m_1 | w_{st}) & \text{if } \tilde{L}(m_{K-1} | z_{st}) \in [L_t(m_{K-1} | z_{st}), U_t(m_1 | z_{st})]. \end{cases} \end{aligned}$$

Condition (16) is satisfied since $U_t(m_j | w_{st}) = P_t(w_{st})$ when $x_t = a_{K-1}$ for any $1 \leq j \leq K-1$.

Now we still need to show that the constructed distribution satisfies monotonicity. If $\tilde{L}(m_{K-1} | z_{st}) \in [L_t(m_{K-1} | z_{st}), U_t(m_1 | z_{st})]$, then $F_{v_t|W_{st}}^*(m_{K-1} | w_{st}) \leq F_t^U(m_1 | w_{st}) \leq F_{v_t|W_{st}}^*(m_K | w_{st})$. Otherwise, by the monotonicity of $\tilde{L}(a | z_{st})$ in a , we know that $j_2 \leq j_1$. If $j_2 \leq j_1 - 1$, then monotonicity is satisfied since $F_{v_t|W_{st}}^*(m_{K-1} | w_{st}) \leq F_t^U(m_{j_2} | w_{st}) \leq F_t^U(m_{j_1-1} | w_{st}) \leq F_{v_t|W_{st}}^*(m_K | w_{st})$. If $j_1 - 1 < j_2 = j_1$, then it implies that $\alpha_2 \leq \alpha_1$ since $\tilde{L}(m_{K-1} | z_{st}) \leq \tilde{L}(m_K | z_{st})$. Therefore, we have $F_{v_t|W_{st}}^*(m_{K-1} | w_{st}) \leq F_{v_t|W_{st}}^*(m_K | w_{st})$.

The distribution $F_{v_t|W_{st}}^*(a | w_{st})$ for the remaining points $a \in \{m_1, \dots, m_{K-2}\}$ can be constructed in the same way as $F_{v_t|W_{st}}^*(m_{K-1} | w_{st})$, so it is omitted here.

Step 3: we show the construction at the points $a = n_k$ for $1 \leq k \leq K$. For points $a = n_k$, we only need to construct a proper distribution to be consistent with Equation (17). We still start with the largest point n_K . The distribution $\tilde{L}(n_K | z_{st})$ can be expressed as

follows in different scenarios: for some $2 \leq j_3 \leq K$

$$\tilde{L}(n_K | z_{st}) = \begin{cases} \alpha_3 \tilde{L}(m_K | z_{st}) + (1 - \alpha_3) * 1 & \text{if } n_K > m_K, \\ \alpha_4 \tilde{L}(m_{j_3-1} | z_{st}) + (1 - \alpha_4) \tilde{L}(m_{j_3} | z_{st}) & \text{if } m_{j_3-1} < n_K < m_{j_3}, \\ \alpha_5 * 0 + (1 - \alpha_5) \tilde{L}(m_1 | z_{st}) & \text{if } n_K < m_1, \end{cases}$$

where $\alpha_3, \alpha_4, \alpha_5 \in [0, 1]$.

Accordingly, the distribution $F_{v_t|W_{st}}^*(n_k | w_{st})$ is constructed as follows,

$$F_{v_t|W_{st}}^*(n_K | w_{st}) = \begin{cases} \alpha_3 F_{v_t|W_{st}}^*(m_K | w_{st}) + (1 - \alpha_3) * 1 & \text{if } n_K > m_K, \\ \alpha_4 F_{v_t|W_{st}}^*(m_{j_3-1} | w_{st}) + (1 - \alpha_4) F_{v_t|W_{st}}^*(m_{j_3} | w_{st}) & \text{if } m_{j_3-1} < n_K < m_{j_3}, \\ \alpha_5 * 0 + (1 - \alpha_5) F_{v_t|W_{st}}^*(m_{j_1} | w_{st}) & \text{if } n_K < m_1. \end{cases}$$

When $a = n_{K-1}$, the distribution $\tilde{L}(n_{K-1} | z_{st})$ can be expressed as follows in different scenarios: for some $1 \leq j_4 \leq j_3 - 1$,

$$\tilde{L}(n_{K-1} | z_{st}) = \begin{cases} \alpha_6 \tilde{L}(m_K | z_{st}) + (1 - \alpha_6) \tilde{L}(n_K | z_{st}) & \text{if } m_K < n_{K-1} < n_K, \\ \alpha_7 \tilde{L}(m_{j_3-1} | z_{st}) + (1 - \alpha_7) \tilde{L}(n_K | z_{st}) & \text{if } m_{j_3-1} < n_{K-1} < n_K < m_{j_3}, \\ \alpha_8 \tilde{L}(m_{j_4-1} | z_{st}) + (1 - \alpha_8) \tilde{L}(m_{j_4} | z_{st}) & \text{if } m_{j_4-1} < n_{K-1} < m_{j_4} \leq m_{j_3-1}, \\ \alpha_9 * 0 + (1 - \alpha_9) \tilde{L}(n_K | z_{st}) & \text{if } n_{K-1} < n_K < m_1, \end{cases}$$

where $\alpha_6, \alpha_7, \alpha_8, \alpha_9 \in [0, 1]$.

The distribution $F_{v_t|W_{st}}^*(n_{K-1} | w_{st})$ is constructed accordingly:

$$F_{v_t|W_{st}}^*(n_K | w_{st}) = \begin{cases} \alpha_6 F_{v_t|W_{st}}^*(m_K | w_{st}) + (1 - \alpha_6) F_{v_t|W_{st}}^*(n_K | w_{st}) & \text{if } m_K < n_{K-1} < n_K, \\ \alpha_7 F_{v_t|W_{st}}^*(m_{j_3-1} | w_{st}) + (1 - \alpha_7) F_{v_t|W_{st}}^*(n_K | w_{st}) & \text{if } m_{j_3-1} < n_{K-1} < n_K < m_{j_3}, \\ \alpha_8 F_{v_t|W_{st}}^*(m_{j_4-1} | w_{st}) + (1 - \alpha_8) F_{v_t|W_{st}}^*(m_{j_4} | w_{st}) & \text{if } m_{j_4-1} < n_{K-1} < m_{j_4} \leq m_{j_3-1}, \\ (1 - \alpha_9) F_{v_t|W_{st}}^*(n_K | w_{st}) & \text{if } n_{K-1} < n_K < m_1. \end{cases}$$

The monotonicity of the distribution $F_{v_t|W_{st}}^*(a | w_{st})$ among the $2K$ points $\{m_1, \dots, m_K, n_1, \dots, n_K\}$ is satisfied by construction. Furthermore, the construction for the rest of points $\{n_1, \dots, n_{K-2}\}$ are similar so it is omitted here.

Step 4: For any $a \notin \{m_1, \dots, m_K, n_1, \dots, n_K\}$, The distribution is constructed as $F_{v_t|W_{st}}^*(a | w_{st}) = 0$ if $a < m_1$ and $F_{v_t|W_{st}}^*(a | w_{st}) = 1$ if $a \geq \bar{m}$. When $a \in [m_1, \bar{m}]$, there must exist some $q_j \in \{m_1, \dots, m_K, n_1, \dots, n_K\}$ such that $q_j < a$ and $\tilde{L}(a | z_{st}) = \tilde{L}(q_j | z_{st})$. Then we set $F_{v_t|W_{st}}^*(a | w_{st}) = F_{v_t|W_{st}}^*(m_j | w_{st})$. This completes the construction for the distribution $F_{v_t|W_{st}}^*(a | w_{st})$ for any a .

A.3 Proof of Proposition 2 and 3

Proof. The proof for the point identification of β_0 consists of two steps: we first show that when $\Delta z \in \Delta \mathcal{Z}$, the sign of $\Delta z' \beta_0$ is identified from the identifying condition (7) in Proposition 1. Then, the large support condition in Assumption 2 ensures that β_0 is point identified up to scale.

When X_{it} is discrete and there are two periods $T = 2$, the identifying condition (7) is given as

$$1 - \Pr(Y_{i1} = 0, z'_1 \beta_0 + X'_{i1} \gamma_0 \geq a \mid z_{12}) \geq \Pr(Y_{i2} = 1, z'_2 \beta_0 + X'_{i2} \gamma_0 \leq a \mid z_{12}),$$

for $a \in \{z'_1 \beta_0 + a'_1 \gamma_0, \dots, z'_1 \beta_0 + a'_K \gamma_0, z'_2 \beta_0 + a'_1 \gamma_0, \dots, z'_2 \beta_0 + a'_K \gamma_0\}$, and another identifying condition switches the order of period 1 and 2.

Let $a = z'_1 \beta_0 + a'_k \gamma_0$,¹⁰ then the above upper bound can be further bounded as

$$1 - \Pr(Y_{i1} = 0, z'_1 \beta_0 + X'_{i1} \gamma_0 \geq z'_1 \beta_0 + a'_k \gamma_0 \mid z_{12}) \leq 1 - \Pr(Y_{i1} = 0, X_{i1} = a_k \mid z_{12}).$$

When $z'_1 \beta_0 - z'_2 \beta_0 \geq 0$ which implies $z'_1 \beta_0 + a'_k \gamma_0 \geq z'_2 \beta_0 + a'_k \gamma_0$, then the lower bound can be bounded below as

$$\Pr(Y_{i2} = 1, z'_2 \beta_0 + X'_{i2} \gamma_0 \leq z'_1 \beta_0 + a'_k \gamma_0 \mid z_{12}) \leq \Pr(Y_{i2} = 1, X_{i2} = a_k \mid z_{12}).$$

Combining the above results leads to

$$\text{If } z'_1 \beta_0 - z'_2 \beta_0 \geq 0 \implies 1 - \Pr(Y_{i1} = 0, X_{i1} = a_k \mid z_{12}) \geq \Pr(Y_{i2} = 1, X_{i2} = a_k \mid z_{12}).$$

The contraposition of the above inequality yields

$$1 - \Pr(Y_{i1} = 0, X_{i1} = a_k \mid z_{12}) < \Pr(Y_{i2} = 1, X_{i2} = a_k \mid z_{12}) \implies \Delta z' \beta_0 > 0.$$

Switching the order of the time period leads to another identifying restriction as follows:

$$1 - \Pr(Y_{i1} = 1, X_{i1} = a_k \mid z_{12}) < \Pr(Y_{i2} = 0, X_{i2} = a_k \mid z_{12}) \implies \Delta z' \beta_0 < 0.$$

Therefore, when $\Delta z \in \Delta \mathcal{Z}$, the sign of $\Delta z' \beta_0$ is identified.

Next, we show that β_0 is point identified under the large support assumption. To prove it, we will show that for any $\beta \neq k\beta_0$ for some k , there exists some value Δz such that $\Delta z' \beta$ has different signs from $\Delta z' \beta_0$.

From Assumption 2, the conditional support of Δz^{j^*} is \mathcal{R} and $\beta_0^{j^*} \neq 0$. We focus on the case where $\beta_0^{j^*} > 0$, and the analysis also applies to the other case. Let $\Delta \tilde{z} := \Delta z \setminus \Delta z^{j^*}$ denote the remaining covariates in Δz and $\tilde{\beta}_0$ denote its coefficient. For any candidate b , we

¹⁰The value of $a = z'_2 \beta_0 + a'_k \gamma_0$ leads to the same identifying condition.

discuss three cases: $b^{j^*} < 0$, $b^{j^*} = 0$, and $b^{j^*} > 0$.

Case 1: $b^{j^*} < 0$. When the covariate Δz^{j^*} takes a large positive value $\Delta z^{j^*} \rightarrow +\infty$ and the remaining covariates take bounded values in their support, it implies that $\Delta z' \beta_0 > 0$ and $\Delta z' b < 0$.

Case 2: $b^{j^*} = 0$. For any value Δz , the value of $\Delta z' b$ is either positive or nonpositive. When $\Delta z' b > 0$ is positive, then let Δz^{j^*} take a large negative value $\Delta z^{j^*} \rightarrow -\infty$ such that $\Delta z' \beta_0 < 0$, which has a different sign from $\Delta z' b$. Similarly, if $\Delta z' b \leq 0$, there exists $\Delta z^{j^*} \rightarrow +\infty$ such that $\Delta z' \beta_0 > 0$.

Case 3: $b^{j^*} > 0$. Assumption 2 requires that $\Delta \mathcal{Z}$ is not contained in any proper linear subspace, so there exists Δz such that $\Delta \tilde{z}' \tilde{\beta}_0 / \beta_0^{j^*} \neq \Delta \tilde{z}' \tilde{b} / b^{j^*}$. Suppose that $\Delta \tilde{z}' \tilde{\beta}_0 / \beta_0^{j^*} - \Delta \tilde{z}' \tilde{b} / b^{j^*} = k > 0$, then when the covariate takes the value $\Delta Z_i = -\Delta \tilde{z}' \tilde{b} / b^{j^*} - \epsilon$ with $0 < \epsilon < k$. The sign of the covariate index satisfies: $\Delta z' \beta_0 = \beta_0^{j^*} (k - \epsilon) > 0$ and $\Delta z' b = -b^{j^*} \epsilon < 0$. The construction is similar when $k < 0$.

For the identification of γ_0 , under the similar analysis for β_0 , we have

$$\begin{aligned} (z_1, z_2) \in \mathcal{Z}_3^j &\implies (x_1^j - x_2^j) \gamma_0^j < \Delta z' \beta_0, \\ (z_1, z_2) \in \mathcal{Z}_4^j &\implies (x_1^j - x_2^j) \gamma_0^j > \Delta z' \beta_0. \end{aligned}$$

As previously shown, when $(z_1, z_2) \in \mathcal{Z}_2$, it implies that $\Delta z' \beta_0 < 0$. Therefore, when $(z_1, z_2) \in \mathcal{Z}_2 \cap \mathcal{Z}_3^j$, we have $(x_1^j - x_2^j) \gamma_0^j < \Delta z' \beta_0 < 0$ and the sign of γ_0^j is identified given $x_1^j \neq x_2^j$. Similarly, when $(z_1, z_2) \in \mathcal{Z}_1 \cap \mathcal{Z}_4^j$, the sign of γ_0^j is also identified given $(x_1^j - x_2^j) \gamma_0^j > \Delta z' \beta_0 > 0$. Proposition 3 requires that for any $j \leq d_x$, either $\mathcal{Z}_2 \cap \mathcal{Z}_3^j \neq \emptyset$ or $\mathcal{Z}_1 \cap \mathcal{Z}_4^j \neq \emptyset$ so that the sign of γ_0^j is identified for any j . □

A.4 Relationship to Khan, Ponomareva, and Tamer (2023)

Lemma 3. Under Assumption 1 and $X_{it} = Y_{i,t-1}$, condition (7) implies the identifying restriction Khan, Ponomareva, and Tamer (2023), presented as follows:

- (i) $P(y_{it} = 1 | z_{st}) > P(y_{is} = 1 | z_{st}) \implies (z_t - z_s)' \beta_0 + |\gamma_0| > 0$,
- (ii) $P(y_{it} = 1 | z_{st}) > 1 - P(y_{i,s-1} = 1, y_{is} = 0 | z_{st}) \implies (z_t - z_s)' \beta_0 - \min\{0, \gamma_0\} > 0$,
- (iii) $P(y_{it} = 1 | z_{st}) > 1 - P(y_{i,s-1} = 0, y_{is} = 0 | z_{st}) \implies (z_t - z_s)' \beta_0 + \max\{0, \gamma_0\} > 0$,
- (iv) $P(y_{i,t-1} = 1, y_{it} = 1 | z_{st}) > P(y_{is} = 1 | z_{st}) \implies (z_t - z_s)' \beta_0 + \max\{0, \gamma_0\} > 0$,
- (v) $P(y_{i,t-1} = 1, y_{it} = 1 | z_{st}) > 1 - P(y_{i,s-1} = 1, y_{is} = 0 | z_{st}) \implies (z_t - z_s)' \beta_0 > 0$,
- (vi) $P(y_{i,t-1} = 1, y_{it} = 1 | z_{st}) > 1 - P(y_{i,s-1} = 0, y_{is} = 0 | z_{st}) \implies (z_t - z_s)' \beta_0 + \gamma_0 > 0$,
- (vii) $P(y_{i,t-1} = 0, y_{it} = 1 | z_{st}) > P(y_{is} = 1 | z_{st}) \implies (z_t - z_s)' \beta_0 - \min\{0, \gamma_0\} > 0$,
- (viii) $P(y_{i,t-1} = 0, y_{it} = 1 | z_{st}) > 1 - P(y_{i,s-1} = 1, y_{is} = 0 | z_{st}) \implies (z_t - z_s)' \beta_0 - \gamma_0 > 0$,
- (ix) $P(y_{i,t-1} = 0, y_{it} = 1 | z_{st}) > 1 - P(y_{i,s-1} = 0, y_{is} = 0 | z_{st}) \implies (z_t - z_s)' \beta_0 > 0$.

Proof. When the endogenous regressor is the lagged dependent variable $X_{it} = Y_{i,t-1}$, the identifying condition in Proposition 1 is given as

$$\begin{aligned} U_s(a) &:= 1 - \Pr(Y_{is} = 0, z'_s \beta_0 + Y_{i,s-1} \gamma_0 \geq a | z_{st}) \\ &\geq \Pr(Y_{it} = 1, z'_t \beta_0 + Y_{i,t-1} \gamma_0 \leq a | z_{st}) := L_t(a). \end{aligned}$$

Since $Y_{i,t-1}$ only takes two values $Y_{i,t-1} \in \{0, 1\}$, we can express the probability in the above bounds as a mixture of $Y_{i,t-1} = 1$ and $Y_{i,t-1} = 0$:

$$\begin{aligned} U_s(a) &= 1 - \Pr(Y_{is} = 0, Y_{i,s-1} = 1 | z_{st}) \mathbb{1}\{z'_s \beta_0 + \gamma_0 \geq a\} \\ &\quad - \Pr(Y_{is} = 0, Y_{i,s-1} = 0 | z_{st}) \mathbb{1}\{z'_s \beta_0 \geq a\}. \end{aligned}$$

Similarly, the lower bound is given as

$$\begin{aligned} L_t(a) &= \Pr(Y_{it} = 1, Y_{i,t-1} = 1 | z_{st}) \mathbb{1}\{z'_t \beta_0 + \gamma_0 \leq a\} \\ &\quad + \Pr(Y_{it} = 1, Y_{i,t-1} = 0 | z_{st}) \mathbb{1}\{z'_t \beta_0 \leq a\}. \end{aligned}$$

Now we discuss different scenarios for the value of $(z'_t \beta_0, z'_s \beta_0, \gamma_0)$, leading to various identifying restrictions in Khan, Ponomareva, and Tamer (2023) described in Lemma 3.

Case 1: consider that $z'_t \beta_0 + \max\{0, \gamma_0\} \leq z'_s \beta_0 + \min\{0, \gamma_0\}$. Let $a = z'_t \beta_0 + \max\{0, \gamma_0\}$, the bounds can be simplified as

$$\begin{aligned} U_s(a) &= 1 - \Pr(Y_{is} = 0, Y_{i,s-1} = 1 | z_{st}) - \Pr(Y_{is} = 0, Y_{i,s-1} = 0 | z_{st}) = \Pr(Y_{is} = 1 | z_{st}), \\ L_t(a) &= \Pr(Y_{it} = 1, Y_{i,t-1} = 1 | z_{st}) + \Pr(Y_{it} = 1, Y_{i,t-1} = 0 | z_{st}) = \Pr(Y_{it} = 1 | z_{st}). \end{aligned}$$

By exploiting the condition $\max\{0, \gamma_0\} - \min\{0, \gamma_0\} = |\gamma_0|$, the identifying condition is given as

$$(z'_t\beta_0 - z'_s\beta_0) + |\gamma_0| \leq 0 \implies U_s(a) = \Pr(Y_{is} = 1 \mid z_{st}) \geq L_t(a) = \Pr(Y_{it} = 1 \mid z_{st}).$$

The contraposition of the above condition yields condition (i) in Lemma 3:

$$\Pr(Y_{it} = 1 \mid z_{st}) > \Pr(Y_{is} = 1 \mid z_{st}) \implies (z'_t\beta_0 - z'_s\beta_0) + |\gamma_0| > 0.$$

Case 2: consider that $z'_t\beta_0 + \max\{0, \gamma_0\} \leq z'_s\beta_0 + \gamma_0$. Let $a = z'_s\beta_0 + \gamma_0$, the bounds are simplified as

$$U_s(a) \leq 1 - \Pr(Y_{is} = 0, Y_{i,s-1} = 1 \mid z_{st}), \quad L_t(a) = \Pr(Y_{it} = 1 \mid z_{st}).$$

Given that $\max\{0, \gamma_0\} - \gamma_0 = -\min\{0, \gamma_0\}$, we have the following identifying condition:

$$z'_t\beta_0 - z'_s\beta_0 - \min\{0, \gamma_0\} \leq 0 \implies 1 - \Pr(Y_{is} = 0, Y_{i,s-1} = 1 \mid z_{st}) \geq \Pr(Y_{it} = 1 \mid z_{st}).$$

The contraposition of this condition leads to condition (ii) in Lemma 3:

$$\Pr(Y_{it} = 1 \mid z_{st}) > 1 - \Pr(Y_{is} = 0, Y_{i,s-1} = 1 \mid z_{st}) \implies z'_t\beta_0 - z'_s\beta_0 - \min\{0, \gamma_0\} > 0.$$

The rest of conditions in Lemma 3 can be derived in a similar way so we omit the analysis here. □

A.5 Proof of Proposition 2

Proof. The identifying restriction (9) in Proposition 4 is given as follows:

$$\begin{aligned} U_{t,order}(a \mid z_{st}) &:= 1 - \sum_{j=1}^J \Pr(Y_{is} = y_j, b_j - z'_s\beta - X'_{is}\gamma \geq a \mid z_{st}) \\ &\geq \sum_{j=1}^J \Pr(Y_{it} = y_j, b_{j+1} - z'_t\beta - X'_{it}\gamma \leq a \mid z_{st}) := L_{t,order}(a \mid z_{st}). \end{aligned} \tag{18}$$

The proof idea of Corollary 2 is the same as Appendix 1 for Corollary 1. Let $q_k \in \mathcal{Q}_2$ denote a point in \mathcal{Q}_2 , and we can rank the $2K(J-1)$ points as $q_1 \leq q_2 \dots \leq q_{2K(J-1)}$.

When $a > q_{2K(J-1)}$, then $\Pr(Y_{is} = y_j, b_j - z'_s\beta - X'_{is}\gamma \geq a \mid z_{st}) = 0$ for any j such that the upper bound is one $U_{t,order}(a \mid z_{st}) = 1$. Therefore, condition (18) holds for any θ and it is not informative for θ .

When $a < q_1$, then $\Pr(Y_{it} = y_j, b_{j+1} - z'_t\beta - X'_{it}\gamma \leq a \mid z_{st}) = 0$ for any $2 \leq j \leq K$ so that the lower bound is zero $L_{t,order}(a \mid z_{st}) = 0$ and it is not informative for θ .

When $a \in (q_k, q_{k+1})$, we show that the identifying condition (18) at q_{k+1} contains more

information than the one at a . The lower bound $L_{t,order}(a | z_{st})$ is increasing in a , therefore the lower bound evaluated at q_{k+1} is larger than the one evaluated at a . The upper bound is the same when evaluated at q_{k+1} and at a , as there does not exist any point $b_j - z'_s \beta - a'_k \gamma$ that lies in $[a, q_{k+1})$ from the definition of q_k . Therefore, if the parameter θ satisfies condition (18) evaluated at q_{k+1} , it must satisfy condition (18) evaluated at a . The identifying condition at a does not provide additional information for θ . \square

A.6 Proof of Lemma 1

Proof. When all regressors are exogenous in the sense $\epsilon_{is} | (W_{ist}, \alpha_i) \stackrel{d}{\sim} \epsilon_{it} | (W_{ist}, \alpha_i)$, the identifying restriction in Proposition 4 is given as

$$1 - \sum_{j=1}^J \Pr(Y_{is} = y_j | w_{st}) \mathbb{1}\{b_j - w'_s \theta_0 \geq a\} \geq \sum_{j=1}^J \Pr(Y_{it} = y_j | w_{st}) \mathbb{1}\{b_{j+1} - w'_t \theta_0 \leq a\}. \quad (19)$$

For any $a \in \mathcal{R}$, the above identifying restriction is only informative when there exists $2 \leq j_1 \leq J$ and $1 \leq j_2 \leq J - 1$ such that

$$b_{j_1} - w'_s \theta_0 \geq a, \quad b_{j_2+1} - w'_t \theta_0 \leq a. \quad (20)$$

It is because if the above condition does not hold, then either the upper bound in condition (19) becomes one or the lower bound becomes zero. Then condition (19) holds for any θ and there is no identifying power.

Condition (20) can be satisfied for some a if and only if there exists $2 \leq j_1 \leq J$ and $1 \leq j_2 \leq J - 1$ such that $b_{j_1} - w'_s \theta_0 \geq b_{j_2+1} - w'_t \theta_0$. In this case, the identifying condition (19) becomes

$$\begin{aligned} \text{if } b_{j_1} - w'_s \theta_0 \geq b_{j_2+1} - w'_t \theta_0 \implies & \sum_{j=1}^{j_1-1} \Pr(Y_{is} = y_j | w_{st}) = 1 - \sum_{j=j_1}^J \Pr(Y_{is} = y_j | w_{st}) \\ & \geq \sum_{j=1}^{j_2} \Pr(Y_{it} = y_j | w_{st}). \end{aligned}$$

Replacing j_1 with $\tilde{j}_1 = j_1 - 1$ yielding the results in Proposition 4. \square

A.7 Proof of Proposition 5 and Proposition 6

Proof. We first prove that β_0 is point identified up to scale by showing that for any $\Delta z \in \Delta \mathcal{Z}_{order}$, the sign of $\Delta z' \beta_0$ is identified. The identifying condition for θ_0 in Proposition 4 is

given as

$$1 - \sum_{j=1}^J \Pr(Y_{i1} = y_j, b_j - z'_1 \beta_0 - X'_{i1} \gamma_0 \geq a \mid z_{12}) \geq \sum_{j=1}^J \Pr(Y_{i2} = y_j, b_{j+1} - z'_2 \beta_0 - X'_{i2} \gamma_0 \leq a \mid z_{12}),$$

and the same condition also holds when changing the order of period 1 and 2.

Let $a = b_k - z'_1 \beta_0 - x' \gamma_0$ for $2 \leq k \leq J$, then we know that $b_j - z'_1 \beta_0 - x' \gamma_0 \geq a$ if $j \geq k$. Therefore, the upper bound becomes

$$1 - \sum_{j=1}^J \Pr(Y_{i1} = y_j, b_j - z'_1 \beta_0 - X'_{i1} \gamma_0 \geq a \mid z_{st}) \geq 1 - \sum_{j=k}^J \Pr(Y_{i1} = y_j, X_{i1} = x).$$

Moreover, if $-z'_2 \beta_0 \leq -z'_1 \beta_0$, then $b_{j+1} - z'_2 \beta_0 - x' \gamma_0 \leq a$ for $j+1 \leq k$. The lower bound can be bounded below by

$$\sum_{j=1}^J \Pr(Y_{i2} = y_j, b_{j+1} - z'_2 \beta_0 - X'_{i2} \gamma_0 \leq a \mid z_{12}) \geq \sum_{j=1}^{k-1} \Pr(Y_{i2} = y_j, X_{i2} = x \mid z_{12}).$$

Combining the results yields the following conditions:

$$\Delta z' \beta_0 \geq 0 \implies 1 - \sum_{j=k}^J \Pr(Y_{i1} = y_j, X_{i1} = x) \geq \sum_{j=1}^{k-1} \Pr(Y_{i2} = y_j, X_{i2} = x \mid z_{12}).$$

The contraposition of the above condition generates

$$1 - \sum_{j=k}^J \Pr(Y_{i1} = y_j, X_{i1} = x) \geq \sum_{j=1}^{k-1} \Pr(Y_{i2} = y_j, X_{i2} = x \mid z_{12}) \implies \Delta z' \beta_0 < 0.$$

Therefore, when $z \in \mathcal{Z}_{1,order}$, the sign of $\Delta z' \beta_0 < 0$ is identified. Similarly, the sign of $\Delta z' \beta_0 > 0$ is also identified when $z \in \mathcal{Z}_{2,order}$. Given that the sign of $\Delta z' \beta_0$ is identified, the parameter β_0 is point identified up to scale under the large support condition Δz . The analysis is the same as proof [A.3](#) for Proposition [2](#), so it is omitted here.

Now we look at the result for γ_0 in Proposition [6](#). Let $a = b_k - z'_1 \beta_0 - x'_1 \gamma_0$, then the upper bound becomes

$$1 - \sum_{j=1}^J \Pr(Y_{i1} = y_j, b_j - z'_1 \beta_0 - X'_{i1} \gamma_0 \geq a \mid z_{st}) \geq 1 - \sum_{j=k}^J \Pr(Y_{i1} = y_j, X_{i1} = x_1).$$

If $b_k - z'_2 \beta_0 - x'_2 \gamma_0 \leq a$, the lower bound becomes

$$\sum_{j=1}^J \Pr(Y_{i2} = y_j, b_{j+1} - z'_2 \beta_0 - X'_{i2} \gamma_0 \leq a \mid z_{12}) \geq \sum_{j=1}^{k-1} \Pr(Y_{i2} = y_j, X_{i2} = x_2 \mid z_{12}).$$

Combining the results leads to

if $-z'_2\beta_0 - x'_2\gamma_0 \leq -z'_1\beta_0 - x'_1\gamma_0 \implies$

$$1 - \sum_{j=k}^J \Pr(Y_{i1} = y_j, X_{i1} = x_1) \geq \sum_{j=1}^{k-1} \Pr(Y_{i2} = y_j, X_{i2} = x \mid z_{12}).$$

Similarly, the contraposition of the above condition gives

$$1 - \sum_{j=k}^J \Pr(Y_{i1} = y_j, X_{i1} = x_1) < \sum_{j=1}^{k-1} \Pr(Y_{i2} = y_j, X_{i2} = x_2 \mid z_{12}) \implies \Delta z' \beta_0 < (x_1 - x_2)' \gamma_0.$$

Therefore, when $\Delta z \in \mathcal{Z}_{3,order}^j \cap \mathcal{Z}_{2,order}$, it implies that

$$0 < \Delta z' \beta_0 < (x_1^j - x_2^j) \gamma_0^j.$$

Given that $x_1^j - x_2^j \neq 0$, it implies that the sign of γ_0^j is identified. Similarly when $\Delta z \in \mathcal{Z}_{4,order}^j \cap \mathcal{Z}_{1,order}$, we have $0 < \Delta z' \beta_0 < (x_1^j - x_2^j) \gamma_0^j$. The sign of γ_0^j is also identified. For any $j \leq d_x$, Proposition 6 requires that either $\mathcal{Z}_{3,order}^j \cap \mathcal{Z}_{2,order} \neq \emptyset$ or $\mathcal{Z}_{4,order}^j \cap \mathcal{Z}_{1,order} \neq \emptyset$, so the sign of γ_0^j is identified for any j . □

A.8 Proof of Proposition 7

Proof. For any set $K \subset \mathcal{J}$, the conditional probability of selecting a choice $j \in K$ given $W_{ist} = w_{st}$ is:

$$\Pr(Y_{it} \in K \mid w_{st}) = \Pr(\exists j \in K \text{ s.t. } w'_{ijt}\theta_0 + v_{ijt} \geq w'_{ikt}\theta_0 + v_{ikt} \forall k \in K^c \mid w_{st}).$$

The above observed probability provides information regarding the distribution of $v_{ikt} - v_{ijt}$. Therefore, we can use observed data to bound the following distribution: for $a_{jk} \in \mathcal{R}$,

$$Q_t(a_{jk} \mid w_{st}) := \Pr(\exists j \in K \text{ s.t. } v_{ikt} - v_{ijt} \leq a_{jk} \forall k \in K^c \mid w_{st}).$$

When a_{jk} satisfies $a_{jk} \geq (w_{ijt} - w_{ikt})' \theta_0$ for any (j, k) , then the above probability can be bounded below as

$$\begin{aligned} Q_t(a_{jk} \mid w_{st}) &\geq \Pr(\exists j \in K \text{ s.t. } v_{ikt} - v_{ijt} \leq (w_{ijt} - w_{ikt})' \theta_0 \forall k \in K^c \mid w_{st}) \\ &= \Pr(Y_{it} \in K \mid w_{st}). \end{aligned}$$

Therefore, the lower bound for $Q_t(a_{jk} \mid w_{st})$ is established as

$$Q_t(a_{jk} \mid w_{st}) \geq \Pr(Y_{it} \in K \mid w_{st}) \mathbb{1}\{a_{jk} \geq (w_{ijt} - w_{ikt})' \theta_0 \forall j \in K, k \in K^c\}.$$

The above inequality holds since either $a_{jk} \geq (w_{ijt} - w_{ikt})' \theta_0$ or the lower bound is zero.

Since Assumption 1 is only conditioning on the exogenous covariate z_{st} , we can bound the conditional distribution $Q_t(a_{jk} | z_{st})$ by taking expectation of X_{it} given z_{st} :

$$Q_t(a_{jk} | z_{st}) \geq \Pr(Y_{it} \in K, a_{jk} \geq (z_{ijt} - z_{ikt})'\beta_0 + (X_{ijt} - X_{ikt})'\gamma_0 \forall j \in K, k \in K^c | z_{st}).$$

Similarly, an upper bound for the conditional probability $Q_t(a_{jk} | w_{st})$ is derived as follows:

$$Q_t(a_{jk} | w_{st}) \leq \Pr(Y_{it} \in K | w_{st}) \mathbb{1}\{a_{jk} \leq (w_{ijt} - w_{ikt})'\theta_0 \forall j \in K, k \in K^c\} + 1 - \mathbb{1}\{a_{jk} \leq (w_{ijt} - w_{ikt})'\theta_0 \forall j \in K, k \in K^c\}.$$

The above inequality holds since either $a_{jk} \leq (w_{ijt} - w_{ikt})'\theta_0$ and the probability $Q_t(a_{jk} | w_{st})$ is weakly increasing or the upper bound is one.

After taking expectation of X_{ist} given z_{st} , the upper bound for $Q_t(a_{jk} | z_{st})$ is obtained as

$$Q_t(a_{jk} | z_{st}) \leq \Pr(Y_{it} \in K, a_{jk} \leq (z_{ijt} - z_{ikt})'\beta_0 + (X_{ijt} - X_{ikt})'\gamma_0 \forall j \in K, k \in K^c | z_{st}) + 1 - \Pr(a_{jk} \leq (z_{ijt} - z_{ikt})'\beta_0 + (X_{ijt} - X_{ikt})'\gamma_0 \forall j \in K, k \in K^c | z_{st}).$$

Rearranging the above formula yields

$$Q_t(a_{jk} | z_{st}) \leq 1 - \Pr(Y_{it} \in K^c, a_{jk} \leq (z_{ijt} - z_{ikt})'\beta_0 + (X_{ijt} - X_{ikt})'\gamma_0 \forall j \in K, k \in K^c | z_{st}).$$

Under Assumption 1, the conditional probability $Q_t(a_{jk} | z_{st})$ is the same over the two periods (s, t). Therefore, the upper bound at period s of $Q_t(a_{jk} | z_{st})$ should be larger than the lower bound at period t , yielding the identifying condition (10) as follows:

$$1 - \Pr(Y_{is} \in K^c, (z_{js} - z_{ks})'\beta_0 + (X_{ijs} - X_{iks})'\gamma_0 \geq a_{jk} \forall j \in K, k \in K^c | z_{st}) \geq \Pr(Y_{it} \in K, (z_{jt} - z_{kt})'\beta_0 + (X_{ijt} - X_{ikt})'\gamma_0 \leq a_{jk} \forall j \in K, k \in K^c | z_{st}).$$

□

A.9 Proof of Proposition 8

Proof. Since the observed outcome Y_{it} is censored, we either observe $Y_{it} = y > 0$ or $Y_{it} = 0$. Let $v_{it} = -(\alpha_i + \epsilon_{it})$, the conditional probability of $Y_{it} = 0$ is given as,

$$\Pr(Y_{it} = 0 | w_{st}) = \Pr(Y_{it}^* \leq 0 | w_{st}) = \Pr(v_{it} \geq z'_t \beta_0 + x'_t \gamma_0 | w_{st}).$$

When $y > 0$, the conditional distribution is given as

$$\begin{aligned}
\Pr(Y_{it} \leq y \mid w_{st}) &= \Pr(Y_{it}^* \leq 0, Y_{it} \leq y \mid w_{st}) + \Pr(0 < Y_{it}^*, Y_{it} \leq y \mid w_{st}) \\
&= \Pr(Y_{it}^* \leq 0 \mid w_{st}) + \Pr(0 < Y_{it}^* \leq y \mid w_{st}) \\
&= \Pr(Y_{it}^* \leq y \mid w_{st}) \\
&= \Pr(v_{it} \geq z'_t \beta_0 + x'_t \gamma_0 - y \mid w_{st}).
\end{aligned}$$

Combining the two scenarios, the full conditional distribution of $Y_{it} \mid W_{ist}$ is characterized as follows:

$$\Pr(Y_{it} \leq y \mid w_{st}) = \begin{cases} \Pr(v_{it} \geq z'_t \beta_0 + x'_t \gamma_0 - y \mid w_{st}) & \text{if } y \geq 0, \\ 0 & \text{if } y < 0. \end{cases}$$

Given observed distribution of $Y_{it} \mid W_{ist}$, the identification approach is to bound the distribution $\Pr(v_{it} \leq a \mid z_{st})$. We first look at the upper bound for the conditional distribution $v_{it} \mid w_{st}$ given all covariates $W_{ist} = w_{st}$.

$$\begin{aligned}
\Pr(v_{it} \geq a \mid w_{st}) &\leq \Pr(Y_{it} \leq z'_t \beta_0 + x'_t \gamma_0 - a) \mathbb{1}\{z'_t \beta_0 + x'_t \gamma_0 \geq a\} + \\
&\quad \Pr(Y_{it} = 0 \mid z_{st}) \mathbb{1}\{z'_t \beta_0 + x'_t \gamma_0 < a\},
\end{aligned}$$

where the above condition holds since either $z'_t \beta_0 + x'_t \gamma_0 - a \geq 0$ so that there exists $y = z'_t \beta_0 + x'_t \gamma_0 - a \geq 0$ such that $\Pr(Y_{it} \leq y \mid w_{st}) = \Pr(v_{it} \geq a \mid w_{st})$, or $\Pr(v_{it} \geq a \mid w_{st}) \leq \Pr(v_{it} \geq z'_t \beta_0 + x'_t \gamma_0 \mid w_{st}) = \Pr(Y_{it} = 0 \mid w_{st})$ when $z'_t \beta_0 + x'_t \gamma_0 < a$.

Taking expectation over the endogenous covariate X_{ist} yields the upper bound for the distribution $v_{it} \mid z_{st}$:

$$\begin{aligned}
\Pr(v_{it} \geq a \mid z_{st}) &\leq \Pr(Y_{it} \leq z'_t \beta_0 + X'_{it} \gamma_0 - a, z'_t \beta_0 + X'_{it} \gamma_0 \geq a \mid z_{st}) + \\
&\quad \Pr(Y_{it} = 0, z'_t \beta_0 + X'_{it} \gamma_0 < a \mid z_{st}).
\end{aligned}$$

The above upper bound can be also expressed as

$$\begin{aligned}
&\Pr(Y_{it} \leq z'_t \beta_0 + X'_{it} \gamma_0 - a, z'_t \beta_0 + X'_{it} \gamma_0 \geq a \mid z_{st}) + \Pr(Y_{it} = 0, z'_t \beta_0 + X'_{it} \gamma_0 < a \mid z_{st}) \\
&= \Pr(0 < Y_{it} \leq z'_t \beta_0 + X'_{it} \gamma_0 - a, z'_t \beta_0 + X'_{it} \gamma_0 \geq a \mid z_{st}) + \Pr(Y_{it} = 0 \mid z_{st}) \\
&= \Pr(0 < Y_{it} \leq z'_t \beta_0 + X'_{it} \gamma_0 - a \mid z_{st}) + \Pr(Y_{it} = 0 \mid z_{st}).
\end{aligned}$$

Similarly, the conditional distribution $v_{it} \mid w_{st}$ can be bounded below

$$\Pr(v_{it} \geq a \mid w_{st}) \geq \Pr(Y_{it} \leq z'_t \beta_0 + x'_t \gamma_0 - a),$$

where the above condition holds since either $z'_t \beta_0 + x'_t \gamma_0 - a \geq 0$ so that there exists $y = z'_t \beta_0 + x'_t \gamma_0 - a \geq 0$ such that $\Pr(Y_{it} \leq y \mid w_{st}) = \Pr(v_{it} \geq a \mid w_{st})$, or the lower bound is zero when $z'_t \beta_0 + x'_t \gamma_0 < a$.

Taking expectation over X_{ist} leads to the following lower bound:

$$\Pr(v_{it} \geq a \mid z_{st}) \geq \Pr(Y_{it} \leq z'_t \beta_0 + X'_{it} \gamma_0 - a \mid z_{st}).$$

The conditional stationarity condition requires $v_{is} \mid Z_{ist} \stackrel{d}{\sim} v_{it} \mid Z_{ist}$, which implies that the bounds for $v_{it} \mid Z_{ist}$ must have intersections over any pair of periods (s, t) . This restriction generates the following identifying condition for θ_0 :

$$\Pr(0 < Y_{is} \leq z'_s \beta + X'_{is} \gamma - a \mid z_{st}) + \Pr(Y_{is} = 0 \mid z_{st}) \geq \Pr(Y_{it} \leq z'_t \beta_0 + X'_{it} \gamma_0 - a \mid z_{st})$$

for any $a \in \mathcal{R}$, any s, t , and any z_{st} .

□

A.10 Dynamic Censored Models with Latent Dependent Variables

Consider the following dynamic censored models with the latent lagged outcome $Y_{i,t-1}^*$:

$$\begin{aligned} Y_{it}^* &= Z'_{it} \beta_0 + Y_{i,t-1}^* \gamma_0 + \alpha_i + \epsilon_{it}, \\ Y_{it} &= \max\{Y_{it}^*, 0\}, \end{aligned}$$

In this model, the endogenous variable X_{it} is the lagged outcome: $X_{it} = Y_{i,t-1}^*$. However, the variable $Y_{i,t-1}^*$ is not observed in data, so the results in Proposition 8 cannot be directly applied here. Due to this feature in the dynamic model, we need to adjust the results in Proposition 8. Given that $Y_{i,t-1}^* = Y_{i,t-1}$ when $Y_{i,t-1} > 0$, we can further relax the lower and upper bounds in (8) to identify θ_0 .

The lower bound in condition (8) can be bounded below as follows:

$$\begin{aligned} \Pr(Y_{it} \leq z'_t \beta + Y_{i,t-1}^* \gamma - a \mid z_{st}) \\ \geq \Pr(Y_{it} \leq z'_t \beta + Y_{i,t-1} \gamma - a, Y_{i,t-1} > 0 \mid z_{st}) := L_{t,cen}(a \mid z_{st}; \theta). \end{aligned}$$

Similarly, the upper bound in condition (8) can be further bounded above

$$\Pr(0 < Y_{is} \leq z'_s \beta + Y_{i,s-1}^* \gamma - a \mid z_{st}) + \Pr(Y_{is} = 0 \mid z_{st}) \leq U_{s,cen}(a \mid z_{st}; \theta),$$

where $U_{s,cen}(a \mid z_{st}; \theta)$ is defined as

$$\begin{aligned} U_{s,cen}(a \mid z_{st}; \theta) := \Pr(0 < Y_{is} \leq z'_s \beta + Y_{i,s-1} \gamma - a, Y_{i,s-1} > 0 \mid z_{st}) \\ + \Pr(Y_{is} > 0, Y_{i,s-1} = 0 \mid z_{st}) + \Pr(Y_{is} = 0 \mid z_{st}). \end{aligned}$$

For the dynamic model, an identified set for θ_0 is characterized by the following lemma:

Lemma 4. *Under Assumption 1 and $X_{it} = Y_{i,t-1}^*$, the identified set for θ_0 is characterized*

by the set of parameters θ that satisfy the following conditions:

$$U_{s,cen}(a \mid z_{st}; \theta) \geq L_{t,cen}(a \mid z_{st}; \theta).$$

for any $a \in \mathcal{R}$, $s, t \leq T$, and z_{st} .

References

- ANDREWS, D. W. AND X. SHI (2013): “Inference based on conditional moment inequalities,” *Econometrica*, 81, 609–666.
- ARISTODEMOU, E. (2021): “Semiparametric identification in panel data discrete response models,” *Journal of Econometrics*, 220, 253–271.
- BONHOMME, S., K. DANO, AND B. S. GRAHAM (2023): “Identification in a Binary Choice Panel Data Model with a Predetermined Covariate,” Tech. rep., National Bureau of Economic Research.
- BOTOSARU, I. AND C. MURIS (2022): “Identification of time-varying counterfactual parameters in nonlinear panel models,” *arXiv preprint arXiv:2212.09193*.
- CHAMBERLAIN, G. (1980): “Analysis of covariance with qualitative data,” *The review of economic studies*, 47, 225–238.
- (1984): “Panel data,” *Handbook of econometrics*, 2, 1247–1318.
- (2010): “Binary response models for panel data: Identification and information,” *Econometrica*, 78, 159–168.
- CHEN, L.-Y. AND S. LEE (2019): “Breaking the curse of dimensionality in conditional moment inequalities for discrete choice models,” *Journal of Econometrics*, 210, 482–497.
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): “Estimation and confidence regions for parameter sets in econometric models,” *Econometrica*, 75, 1243–1284.
- CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): “Intersection bounds: estimation and inference,” *Econometrica*, 81, 667–737.
- CHESHER, A., A. ROSEN, AND Y. ZHANG (2023): “Identification analysis in models with unrestricted latent variables: Fixed effects and initial conditions,” Tech. rep., Institute for Fiscal Studies.

- CHESHER, A. AND A. M. ROSEN (2017): “Generalized instrumental variable models,” *Econometrica*, 85, 959–989.
- (2020): “Generalized instrumental variable models, methods, and applications,” in *Handbook of Econometrics*, Elsevier, vol. 7, 1–110.
- DANO, K. (2023): “Transition Probabilities and Moment Restrictions in Dynamic Fixed Effects Logit Models,” *Working Paper*.
- DAVEZIES, L., X. D’HAULTFOEUILLE, AND L. LAAGE (2021): “Identification and estimation of average marginal effects in fixed effects logit models,” *arXiv preprint arXiv:2105.00879*.
- DOBRONYI, C., J. GU, AND K. I. KIM (2021): “Identification of Dynamic Panel Logit Models with Fixed Effects,” *Working Paper*.
- GAO, W. Y. AND M. LI (2020): “Robust semiparametric estimation in panel multinomial choice models,” *Available at SSRN 3282293*.
- HONORÉ, B. E. AND E. KYRIAZIDOU (2000): “Panel data discrete choice models with lagged dependent variables,” *Econometrica*, 68, 839–874.
- HONORÉ, B. E., C. MURIS, AND M. WEIDNER (2021): “Dynamic ordered panel logit models,” *arXiv preprint arXiv:2107.03253*.
- HONORÉ, B. E. AND E. TAMER (2006): “Bounds on parameters in panel dynamic discrete choice models,” *Econometrica*, 74, 611–629.
- HONORÉ, B. E. AND M. WEIDNER (2020): “Moment conditions for dynamic panel logit models with fixed effects,” *arXiv preprint arXiv:2005.05942*.
- 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.
- MANSKI, C. F. (1987): “Semiparametric analysis of random effects linear models from binary panel data,” *Econometrica: Journal of the Econometric Society*, 357–362.

- MBAKOP, E. (2023): “Identification in Some Discrete Choice Models: A Computational Approach,” *arXiv preprint arXiv:2305.15691*.
- PAKES, A. AND J. PORTER (2022): “Moment Inequalities for Multinomial Choice with Fixed Effects,” *forthcoming in Quantitative Economics*.
- SHI, X., M. SHUM, AND W. SONG (2018): “Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity,” *Econometrica*, 86, 737–761.
- SHIU, J.-L. AND Y. HU (2013): “Identification and estimation of nonlinear dynamic panel data models with unobserved covariates,” *Journal of Econometrics*, 175, 116–131.
- WANG, R. (2022): “Semiparametric Identification and Estimation of Substitution Patterns,” *Available at SSRN 4157978*.