Identification of Dynamic Nonlinear Panel Models under Partial Stationarity*

Wayne Yuan Gao† and Rui Wang‡

April 26, 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, Xiaohong Chen, Jiaying Gu, Robert de Jong, Shakeeb Khan, Kyoo il Kim, Louise Laage, Xiao Lin, Eric Mbakop, Adam Rosen, Liyang Sun, Elie Tamer, Valentin Verdier 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 provides a unified identification approach for a wide range of panel data models with limited dependent variables, including various discrete (binary, multinomial, and ordered) choice models and censored dependent variable models. In particular, our approach accommodates dynamic models with any number of lagged dependent variables as well as contemporarily 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. Section 3 generalizes the approach to other limited dependent variable models. Specifically, consider

$$Y_{it} = \mathbb{1}\{Z_{it}'\beta_0 + X_{it}'\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\}, \tag{1}$$

where $Y_{it} \in \{0, 1\}$ denotes a binary outcome variable for individual $i = 1, 2, ...$ and time $t = 1, ..., T$, 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 $\theta_0 := \left(\beta_0', \gamma_0'\right)'$ using a panel of observed variables $(Z_i, X_i, Y_i)_{i=1}^{n}$, where $Z_i := (Z_{i1}, ..., Z_{iT})$, and simiarly X_i, Y_i are vector representations of X_{it}, Y_{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:

$$\epsilon_{is} \sim \epsilon_{it} \mid \alpha_i, Z_i, X_i \quad \forall s, t = 1, ..., T \tag{2}$$

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).[1] 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_i, X_i). However, condition (2) does impose substantial restriction on the dependence between (Z_i, X_i) and the time-varying error term ϵ_{it}: it effectively requires that all covariates

[1] To be precise, condition (2) is often stated in the following weaker "pairwise" version in the literature,

$$\epsilon_{is} \sim \epsilon_{it} \mid \alpha_i, Z_{is}, Z_{it}, X_{is}, X_{it}, \quad \forall s, t = 1, ..., T,$$

where only covariate realizations from the two periods (s, t) are conditioned on. However, the difference between condition (2) and the pairwise version above usually only leads to minor adaption of the results in the aforementioned papers (as well as in the current one). See Remark 4 for a follow-up discussion.

in (Z_i, X_i) are exogeneous with respect to the time varying error ϵ_{it}.[2]

In many economic applications, certain components of the observable covariates, namely X_i, 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 $\epsilon_{i,t-1}$ (and all $\epsilon_{i,s}$ with $s \geq t$) arises immediately. For another example, if X_{it} includes "price" or other variables that may be endogenously chosen by economic agents after observing ϵ_{it}, then X_{it} would be correlated with contemporary ϵ_{it}, so the exogeneity restriction imposed by condition (2) will again fail to hold.

In this paper, we instead impose and exploit the following "partial stationarity" condition, which can be viewed as a weaker verision of the condition (2) above:

$$\epsilon_{is} \sim \epsilon_{it} \mid \alpha_i, Z_i, \quad \forall s, t = 1, ..., T. \tag{3}$$

Our partial stationary condition (3), as its name suggests, only requires that the errors are stationary conditional on the realizations of a subvector of the covariates (i.e., the exogenous covariates denoted by Z_i) while allowing the remaining covariates (denoted by X_i) to be endogenous in arbitrary manners.[3] 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.[4]

We describe how to exploit the partial stationarity condition (3) to derive the identified set on the model parameters θ_0 through a class of conditional moment inequalities, which take the form of lower and upper bounds for the conditional distribution $\epsilon_{it} + \alpha_i \mid Z_i$, 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 a class of identifying restrictions for the parameter θ_0. Conditional on the exogenous covariates Z_i, our class of inequalities is indexed by a scalar $c \in \mathcal{R}$, which implicitly traces out all possible values that the parametrix index $Z_{it}'\beta_0 + X_{it}'\gamma_0$ can take. That said, we show how the *effective* number of identifying restrictions can be reduced to be finite when X_{it} has finite support, a condition naturally satisfied in the important special case of "p-th order autorgressive" dynamic binary choice models, where X_{it} consists of lagged outcome variables $Y_{i,t-1}, Y_{i,t-2}, ..., Y_{i,t-p}$ that are by construction discrete.

We demonstrate the sharpness of the identified set we derived for binary choice models. More precisely, we show that, for any θ that satisfies all the conditional moment inequalities

[2] For instance, suppose $\mathbb{E}[\epsilon_{it} | Z_{is}, Z_{it}, X_{is}, X_{it}] = X_{it}'\eta$, 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.

[3] Our identification strategy and results can be easily adapted under the alternative "pairwise parital stationarity" condition $\epsilon_{is} \sim \epsilon_{it} \mid \alpha_i, Z_{is}, Z_{it}$. See Remarks 4 and 5 for follow-up discussions.

[4] Condition (3) also accommodates the standard stationarity assumption conditional on all covariates.

we derived, we can construct an observationally equivalent joint distribution of the observed and unobserved variables in our model. Our proof of sharpness consists of three main lemmas in a chain: essentially, we begin demonstrating "per-period" sharpness under discreteness of X_{it} and then progressively generalize the result from "per-period" to "all-period" sharpness, and from discrete X_{it} to general X_{it}. A key innovation in our proof technique is using an explicit, simple and general construction that shows how marginal/aggregate stationarity restrictions and joint choice probability restrictions can be satisfied simultaneously, which might be of independent and wider use. While our main result is about set identification, we also provide sufficient conditions for the point identification of the coefficients on exogenous covariates (under scale normalization) as well as the signs of the coefficients on endogenous covariates.

Our identification strategy based on partial stationarity applies more broadly beyond the context of dynamic binary choice models. In Section 3, we demonstrate its applicability in a general nonseparable semiparametric model under monotonicity, and shows how it can be applied to a range of alternative 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.

While extensive work exists on panel discrete choice or other nonlinear panel data models under stationarity-type conditions, 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 the specific forms of endogeneity (lagged/contemporary endogenous regressors).

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, Lee, and Rosen (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

$$Y_{it} = \mathbb{1}\{Z_{it}'\beta_0 + Y_{i,t-1}\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\}, \tag{4}$$

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 exactly imposes the "partial stationarity" condition (3) 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. $Y_{i,t-1}, Y_{i,t-2}$, ...) 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.[5] More differently, some other papers work with sequential exogeneity in various dynamic panel models 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-varing errors. Honoré, Muris, and Weidner (2021) extends the approach in Honoré and Weidner (2020) to dynamic ordered logit mod-

[5] Our 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. 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 panel binary choice models with endogenous covariates. Sections 3 demonstrates how our key identification strategy generalizes to a wide range of dynamic nonlinear panel data models, such as ordered response models, multinomial choice models, and censored outcome models. Section 4 presents simulation results about the finite-sample performances of our approach and Section 5 explores the empirical application of income categories using various ordered response models. We conclude with Section 6.

2 Binary Choice Model

2.1 Model

We start with an illustration of our general identification strategy in the binary choice setting. Specifically, consider the following binary choice model:

$$Y_{it} = \mathbb{1}\left\{Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\right\}, \tag{5}$$

where a panel of data (Y_{it}, Z_{it}, X_{it}) are observed across individuals $i = 1, ..., n$ and time periods $t = 1, ..., T$. Here Y_{it} denotes the binary outcome variable, Z_{it} and X_{it} denote two types of observed covariates (the difference between Z_{it} and X_{it} will be clarified below), α_i 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 unknown parameter $\theta_0 := \left(\beta_0^{'}, \gamma_0^{'}\right)^{'}$, with $\beta_0 \in \mathcal{R}^{d_z}$ and $\gamma_0 \in \mathcal{R}^{d_x}$. We focus on the short panel setting, where the number of individuals n is considered to be larg while the number of time periods $T \geq 2$ is fixed and finite.

Write $Z_i := (Z_{i1}, ..., Z_{iT})$ and $X_i := (X_{i1}, ..., X_{iT})$. Throughout this paper, we will refer to Z_i as "exogenous covariates", and refer to X_i as "endogenous covariates". The exact difference between Z_i and X_i, as well as the precise meaning of "exogeneity" versus "endogeneity" in our context, are formalized through the following "*partial stationarity*" assumption: the word "partial" emphasizes that this assumption is imposed on only a part of the observed covariates Z_i, but not on X_i.

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

$$\epsilon_{it} \mid Z_i, \alpha_i \overset{d}{\sim} \epsilon_{is} \mid Z_i, \alpha_i \quad \forall t, s = 1, ..., T.$$

Assumption 1 essentially requires that the (conditional) distribution of ϵ_{it} stays the same across all time periods $t = 1, ..., T$ even if Z_i realize to different values. To illustrate, suppose that there are only two periods $t = 1, 2$, and that Z_{i1}, Z_{i2} realize to two values z_1, z_2, respectively, with $z_1 < z_2$. Then Assumption 1 requires that ϵ_{i1} and ϵ_{i2} still have the same (conditional) distributions: in particular, ϵ_{i1} cannot be stochastically smaller (or larger) than ϵ_{i2} because of $z_1 < z_2$. Hence, Assumption 1 can be thought as a definition of the "exogeneity" of the covariates Z_{it} in our context.

In contrast, Assumption 1 imposes no such restrictions on the (potentially) endogenous covariates X_i. In fact, since X_i does not appear in Assumption 1 at all, here we are completely agnostic about the dependence structure between $\epsilon_{i1}, ..., \epsilon_{iT}$ and X_i: in particular, the conditional distribution of ϵ_{it} is allowed to vary across t arbitrarily for any particular realization of X_i. As a result, different forms of endogeneity in X_i can be incorporated under our framework in a unified manner, as we illustrate in the examples below.

Example 1 (Dynamic Effects via Lagged Outcomes). Consider the following "AR(1)-type" dynamic binary choice model studied in Khan, Ponomareva, and Tamer (2023, KPT thereafter):

$$Y_{it} = \mathbb{1}\left\{Z_{it}'\beta_0 + Y_{i,t-1}\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\right\},$$

which is a special case of our model with X_{it} set to be the one-period lagged binary outcome variable $Y_{i,t-1}$. Here, X_{it} is endogenous since $X_{it} \equiv Y_{i,t-1}$ and $\epsilon_{i,t-1}$ is by construction positively correlated with $Y_{i,t-1}$ for any t, and thus the distribution of ϵ_{it} cannot be stationary across t when conditioned on the realizations of $Y_{i0}, ..., Y_{i,T-1}$. For example, given $Y_{i0} = Y_{i1} = 1$, $Y_{i2} = 0$ (and Z_i, α_i), the conditional distribution of ϵ_{i1} will naturally be different from that of ϵ_{i2}. To obtain identification under the endogeneity of $Y_{i,t-1}$, KPT imposes the stationarity of ϵ_{it} conditional on the exogenous covariates Z_i only, which coincides with our "partial stationarity" condition (Assumption 1) when specialized to their setting.

A natural generalization of the AR(1) model above in KPT is the following "AR(p)" model, which is again a special case of our model with X_{it} taken to be the vector of p lagged outcomes $Y_{i,t-1}, ..., Y_{i,t-p}$:

$$Y_{it} = \mathbb{1}\left\{Z_{it}'\beta_0 + \sum_{j=1}^{p} Y_{i,t-j}\gamma_j + \alpha_i + \epsilon_{it} \geq 0\right\}.$$

Similarly, X_{it} is endogenous here due to dependence on $\epsilon_{i,t-1}, ..., \epsilon_{t-p}$, which can again be handled in our framework under the "partial stationarity" assumption. While it is not clear how the identification results in KPT can be easily generalized to the AR(p) model above, we show in the next subsection how our identification strategy provides a simple and unified approach to derive moment inequalities regardless of the exact specifications of X_{it}.

Example 2 (Contemporarily Endogenous Covariates)**.** Alternatively, consider the following binary choice model with contemporarily endogenous covariates:

$$Y_{it} = \mathbb{1}\left\{Z_{it}'\beta_0 + X_{it}'\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\right\},$$
$$X_{it} = \phi\left(Z_{it}, u_{it}\right)$$

where ϕ is an unknown "first-stage" function and the "first-stage error" u_{it} is allowed to be arbitrarily correlated with ϵ_{it}. For example, X_{it} may be a "price-type" variable that is strategically chosen by a decision maker after observing the current-period error ϵ_{it}, which generates contemporary dependence between X_{it} and ϵ_{it}. Even though contemporary endogeneity of this type is very different in nature from the dynamic endogeneity discussed in the previous example, it also induces non-stationarity of ϵ_{it} when conditioned on X_i: for example, if X_{it} and ϵ_{it} are positively correlated, then, conditional on $X_{i1} < X_{i2}$, it is unreasonable to assume the distribution of ϵ_{i1} is the same as ϵ_{i2}. That said, such type of contemporary endogeneity can also be handled in our framework under the "partial stationarity" condition (Assumption 1).

Remark 1 (Combination of Dynamic and Contemporary Endogeneity)**.** *We separately discussed two types of endogenous covariates, dynamic covariates (lagged outcome variables) and contemporarily endogenous covariates, in the two examples above, but our identification strategy also applies if both types of endogenous covariates are present together, since our identification strategy works generally under "partial stationarity", which does not impose or exploit any restrictions on the form of endogeneity between ϵ_{it} and X_i.*

Remark 2 (Full Stationarity as Special Case)**.** *Obviously, the standard "full stationarity" condition* (2) *is nested under "partial stationarity" condition (Assumption 1) as a special case, where the endogenous covariate X_{it} contains no variables. Hence, "full stationarity" is in general stronger than "partial stationarity".*

Remark 3 (Focus on Time-Varying Endogeneity). *Technically, our partial stationarity condition also allows some endogeneity between ϵ_{it} and Z_i, as long as such endogeneity is time-invariant. This is because Assumption 1 is stated conditional on the full vector $Z_i = (Z_{i1}, ..., Z_{iT})$ and the time-invariant fixed effect α_i. Hence, as long as the conditional distribution of ϵ_{it} depends on $Z_{i1}, ..., Z_{iT}$ and α_i in a time-invariant manner, the stationarity of ϵ_{it} can still hold. That said, since in empirical applications we are mostly interested in "time-varying endogeneity", such as the dynamic and contemporary endogeneity discussed in the examples above, in this paper we refer to Z_i as "exogenous" even though it may be endogenous in a time-invariant manner, and only call X_i, which features time-varying endogeneity, the "endogenous" covariates.*

Remark 4 (Pairwise Version of Partial Stationarity). *In Assumption 1, we impose partial stationarity of ϵ_{it} conditional on Z_{it} from all periods $t = 1, ..., T$. Alternatively, we could impose partial stationarity in a "pairwise" version, conditional on (Z_{it}, Z_{is}) from any pair of time periods (t, s) only:*

$$\textit{Pairwise Partial Stationarity:} \quad \epsilon_{it} \mid Z_{it}, Z_{is}, \alpha_i \overset{d}{\sim} \epsilon_{is} \mid Z_{it}, Z_{is}, \alpha_i, \quad \forall t, s = 1, ..., T. \tag{6}$$

Clearly, the "pairwise" version is equivalent to the "all-periods" version when $T = 2$, but is weaker when $T \geq 3$. Our identification strategy applies under both versions of partial stationarity, though the identification results and the corresponding proofs have slightly different representations. Essentially, conditioning on all-period covariate realizations would be replaced with conditioning the realizations in any specific pair of period. See Remark 5 at the end of Section 2.2 for a follow-up discussion.

2.2 Key Identification Strategy

We now explain our key identification strategy based on partial stationarity.

Write $v_{it} := -(\epsilon_{it} + \alpha_i)$ so that model (5) can be rewritten as

$$Y_{it} = \mathbb{1}\left\{v_{it} \leq Z_{it}'\beta_0 + X_{it}'\gamma_0\right\}.$$

For any constant $c \in \mathcal{R}$, consider first the event

$$Y_{it} = 1 \text{ and } Z_{it}'\beta_0 + X_{it}'\gamma_0 \leq c.$$

Whenever the event above happens, we must have $v_{it} \leq Z_{it}'\beta_0 + X_{it}'\gamma_0 \leq c$, implying that $v_{it} \leq c$. Formally, the above can be summarized by the following inequality:

$$\begin{aligned} Y_{it}\mathbb{1}\left\{Z_{it}'\beta_0 + X_{it}'\gamma_0 \leq c\right\} &= \mathbb{1}\left\{v_{it} \leq Z_{it}'\beta_0 + X_{it}'\gamma_0\right\}\mathbb{1}\left\{Z_{it}'\beta_0 + X_{it}'\gamma_0 \leq c\right\} \\ &\leq \mathbb{1}\left\{v_{it} \leq c\right\}. \end{aligned} \tag{7}$$

Symmetrically, we can also consider the "flipped" event

$$Y_{it} = 0 \text{ and } Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \geq c,$$

which implies $v_{it} > c$, and similarly

$$\begin{aligned}(1 - Y_{it})\,\mathbb{1}\left\{Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \geq c\right\} &= \mathbb{1}\left\{v_{it} > Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0\right\}\mathbb{1}\left\{Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \geq c\right\} \\ &\leq \mathbb{1}\left\{v_{it} > c\right\} \equiv 1 - \mathbb{1}\left\{v_{it} \leq c\right\}\end{aligned}$$

which is equivalent to

$$\mathbb{1}\left\{v_{it} \leq c\right\} \leq 1 - (1 - Y_{it})\,\mathbb{1}\left\{Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \geq c\right\}. \tag{8}$$

Taking conditional expectations of (7) and (8) given $Z_i = z \equiv (z_t)_{t=1}^{T}$,

$$\begin{aligned}&\mathbb{P}\left(Y_{it} = 1,\ z_t^{'}\beta_0 + X_{it}^{'}\gamma_0 \leq c \middle| Z_i = z\right) \\ &\leq \mathbb{P}\left(v_{it} \leq c \middle| Z_i = z\right) \\ &= \mathbb{P}\left(v_{is} \leq c \middle| Z_i = z\right) \\ &\leq 1 - \mathbb{P}\left(Y_{is} = 0,\ z_s^{'}\beta_0 + X_{is}^{'}\gamma_0 \geq c \middle| Z_i = z\right),\end{aligned} \tag{9}$$

where the middle equality follows from the partial stationarity condition (Assumption 1). Specifically, observe that Assumption 1 implies the partial stationarity of v_{it} given Z_i, i.e.,

$$v_{it} \mid Z_i \overset{d}{\sim} v_{is} \mid Z_i,$$

so that $\mathbb{P}\left(v_{it} \leq c \middle| Z_i = z\right) = \mathbb{P}\left(v_{is} \leq c \middle| Z_i = z\right)$ for any $c \in \mathcal{R}$.

Essentially, in the above we exploit the joint occurrence of $v_{it} \leq Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0$ and $Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \leq c$ to deduce an implication on the composite error $v_{it} \leq c$ that is free of the endogenous covariates X_{it}, and then leverage the partial stationarity of v_{it} given Z_i for intertemporal comparisons.

Since the lower and upper bounds in (9) hold for any t and s, we summarize the identifying restrictions (9) across all time periods in the following proposition.

Proposition 1 (Identified Set)**.** *Under model* (5) *and Assumption 1,* $\theta_0 \in \Theta_I$*, where the identified set* Θ_I *consists of all* $\theta = \left(\beta^{'}, \gamma^{'}\right)^{'} \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_x}$ *such that*

$$\max_{t=1,\ldots,T} \mathbb{P}\left(Y_{it} = 1,\ z_t^{'}\beta + X_{it}^{'}\gamma \leq c \middle| Z_i = z\right) \leq 1 - \max_{s=1,\ldots,T} \mathbb{P}\left(Y_{is} = 0,\ z_s^{'}\beta + X_{is}^{'}\gamma \geq c \middle| Z_i = z\right) \tag{10}$$

for any $c \in \mathcal{R}$ *and any realization* $z = (z_1, \ldots, z_T)$ *in the support of* Z_i*.*

Proposition 1 characterizes the identified set Θ_I for θ_0 as restrictions on the conditional joint distribution of Y_{it} and X_{it} given $Z_i = z$. More specifically, the restrictions in (10) can be regarded as a collection of conditional moment inequalities that relate $\mathbb{1}\left\{Y_{it}=1,\ z_t'\beta+X_{it}'\gamma\le c\right\}$ and $\mathbb{1}\left\{Y_{it}=1,\ z_t'\beta+X_{it}'\gamma\le c\right\}$ conditional on $Z_i=z$.

Proposition 1 holds regardless of whether the endogenous covariates X_{it} are discrete or continuous. When X_{it} are continuous (taking a continuum of values), then Proposition 1 requires that condition (10) hold for a continuum of constants $c\in\mathcal{R}$, so that (the information in) the whole joint distribution of the binary variable Y_{it} and the continuous variable $z_t'\beta+X_{it}'\gamma$ can be captured by the collection of joint distributions of $\left(Y_{it},\mathbb{1}\left\{z_t'\beta+X_{it}'\gamma\le c\right\}\right)$ across all possible cutoff values c.

However, when X_{it} are discrete, such as in the AR(p) dynamic model where X_{it} consists of p lagged binary outcome variables, there is no need to evaluate (10) at all possible values of $c\in\mathcal{R}$, since the inequalities in (10) can only bind at finitely many values of c. We formalize this observation via the following Proposition.

Proposition 2 (Identified Set with Discrete Endogenous Covariates)**.** *Suppose that the endogenous covariate X_{it} can only take finite number of values in $\{\overline{x}_1,...,\overline{x}_K\}$ across all time periods $t=1,...,T$, then $\Theta_I=\Theta_I^{disc}$, where Θ_I^{disc} consists of all $\theta=\left(\beta',\gamma'\right)'\in\mathcal{R}^{d_z}\times\mathcal{R}^{d_x}$ that satisfy condition* (10) *for any*

$$c\in\left\{z_t'\beta+\overline{x}_k'\gamma:k=1,...,K,t=1,...,T\right\},\tag{11}$$

for any $z=(z_1,...,z_T)$ in the support of Z_i.

Proposition 2 shows that the discreteness of the endogenous covariates X_{it} help reduce the infinite number of inequality restrictions in Proposition 1 to finitely many, or more precisely, KT ones (conditional on z).

The case of discrete X_{it} is conceptually important, since it nests the dynamic AR(p) model widely studied in the literature as a special case. Clearly, when X_{it} consists of p (finitely many) lagged binary outcome variables $Y_{i,t-1},...,Y_{i,t-p}$, then X_{it} by construction can only take $K=2^p$ discrete values. Specialized further to the AR(1) model in KPT, Proposition 2 shows that the identified set Θ_I is characterized by $2T$ conditional restrictions, which is drastically smaller than the $9T\left(T-1\right)$ conditional restrictions listed out in KPT (even when T is small).

Remark 5. *Following up on Remark 4, if pairwise partial stationarity is adopted, then Propositions 1 and 2 continue to hold with* (10) *adapted to the following "pairwise" version:*

$$\mathbb{P}\left(Y_{it}=1,\ z_t'\beta+X_{it}'\gamma\le c\,\middle|\,Z_{i,ts}=z_{ts}\right)\le 1-\mathbb{P}\left(Y_{is}=0,\ z_s'\beta+X_{is}'\gamma\ge c\,\middle|\,Z_{i,ts}=z_{ts}\right),\forall\,(t,s),\tag{12}$$

where $Z_{i,ts}:=(Z_{it},Z_{is})$ and $z_{ts}:=(z_t,z_s)$. Relative to (10)*, the statement in* (12) *reflects the fact that pairwise partial stationarity is imposed on all pairs of time periods separately instead* *of all T time periods jointly. It is straightforward to verify that the identification arguments above, in particular* (7)-(9)*, carry over with all conditional probabilities/expectations taken conditional on $Z_{i,ts} = z_{ts}$ instead of $Z_i = z$.*

2.3 Sharpness of Identified Set

So far we have only shown that Θ_I, which equals to Θ_I^{disc} under discreteness of X_{it}, is a valid identified set for θ_0. However, it is not yet clear whether it has incorporated all the available information for θ_0. The following theorem establishes the sharpness of the identified set Θ_I.

Theorem 1 (Sharpness)**.** *Under model* (5) *and Assumption 1, the identified set Θ_I is sharp.*

The formal definition of sharpness, along with the complete proof of Theorem 1, are available in Appendix A.2. In short, we show (by construction) that, for each $\theta \in \Theta_I \setminus \{\theta_0\}$, there exists a data generating process (DGP) that satisfies Assumption 1 and produces the same joint distribution of observable data (Y_i, X_i, Z_i) under model (5) with parameter θ. Theorem 1 demonstrates that our key identification strategy based on the bounding of (endogenous) parametrix index by arbitrary constants, as described in Section 2.2, is able to extract all the available information for θ_0 from the model and the observable data, and thus it is impossible to further differentiate θ_0 from alternatives in the identified set Θ_I under model (5) and our assumption of partial stationarity (without further restrictions).

Theorem 1 immediately implies that, in the special case of dynamic AR(p) models where X_{it} consists of discrete lagged outcomes, our characterization of the identified set Θ_I^{disc} in Proposition 2 is sharp. In particular, our result generalizes the corresponding result in KPT, which focuses on the AR(1) model. Furthermore, KPT characterizes the sharp identified set via $9T(T-1)$ conditional restrictions, the derivation of which is based on an exhaustive enumeration of lagged outcome realizations $Y_{i,t-1}$. In this paper we adopt an entirely different (and much more general) identification strategy, and arrive at a characterization of the identified set by $2T$ conditional restrictions, which we also show to be sharp by Theorem 1. Since our model and assumption specialize exactly to that in KPT under the AR(1) specification, it follows that our $2T$ restrictions must be able to reproduce all the $9T(T-1)$ restrictions in KPT. This demonstrates that our identification strategy not only applies more generally than the one in KPT, but also leads to a more elegant characterization of the sharp identified set with much fewer restrictions. We provide a more detailed explanation about this point in the next subsection.

Another conceptually remarkable, or surprising, feature of Proposition 2 and Theorem 1 is that they are established without reference to the exact nature, or interpretation, of the endogenous covariates X_{it}. The identified set Θ_I (and $\Theta_{I,disc}$ under discreteness of X_{it}) we characterized is valid and sharp regardless of whether X_{it} are specified as lagged outcome variables, contemporarily endogenous covariates, or a combination of the both.

Our proof of sharpness consists of three main lemmas. First, we start with the simple case where the endogenous covariates X_{it} only take finitely many values (referred to as the "discrete case" therafter), and show for each $\theta \in \Theta_I \backslash \{\theta_0\}$ how to construct the per-period marginal distributions of errors that match the per-period marginal choice probabilities. Second, we show (in the discrete case) how to combine the T per-period marginal distributions into an all-period joint distribution that matches the all-period joint choice probabilities, so that observational equivalence holds. Lastly, we show that how the sharpness in the discrete case generalizes to the continuou case, by taking appropriate limits of an increasing (finite) set of discretized points in the potentially continuous (or mixed) support of X_{it} that becomes dense in the limit.

The proof techniques we exploited are also different from, and thus novel relative to, those used in the related work that leverages stationarity-type conditions for partial identification, such as Pakes and Porter (2022) for static multinomial choice model and KPT for dynamic AR(1) model. Instead of showing existence only, we provide a more explicit construction of the joint distribution of the latent variables, which is valid regardless of the exact type of endogeneity in X_{it}. In particular, a key challenge in proving sharpness based on stationarity-type conditions lies in that stationarity imposes only aggregate restrictions (via integrals/sums) of the joint distribution of errors, which is rather implicit to work with. A key innovation in our proof technique is to show how marginal/aggregate stationarity restrictions and joint choice probability restrictions can be satisfied simultaneously by an explicit, simple and general construction, which might be of independent and wider use.

2.4 Reconciliation with Related Work

Our identifying restrictions in (10) and (11) have a somewhat "nonstandard" representation in terms of (conditional) joint probabilities instead of Y_{it} and X_{it} (given Z_i), instead of conditional probabilities of Y_{it} given X_{it} (such as lagged outcomes), which are more usually found in the related literature. Hence, we provide a more detailed discussion about the content and interpretation of our identifying restrictions, as well as a more explicit explanation of how they relate to exisiting results in the related literature.

Reconciliation with Manski (1987)

Consider first the special case where there are *no* endogenous covariates X_{it}, or in other words, X_{it} is degenerate. In this case, our "partial stationarity" condition specializes to the "full stationarity" condition (2) as in Manski (1987). However, our identifying restriction (10) still has a different form than the identifying restriction in Manski (1987). To illustrate, focus on any two periods (t, s), and observe that in this case our identifying restriction takes the form of

$$\mathbb{P}\left(Y_{it} = 1,\ z_t^{'}\beta_0 \leq c \middle| z\right) \leq 1 - \mathbb{P}\left(Y_{is} = 0,\ z_s^{'}\beta_0 \geq c \middle| z\right),\ \forall c, \tag{13}$$

while the "maximum-score-type" identifying restrictions in Manski (1987) are of the form

$$z_s^{'}\beta_0 \geq z_t^{'}\beta_0 \ \Leftrightarrow\ \mathbb{P}\left(Y_{is} = 1 | z\right) \geq \mathbb{P}\left(Y_{it} = 1 | z\right). \tag{14}$$

The "maximum-score-type" identifying restriction (14) has a quite intuitive and interpretable representation: across two periods (t, s) under full stationarity, the conditional choice probability at period s is larger if and only if the index $z_s^{'}\beta_0$ is larger. In contrast, our restriction (13) has a somewhat twisted representation even in this simple setting.

However, a closer look reveals that our (13) is exactly equivalent to Manski's "maximum-score-type" identifying restrictions in the current context. To see this, notice that, by setting $c = z_t^{'}\beta_0$ in (13), we obtain

$$\begin{aligned}\mathbb{P}\left(Y_{it} = 1 | z\right) =& \mathbb{P}\left(Y_{it} = 1 | z\right) \mathbb{1}\left\{z_t^{'}\beta_0 \leq z_t^{'}\beta_0\right\} \\ \leq& 1 - \mathbb{P}\left(Y_{is} = 0 | z\right) \mathbb{1}\left\{z_s^{'}\beta_0 \geq z_t^{'}\beta_0\right\}.\end{aligned}$$

Hence, if $z_s^{'}\beta_0 \geq z_t^{'}\beta_0$, i.e., the left-hand side of (14) holds, then the above implies that

$$\mathbb{P}\left(Y_{it} = 1 | z\right) \leq 1 - \mathbb{P}\left(Y_{is} = 0 | z\right) = \mathbb{P}\left(Y_{is} = 1 | z\right),$$

which becomes exactly the right-hand side of (14). Switching t with s in the argument above produces the other implication $z_s^{'}\beta_0 \leq z_t^{'}\beta_0 \Rightarrow \mathbb{P}\left(Y_{is} = 1 | z\right) \leq \mathbb{P}\left(Y_{it} = 1 | z\right)$. Together these exactly constitute the "if-and-only-if" restriction in (14). Hence, even though our inequality restriction (13) looks different from the more intuitive "maximum-score-type" restriction, they both incorporate the same information.

Reconciliation with KPT (Khan, Ponomareva, and Tamer, 2023)

Now, consider the dynamic AR(1) model as studied in KPT, where the only endogenous covariate is the one-period lagged outcome variable, i.e., $X_{it} := Y_{i,t-1}$.

To illustrate, first focus on any two periods (t, s) only, and observe that in this case our identifying restriction becomes

$$\mathbb{P}\left(Y_{it}=1,\ z_t'\beta_0+Y_{i,t-1}\gamma_0\leq c\middle|\, z\right)\leq 1-\mathbb{P}\left(Y_{is}=0,\ z_s'\beta_0+Y_{i,t-1}\gamma_0\geq c\middle|\, z\right),\ \forall c. \tag{15}$$

Under the same model and assumption, KPT derives the following 9 inequality implications for (t,s):[6]

KPT(i): $\mathbb{P}(Y_{it}=1|\, z)>\mathbb{P}(Y_{is}=1|\, z)\Rightarrow(z_t-z_s)'\beta_0+|\gamma_0|>0$.
KPT(ii): $\mathbb{P}(Y_{it}=1|\, z)>1-\mathbb{P}(Y_{i,s}=0,Y_{i,s-1}=1|\, z)\Rightarrow(z_t-z_s)'\beta_0-\min\{0,\gamma_0\}>0$.
KPT(iii): $\mathbb{P}(Y_{it}=1|\, z)>1-\mathbb{P}(Y_{i,s}=0,Y_{i,s-1}=0|\, z)\Rightarrow(z_t-z_s)'\beta_0+\max\{0,\gamma_0\}>0$.
KPT(iv): $\mathbb{P}(Y_{it}=1,Y_{it-1}=1|\, z)>\mathbb{P}(Y_{is}=1|\, z)\Rightarrow(z_t-z_s)'\beta_0+\max\{0,\gamma_0\}>0$.
KPT(v): $\mathbb{P}(Y_{it}=1,Y_{it-1}=1|\, z)>1-\mathbb{P}(Y_{is}=0,Y_{i,s-1}=1|\, z)\Rightarrow(z_t-z_s)'\beta_0>0$.
KPT(vi): $\mathbb{P}(Y_{it}=1,Y_{it-1}=1|\, z)>1-\mathbb{P}(Y_{is}=0,Y_{i,s-1}=0|\, z)\Rightarrow(z_t-z_s)'\beta_0+\gamma_0>0$.
KPT(vii): $\mathbb{P}(Y_{it}=1,Y_{it-1}=0|\, z)>1-\mathbb{P}(Y_{is}=0|\, z)\Rightarrow(z_t-z_s)'\beta_0-\min\{0,\gamma_0\}>0$.
KPT(viii): $\mathbb{P}(Y_{it}=1,Y_{it-1}=0|\, z)>1-\mathbb{P}(Y_{is}=0,Y_{i,s-1}=1|\, z)\Rightarrow(z_t-z_s)'\beta_0-\gamma_0>0$.
KPT(ix): $\mathbb{P}(Y_{it}=1,Y_{it-1}=0|\, z)>1-\mathbb{P}(Y_{is}=0,Y_{i,s-1}=0|\, z)\Rightarrow(z_t-z_s)'\beta_0>0$.

In a way, the 9 inequality restrictions in KPT above are similar to the "maximum-score restrictions", in the sense that all of them take the form of logical implications between intertemporal comparisons of various conditional probabilities and intertemporal differences of covariate indexes.

Using a very different identification strategy than the one in KPT, we arrived at our inequality restriction (15), which looks very different from the collection of 9 inequality restrictions in KPT. At first sight it is not clear how (15) relates to and compares with the 9 KPT restrictions. However, a closer look again reveals that our restriction (15) can reproduce all the 9 restrictions in KPT, and thus incorporate all the information therein in a unified format.

Take KPT(v) as an illustration and suppose that the left-hand side of KPT(v) holds, then it implies

$$\mathbb{P}(Y_{it}=1,Y_{it-1}=1|\, z)>1-\mathbb{P}(Y_{is}=0,Y_{i,s-1}=1|\, z). \tag{16}$$

With $X_{it}=Y_{i,t-1}$, our inequality restriction (15) can be equivalently rewritten as follows,

$$\mathbb{P}(Y_{it}=1,\ Y_{i,t-1}=1|\, z)\,\mathbb{1}\left\{z_t'\beta_0+\gamma_0\leq c\right\}+\mathbb{P}(Y_{it}=1,\ Y_{i,t-1}=0|\, z)\,\mathbb{1}\left\{z_t'\beta_0\leq c\right\}$$
$$\leq 1-\mathbb{P}(Y_{is}=0,\ Y_{i,s-1}=1|\, z)\,\mathbb{1}\left\{z_s'\beta_0+\gamma_0\geq c\right\}-\mathbb{P}(Y_{is}=0,\ Y_{i,s-1}=0|\, z)\,\mathbb{1}\left\{z_s'\beta_0\geq c\right\}, \tag{17}$$

[6] We adapt the notation in KPT to our current notation, and state these 9 inequalities as strict inequalities, which lead to a simpler and more focused explanation. The equivalence between our restriction and the KPT restrictions still hold if their inequalities are stated in the weak form.

where the realization of $Y_{i,t-1}$ is explicitly enumerated as in KPT.

Note that we can further relax condition (17) by dropping the two probabilities $\mathbb{P}\left(Y_{it}=1,\ Y_{i,t-1}=0|\,z\right)\mathbb{1}\left\{z_t'\beta_0\leq c\right\}$ and $\mathbb{P}\left(Y_{is}=0,\ Y_{i,s-1}=0|\,z\right)\mathbb{1}\left\{z_s'\beta_0\geq c\right\}$ as it makes the lower bound smaller and the upper bound larger:

$$\mathbb{P}\left(Y_{it}=1,\ Y_{i,t-1}=1|\,z\right)\mathbb{1}\left\{z_t'\beta_0+\gamma_0\leq c\right\}\leq$$
$$1-\mathbb{P}\left(Y_{is}=0,\ Y_{i,s-1}=1|\,z\right)\mathbb{1}\left\{z_s'\beta_0+\gamma_0\geq c\right\}.$$

Then, the statement that $\mathbb{1}\left\{z_t'\beta_0+\gamma_0\leq c\right\}$ and $\mathbb{1}\left\{z_s'\beta_0+\gamma_0\geq c\right\}$ both holds is precisely equivalent to the following statement of

$$z_t'\beta_0\leq z_s'\beta_0\Longrightarrow\mathbb{P}\left(Y_{it}=1,\ Y_{i,t-1}=1|\,z\right)\leq\ 1-\mathbb{P}\left(Y_{is}=0,\ Y_{i,s-1}=1|\,z\right).$$

By contraposition, it leads to exactly the same implication of KPT(v):

$$\mathbb{P}\left(Y_{it}=1,\ Y_{i,t-1}=1|\,z\right)>\ 1-\mathbb{P}\left(Y_{is}=0,\ Y_{i,s-1}=1|\,z\right)\Longrightarrow z_t'\beta_0>z_s'\beta_0.$$

Hence, we have shown that (17) implies KPT(v).

Similarly, it is shown in Appendix A.3 that (17) implies all 9 restrictions in KPT. In fact, the representation (17) reveals why there are precisely 9 KPT-type restrictions. The two period-t indicators $\mathbb{1}\left\{z_t'\beta_0+\gamma_0\leq c\right\}$ and $\mathbb{1}\left\{z_t'\beta_0\leq c\right\}$ in the upper expression of (17) may take 3 "useful"[7] combinations $(1,0),(0,1)$ and $(1,1)$, while the two period-s indicators $\mathbb{1}\left\{z_s'\beta_0+\gamma_0\geq c\right\}$ and $\mathbb{1}\left\{z_s'\beta_0\geq c\right\}$ in the lower expression of (17) may also take 3 useful combinations. Consequently, in total there are $3\times3=9$ useful combinations, which exactly correspond to the 9 left-hand-side suppositions in the 9 KPT restrictions.

Hence, while our restriction (15) appears very different from the 9 KPT restrictions, it actually automatically incorporates all the KPT restrictions. In particular, by treating the endogenous covariate $X_{it}=Y_{i,t-1}$ as a random variable, our restriction (15) automatically aggregates the identifying information across all possible realizations of $Y_{i,t-1}$, without the need to explicitly consider each possibility separately.

Now, consider a general setting with $T\geq2$ periods. By our Proposition 2 and Theorem 1, the sharp identified set can be characterized by $2T$ restrictions, which are generated by evaluating (10) at each c of the $2T$ points in $\left\{z_t'\beta,z_t'\beta+\gamma:t=1,...,T\right\}$. In contrast, across T periods the KPT approach produces $9T\left(T-1\right)$ restrictions, which are generated by imposing the 9 KPT restrictions across all ordered time pairs (t,s). Hence, our approach provides a much simpler characterization of the sharp identified set, using a significantly smaller number of restrictions. For example, with $T=2$ periods, we have 4 restrictions

[7] The 4th combination, $\mathbb{1}\left\{z_t'\beta_0+\gamma_0\leq c\right\}=\mathbb{1}\left\{z_t'\beta_0+\gamma_0\leq c\right\}=0$, will make the upper expression of (17) equal to 0, so that the inequality (17) holds trivially. Hence, this $(0,0)$ combination is not useful.

while KPT has 18; with $T = 3$, we have 6 restrictions while KPT has 54. Hence, the reduction in the number of restrictions relative to KPT is quite remarkable.

In summary, while the representation of our identifying restrictions in (10) and (11) may appear somewhat unnatural in the first place, it actually becomes equivalent to more familiar (and intuitive) representations in the specialized settings of Manski (1987) and KPT.

2.5 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 $X_{it} \in \mathcal{X} \equiv \{\overline{x}_1, ..., \overline{x}_K\}$ 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:

$$\mathcal{Z}_1 := \Big\{(z_1, z_2) \mid \exists x \in \mathcal{X} \text{ s.t. } 1 - \mathbb{P}(Y_{i1} = 0, X_{i1} = x \mid z) < \mathbb{P}(Y_{i2} = 1, X_{i2} = x \mid z)\Big\},$$
$$\mathcal{Z}_2 := \Big\{(z_1, z_2) \mid \exists x \in \mathcal{X} \text{ s.t. } 1 - \mathbb{P}(Y_{i1} = 1, X_{i1} = x \mid z) < \mathbb{P}(Y_{i2} = 0, X_{i2} = x \mid z)\Big\}.$$

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

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

As shown in Appendix A.4, 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 jth 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 3. *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:

$$\mathcal{Z}_3^j := \Big\{(z_1, z_2) \mid \exists x_1, x_2 \in \mathcal{X} \text{ with } x_1^j \neq x_2^j, x_1^m = x_2^m \ \forall m \neq j \text{ s.t.}$$
$$1 - \mathbb{P}(Y_{i1} = 0, X_{i1} = x_1 \mid z) < \mathbb{P}(Y_{i2} = 1, X_{i2} = x_2 \mid z)\Big\};$$
$$\mathcal{Z}_4^j := \Big\{(z_1, z_2) \mid \exists x_1, x_2 \in \mathcal{X} \text{ with } x_1^j \neq x_2^j, x_1^m = x_2^m, \ \forall m \neq j \text{ s.t.}$$
$$1 - \mathbb{P}(Y_{i1} = 1, X_{i1} = x_1 \mid z) < \mathbb{P}(Y_{i2} = 0, X_{i2} = x_2 \mid z)\Big\}.$$

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:

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

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 4. *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 $\mathcal{Z}_3^2, \mathcal{Z}_4^2$.

2.6 Identification of Counterfactual Parameters

In previous subsections, we have focused on the (partial) identification of the index parameters θ_0. Here we show how our identification results can also be leveraged to (partially) identify counterfactual parameters.

Write $W_i := (Z_i, X_i)$ in short, and correspondingly $w := (z, x)$, and $w_t'\theta = z_t'\beta + x_t'\gamma$. Consider a general counterfactual change in the observable covariates W_i from w to $\tilde{w}$, and the consequent counterfactual period-t conditional choice probability of the form

$$\tilde{p}_t(\tilde{w}) := \mathbb{P}\left(v_{it} \leq \tilde{w}'\theta_0 \middle| W_i = w \right). \tag{18}$$

Importantly, in the definition above, the utility index is changed from $w_t'\theta_0$ to the counterfactual $\tilde{w}_t'\theta_0$, while the conditional distribution of the latent v_{it} is held unchanged at $v_{it}|\, W_i = w$. Hence, $\tilde{p}_t(\tilde{w})$ can be interpreted as a counterfactual CCP induced by an exogenous policy intervention that only changes the characteristics from w to $\tilde{w}$, but leaves all other unobserved individual heterogeneity reflected in the distribution of v_{it} unchanged. In particular, note that the (partial) derivative of $\tilde{p}_t(w)$ can be interpreted as average marginal effects.

Proposition 5 (Bounds on Counterfactual CCP)**.** *Under model 5 and Assumption 1,*

$$\inf_{\theta \in \Theta_I} \mathbb{P}\left(Y_{it} = 1,\ w_t'\theta \leq \tilde{w}_t'\theta \middle| W_i = w \right) \leq \tilde{p}_t(\tilde{w}) \leq 1 - \inf_{\theta \in \Theta_I} \mathbb{P}\left(Y_{it} = 0,\ w_t'\theta \geq \tilde{w}_t'\theta \middle| W_i = w \right). \tag{19}$$

The lower and upper bounds in Proposition 5 above are identified since the involved conditional probabilities are all about observed data (Y_i, W_i) for each $\theta \in \Theta_I$, while the set Θ_I is identified by Proposition 1. Hence, Proposition 5 establishes the partial identification of the counterfactual CCP $\tilde{p}_t(\tilde{w})$.

Note that the (partial) identification of counterfactual CCP $\tilde{p}_t(\tilde{w})$ relies on the identification of the index parameter θ_0 as well as the identification of the latent conditional distribution $v_{it}|\, W_i = w$, which also involves the endogenous covariates X_i. It turns out that, our key identification strategy in Section 2.2 also provides a straightforward way to derive bounds on $F_t(c|\, w)$, the CDF of $v_{it}|\, W_i = w$ at any point c, by taking conditional expectations of (7) and (8) given $W_i = w$ (instead of $Z_i = z$ as in Section 2.2):

$$\mathbb{P}\left(Y_{it} = 1,\ w_t'\theta_0 \leq c \middle| W_i = w \right) \leq F_t(c|\, w) \leq 1 - \mathbb{P}\left(Y_{it} = 0,\ w_t'\theta_0 \geq c \middle| W_i = w \right), \tag{20}$$

which can then be combined with Proposition 1 to derive the bounds in Proposition 5.

3 Generalization

3.1 General Identification Strategy

The key idea underlying our identification strategy generalizes further beyond the binary choice model. In this section, we study a general semiparametric model that can accommodate a wide range of panel data models. This general model allows for various types of dependent variables, including ordered, multinomial, and censored outcomes. Moreover, we explore more flexible specifications for the dependent variable, allowing for multidimensional fixed effects and time-varying errors, as well as nonseparable interactions between them.

To illustrate, consider the following nonseparable semiparametric model:

$$Y_{it} = G\left(W_{it}'\theta_0,\, \alpha_i,\, \epsilon_{it}\right), \tag{21}$$

where $Y_{it} \in \mathcal{Y}$ can be either a discrete or continuous variable, 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 3 (Monotonicity)**.** *The mapping $\delta \longmapsto G(\delta, \alpha, \epsilon)$ 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 $G\left(W_{it}'\theta_0,\, \alpha_i,\, \epsilon_{it}\right) = \mathbb{1}\left\{W_{it}'\theta_0 + \alpha_i + \epsilon_{it} \geq 0\right\}$, we obtain the binary choice model in Section 2, where G is by construction weakly increasing in $W_{it}'\theta_0$. Beyond binary choice model, since the dependent variable Y_{it} is not restricted, we show that this general model can accommodate a wide range of panel models with various types of dependent variables.

Example 3 (Ordered Response Model)**.** Consider that the dependent variable Y_{it} can take J possible ordered values: $Y_{it} \in \{y_1, ..., y_J\}$ with $y_j < y_{j+1}$, and it is generated as follows:

$$Y_{it}^* = W_{it}'\theta_0 + \alpha_i + \epsilon_{it},$$
$$Y_{it} = \sum_{j=1}^{J} y_j \mathbb{1}\{b_j < Y_{it}^* \leq b_{j+1}\},$$

where the threshold parameters satisfy $b_1 = -\infty, b_{J+1} = \infty$, and $b_j \leq b_{j+1}$.

Then the function G is given as

$$G\left(W_{it}'\theta_0,\, \alpha_i,\, \epsilon_{it}\right) = \sum_{j=1}^{J} y_j \mathbb{1}\left\{b_j < W_{it}'\theta_0 + \alpha_i + \epsilon_{it} \leq b_{j+1}\right\}$$

and it is weakly monotone in $W_{it}'\theta_0$ since $y_j < y_{j+1}$ for any $1 \leq j \leq J-1$.

Example 4 (Multinomial Choice Model)**.** Consider that the dependent variable Y_{it} takes J possible unordered choices: $Y_{it} \in \mathcal{J} = \{0, 1, ..., J\}$. The latent utility u_{ijt} for each option j and the dependent variable Y_{it} are generated as follows:

$$u_{ijt} = W_{ijt}'\theta_0 + \alpha_{ij} + \epsilon_{ijt},$$
$$Y_{it} = \arg\max_{j \in \mathcal{J}} u_{ijt}.$$

Although the J choices are unordered and cannot be directly compared, we can construct a new variable $\tilde{Y}_{it}^K = \mathbb{1}\{Y_{it} \in K\}$ for any subset $K \subset \mathcal{J}$, representing individuals' choice belonging to the set K. It is equivalent to the event that there exists one choice from the set K yielding higher utility than other choices:

$$\begin{aligned}\tilde{Y}_{it}^K &= \mathbb{1}\{W_{ijt}'\theta_0 + \alpha_{ij} + \epsilon_{ijt} \geq W_{ikt}'\theta_0 + \alpha_{ik} + \epsilon_{ikt},\ \forall j \in K, k \in \mathcal{J}\backslash K\} \\ &:= G\left((W_{ijt} - W_{ikt})'\theta_0, \alpha_{ij} - \alpha_{ik}, \epsilon_{ijt} - \epsilon_{ikt},\ \forall j \in K, k \in \mathcal{J}\backslash K\right).\end{aligned}$$

Given this new variable $\tilde{Y}_{it}^K$, the function G is weakly monotone in $(W_{ijt} - W_{ikt})'\theta_0$ for $j \in K, k \in \mathcal{J}\backslash K$ and for any set $K \subset \mathcal{J}$, and we have a similar monotonicity structure.

Example 5 (Censored Outcome Model)**.** Consider the dependent variable Y_{it} is censored at 0, given as follows:

$$Y_{it}^* = W_{it}'\theta_0 + \alpha_i + \epsilon_{it},$$
$$Y_{it} = \max\{Y_{it}^*, 0\},$$

and the function $G(W_{it}'\theta_0, \alpha_i, \epsilon_{it}) = \max\{W_{it}'\theta_0 + \alpha_i + \epsilon_{it}, 0\}$ clearly satisfies the monotonicity in $W_{it}'\theta_0$.

We have shown that the semiparametric model with monotonicity can accommodate various types of panel data models. Now we describe our identification strategy for this general model. As before, we decompose the covariate W_{it}, and correspondingly θ_0, into two components, $W_{it} = \left(Z_{it}', X_{it}'\right)'$ and $\theta_0 = \left(\beta_0', \gamma_0'\right)'$ 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. We show how partial stationarity can be exploited in conjunction with weak monotonicity (Assumption 3) to obtain identifying restrictions in the presence of endogeneity.

Let $\mathcal{Y}$ denote the support of Y_{it}. For any $c \in \mathcal{R}$ and $y \in \mathcal{Y}$, conditional on $W_{ist} = w_{st}$, we consider the event that

$$Y_{it} \leq y, \qquad Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \geq c,$$

implying that

$$\begin{aligned}\mathbb{1}\{Y_{it} \leq y,\ Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \geq c\} &= \mathbb{1}\left\{G\left(Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0,\, \alpha_i,\, \epsilon_{it}\right) \leq y,\ Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \geq c\right\} \\ &\leq \mathbb{1}\{G\left(c,\, \alpha_i,\, \epsilon_{it}\right) \leq y\},\end{aligned}$$

where the inequality holds by the monotonicity assumption on the function G.

Symmetrically, we can provide an upper bound for $\mathbb{1}\{G(c,\, \alpha_i,\, \epsilon_{it}) \leq y\}$ by looking at the following event:

$$\begin{aligned}\mathbb{1}\{Y_{it} > y,\ Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \leq c\} &= \mathbb{1}\left\{G\left(Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0,\, \alpha_i,\, \epsilon_{it}\right) > y,\ Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \leq c\right\} \\ &\leq \mathbb{1}\{G\left(c,\, \alpha_i,\, \epsilon_{it}\right) > y\} \\ &= 1 - \mathbb{1}\{G\left(c,\, \alpha_i,\, \epsilon_{it}\right) \leq y\}.\end{aligned}$$

which is equivalent to

$$\mathbb{1}\{G\left(c,\, \alpha_i,\, \epsilon_{it}\right) \leq y\} \leq 1 - \mathbb{1}\{Y_{it} > y,\ Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \leq c\}.$$

The partial stationarity assumption in 1 implies that

$$\begin{aligned}&\mathbb{P}\left(Y_{it} \leq y,\ Z_{it}^{'}\beta_0 + X_{it}^{'}\gamma_0 \geq c \mid Z_i = z\right) \\ =\ &\mathbb{P}\left(G\left(c,\, \alpha_i,\, \epsilon_{it}\right) \leq y \mid Z_i = z\right) \\ =\ &\mathbb{P}\left(G\left(c,\, \alpha_i,\, \epsilon_{is}\right) \leq y \mid Z_i = z\right) \\ \leq\ &1 - \mathbb{P}\left(Y_{is} > y,\ Z_{is}^{'}\beta_0 + X_{is}^{'}\gamma_0 \leq c \mid Z_i = z\right).\end{aligned}$$

The key difference of the above identification strategy in (3.1) relative to the corresponding identifying restrictions in previous sections lies in that the "middle term" in (3.1) is no longer the conditional CDF of $\alpha_i + \epsilon_{it}$, but a conditional distribution of $\mathbb{P}\left(G\left(c,\, \alpha_i,\, \epsilon_{is}\right) \leq y \mid Z_i = z\right)$, 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 6 (Identified Set)**.** *Under Assumptions 1 and 3, $\theta_0 \in \Theta_{I,gen}$, where the identified set $\Theta_{I,gen}$ consists of all $\theta = \left(\beta^{'}, \gamma^{'}\right)^{'} \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_x}$ such that*

$$\max_{t=1,\ldots,T} \mathbb{P}\left(Y_{it} \leq y,\ z_t^{'}\beta + X_{it}^{'}\gamma \geq c \mid Z_i = z\right) \leq 1 - \max_{s=1,\ldots,T} \mathbb{P}\left(Y_{is} > y,\ z_s^{'}\beta + X_{is}^{'}\gamma \leq c \mid Z_i = z\right), \tag{22}$$

for any $c \in \mathcal{R}$, $y \in \mathcal{Y}$, and any realization $z = (z_1, \ldots, z_T)$ in the support of Z_i.

Note that in the binary choice setting of Section 2, it suffices to set $y = 0$ in (22), which then coincides with the identifying results in Proposition 1. More generally, we show that our identification results in Proposition 6 can be adapted to the ordered response model in Section 3.2, the multinomial choice model in Section 3.3, and the censored outcome model in Section 3.4.

The results in Proposition 6 generally hold regardless of whether the dependent variable and the endogenous covariate are discrete or continuous. The next proposition shows that additional discreteness in either the dependent variable or endogenous covariates can further simplify and reduce the number of the identifying conditions in (22).

Proposition 7 (Discreteness)**.** *When $X_{it} \in \{\overline{x}_1, ..., \overline{x}_K\}$ for any t, then $\Theta_{I,gen} = \Theta_{I,gen}^{disc_x}$, where $\Theta_{I,gen}^{disc_x}$ consists of all $\theta = \left(\beta', \gamma'\right)'$ that satisfy condition (22) for any $c \in \left\{z_t'\beta + \overline{x}_k'\gamma : k = 1, ..., K, t = 1, ..., T\right\}$.*

Moreover, when $Y_{it} \in \{\overline{y}_1, ..., \overline{y}_K\}$ with $\overline{y}_j \leq \overline{y}_{j+1}$ for any t, then $\Theta_{I,gen} = \Theta_{I,gen}^{disc_y}$, where $\Theta_{I,gen}^{disc_y}$ consists of all $\theta = \left(\beta', \gamma'\right)'$ that satisfy condition (22) for any $y \in \left\{\overline{y}_1, ..., \overline{y}_{K-1}\right\}$.

Proposition 7 shows that for any discrete choice models with discrete endogenous variables, such as dynamic binary, ordered, and multinomial choice models, the identified set $\Theta_{I,gen}$ is characterized by a finite number of moment inequalities. Sections 3.2-3.4 establish general identification results for panel models with various types of dependent variables and endogeneity, and also explore both the static model without endogeneity and the dynamic model with lagged dependent variables.

3.2 Ordered Response Model

Consider that the outcome variable Y_{it} takes J ordered values: $Y_{it} \in \{y_1, .., y_J\}$ with $y_j \leq y_{j+1}$. Examples of such ordered outcomes include various income categories, health outcomes, or levels of educational attainment. We study the following panel ordered choice model:

$$\begin{aligned} Y_{it}^* &= W_{it}'\theta_0 + \alpha_i + \epsilon_{it}, \\ Y_{it} &= \sum_{j=1}^{J} y_j \mathbb{1}\{b_j < Y_{it}^* \leq b_{j+1}\}, \end{aligned} \tag{23}$$

where Y_{it}^* denotes the latent dependent variable, and Y_{it} denotes the ordered outcome which takes value y_j when $Y_{it}^* \in (b_j, b_{j+1}]$. The threshold parameters satisfy $b_1 = -\infty, b_{J+1} = +\infty$, 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 $b_2 = 0$.

For this ordered choice model, we observe the conditional probability of each choice j:

$$\mathbb{P}\left(Y_{it} \leq y_j \mid w\right) = \mathbb{P}\left(\alpha_i + \epsilon_{it} \leq b_{j+1} - w_t'\theta_0 \mid w\right),$$

and this choice probability is monotone with respect to $b_{j+1} - w_t'\theta_0$. Compared to the binary choice model, here we can exploit variations in the conditional probability across all possible choices to provide more informative results for θ_0.

Following the same identification strategy in Sections 2 and 3, we can bound the conditional probability of $\mathbb{P}(\alpha_i + \epsilon_{it} \le c \mid w)$ below as follows: for any constant c,

$$\mathbb{P}(\alpha_i + \epsilon_{it} \le c \mid w) \ge \mathbb{P}(Y_{it} \le y_j,\ b_{j+1} - w_t'\theta_0 \le c \mid w). \tag{24}$$

Since the above inequality holds for any choice j, we can take the largest lower bound over all possible choices that satisfy $b_{j+1} - w_t'\theta_0 \le c$. We define $\overline{j}_c$ as $\overline{j}_c := \max\{j : b_{j+1} - w_t'\theta_0 \le c\}$, and condition (24) leads to:

$$\begin{aligned}\mathbb{P}(\alpha_i + \epsilon_{it} \le c \mid w) &\ge \sum_{k=1}^{\overline{j}_c} \mathbb{P}\left(Y_{it} = y_k,\ b_{\overline{j}_c+1} - w_t'\theta_0 \le c \mid w\right)\\ &= \sum_{j=1}^{J} \mathbb{P}(Y_{it} = y_j,\ b_{j+1} - w_t'\theta_0 \le c \mid w).\end{aligned}$$

where the last inequality holds since $\mathbb{1}\{b_{j+1} - w_t'\theta_0 \le c\} = 0$ for any choice $j > \overline{j}_c$. Furthermore, the final expression offers the advantage of avoiding the need to search for the maximum choice $\overline{j}_c$ for each constant c.

Similarly, we can derive an upper bound for the probability $\mathbb{P}(\alpha_i + \epsilon_{it} \le c \mid w)$:

$$\mathbb{P}(\alpha_i + \epsilon_{it} \le c \mid w) \le 1 - \mathbb{P}(Y_{it} > y_j,\ b_{j+1} - w_t'\theta_0 \ge c \mid w).$$

Define $\underline{j}_c$ as $\underline{j}_c := \min\{j : b_{j+1} - w_t'\theta_0 \ge c\}$, then we have

$$\begin{aligned}\mathbb{P}(\alpha_i + \epsilon_{it} \le c \mid w) &\le 1 - \mathbb{P}\left(Y_{it} > b_{\underline{j}_c},\ b_{\underline{j}_c+1} - w_t'\theta_0 \ge c \mid w\right)\\ &= 1 - \mathbb{P}\left(Y_{it} \ge b_{\underline{j}_c+1},\ b_{\underline{j}_c+1} - w_t'\theta_0 \ge c \mid w\right).\end{aligned}$$

Furthermore, the above inequality can be equivalently written as

$$\mathbb{P}(\alpha_i + \epsilon_{it} \le c \mid w) \le 1 - \sum_{j=1}^{J} \mathbb{P}(Y_{is} = y_j, b_j - z_s'\beta - X_{is}'\gamma \ge c \mid z),$$

which holds since for $\mathbb{1}\{b_{j+1} - w_t'\theta_0 < c\} = 0$ for any choice $j < \underline{j}_c$.

Given the established lower and upper bounds on the conditional probability $\mathbb{P}(\alpha_i + \epsilon_{it} \le c \mid w)$, we can derive the corresponding bounds for $\mathbb{P}(\alpha_i + \epsilon_{it} \le c \mid z)$ by taking expectation over the endogenous covariate X. Then, the identifying condition for θ_0 is characterized by the restriction that the bounds over different periods must have nonempty intersections, as presented in the following proposition.

Proposition 8. *Under Assumptions 1, $\theta_0 \in \Theta_{I,order}$, where the identified set $\Theta_{I,order}$ consists* *of all* $\theta = \left(\beta', \gamma'\right)' \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_x}$ *such that*

$$\max_{t=1,\ldots,T} \sum_{j=1}^{J} \mathbb{P}\left(Y_{it} = y_j, b_{j+1} - z_t'\beta - X_{it}'\gamma \leq c \mid z\right) \leq 1 - \max_{s=1,\ldots,T} \sum_{j=1}^{J} \mathbb{P}\left(Y_{is} = y_j, b_j - z_s'\beta - X_{is}'\gamma \geq c \mid z\right), \quad (25)$$

for any $c \in \mathcal{R}$ *and any realization* $z = (z_1, \ldots, z_T)$ *in the support of* Z_i.

Proposition 8 characterizes the identified set of θ_0 for the general ordered response model with endogeneity. In contrast to the binary choice model discussed in Section 2, Proposition 8 exploits information from all possible choices across different time periods to identify θ_0. This result accommodates both static and dynamic models, and we derive the simplified identifying results for these two types of models as follows.

Static model: consider that there is no endogeneity and the full stationarity assumption holds conditional on all covariates W_i:

$$\epsilon_{is} \mid W_i, \alpha_i \overset{d}{\sim} \epsilon_{it} \mid W_i, \alpha_i.$$

The identifying restriction in Proposition 8 is given as

$$\max_{t=1,\ldots,T} \sum_{j=1}^{J} \mathbb{P}\left(Y_{it} = y_j,\ b_{j+1} - w_t'\theta_0 \leq c \mid w\right) \leq 1 - \max_{s=1,\ldots,T} \sum_{j=1}^{J} \mathbb{P}(Y_{is} = y_j,\ b_j - w_s'\theta_0 \geq c \mid w).$$

The above condition is informative only if there exists j_1, j_2 such that $b_{j_2+1} - w_t'\theta_0 \leq c \leq b_{j_1} - w_s'\theta_0$, leading to

$$\begin{aligned}\max_{t=1,\ldots,T} \sum_{j=1}^{j_2} \mathbb{P}\left(Y_{it} = y_j \mid w_i\right) &\leq 1 - \max_{s=1,\ldots,T} \sum_{j=j_1}^{J} \mathbb{P}(Y_{is} = y_j \mid w) \\ &= \max_{s=1,\ldots,T} \sum_{j=1}^{j_1-1} \mathbb{P}(Y_{is} = y_j \mid w).\end{aligned}$$

The following proposition presents the identification results for the static ordered choice model by changing $j_1 - 1$ with j_1.

Corollary 1. *Assume that* $\epsilon_{is} \mid (W_i, \alpha_i) \overset{d}{\sim} \epsilon_{it} \mid (W_i, \alpha_i)$, *then* $\Theta_{I,order}$ *consists of all* $\theta = \left(\beta', \gamma'\right)' \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_x}$ *such that*

$$b_{j_1+1} - w_s'\theta \geq b_{j_2+1} - w_t'\theta \Longrightarrow \max_{s=1,\ldots,T} \sum_{j=1}^{j_1} \mathbb{P}(Y_{is} = y_j \mid w) \geq \max_{t=1,\ldots,T} \sum_{j=1}^{j_2} \mathbb{P}(Y_{it} = y_j \mid w),$$

for any $1 \leq j_1, j_2 \leq J - 1$, *and any realization* $w = (w_1, \ldots, w_T)$ *in the support of* W_i.

The results in Corollary 1 are analogous to the maximum-score type result in Manski (1987), with the distinction being that we can exploit variations in the sum of multiple choices rather than investigating a single choice to identify θ_0. Furthermore, with multiple choices, we can utilize variations in the sum of different choices across various periods for identification. Besides the static model, Proposition 8 can also accommodate dynamic ordered choice models with lagged dependent variable.

Dynamic model: consider the following dynamic ordered choice 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 $X_{it} = Y_{i,t-1} \in \{y_1, ..., y_J\}$. Then, the identifying restriction in Proposition 8 holds with $X_{it} = Y_{i,t-1}$ and for any $c \in \{b_j - z_t'\beta - y_1\gamma, ..., b_j - z_t'\beta - y_J\gamma\}_{2\leq j\leq J,\ 1\leq t\leq T}$, and the identified set $\Theta_{I,order}$ is characterized by $TJ(J-1)$ number of conditional inequalities. Moreover, our approach can be easily applied to dynamic models with more than one lagged dependent variable, e.g., $X_{it} = (Y_{i,t-1}, Y_{i,t-2})$.

3.3 Multinomial Choice Model

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

$$u_{ijt} = W_{ijt}'\theta_0 + \alpha_{ij} + \epsilon_{ijt},$$
$$Y_{it} = \arg\max_{j\in\mathcal{J}} u_{ijt},$$

and impose the same partial stationarity assumption:

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

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, and $Z_i := (Z_{i1}, Z_{i2}, ..., Z_{iT})$.

Although the J choices are unordered and not directly comparable, we look at the indicator variable $Y_{it}^j := \mathbb{1}\{Y_{it} = j\}$ of choosing option j, which maintains a similar monotone structure with Assumption 3:

$$Y_{it}^{j} = 1 \Longleftrightarrow W_{ijt}'\theta_0 + \alpha_{ij} + \epsilon_{ijt} \geq W_{ikt}'\theta_0 + \alpha_{ik} + \epsilon_{ikt},\ \forall k \in \mathcal{J},$$

and the new variable Y_{it}^{j} is monotone in $W_{ijt}'\theta_0 - W_{ikt}'\theta_0$ given $(\alpha_i, \epsilon_{it})$.

More generally, for any subset $K \subset \mathcal{J}$, the indicator variable $Y_{it}^{K} := \mathbb{1}\{Y_{it} \in K\}$ represents individual i's choice belonging to the subset K, given as follows:

$$Y_{it}^{K} = 1 \Longleftrightarrow W_{ijt}'\theta_0 + \alpha_{ij} + \epsilon_{ijt} \geq W_{ikt}'\theta_0 + \alpha_{ik} + \epsilon_{ikt},\ \exists j \in K, \forall k \in \mathcal{J} \setminus K,$$

and the variable Y_{it}^{K} is monotone in $W_{ijt}'\theta_0 - W_{ikt}'\theta_0$ for any $j \in K$ and $k \in \mathcal{J} \setminus K$.

Following Proposition 6, the identification results for panel multinomial choice models are presented in the following proposition.

Proposition 9. *Under Assumption 1, $\theta_0 \in \Theta_{I,mul}$, where the identified set $\Theta_{I,mul}$ consists of all $\theta = \left(\beta', \gamma'\right)' \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_x}$ such that*

$$\begin{aligned} &\max_{t=1,\dots,T} \mathbb{P}(Y_{it}^{K} = 0, (z_{js} - z_{ks})'\beta + (X_{ijs} - X_{iks})'\gamma \geq c_{jk}\ \forall j \in K, k \in \mathcal{J} \setminus K \mid z) \\ &\quad \leq 1 - \max_{s=1,\dots,T} \mathbb{P}\left(Y_{is}^{K} = 1, (z_{js} - z_{ks})'\beta + (X_{ijs} - X_{iks})'\gamma \leq c_{jk}\ \forall j \in K, k \in \mathcal{J} \setminus K \mid z\right), \end{aligned} \tag{26}$$

for any subset $K \subset \mathcal{J}$, any $c_{jk} \in \mathcal{R}$, any $j \in K$ and $k \in \mathcal{J} \setminus K$, and any realization $z = (z_1, ..., z_T)$ in the support of Z_i.

Proposition 9 provides general identification results for multinomial choice models with endogenous covariates. It accommodates static, dynamic multinomial choice models with any number of lagged dependent variables, as well as other types of endogeneity. For the dynamic model, similar to Proposition 2, the results can be simplified to hold for a finite number of values of c_{jk}, which is the collection of values the covariate index $(z_{jt} - z_{kt})'\beta + (X_{ijt} - X_{ikt})'\gamma$ can take.

Reconciliation with Pakes and Porter (2022)

Next, we show that our results specialize to those in Pakes and Porter (2022), who focuses on the static panel multinomial choice model without any endogeneity.

Formally, Pakes and Porter (2022) characterizes the sharp identified set for θ_0 under the full stationarity assumption given all covariates:

$$\epsilon_{is} \mid W_i, \alpha_i \overset{d}{\sim} \epsilon_{it} \mid W_i, \alpha_i.$$

Under full stationarity, for any two periods (s, t), our identifying condition in (26) is simplified as

$$\mathbb{P}\left(Y_{it}^K = 0, (w_{jt} - w_{kt})'\theta_0 \geq c_{jk}\ \forall j \in K, k \in \mathcal{J} \setminus K \mid w\right)$$
$$\leq 1 - \mathbb{P}\left(Y_{is}^K = 1, (w_{js} - w_{ks})'\theta_0 \leq c_{jk}\ \forall j \in K, k \in \mathcal{J} \setminus K \mid w\right) \quad (27)$$

The above equation is only informative when $(w_{js} - w_{ks})'\theta_0 \leq c_{jk} \leq (w_{jt} - w_{kt})'\theta_0$ for any $j \in K, k \in k \in \mathcal{J} \setminus K$; otherwise either the upper bound becomes one or the lower bound becomes zero so condition (27) holds for any θ. There exists one value c_{jk} satisfying the condition $(w_{js} - w_{ks})'\theta_0 \leq c_{jk} \leq (w_{jt} - w_{kt})'\theta_0$ is equivalent to $(w_{js} - w_{ks})'\theta_0 \leq (w_{jt} - w_{kt})'\theta_0$, generating the following inequality: for any $K \subset \mathcal{J}$,

$$\text{If } (w_{js} - w_{ks})'\theta_0 \leq (w_{jt} - w_{kt})'\theta_0 \quad \forall j \in K, k \in \mathcal{J} \setminus K$$
$$\Longrightarrow \mathbb{P}\left(Y_{it}^K = 0 \mid w\right) \leq 1 - \mathbb{P}\left(Y_{is}^K = 1 \mid w\right),$$

which becomes the same result in Pakes and Porter (2022) (Proposition 1, P. 12):

$$\text{If } (w_{js} - w_{ks})'\theta_0 \leq (w_{jt} - w_{kt})'\theta_0 \quad \forall j \in K, k \in \mathcal{J} \setminus K$$
$$\Longrightarrow \mathbb{P}\left(Y_{is} \in K \mid w\right) \leq \mathbb{P}\left(Y_{it} \in K \mid w\right),$$

since $Y_{it}^K = 1$ is equivalent to $Y_{it} \in K$ by the definition.

Beyond the static model, Proposition 9 allows for any type of endogeneity including dynamic multinomial models with lagged dependent variable. For example, consider the following dynamic model:

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

where individual i's utility at time t can potentially depend on their choices in the previous period $t-1$. In this model, the endogenous variable X_{ijt} is whether option j is chosen in the previous period $\mathbb{1}\{Y_{i,t-1} = j\}$. Then, the difference in the endogenous covariate between choices only takes three values: $X_{ijt} - X_{ikt} \in \{1, -1, 0\}$, and the identified set for $\Theta_{I,mul}$ is characterized by the condition in (26) with $c_{jk} \in \{(z_{jt} - z_{kt})'\beta + \gamma, (z_{jt} - z_{kt})'\beta - \gamma, (z_{jt} - z_{kt})'\beta\}_{t=1}^T$.

3.4 Censored Outcome 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 the following panel model with censored outcomes:

$$Y_{it}^* = Z_{it}'\beta_0 + X_{it}'\gamma_0 + \alpha_i + \epsilon_{it},$$
$$Y_{it} = \max\{Y_{it}^*, 0\},$$

where Y_{it}^* denotes the latent outcome which is not observed in the data, and Y_{it} represents the observed outcome, censored at zero. The threshold for censoring can be replaced with other nonzero constants.

The identification strategy is still to exploit the partial stationarity assumption and bound the conditional distribution of $\alpha_i + \epsilon_{it} \mid Z_i = z$. This censored outcome model imposes an additional structure between the outcome and the parametric index: when $Y_{it} > 0$, we have $Y_{it} = Y_{it}^*$ and

$$\alpha_i + \epsilon_{it} \leq c \Longleftrightarrow Y_{it} - Z_{it}'\beta_0 - X_{it}'\gamma_0 \leq c.$$

This specific structure can be exploited to further tighten the identified set for θ_0, and we provide the details of the identification strategy in Appendix A.7. The following proposition presents the identification results of θ_0 with censored outcomes.

Proposition 10. *Under Assumption 1, $\theta_0 \in \Theta_{I,cen}$, where the identified set $\Theta_{I,cen}$ consists of all $\theta = (\beta', \gamma')' \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_x}$ such that*

$$\max_{t=1,\dots,T} \mathbb{P}(Y_{it} \leq z_t'\beta + X_{it}'\gamma - c \mid z) \leq \max_{s=1,\dots,T} \{\mathbb{P}(0 < Y_{is} \leq z_s'\beta + X_{is}'\gamma - c \mid z) + \mathbb{P}(Y_{is} = 0 \mid z)\},$$

for any $c \in \mathcal{R}$ and any realization $z = (z_1, \dots, z_T)$ in the support of Z_i.

Similar to discrete choice models studied in previous sections, Proposition 10 characterizes an identified set for θ_0 by exploiting the variation in the joint distribution of $(Y_{it}, X_{it}) \mid Z_i$ over time and the variation in the exogenous covariates Z_i. The bounds on the distribution $\alpha_i + \epsilon_{it} \mid Z_i = z$ can be derived either from the probability $\mathbb{P}(0 < Y_{it} \leq y \mid z)$ or $\mathbb{P}(Y_{it} = 0 \mid z)$, depending on the value of the covariate index $z_t'\beta_0 + X_{it}'\gamma_0$. This result still accommodates both static and dynamic models with censored outcomes, and we provide the simplified results for each model.

Static model: consider that the full stationarity assumption holds, i.e., $\epsilon_{it} \mid \alpha_i, W_i \overset{d}{\sim} \epsilon_{is} \mid \alpha_i, W_i$. Then, the identifying condition in Proposition 10 is given as

$$\mathbb{P}(Y_{it} \leq w_t'\theta - c \mid w) \leq \mathbb{P}(0 < Y_{is} \leq w_s'\theta - c \mid w) + \mathbb{P}(Y_{is} = 0 \mid w). \tag{28}$$

The above restriction is informative only when $w_t'\theta - c \geq 0$, otherwise the lower bound becomes zero. We discuss two cases for the constant c: $w_s'\theta \leq c \leq w_t'\theta$ and $c \leq \min\{w_s'\theta, w_t'\theta\}$.

When $w_s'\theta \leq c \leq w_t'\theta$, then condition (28) becomes

$$\mathbb{P}(Y_{it} \leq w_t'\theta - c \mid w) \leq \mathbb{P}(Y_{is} = 0 \mid w).$$

When c satisfies $c \le \min\{w_s'\theta, w_t'\theta\}$, condition (28) transforms into

$$\mathbb{P}(Y_{it} \le w_t'\theta - c \mid w) \le \mathbb{P}(0 < Y_{is} \le w_s'\theta - c \mid w) + \mathbb{P}(Y_{is} = 0 \mid w) = \mathbb{P}(Y_{is} \le w_s'\theta - c \mid w).$$

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 summarizes the results for the static model.

Corollary 2. *Assuming that $\epsilon_{is} \mid \alpha_i, W_i \overset{d}{\sim} \epsilon_{it} \mid \alpha_i, W_i$, the identified set $\Theta_{I,cen}$ consists of all $\theta = (\beta', \gamma')' \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_x}$ such that*

$$\begin{cases} \text{If } w_s'\theta \le c \le w_t'\theta \Longrightarrow \mathbb{P}(Y_{it} \le w_t'\theta - c \mid w) \le \mathbb{P}(Y_{is} = 0 \mid w); \\ \text{If } c \le \min\{w_s'\theta, w_t'\theta\} \Longrightarrow \mathbb{P}(Y_{it} \le w_t'\theta - c \mid w) = \mathbb{P}(Y_{is} \le w_s'\theta - c \mid w), \end{cases}$$

for any $c \in \mathcal{R}$, any $(s,t) \le T$, and any realization $w = (w_1, ..., w_T)$ in the support of W_i.

Dynamic model: Proposition 10 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 10. Appendix A.8 also studies dynamic models with the latent lagged outcome $Y_{i,t-1}^*$. Consequently, the results in Proposition 10 need to be adjusted as the endogenous variable $X_{it} = Y_{i,t-1}^*$ is not observed.

4 Simulation

This section examines the finite sample performance of our identification approaches using Monte Carlo simulations. We focus on the static and dynamic ordered choice models explored in Section 3.2 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.

4.1 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 $\Sigma = (1\ \rho; \rho\ 1)$. The true parameter is $\beta_0 := (\beta_{0,1}, \beta_{02})' = (1, 1)'$, 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 $b_3 = 1$.

With $Y_i := (Y_{i1}, Y_{i2})$ and $Z_i := (Z_{i1}, Z_{i2})$, Corollary 1 characterizes the identified set for β_0 using the following conditional moment inequalities: for $s \neq t \leq 2$,

$$E[g(Z_i, Y_i; \beta_0) \mid z] \geq 0,$$

where

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

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.

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 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

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

4.2 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 $\mathbb{P}(Y_{i0} = 1) = 0.6, \mathbb{P}(Y_{i0} = 2) = \mathbb{P}(Y_{i0} = 3) = 0.2$.

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 8 characterizes the identified set for θ_0 with the following conditional moment inequalities:

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

$$1 - \sum_{j=2}^{3} \mathbb{P}\left(Y_{i1} = y_j \mid z, y_0\right) * \mathbb{1}\{b_j - z_1'\beta - y_0\gamma \geq c\} \geq \sum_{j=1}^{2} \mathbb{P}\left(Y_{is} = y_j, b_{j+1} - z_s'\beta - Y_{is-1}\gamma \leq c \mid z, y_0\right),$$

$$1 - \sum_{j=2}^{3} \mathbb{P}(Y_{is} = y_j, b_j - z_s'\beta - Y_{is-1}\gamma \geq c \mid z, y_0) \geq \sum_{j=1}^{2} \mathbb{P}\left(Y_{i1} = y_j \mid z, y_0\right) * \mathbb{1}\{b_{j+1} - z_1'\beta - y_0\gamma \leq c\},$$

for any $c \in \{b_j - z_1'\beta - y_0\gamma, b_j - z_s'\beta - \gamma, b_j - z_s'\beta - 2\gamma, b_j - z_s'\beta - 3\gamma\}_{j=2}^T$;

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

$$1 - \sum_{j=2}^{3} \mathbb{P}(Y_{is} = y_j, b_j - z_s'\beta - Y_{is-1}\gamma \geq c \mid z, y_0)$$
$$\geq \sum_{j=1}^{2} \mathbb{P}\left(Y_{it} = y_j, b_{j+1} - z_t'\beta - Y_{it-1}\gamma \leq c \mid z, y_0\right),$$

for any $c \in \{b_j - z_s'\beta - \gamma, b_j - z_s'\beta - 2\gamma, b_j - z_s'\beta - 3\gamma, b_j - z_t'\beta - \gamma, b_j - z_t'\beta - 2\gamma, b_j - z_t'\beta - 3\gamma\}_{j=2}^{3}$.

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

5 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.[8] 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.

We adopt various ordered response models introduced in Section 3.2 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 8 characterizes the identified set of the model coefficients for these three models using conditional moment inequalities. Similar to Section 4, we exploit the kernel-

[8]This 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.

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

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.

6 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).[9] Additionally, the idea of bounding an endogenous object (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.

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.

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

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.*

Nelsen, R. B. (2006): *An introduction to copulas*, Springer.

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.*

A Appendix

A.1 Proof of Proposition 2

Proof. Clearly, $\Theta_I \subseteq \Theta^{disc}$. Below we show $\Theta^{disc} \subseteq \Theta_I$ when X_{it} is discrete. Suppose that θ satisfies condition (10) at all

$$c \in \mathcal{C}(\theta) := \left\{ z_t'\beta + \overline{x}_k'\gamma : k = 1, ..., K, t = 1, ..., T \right\}$$

for any realization $z = (z_1, ..., z_T)$. We seek to show that θ must also satisfy condition (10) for any $c \in \mathcal{R} \backslash \mathcal{C}(\theta)$. Without loss of generality, we order elements in $\mathcal{C}(\theta)$ from the smallest to the largest as

$$\overline{c}_1 \leq \overline{c}_2 \leq ... \leq \overline{c}_{KT}.$$

For $c < \overline{c}_1$, we must have

$$\mathbb{P}\left(Y_{is} = 0,\ z_s'\beta + X_{is}'\gamma \geq c \middle| Z_i = z\right) \equiv 0,$$

so (10) holds trivially. Similarly, for $c > \overline{c}_{KT}$, we must have

$$\mathbb{P}\left(Y_{is} = 1,\ z_s'\beta + X_{is}'\gamma \leq c \middle| Z_i = z\right) \equiv 0,$$

so (10) again holds trivially. For any c s.t. $\overline{c}_j < c < \overline{c}_{j+1}$ for some j, we have

$$z_t'\beta + X_{it}'\gamma \leq c \quad \Leftrightarrow \quad z_t'\beta + X_{it}'\gamma \leq \overline{c}_j$$

and

$$z_s'\beta + X_{is}'\gamma \geq c \quad \Leftrightarrow \quad z_s'\beta + X_{is}'\gamma \geq \overline{c}_{j+1}.$$

which implies

$$\mathbb{P}\left(Y_{it} = 1,\ z_t'\beta + X_{it}'\gamma \leq c \middle| Z_i = z\right) = \mathbb{P}\left(Y_{it} = 1,\ z_t'\beta + X_{it}'\gamma \leq \overline{c}_j \middle| Z_i = z\right) \tag{29}$$

and

$$\begin{aligned}\mathbb{P}\left(Y_{is} = 0,\ z_s'\beta + X_{is}'\gamma \geq c \middle| Z_i = z\right) &= \mathbb{P}\left(Y_{is} = 0,\ z_s'\beta + X_{is}'\gamma \geq \overline{c}_{j+1} \middle| Z_i = z\right) \\ &\leq \mathbb{P}\left(Y_{is} = 0,\ z_s'\beta + X_{is}'\gamma \geq \overline{c}_j \middle| Z_i = z\right),\end{aligned}$$

or equivalently,

$$1 - \mathbb{P}\left(Y_{is} = 0,\ z_s'\beta + X_{is}'\gamma \geq \overline{c}_j \middle| Z_i = z\right) \leq 1 - \mathbb{P}\left(Y_{is} = 0,\ z_s'\beta + X_{is}'\gamma \geq \overline{c}_{j+1} \middle| Z_i = z\right). \tag{30}$$

Since (10) holds at $\overline{c}_j$, we have

$$\max_t \mathbb{P}\left(Y_{it} = 1,\ z_t'\beta + X_{it}'\gamma \leq \overline{c}_j \middle| Z_i = z\right) \leq 1 - \max_s \mathbb{P}\left(Y_{is} = 0,\ z_s'\beta + X_{is}'\gamma \geq \overline{c}_j \middle| Z_i = z\right).$$

Combining the above with (29) and (30), we have

$$\max_t \mathbb{P}\left(Y_{it} = 1,\ z_t'\beta + X_{it}'\gamma \leq c \middle| Z_i = z\right) \leq 1 - \max_s \mathbb{P}\left(Y_{is} = 0,\ z_s'\beta + X_{is}'\gamma \geq c \middle| Z_i = z\right).$$

□

A.2 Proof of Theorem 1

For shorter notation we write $W_{it} := (Z_{it}, X_{it})$ for the combination of the exogenous covariates Z_{it} and endogenous covariates X_{it}. Correspondingly, we write $W_i \equiv (W_{i1}, ..., W_{iT})$, $W_i'\theta_0 \equiv Z_i'\beta_0 + X_i'\gamma_0$ and the lower cases $w \equiv (w_1, ..., w_T)$ for realizations.

We first clarify the rigorous meaning of "sharpness" in Theorem 1 through the following definition.

Definition 1. *We say that* Θ_I^{disc} *is* sharp *under model* (5) *and Assumption 1 if, for any* $\theta \equiv \left(\beta', \gamma'\right)' \in \Theta_I^{disc} \backslash \{\theta_0\}$, *there exist well-defined latent random variables* $(\epsilon_i^*, \alpha_i^*)$ *such that:*

- *Assumption 1 (partial stationarity) is satisfied, i.e.,*

$$\epsilon_{it}^* \sim \epsilon_{is}^* \,|\, Z_i, \alpha_i^*, \ \forall t, s = 1, ..., T.$$

- *(CCP-J)* $(\theta, \epsilon_i^*, \alpha_i^*)$ *are observationally equivalent to* $(\theta_0, \epsilon_i, \alpha_i)$, *i.e., formally,* $(\theta, \epsilon_i^*, \alpha_i^*)$ *produces the following conditional choice probabilities under model* (5)*:*

$$\mathbb{P}\left(v_{it}^* \leq w_t'\theta \forall t \ s.t. \ y_t = 1, \ v_{is}^* > w_s'\theta \forall s \ s.t. \ y_s = 0 \mid w\right) = p\left(y \mid w\right), \tag{31}$$

where $v_{it}^* := -\left(\epsilon_{it}^* + \alpha_i^*\right)$ *and* $p\left(\cdot \mid w\right)$ *denotes the* true *conditional probability*

$$\begin{aligned} p\left(y \mid w\right) &:= \mathbb{P}\left(Y_{it} = y_t \, \forall t = 1, ..., T \mid W_i = w\right) \\ &\equiv \mathbb{P}\left(v_{it} \leq w_t'\theta_0 \forall t \ s.t. \ y_t = 1, \ v_{is} > w_s'\theta \forall s \ s.t. \ y_s = 0 \mid W_i = w\right), \end{aligned}$$

for any outcome realization $y \equiv (y_1, ..., y_T) \in \{0, 1\}^T$, *given any realization* $W_i = w$.

We prove Theorem 1, i.e., the sharpness of Theorem 1 under discreteness of X_{it} by, for any candidate parameter $\theta \in \Theta_I^{disc} \backslash \{\theta_0\}$, we construct the $(\epsilon_i^*, \alpha_i^*)$.

Proof. Set $\alpha_i^* \equiv 0$ and $\epsilon_i^* := -v_i^*$. Then the conclusion follows from Lemma 1, 2 and 3 below. ☐

Lemma 1 (Per-Period Construction with Discrete X)**.** *Suppose that* $\bigcup_{t=1}^T Supp\left(X_{it}\right)$ *is finite. For any* $\theta \equiv \left(\beta', \gamma'\right)' \in \Theta_I^{disc} \backslash \{\theta_0\}$, *there exist well-defined latent random variables* $v_{i1}^*, ..., v_{iT}^*$ *with marginal CDFs* $F_1^*, ..., F_T^*$ *such that*

$$F_t^*\left(\cdot \mid Z_i = z\right) = F_s^*\left(\cdot \mid Z_i = z\right) \tag{32}$$

and

$$F_t^*\left(w_t'\theta \,\middle|\, W_i = w\right) = p_t\left(w\right), \quad \forall t, \forall w, \tag{33}$$

where

$$p_t\left(w\right) := \mathbb{P}\left(Y_{it} = 1 \mid W_i = w\right).$$

Proof. For any $\theta \equiv \left(\beta', \gamma'\right)' \in \Theta_I^{disc} \backslash \{\theta_0\}$, below we show how to construct $v_{i1}^*, ..., v_{iT}^*$, or equivalently, the conditional CDFs $F_1^*\left(c \mid W_i = w\right), ..., F_T^*\left(c \mid W_i = w\right)$ for each realization w and each $c \in \mathcal{R}$ so that (i) condition (32) is satisfied so that partial stationarity holds; and (ii) condition (33) is satisfied so that per-period marginal CCPs are matched.

Fix a specific realization of the exogenous covariates at $z \equiv (z_1, ..., z_T)$. We construct the (conditional) CDF F_t^* of v_{it}^* for each $t = 1, ..., T$ and each given z in the following manner.

Define

$$L_t(c|z) := \mathbb{P}\left(Y_{it}=1,\ z_t'\beta + X_{it}'\gamma \le c \middle| Z_i = z\right),$$
$$U_t(c|z) := 1 - \mathbb{P}\left(Y_{it}=0,\ z_t'\beta + X_{it}'\gamma \ge c \middle| Z_i = z\right),$$

and

$$\overline{L}(c|z) := \max_s L_s(c|z), \quad \underline{U}(c|z) := \min_s U_s(c|z).$$

Since $\theta \equiv (\beta', \gamma')' \in \Theta_I^{disc} \backslash \{\theta_0\}$, by 10 we have,

$$\overline{L}(c|z) \le \underline{U}(c|z), \quad \forall c \in \mathcal{R}.$$

Observe that both $\overline{L}(c|z)$ and $\underline{U}(c|z)$ are weakly increasing in c.

Conditional on z, sinceX_{it} can only take K values $\overline{x}_1, ..., \overline{x}_K$, the parametric index $w_t'\theta \equiv z_t'\beta + x_t'\gamma$ can only take values in the set

$$\mathcal{C} := \left\{ z_t'\beta + \overline{x}_k'\gamma : t = 1, ..., T, k = 1, ..., K \right\}.$$

Let $\delta > 0$ be a sufficiently small constant[10], and define

$$\underline{c} := \min \mathcal{C}, \quad \overline{c} := \max \mathcal{C} + \delta.$$

The parametrix index $w_t'\theta$ must lie within the interval $[\underline{c}, \overline{c})$ across all t for all possible realization of x.

For each $t = 1, ..., T$, we show how to construct v_t^* with CDF F_t^* so that

$$F_t^*(c|z) \equiv \begin{cases} 0, & \text{if } c < \underline{c}, \\ \overline{L}(c|z), & \text{if } \underline{c} \le c < \overline{c}, \\ 1, & \text{if } c \ge \overline{c}, \end{cases} \tag{34}$$

and

$$F_t^*(w_t'\theta | w) = p_t(w). \tag{35}$$

Clearly, partial stationarity (32) will be satisfied under (34), the right-hand side of which does not depend on the time index t. Furthermore, (35) is the same as (33), i.e., the marginal CCPs will be matched for each t. □

Proof. **Step 1:**

We construct the conditional CDF of $v^*|W_i = w$ using two auxillary CDFs F_t^L and F_t^U,

[10] The small positive constant $\delta > 0$ is used to ensure the right continuity of CDFs defined afterwards. Let $\underline{\delta}$ to be smallest distance between two distinct points in $\mathcal{C}$. If $\underline{\delta} > 0$, then we may set $\delta := \underline{\delta}/2$. If $\underline{\delta} = 0$, then δ can be set as any positive number, say, $\delta := 1$.

defined by

$$F_t^L\left(c|\, w\right) = \begin{cases} 0, & c < w_t'\theta, \\ p_t\left(w\right), & w_t'\theta \leq c < \overline{c}_t, \\ 1, & c \geq \overline{c}_t, \end{cases}$$

and

$$F_t^U\left(c|\, w\right) = \begin{cases} 0, & c < \underline{c}_t, \\ p_t\left(w\right), & \underline{c}_t \leq c < w_t'\theta + \delta, \\ 1, & c \geq w_t'\theta + \delta. \end{cases}$$

where

$$\overline{c}_t := \max \mathcal{C}_t + \delta, \quad \underline{c}_t := \min \mathcal{C}_t, \quad \mathcal{C}_t := \left\{ z_t'\beta + \overline{x}_k'\gamma : k = 1, ..., K \right\}.$$

Clearly, by construction we have

$$F_t^L\left(w_t'\theta \middle|\, w\right) = F_t^U\left(w_t'\theta \middle|\, w\right) = p_t\left(w\right). \tag{36}$$

Furthermore, for any $c < \overline{c}_t$, we have

$$\begin{aligned} F_t^L\left(c|\, z\right) &= \sum_x \mathbb{P}\left(X_i = x|\, Z_i = z\right) F_t^L\left(c|\, w\right) \\ &= \sum_x \mathbb{P}\left(X_i = x|\, Z_i = z\right) \mathbb{1}\left\{ w_t'\theta \leq c \right\} p_t\left(w\right) \\ &= \sum_x \mathbb{P}\left(X_i = x|\, Z_i = z\right) \mathbb{P}\left(Y_i = 1 \text{ and } w_t'\theta \leq c \middle|\, W_i = w\right) \\ &= \mathbb{P}\left(Y_i = 1 \text{ and } z_t'\beta + X_{it}'\gamma \leq c \middle|\, Z_i = z\right) \\ &= L_t\left(c|\, z\right), \end{aligned}$$

while for $c \geq \overline{c}_t$, we have

$$F_t^L\left(c|\, z\right) = 1 = U_t\left(c|\, z\right).$$

Similarly, for each $c \geq \underline{c}_t$, we have

$$
\begin{aligned}
F_t^U(c|z) &= \sum_x \mathbb{P}(X_i = x|Z_i = z) F_t^U(c|w) \\
&= \sum_x \mathbb{P}(X_i = x|Z_i = z) \left[1 - \mathbb{P}\left(Y_i = 0 \text{ and } w_t'\theta \geq c - \delta \middle| W_i = w\right)\right] \\
&= \sum_x \mathbb{P}(X_i = x|Z_i = z) \left[1 - \mathbb{P}\left(Y_i = 0 \text{ and } w_t'\theta \geq c - \delta \middle| W_i = w\right)\right] \\
&= 1 - \mathbb{P}\left(Y_i = 0 \text{ and } z_t'\beta + X_{it}'\gamma \geq c - \delta \middle| Z_i = z\right) \\
&= 1 - \mathbb{P}\left(Y_i = 0 \text{ and } z_t'\beta + X_{it}'\gamma \geq c \middle| Z_i = z\right) \\
&= U_t(c|z),
\end{aligned}
$$

where the second last equality holds for sufficiently small $\delta > 0$ due to the discreteness of $\mathcal{C}$. Lastly, for $c < \underline{c}_t$, we have

$$F_t^U(c|z) = 0 = L_t(c|z).$$

In summary, we have

$$
\begin{aligned}
F_t^L(c|z) &= \begin{cases} L_t(c|z), & \forall c < \overline{c}_t, \\ U_t(c|z) = 1, & \forall c \geq \overline{c}_t, \end{cases} \\
F_t^U(c|z) &= \begin{cases} L_t(c|z) = 0, & \forall c < \underline{c}_t, \\ U_t(c|z), & \forall c \geq \underline{c}_t, \end{cases}
\end{aligned}
\tag{37}
$$

Furthermore, observe that

$$L_t(\cdot|z) \leq F_t^L(\cdot|z) \leq F_t^U(\cdot|z) \leq U_t(\cdot|z).$$

Step 2:

Now, we construct $F_t^*(c|w)$ for $c \in \mathcal{C}_t$ using the two auxillary CDFs $F_t^L(c|w)$ and $F_t^U(c|w)$. We rank-order elements in $\mathcal{C}_t$ in ascending order

$$\underline{c}_t \equiv c_{t1} \leq c_{t2} \leq ... \leq c_{tK} < \overline{c}_t \equiv c_{tK} + \delta.$$

(i) We start with the largest element c_{tK}. By the definition of L_t, U_t, $\overline{L}$ and (10), we know that

$$U_t(c_{t1}|z) = L_t(c_{tK}|z) \leq \overline{L}(c_{tK}|z) \leq U_t(c_{tK}|z).$$

Hence, we can find $2 \leq j_1 \leq K$ such that

$$U_t(c_{t,j_1-1}|z) \leq \overline{L}(c_{tK}|z) \leq U_t(c_{t,j_1}|z),$$

so that there exists $\alpha_1 \in [0,1]$ such that

$$\overline{L}\left(c_{tK}\right| z) = \alpha_1 U_t \left(c_{t,j_1-1}\right| z) + (1-\alpha) U_t \left(c_{t,j_1}\right| z) . \tag{38}$$

Then, we set

$$F_t^* \left(c_{tK}\right| w) := \alpha_1 F_t^U \left(c_{t,j_1-1}\right| w) + (1-\alpha) F_t^U \left(c_{t,j_1}\right| w) , \tag{39}$$

which ensures that

$$F_t^* \left(c_{tK}\right| z) = \overline{L}\left(c_{tK}\right| z) . \tag{40}$$

Furthermore, whenever w is such that $w_t'\theta = c_{tK}$, we have

$$F_t^* \left(w_t'\theta \middle| w \right) = F_t^* \left(c_{tK}\right| w) = p_t(w) .$$

(ii) Next, we consider the point $c_{t,K-1}$. Given that

$$L_t \left(c_{t,K-1}\right| z) \leq \overline{L}\left(c_{t,K-1}\right| z) \leq U_t \left(c_{t,K-1}\right| z) ,$$

then either there exists some $1 \leq j_2 \leq K-1$ such that

$$\overline{L}\left(c_{t,K-1}\right| z) \in \left[U_t \left(c_{t,j_2-1}\right| z) , U_t \left(c_{t,j_2}\right| z)\right]$$

or

$$\overline{L}\left(c_{t,K-1}\right| z) \in \left[L_t \left(c_{t,K-1}\right| z) , L_t \left(c_{t,K}\right| z) \equiv U_t \left(c_{t,1}\right| z)\right] .$$

Hence, either there exists $\alpha_2 \in [0,1]$ s.t.

$$\overline{L}\left(c_{t,K-1}\right| z) = \alpha_2 U_t \left(c_{t,j_2-1}\right| z) + (1-\alpha_2) U_t \left(c_{t,j_2}\right| z) \tag{41}$$

or there exists $\tilde{\alpha}_2 \in [0,1]$ s.t.

$$\overline{L}\left(c_{t,K-1}\right| z) = \tilde{\alpha}_2 L_t \left(c_{t,K-1}\right| z) + (1-\alpha_2) \tilde{\alpha}_2 L_t \left(c_{t,K}\right| z) , \tag{42}$$

so that we can set

$$F_t^* \left(c_{t,K-1}\right| w) := \begin{cases} \alpha_2 F_t^U \left(c_{t,j_2-1}\right| w) + (1-\alpha_2) F_t^U \left(c_{t,j_2}\right| w) \\ \qquad \text{if } \overline{L}\left(c_{t,K-1}\right| z) \in \left[U_t \left(c_{t,j_2-1}\right| z) , U_t \left(c_{t,j_2}\right| z)\right] , \\ \tilde{\alpha}_2 F_t^L \left(c_{t,K-1}\right| w) + (1-\tilde{\alpha}_2) F_t^L \left(c_{t,K}\right| w) \\ \qquad \text{if } \left[L_t \left(c_{t,K-1}\right| z) , L_t \left(c_{t,K}\right| z)\right] . \end{cases} \tag{43}$$

which again ensures

$$F_t^* \left(c_{t,K-1}\right| z) = \overline{L}\left(c_{t,K-1}\right| z) .$$

We now show that the contructions of $F_t^* \left(c\right| w)$ at $c = c_{t,K}$ and $c_{t,K-1}$ in (39) and (43) satisfy the monotonicity requirement of a CDF, i.e.,

$$F_t^* \left(c_{t,K-1}\right| w) \leq F_t^* \left(c_{t,K}\right| w) . \tag{44}$$

To see this, first notice from (39) and the monotonicity of F_t^U that

$$F_t^*\left(c_{tK}\right|w) \geq F_t^U\left(c_{t,j_1-1}\right|w) \tag{45}$$

Next, consider the two cases in (43). If $\overline{L}\left(c_{t,K-1}\right|z) \in \left[L_t\left(c_{t,K-1}\right|z), L_t\left(c_{t,K}\right|z)\right]$,then by the construction and the monotonicity of F_t^L, F_t^U, we have

$$F_t^*\left(c_{t,K-1}\right|w) \leq F_t^L\left(c_{t,K-1}\right|w) \leq p_t\left(w\right) = F_t^U\left(c_{t,1}\right|w),$$

which together with (45) imply (44). If $\overline{L}\left(c_{t,K-1}\right|z) \in \left[U_t\left(c_{t,j_2-1}\right|z), U_t\left(c_{t,j_2}\right|z)\right]$, then by the monotonicity of $U_t\left(\cdot\right|z)$ we know that $j_2 \leq j_1$. If in addition $j_2 \leq j_1 - 1$, then by the monotonicity of F_t^U we have

$$F_t^*\left(c_{t,K-1}\right|w) \leq F_t^U\left(c_{t,j_2}\right|w) \leq F_t^U\left(c_{t,j_1-1}\right|w),$$

which together with (45) imply (44). Otherwise, we must have $j_2 = j_1$, then by (38), (41) and the monotonicity of $\overline{L}\left(\cdot\right|z)$, we must have

$$\alpha_2 \geq \alpha_1$$

since $\overline{L}\left(c_{t,K-1}\right|z) \leq \overline{L}\left(c_{t,K}\right|z)$ and $U_t\left(c_{t,j_2-1}\right|z) \leq U_t\left(c_{t,j_2}\right|z)$. Therefore, by (39), (43), and the monotonicity of F_t^U, we have

$$\begin{aligned} F_t^*\left(c_{t,K-1}\right|w) &= \alpha_2 F_t^U\left(c_{t,j_2-1}\right|w) + (1-\alpha_2) F_t^U\left(c_{t,j_2}\right|w) \\ &\leq \alpha_1 F_t^U\left(c_{t,j_1-1}\right|w) + (1-\alpha_1) F_t^U\left(c_{t,j_2}\right|w) = F_t^*\left(c_{t,K}\right|w). \end{aligned}$$

(iii) The construction of $F_t^*\left(c\right|w)$ at $c = c_{t,1}, ..., c_{t,K-2}$ can be carried out in the same way as $F_t^*\left(c_{t,K-1}\right|w)$, and is thus omitted here. In summary, F_t^* is constructed so that

$$F_t^*\left(c\right|z) = \overline{L}\left(c\right|z), \quad \forall c \in \mathcal{C}_t$$

Furthermore, since for each $c \in \mathcal{C}_t$, the value $F_t^*\left(c\right|w)$ is always constructed as a weighted average of (two) values in

$$\left\{F^L\left(c\right|w), F^U\left(c\right|w): \; c \in \mathcal{C}\right\},$$

by (35), we have

$$F_t^*\left(c\right|w) = p_t\left(w\right), \; \forall c \in \mathcal{C}_t.$$

Hence, for any w, we must have $w_t'\theta \in \mathcal{C}_t$ and thus

$$F_t^*\left(w_t'\theta\middle| w\right) = p_t\left(w\right),$$

which ensures (33).

Step 3:

We now show how to construct $F_t^*\left(c\right|w)$ at $c \in \mathcal{C}_{-t} := \mathcal{C} \backslash \mathcal{C}_t$ to ensure (34), using based on the previously assigned values of $F_t^*(c|w)$ for $c \in \mathcal{C}_t$ in Step 2. We rank-order elements in $\mathcal{C}_{-t}$ in strict ascending order as follows

$$c_{-t,1} < ... < c_{-t,\overline{K}} \text{ for some } \overline{K} \geq 1.$$

(i) We start with the largest element $c_{-t,\overline{K}}$. Then $\overline{L}\left(c_{-t,\overline{K}}\middle| z\right)$ can be expressed as follows in three different scenarios:

$$\overline{L}\left(c_{-t,\overline{K}}\middle| z\right) := \begin{cases} \alpha_3 \overline{L}(c_{t,K}|z) + (1-\alpha_3) \cdot 1, & \text{if } c_{-t,\overline{K}} > c_{t,K}, \\ \alpha_4 \overline{L}(c_{t,j-1}|z) + (1-\alpha_3) \overline{L}(c_{t,j}|z), & \text{if } c_{-t,\overline{K}} \in (c_{t,j-1}, c_{t,j}) \text{ for some } j, \\ \alpha_5 \cdot 0 + (1-\alpha_5) \overline{L}(c_{t,1}|z) & \text{if } c_{-t,\overline{K}} < c_{t,1}, \end{cases}$$

for some $\alpha_3, \alpha_4, \alpha_5 \in [0,1]$. Accordingly, we can set

$$F_t^*\left(c_{-t,\overline{K}}\middle| w\right) := \begin{cases} \alpha_3 F_t^*(c_{t,K}|w) + (1-\alpha_3) \cdot 1, & \text{if } c_{-t,\overline{K}} > c_{t,K}, \\ \alpha_4 F_t^*(c_{t,j-1}|z) + (1-\alpha_3) F_t^*(c_{t,j}|z), & \text{if } c_{-t,\overline{K}} \in (c_{t,j-1}, c_{t,j}), \\ \alpha_5 \cdot 0 + (1-\alpha_5) F_t^*(c_{t,1}|z). & \text{if } c_{-t,\overline{K}} < c_{t,1}. \end{cases} \tag{46}$$

(ii) If $\overline{K} > 1$, we now move to the second largest element $c_{-t,\overline{K}-1}$. Then $\overline{L}\left(c_{-t,\overline{K}-1}\middle| z\right)$ can be expressed as follows:

$$\overline{L}\left(c_{-t,\overline{K}}\middle| z\right) := \begin{cases} \alpha_6 \overline{L}(c_{t,K}|z) + (1-\alpha_6) \overline{L}\left(c_{-t,\overline{K}}\middle| z\right), & \text{if } c_{t,K} < c_{-t,\overline{K}-1} < c_{-t,\overline{K}} \\ \alpha_7 \overline{L}(c_{t,j_1-1}|z) + (1-\alpha_7) \overline{L}\left(c_{-t,\overline{K}}\middle| z\right), & \text{if } c_{t,j_1-1} < c_{-t,\overline{K}-1} < c_{-t,\overline{K}} < c_{t,j_1} \\ \alpha_8 \overline{L}(c_{t,j_2-1}|z) + (1-\alpha_8) \overline{L}(c_{t,j_2}|z), & \text{if } c_{t,j_2-1} < c_{-t,\overline{K}-1} < c_{t,j_2} < c_{-t,\overline{K}} \\ \alpha_9 \cdot 0 + (1-\alpha_9) \overline{L}(c_{t,1}|z), & \text{if } c_{-t,\overline{K}-1} < c_{-t,\overline{K}} < c_{t,1}, \end{cases}$$

Accordingly, we can set

$$F_t^*\left(c_{-t,\overline{K}-1}\middle| w\right) := \begin{cases} \alpha_6 F_t^*(c_{t,K}|w) + (1-\alpha_6) F_t^*\left(c_{-t,\overline{K}}\middle| w\right), & \text{if } c_{t,K} < c_{-t,\overline{K}-1} < c_{-t,\overline{K}}, \\ \alpha_7 F_t^*(c_{t,j_1-1}|w) + (1-\alpha_7) F_t^*\left(c_{-t,\overline{K}}\middle| w\right), & \text{if } c_{t,j_1-1} < c_{-t,\overline{K}-1} < c_{-t,\overline{K}} < c_{t,j_1}, \\ \alpha_8 F_t^*(c_{t,j_2-1}|w) + (1-\alpha_8) F_t^*(c_{t,j_2}|w), & \text{if } c_{t,j_2-1} < c_{-t,\overline{K}-1} < c_{t,j_2} < c_{-t,\overline{K}}, \\ \alpha_9 \cdot 0 + (1-\alpha_9) F_t^*(c_{t,1}|w), & \text{if } c_{-t,\overline{K}-1} < c_{-t,\overline{K}} < c_{t,1}. \end{cases}$$

(iii) Iteratively, $F_t^*(c|w)$ can be constructed in the same way at all $c \in \mathcal{C}_{-t}$. By construction, partial stationarity (34) and the monotonicity of $F_t^*(\cdot|w)$ are both satisfied on $\mathcal{C} = \mathcal{C}_t \cup \mathcal{C}_{-t}$.

Step 4:

Finally, we construct $F_t^*(c|w)$ for $c \in \mathcal{R}\backslash\mathcal{C}$. We set $F_t^*(c|w) = 0$ for $c < c_{t,1}$ and $F_t^*(c|w) = 1$ for $c \geq \overline{c} = \max \mathcal{C} + \delta$. For $c \in [c_{t,1}, \overline{c}]$, there must exists some $\tilde{c} \in \mathcal{C}$ s.t.

$c > \tilde{c} \in \mathcal{C}$ and $\overline{L}(c|\, z) = \overline{L}(\tilde{c}|\, z)$, and we then set

$$F_t^*(c|\, w) := F^*(\tilde{c}|\, w).$$

This guarantees (34) at any $c \in \mathcal{R} \backslash \mathcal{C}$.

This completes the construction $F_t^*(c|\, w)$ for all $c \in \mathcal{R}$ at each $t = 1, ..., T$. Together, we have ensured that:

(a) $F_t^*(\cdot|\, w)$ is a proper conditional CDF;

(b) partial stationarity holds since (34) is satisfied for all $c \in \mathcal{R}$;

(c) period-t marginal CCPs are matched since (33) holds for all $c \in \mathcal{C}_t$ (in Step 2).

Observe also that each $F_t^*(\cdot|\, w)$ defines a discrete distribution with finite support points. □

Lemma 2 (From Per-Period to All-Period Construction)**.** *There exists a well-defined joint distribution of* $(v_{i1}^*, ..., v_{iT}^*)$ *with period−*t *marginal CDF (conditional on* w*) given by*

$$F_t^*(\cdot|\, w)$$

as constructed in Lemma 1, such that (31) *holds.*

Proof. Recall from Lemma 1 that each $F_t^*(\cdot|\, w)$ defines a discrete distribution with finite support points. Let $\overline{\mathcal{C}}$ denote the union of support points of $F_t^*(\cdot|\, w)$ across all $t = 1, ..., T$, and let $f_t^*(\cdot|\, w)$ denote the corresponding probability mass function for $F_t^*(\cdot|\, w)$. Then, by definition,

$$F_t^*(c|\, w) = \sum_{\tilde{c} \in \overline{\mathcal{C}} : \tilde{c} \leq c} f_t^*(\tilde{c}|\, w), \quad \forall c.$$

We now show how to construct a joint pmf $f^*(\cdot|\, w)$ whose period-t marginals are given by $f_t^*(\cdot|\, w)$.

For each t, define

$$c_t^* := \max\left\{c \in \overline{\mathcal{C}} : \ F_t^*(c|\, w) = F_t^*\left(w_t'\theta \middle|\, w\right)\right\}, \tag{47}$$

which exists and is unique by the construction in Lemma 1.

For each $\mathbf{c} \equiv (c_1, ..., c_T) \in \overline{\mathcal{C}}^T$, write

$$\begin{aligned} y_t(c_t) &:= \mathbb{1}\{c_t \leq c_t^*\}, \\ y(\mathbf{c}) &:= (y_1(c_1), ..., y_T(c_T))'. \end{aligned}$$

and define

$$f^*(\mathbf{c}|\, w) := p(y(\mathbf{c})|\, w) \prod_{t=1}^{T} \frac{f_t^*(c_t|\, w)}{p_t(w)^{y_t(c_t)}(1 - p_t(w))^{1 - y_t(c_t)}}, \tag{48}$$

under the convention $0^0 = 1$.

We show that $f^*(\cdot|\, w)$ is a probability mass function that characterizes a well-defined joint distribution of $(v_{i1}^*, ..., v_{iT}^*)$ and satisfies the requirements in Lemma 2.

Step 1:

First, note that the right-hand (48) only involves known (observed or constructed) quantities. In particular:

- $p(y|\, w) := \mathbb{P}(Y_{it} = y_t \forall t = 1, ..., T|\, W_i = w)$ is the (observed) joint CCP of observing a particular path of outcomes y across all periods, given $W_i = w$.
- $f_t^*(c|\, w)$ is the period-t marginal pmf corresponding to $F_t^*(c|\, w)$ defined in Lemma 1.
- $f_t(w) = \mathbb{P}(Y_{it} = 1|\, W_i = w)$ is the observed period-t marginal CCP, with

$$p_t(w) = F_t(c_t^*|\, w) = \sum_{\tilde{c} \in \overline{\mathcal{C}}: \tilde{c} \leq c_t^*} f_t^*(\tilde{c}|\, w). \tag{49}$$

Step 2:

We show that the period-t marginal pmf implied by $f^*(\cdot|\, w)$ coincides with $f_t^*(\cdot|\, w)$. To see this, observe that, for any t and $y_t \in \{0, 1\}$, we have

$$\begin{aligned}
&\sum_{c_t \in \overline{\mathcal{C}}: y_t(c_t) = y_t} \frac{f_t^*(c_t|\, w)}{p_t(w)^{y_t(c_t)} (1 - p_t(w))^{1 - y_t(c_t)}} \\
=& y_t \sum_{c_t \leq c_t^*} \frac{f_t^*(c_t|\, w)}{p_t(w)} + (1 - y_t) \sum_{c_t > c_t^*} \frac{f_t^*(c_t|\, w)}{1 - p_t(w)} \\
=& y_t \frac{\sum_{c_t \leq c_t^*} f_t^*(c_t|\, w)}{\sum_{c_t \leq c_t^*} f_t^*(c_t|\, w)} + (1 - y_t) \frac{\sum_{c_t > c_t^*} f_t^*(c_t|\, w)}{\sum_{c_t > c_t^*} f_t^*(c_t|\, w)} \text{ by (49)} \\
=& y_t \cdot 1 + (1 - y_t) \cdot 1 \\
=& 1,
\end{aligned} \tag{50}$$

Hence, for any $c_t \in \overline{\mathcal{C}}$, the period-$t$ marginal implied by $f^*(\cdot|\, w)$ is

$$
\begin{aligned}
&\sum_{c_{-t}\in\overline{\mathcal{C}}^{T-1}} f^*(c_t, c_{-t}|\, w) \\
=&\frac{f_t^*(c_t|\, w)}{p_t(w)^{y_t(c_t)}(1-p_t(w))^{1-y_t(c_t)}}\sum_{c_{-t}} p(y(c_t,c_{-t})|\, w)\prod_{s\neq t}\frac{f_s^*(c_s|\, w)}{p_s(w)^{y_s(c_s)}(1-p_s(w))^{1-y_s(c_s)}} \\
=&\frac{f_t^*(c_t|\, w)}{p_t(w)^{y_t(c_t)}(1-p_t(w))^{1-y_t(c_t)}}\sum_{y_{-t}} p(y_t(c_t), y_{-t}|\, w)\sum_{c_{-t}:y_{-t}(c_{-t})=y_{-t}}\prod_{s\neq t}\frac{f_s^*(c_s|\, w)}{p_s(w)^{y_s(c_s)}(1-p_s(w))^{1-y_s(c_s)}} \\
=&\frac{f_t^*(c_t|\, w)}{p_t(w)^{y_t(c_t)}(1-p_t(w))^{1-y_t(c_t)}}\sum_{y_{-t}} p(y_t(c_t), y_{-t}|\, w)\prod_{s\neq t}\sum_{c_s:y_s(c_s)=y_s}\frac{f_s^*(c_s|\, w)}{p_s(w)^{y_s(c_s)}(1-p_s(w))^{1-y_s(c_s)}} \\
=&\frac{f_t^*(c_t|\, w)}{p_t(w)^{y_t(c_t)}(1-p_t(w))^{1-y_t(c_t)}}\sum_{y_{-t}} p(y_t(c_t), y_{-t}|\, w)\prod_{s\neq t} 1 \text{ by } (50) \\
=&\frac{f_t^*(c_t|\, w)}{p_t(w)^{y_t(c_t)}(1-p_t(w))^{1-y_t(c_t)}} p_t(w)^{y_t(c_t)}(1-p_t(w))^{1-y_t(c_t)} \\
=&f_t^*(c_t|\, w).
\end{aligned}
$$

Step 3:

We show that $f^*(\cdot|\, w)$ is a valid joint pmf. Clearly, $f^*(\mathbf{c}|\, w) \geq 0$, since all quantities on the right-hand side of (48) are nonnegative. In addition, since the period-t marginal of $f^*(\cdot|\, w)$ coincides with $f_t^*(\cdot|\, w)$ as established in (2), we must have

$$\sum_{\mathbf{c}} f^*(\mathbf{c}|\, w) = \sum_{c_t} f_t^*(c_t|\, w) = 1.$$

Hence, $f^*(c|\, w)$ is a valid pmf and thus characterizes a well-defined joint distribution of $(v_{i1}^*, ..., v_{iT}^*)$.

Step 4:

Lastly, we show that (31) holds under $f^*(\cdot|\, w)$. For any $y \in \{0,1\}^T$,

$$
\begin{aligned}
&\mathbb{P}\left(v_{it}^* \leq w_t'\theta \forall t \text{ s.t. } y_t = 1,\ v_{is}^* > w_s'\theta \forall s \text{ s.t. } y_s = 0 \mid w\right), \\
=& \sum_{\mathbf{c}} f^*(\mathbf{c}|\, w)\, \mathbb{1}\{c_t \leq c_t^* \,\forall t \text{ s.t. } y_t = 1,\ c_s > c_s^* \,\forall s \text{ s.t. } y_s = 0\} \\
=& \sum_{\mathbf{c}:y(\mathbf{c})=y} f^*(\mathbf{c}|\, w) \\
=& \sum_{\mathbf{c}:y(\mathbf{c})=y} p(y(\mathbf{c})|\, w) \prod_{t=1}^{T} \frac{f_t^*(c_t|\, w)}{p_t(w)^{y_t(c_t)} (1-p_t(w))^{1-y_t(c_t)}}, \\
=& p(y|\, w) \sum_{\mathbf{c}:y(\mathbf{c})=y} \prod_{t=1}^{T} \frac{f_t^*(c_t|\, w)}{p_t(w)^{y_t(c_t)} (1-p_t(w))^{1-y_t(c_t)}} \\
=& p(y|\, w) \prod_{t=1}^{T} \left(\sum_{c_t:y_t(c_t)=y_t} \frac{f_t^*(c_t|\, w)}{p_t(w)^{y_t} (1-p_t(w))^{1-y_t}} \right) \\
=& p(y|\, w) \prod_{t=1}^{T} 1 \text{ by (50)} \\
=& p(y|\, w).
\end{aligned}
$$

□

Lemma 3. *[From Discrete to Gegeneral X] Suppose that the support of $(X_{it})_{t=1}^T$ is bounded. Then Θ_I is sharp (regardless of whether X_{it} is discrete, continuous, or mixed).*

Proof. For each $M \in \mathbb{N}$, define

$$X_i^{(M)} := \frac{1}{2^M} \lceil 2^M X_i \rceil$$

where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x. Since $\mathcal{X} := \text{Supp}(X_i)$ is bounded,

$$\mathcal{X}^{(M)} := \text{Supp}\left(X_i^{(M)}\right) \subseteq \left\{ \frac{m}{2^M} : m \in \mathbb{Z}, \inf \lceil 2^M \mathcal{X} \rceil \leq m \leq \sup \lceil 2^M \mathcal{X} \rceil \right\}$$

is by construction finite for each M. Furthermore,

$$\mathcal{X}^{(M)} \subseteq \mathcal{X}^{(M+1)} \subseteq ...$$

is an increasing sequence of subsets that are becoming dense in $\mathcal{X}$. In addition, observe that, by construction, for each realization $x^{(M)} \in \mathcal{X}^{(M)}$, we have

$$X_i^{(M)} \leq x^{(M)} \quad \Leftrightarrow \quad X_i \leq x^{(M)}. \tag{51}$$

Fix any $\theta \equiv (\beta', \gamma')' \in \Theta_I \backslash \{\theta_0\}$. For each realization z, since θ satisfies the identifying inequality (10) for each $c \in \mathcal{R}$, it must in particular satisfy inequality (10) for

$$c \in \mathcal{C}^M := \bigcup_{t=1}^{T} \text{Supp}\left(z_t'\beta + X_{it}^{(M)'}\gamma\right). \tag{52}$$

For each $x^{(M)} \in \mathcal{X}^{(M)} := \text{Supp}\left(X_i^{(M)}\right)$, define

$$\begin{aligned} p_t^{(M)}\left(x^{(M)}, z\right) &:= \mathbb{P}\left(Y_{it} = 1 \middle| X_i^{(M)} = x^{(M)}, Z_i = z\right) \\ &\equiv \mathbb{P}\left(Y_{it} = 1 \middle| 2^M x^{(M)} - 1 < X_i \leq 2^M x^{(M)}, Z_i = z\right) \end{aligned}$$

whenever

$$\mathbb{P}\left(X_i^{(M)} = x^{(m)} \middle| Z_i = z\right) \equiv \mathbb{P}\left(2^M x^{(M)} - 1 < X_i \leq 2^M x^{(M)} \middle| Z_i = z\right) > 0.$$

In the (irrelevant) case where $\mathbb{P}\left(X_i^{(M)} = x^{(m)} \middle| Z_i = z\right) = 0$, we can set $p_t^{(M)}\left(x^{(M)}, z\right)$ arbitrarily, say, to be zero. Note that $p_t^{(M)}(\cdot,\cdot)$ is a well-defined family of CCPs for $Y_{it} = 1$ given $\left(X_i^{(M)}, Z_i\right) = \left(x^{(M)}, z\right)$.

The finiteness of $\mathcal{X}^{(M)}$ and (52) imply that the conditions for Lemma 1 is satisfied with $X_i^{(M)}$ in lieu of X_i. Hence, by Lemma 1, there exists $\epsilon_{it}^{*(M)}$ with CDF $F_{\epsilon_t}^{*(M)}$ such that partial stationarity holds, i.e.,

$$F_{\epsilon_1}^{*(M)}\left(c \middle| z\right) = ... = F_{\epsilon_T}^{*(M)}\left(c \middle| z\right), \ \forall c \in \mathcal{R} \tag{53}$$

and that

$$F_t^{*(M)}\left(z_t'\beta + x_t^{(M)'}\gamma \middle| x^{(M)}, z\right) = p_t^{(M)}\left(x^{(M)}, z\right).$$

Define

$$Y_{it}^{(M)} := \mathbb{1}\left\{Z_{it}'\beta + X_{it}^{(M)'}\gamma + \epsilon_{it}^{*(M)} \geq 0\right\}. \tag{54}$$

By Lemma 2, there exists a well-defined joint distribution of $\epsilon_i^{*(M)} := \left(\epsilon_{it}^{*(M)}\right)_{t=1}^{T}$ such that partial stationarity (53) holds and that

$$\mathbb{P}\left(Y_i^{(M)} = y \middle| X_i^{(M)} = x^{(m)}, Z_i = z\right) = \mathbb{P}\left(Y_i = y \middle| X_i^{(M)} = x^{(m)}, Z_i = z\right)$$

whenever $\mathbb{P}\left(X_i^{(M)} = x^{(m)} \middle| Z_i = z\right) > 0$. Consequently, the joint distribution of

$$\left(Y_i^{(M)}, X_i^{(M)}, Z_i\right)$$

is produced under (θ, ϵ_i^*) and (54) coincides with the true joint distribution of $\left(Y_i, X_i^{(M)}, Z_i\right)$.

To summarize the above, let F denote the joint CDF of $(Y_i, X_i, Z_i, \epsilon_i)$ in the true DGP, and let $F^{*(M)}$ denote the joint CDF of $\left(Y_i^{(M)}, X_i^{(M)}, Z_i, \epsilon_i^{*(M)}\right)$ as constructed above.

Since (i) the joint distribution of $\left(Y_i^{(M)}, X_i^{(M)}, Z_i\right)$ coincides with the joint distribution of $\left(Y_i, X_i^{(M)}, Z_i\right)$, and (ii) $X_i^{(M)} \leq x^{(M)}$ if and only if $X_i \leq x^{(M)}$ for each $x^{(M)} \in \mathcal{X}^{(M)}$ by (51), we must have

$$F^{*(M)}\left(y, x^{(M)}, z, \infty\right) = F\left(y, x^{(M)}, z, \infty\right) \quad \forall x^{(M)} \in \mathcal{X}^{(M)}, \tag{55}$$

since $F\left(y, x^{(M)}, z, \infty\right)$ represents the joint probability of $\left(Y_i \leq y, X_i \leq x^{(M)}, Z_i \leq z\right)$ and similarly for $F^{*(M)}\left(y, x^{(M)}, z, \infty\right)$.

Now, fix any $x \in \mathcal{X}$ (which may contain a continuum). Clearly, there eixsts an $\underline{M} \in \mathbb{N}$ s.t. $x < 2^M$ for all $M \geq \underline{M}$. Then, for all $M > \underline{M}$, define

$$x^{(M)} := \min\left\{\tilde{x}^{(m)} \in \mathcal{X}^{(M)} : \tilde{x}^{(m)} \geq x\right\}.$$

Then $x^{(M)}$ is well-defined and weakly decreasing in M since $\mathcal{X}^{(M)} \subseteq \mathcal{X}^{(M+1)}$ for all $M \geq \underline{M}$. Since $\mathcal{X}^{(M)}$ becomes dense in $\mathcal{X}$ as $M \to \infty$, we have

$$x^{(M)} \searrow x. \tag{56}$$

Now, for any $y \in \{0,1\}^T, z \in \mathcal{R}^{d_z \times T}, \epsilon \in \mathcal{R}^T$, the sequence of numbers lie within the compact set $[0,1]$, so there must be a convergent subsequence, say, indexed namely by m_M such that

$$\lim_{M \to \infty} F^{*(m_M)}\left(y, x^{(m_M)}, z, \epsilon\right) \text{ exists.}$$

We then define

$$F^{*\infty}(y, x, z, \epsilon) := \lim_{M \to \infty} F^{*(m_M)}\left(y, x^{(m_M)}, z, \epsilon\right), \forall y, z, \epsilon. \tag{57}$$

It is known, e.g., by Chapter 2.10 of Nelsen (2006), that a multivariate function $F : \mathcal{R}^d \to [0,1]$ is a valid CDF if and only if the following defining properties hold: (1) $F(..., u_j, ...) = 0$ if $u_j = -\infty$ for any $j = 1, ..., d$, (2) $F(\infty, ..., \infty) = 1$, (3) F is weakly d-increasing, i.e., for any hyper-rectangle $B = \prod_{j=1}^d [a_j, b_j]$,

$$\sum_{z \in \prod_{j=1}^d \{a_j, b_j\}} (-1)^{\#\{k : z_k = a_k\}} F(z) \geq 0,$$

and (4) F is right-continuous.

Clearly, properties (1)-(3) are preserved under the operation of taking limits, and thus $F^{*\infty}$ satisfies (1)-(3). Furthermore, if $F^{*\infty}$ is not right-continuous at any point (y, x, z, ϵ), there exists a right-continuous modification F^* of $F^{*\infty}$, which sets the value of F^* at any point as the right limit of $F^{*\infty}$ at that point. Note that the right-continuous modification of $F^{*\infty}$ described above does not affect properties (1)-(3), and thus F^* by construction satisfies (1)-(4).

Hence, F^* is a valid (multivariate) CDF that defines a well-defined joint distribution of $(Y_i^*, X_i, Z_i, \epsilon_i^*)$. Furthermore, for any (y, x, z), the joint CDF of (Y_i^*, X_i, Z_i) at (y, x, z) can be obtained by evaluating $F^*(y, x, z, \infty)$, i.e., with ϵ set to be infinity. Then, for any (y, x, z) at which $F^*(\cdot, \cdot, \cdot, \infty)$ is continuous,[11] we have

$$\begin{aligned} F^*(y, x, z) &= F^*(y, x, z, \infty) \\ &= F^{*\infty}(y, x, z, \infty) \\ &= \lim_{M\to\infty} F^{*(m_M)}\left(y, x^{(m_M)}, z, \infty\right) \text{ by (57)} \\ &= \lim_{M\to\infty} F\left(y, x^{(m_M)}, z\right) \quad \text{by (55)} \\ &= F(y, x, z) \text{ since } x^{(m_M)} \searrow x \text{ and } F \text{ is right-continuous.} \end{aligned}$$

Hence, the distribution of observable (Y_i, X_i, Z_i) under F^* coincides with that under F, i.e., F^* is observationally equivalent to F. □

A.3 Reconciliation with Khan, Ponomareva, and Tamer (2023)

We show that under Assumption 1 and $X_{it} = Y_{i,t-1}$, our identifying condition (10) implies the following result in Khan, Ponomareva, and Tamer (2023):

KPT(i): $\mathbb{P}(Y_{it} = 1 | z) > \mathbb{P}(Y_{is} = 1 | z) \Rightarrow (z_t - z_s)' \beta_0 + |\gamma_0| > 0$.

KPT(ii): $\mathbb{P}(Y_{it} = 1 | z) > 1 - \mathbb{P}(Y_{i,s} = 0, Y_{i,s-1} = 1 | z) \Rightarrow (z_t - z_s)' \beta_0 - \min\{0, \gamma_0\} > 0$.

KPT(iii): $\mathbb{P}(Y_{it} = 1 | z) > 1 - \mathbb{P}(Y_{i,s} = 0, Y_{i,s-1} = 0 | z) \Rightarrow (z_t - z_s)' \beta_0 + \max\{0, \gamma_0\} > 0$.

KPT(iv): $\mathbb{P}(Y_{it} = 1, Y_{it-1} = 1 | z) > \mathbb{P}(Y_{is} = 1 | z) \Rightarrow (z_t - z_s)' \beta_0 + \max\{0, \gamma_0\} > 0$.

KPT(v): $\mathbb{P}(Y_{it} = 1, Y_{it-1} = 1 | z) > 1 - \mathbb{P}(Y_{is} = 0, Y_{i,s-1} = 1 | z) \Rightarrow (z_t - z_s)' \beta_0 > 0$.

KPT(vi): $\mathbb{P}(Y_{it} = 1, Y_{it-1} = 1 | z) > 1 - \mathbb{P}(Y_{is} = 0, Y_{i,s-1} = 0 | z) \Rightarrow (z_t - z_s)' \beta_0 + \gamma_0 > 0$.

KPT(vii): $\mathbb{P}(Y_{it} = 1, Y_{it-1} = 0 | z) > 1 - \mathbb{P}(Y_{is} = 0 | z) \Rightarrow (z_t - z_s)' \beta_0 - \min\{0, \gamma_0\} > 0$.

KPT(viii): $\mathbb{P}(Y_{it} = 1, Y_{it-1} = 0 | z) > 1 - \mathbb{P}(Y_{is} = 0, Y_{i,s-1} = 1 | z) \Rightarrow (z_t - z_s)' \beta_0 - \gamma_0 > 0$.

KPT(ix): $\mathbb{P}(Y_{it} = 1, Y_{it-1} = 0 | z) > 1 - \mathbb{P}(Y_{is} = 0, Y_{i,s-1} = 0 | z) \Rightarrow (z_t - z_s)' \beta_0 > 0$.

Proof. With $X_{it} = Y_{i,t-1}$, our inequality restriction (15) can be equivalently rewritten as follows:

$$\begin{aligned} &\mathbb{P}(Y_{it} = 1,\ Y_{i,t-1} = 1 | z)\, \mathbb{1}\left\{z_t' \beta_0 + \gamma_0 \le c\right\} + \mathbb{P}(Y_{it} = 1,\ Y_{i,t-1} = 0 | z)\, \mathbb{1}\left\{z_t' \beta_0 \le c\right\} \\ &\le 1 - \mathbb{P}(Y_{is} = 0,\ Y_{i,s-1} = 1 | z)\, \mathbb{1}\left\{z_s' \beta_0 + \gamma_0 \ge c\right\} - \mathbb{P}(Y_{is} = 0,\ Y_{i,s-1} = 0 | z)\, \mathbb{1}\left\{z_s' \beta_0 \ge c\right\}, \end{aligned} \quad (58)$$

[11] Note that a distribution is completely pinned down by the values of its CDF at continuous points, given the right-continuity of CDFs.

by enumerating the realization of $Y_{i,t-1}$.

Note that the lower and upper expressions in the inequality (58) both have three possible (informative) outcomes depending on the value of c, leading to the 9 inequalities in KPT. We derive the first two inequalities KPT(i) and KPT(ii), and the rest of inequalities can be derived in the same way.

KPT(i): consider the event that all indicators in condition (58) are equal to one, saying that

$$\max\{z_t^{'}\beta_0 + \gamma_0, z_t^{'}\beta_0\} \le c \le \min\{z_s^{'}\beta_0 + \gamma_0, z_s^{'}\beta_0\},$$

which is equivalent to

$$z_t^{'}\beta_0 + \max\{0, \gamma_0\} - (z_s^{'}\beta_0 + \min\{0, \gamma_0\}) = (z_t - z_s)^{'}\beta_0 + |\gamma_0| \le 0.$$

Then, when $(z_t - z_s)^{'}\beta_0 + |\gamma_0| \le 0$, condition (58) becomes

$$\begin{aligned}
\mathbb{P}\left(Y_{it} = 1 \middle| z\right) &= \mathbb{P}\left(Y_{it} = 1,\ Y_{i,t-1} = 1 \middle| z\right) + \mathbb{P}\left(Y_{it} = 1,\ Y_{i,t-1} = 0 \middle| z\right) \\
&\le\ 1 - \mathbb{P}\left(Y_{is} = 0,\ Y_{i,s-1} = 1 \middle| z\right) - \mathbb{P}\left(Y_{is} = 0,\ Y_{i,s-1} = 0 \middle| z\right) \\
&= 1 - \mathbb{P}(Y_{is} = 0 \mid z) = \mathbb{P}\left(Y_{is} = 1 \middle| z\right).
\end{aligned}$$

By contraposition, it implies the same restriction in KPT(i):

$$\mathbb{P}\left(Y_{it} = 1 \middle| z\right) > \mathbb{P}\left(Y_{is} = 1 \middle| z\right) \Longrightarrow (z_t - z_s)^{'}\beta_0 + |\gamma_0| > 0.$$

KPT(ii): we first relax condition (58) by dropping the last term in the upper expression $\mathbb{P}\left(Y_{is} = 0,\ Y_{i,s-1} = 0 \middle| z\right) \mathbb{1}\left\{z_s^{'}\beta_0 \ge c\right\}$ and have the following relaxed inequality:

$$\begin{aligned}
&\mathbb{P}\left(Y_{it} = 1,\ Y_{i,t-1} = 1 \middle| z\right) \mathbb{1}\left\{z_t^{'}\beta_0 + \gamma_0 \le c\right\} + \mathbb{P}\left(Y_{it} = 1,\ Y_{i,t-1} = 0 \middle| z\right) \mathbb{1}\left\{z_t^{'}\beta_0 \le c\right\} \\
&\le 1 - \mathbb{P}\left(Y_{is} = 0,\ Y_{i,s-1} = 1 \middle| z\right) \mathbb{1}\left\{z_s^{'}\beta_0 + \gamma_0 \ge c\right\}.
\end{aligned} \tag{59}$$

Now, consider the event that the indicators in the above restriction are all equal to one, which implies that

$$\max\{z_t^{'}\beta_0 + \gamma_0, z_t^{'}\beta_0\} \le c \le z_s^{'}\beta_0 + \gamma_0,$$

and it is equivalent to the following condition:

$$(z_t - z_s)^{'}\beta_0 + \max\{0, \gamma_0\} - \gamma_0 = (z_t - z_s)^{'}\beta_0 - \min\{0, \gamma_0\} \le 0.$$

Given the above event, condition (59) becomes

$$\begin{aligned}
\mathbb{P}\left(Y_{it} = 1 \middle| z\right) &= \mathbb{P}\left(Y_{it} = 1,\ Y_{i,t-1} = 1 \middle| z\right) + \mathbb{P}\left(Y_{it} = 1,\ Y_{i,t-1} = 0 \middle| z\right) \\
&\le\ 1 - \mathbb{P}\left(Y_{is} = 0,\ Y_{i,s-1} = 1 \middle| z\right).
\end{aligned}$$

Similarly, we can derive the same restriction in KPT(ii) by contraposition:

$$\mathbb{P}\left(Y_{it}=1|\,z\right)>1-\mathbb{P}\left(Y_{is}=0,\ Y_{i,s-1}=1|\,z\right)\Longrightarrow\left(z_t-z_s\right)'\beta_0-\min\{0,\gamma_0\}>0.$$

□

A.4 Proof of Propositions 3 and 4

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 (10) 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 (10) is given as

$$1-\mathbb{P}(Y_{i1}=0,z_1'\beta_0+X_{i1}'\gamma_0\geq c\mid z)\geq\mathbb{P}(Y_{i2}=1,z_2'\beta_0+X_{i2}'\gamma_0\leq c\mid z),$$

for $c\in\{z_t'\beta_0+x_k'\gamma_0,\ t=1,2,k=1,...,K\}$, and another identifying condition switches the order of period 1 and 2.

Let $c=z_1'\beta_0+x_k'\gamma_0$,[12], then the above upper bound can be further bounded as

$$1-\mathbb{P}(Y_{i1}=0,z_1'\beta_0+X_{i1}'\gamma_0\geq z_1'\beta_0+x_k'\gamma_0\mid z)\leq1-\mathbb{P}(Y_{i1}=0,X_{i1}=x_k\mid z).$$

When $z_1'\beta_0-z_2'\beta_0\geq0$ which implies $z_1'\beta_0+x_k'\gamma_0\geq z_2'\beta_0+x_k'\gamma_0$, then the lower bound can be bounded below as

$$\mathbb{P}(Y_{i2}=1,z_2'\beta_0+X_{i2}'\gamma_0\leq z_1'\beta_0+x_k'\gamma_0\mid z)\leq\mathbb{P}(Y_{i2}=1,X_{i2}=x_k\mid z).$$

Combining the above results leads to

$$\text{If } z_1'\beta_0-z_2'\beta_0\geq0\Longrightarrow1-\mathbb{P}(Y_{i1}=0,X_{i1}=x_k\mid z)\geq\mathbb{P}(Y_{i2}=1,X_{i2}=x_k\mid z).$$

The contraposition of the above inequality yields

$$1-\mathbb{P}(Y_{i1}=0,X_{i1}=x_k\mid z)<\mathbb{P}(Y_{i2}=1,X_{i2}=x_k\mid z)\Longrightarrow\Delta z'\beta_0>0.$$

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

$$1-\mathbb{P}(Y_{i1}=1,X_{i1}=x_k\mid z)<\mathbb{P}(Y_{i2}=0,X_{i2}=x_k\mid z)\Longrightarrow\Delta z'\beta_0<0.$$

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

[12]The value of $c=z_2'\beta_0+x_k'\gamma_0$ leads to the same identifying condition.

it, we will show that for any $\beta \neq k\beta_0$ for some k, there exists some value Δz such that $\Delta z'b$ 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 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 $\Delta z^{j^*} \to +\infty$ 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 $\Delta z^{j*} \to -\infty$ 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 $\Delta z^{j^*} \to +\infty$ 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

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

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 4 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.5 Proof of Proposition 5

Proof. By (20), we have

$$\mathbb{P}\left(Y_{it} = 1,\ w_t'\theta_0 \leq c \middle| W_i = w\right) \leq F_t\left(c|\, w\right) \leq 1 - \mathbb{P}\left(Y_{it} = 0,\ w_t'\theta_0 \geq c \middle| W_i = w\right).$$

Since $\tilde{p}_t\left(\tilde{w}\right) = F_t\left(\tilde{w}_t'\theta_0 \middle|\, w\right)$, we have

$$\mathbb{P}\left(Y_{it} = 1,\ w_t'\theta_0 \leq \tilde{w}_t'\theta_0 \middle| W_i = w\right) \leq \tilde{p}_t\left(\tilde{w}\right) \leq 1 - \mathbb{P}\left(Y_{it} = 0,\ w_t'\theta_0 \geq \tilde{w}_t'\theta_0 \middle| W_i = w\right),$$

and hence

$$\inf_{\theta\in\Theta_I}\mathbb{P}\left(Y_{it}=1,\ w_t'\theta\leq\tilde{w}_t'\theta\middle|W_i=w\right)\leq\tilde{p}_t\left(\tilde{w}\right)\leq 1-\inf_{\theta\in\Theta_I}\mathbb{P}\left(Y_{it}=0,\ w_t'\theta\geq\tilde{w}_t'\theta\middle|W_i=w\right).$$

□

A.6 Proof of Proposition 9

Proof. Let $v_{ijt} := \alpha_{ij} + \epsilon_{ijt}$, for any set $K \subset \mathcal{J}$, the probability of selecting a choice $j \in K$ conditional on $W_i = w$ is given as:

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

The above observed probability restricts the conditional distribution of $v_{ikt} - v_{ijt} \mid w$ and can be exploited to bound this distribution.

We define $Q_t(c_{jk} \mid w)$ as follows: for $c_{jk} \in \mathcal{R}$,

$$Q_t(c_{jk} \mid w) := \mathbb{P}\left(\exists j \in K \text{ s.t. } v_{ikt} - v_{ijt} \leq c_{jk}\ \forall k \in \mathcal{J} \setminus K \mid w\right).$$

Then, we can derive lower and upper bounds for the above probability using variations in observed choice probabilities. When c_{jk} satisfies $c_{jk} \geq (w_{ijt} - w_{ikt})'\theta_0$ for any $j \in K$ and $k \in \mathcal{J} \setminus K$, then $Q_t(c_{jk} \mid w)$ can be bounded below as

$$\begin{aligned} Q_t(c_{jk} \mid w) &\geq \mathbb{P}\left(\exists j \in K \text{ s.t. } v_{ikt} - v_{ijt} \leq (w_{ijt} - w_{ikt})'\theta_0\ \forall k \in \mathcal{J} \setminus K \mid w\right) \\ &= \mathbb{P}(Y_{it} \in K \mid w). \end{aligned}$$

Therefore, the lower bound for $Q_t(c_{jk} \mid w)$ is established as

$$Q_t(c_{jk} \mid w) \geq \mathbb{P}(Y_{it} \in K,\ c_{jk} \geq (w_{ijt} - w_{ikt})'\theta_0\ \forall j \in K, k \in k \in \mathcal{J} \setminus K \mid w).$$

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

By taking expectation of X_i given z, we can bound the conditional distribution $Q_t(c_{jk} \mid z)$ as

$$\begin{aligned} Q_t(c_{jk} \mid z) &\geq \mathbb{P}\left(Y_{it} \in K, c_{jk} \geq (z_{ijt} - z_{ikt})'\beta_0 + (X_{ijt} - X_{ikt})'\gamma_0\ \forall j \in K, k \in \mathcal{J} \setminus K \mid z\right) \\ &= \mathbb{P}\left(Y_{it}^K = 1, c_{jk} \geq (z_{ijt} - z_{ikt})'\beta_0 + (X_{ijt} - X_{ikt})'\gamma_0\ \forall j \in K, k \in \mathcal{J} \setminus K \mid z\right). \end{aligned}$$

Similarly, the conditional probability $Q_t(c_{jk} \mid w)$ can be bounded above as

$$\begin{aligned} Q_t(c_{jk} \mid w) \leq \mathbb{P}(Y_{it}^K = 1 \mid w)\mathbb{1}\{c_{jk} \leq (w_{ijt} - w_{ikt})'\theta_0\ \forall j \in K, \mathcal{J} \setminus K\} + \\ 1 - \mathbb{1}\{c_{jk} \leq (w_{ijt} - w_{ikt})'\theta_0\ \forall j \in K, k \in \mathcal{J} \setminus K\}. \end{aligned}$$

The above inequality holds since either $c_{jk} \leq (w_{ijt} - w_{ikt})'\theta_0$ or the upper bound is one with $c_{jk} > (w_{ijt} - w_{ikt})'\theta_0$. After taking expectation of X_i given z, the upper bound for $Q_t(c_{jk} \mid z)$

is obtained as

$$
\begin{aligned}
Q_t(c_{jk} \mid z) \leq \mathbb{P}\left(Y_{it}^K = 1, c_{jk} \leq (z_{ijt} - z_{ikt})'\beta_0 + (X_{ijt} - X_{ikt})'\gamma_0 \ \forall j \in K, k \in K^c \mid z\right) \\
+ 1 - \mathbb{P}\left(c_{jk} \leq (z_{ijt} - z_{ikt})'\beta_0 + (X_{ijt} - X_{ikt})'\gamma_0 \ \forall j \in K, k \in \mathcal{J} \setminus K \mid z\right).
\end{aligned}
$$

Rearranging the above formula yields

$$
Q_t(c_{jk} \mid z) \leq 1 - \mathbb{P}\left(Y_{it}^K = 0, c_{jk} \leq (z_{ijt} - z_{ikt})'\beta_0 + (X_{ijt} - X_{ikt})'\gamma_0 \ \forall j \in K, k \in \mathcal{J} \setminus K \mid z\right).
$$

Under Assumption 1, the conditional probability $Q_t(c_{jk} \mid z)$ is the same for any t. Therefore, the smallest upper bound of $Q_t(c_{jk} \mid z)$ should be larger than the largest lower bound over all periods, yielding the identifying condition (26) as follows:

$$
\begin{aligned}
1 - \max_{s=1,\ldots,T} \mathbb{P}(Y_{is}^K = 0, (z_{js} - z_{ks})'\beta_0 + (X_{ijs} - X_{iks})'\gamma_0 \geq c_{jk} \ \forall j \in K, k \in \mathcal{J} \setminus K \mid z) \\
\geq \max_{t=1,\ldots,T} \mathbb{P}(Y_{it}^K = 1, (z_{jt} - z_{kt})'\beta_0 + (X_{ijt} - X_{ikt})'\gamma_0 \leq c_{jk} \ \forall j \in K, k \in \mathcal{J} \setminus K \mid z).
\end{aligned}
$$

□

A.7 Proof of Proposition 10

Proof. Since the observed outcome Y_{it} is censored at 0, 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,

$$
\mathbb{P}(Y_{it} = 0 \mid w) = \mathbb{P}(Y_{it}^* \leq 0 \mid w) = \mathbb{P}(v_{it} \geq z_t'\beta_0 + x_t'\gamma_0 \mid w).
$$

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

$$
\begin{aligned}
\mathbb{P}(Y_{it} \leq y \mid w) &= \mathbb{P}(Y_{it}^* \leq 0,\ Y_{it} \leq y \mid w) + \mathbb{P}(0 < Y_{it}^*,\ Y_{it} \leq y \mid w) \\
&= \mathbb{P}(Y_{it}^* \leq 0 \mid w) + \mathbb{P}(0 < Y_{it}^* \leq y \mid w) \\
&= \mathbb{P}(Y_{it}^* \leq y \mid w) \\
&= \mathbb{P}(v_{it} \geq z_t'\beta_0 + x_t'\gamma_0 - y \mid w).
\end{aligned}
$$

Combining the two scenarios, the conditional distributional of $Y_{it} \mid W_i$ is characterized as follows:

$$
\mathbb{P}(Y_{it} \leq y \mid w) = \begin{cases} \mathbb{P}(v_{it} \geq z_t'\beta_0 + x_t'\gamma_0 - y \mid w) & \text{if } y \geq 0, \\ 0 & \text{if } y < 0. \end{cases}
$$

Given observed distribution of $Y_{it} \mid W_i$, we can bound the distribution $\mathbb{P}(v_{it} \leq c \mid w)$ above as

$$
\begin{aligned}
\mathbb{P}(v_{it} \geq c \mid w) \leq \mathbb{P}(Y_{it} \leq z_t'\beta_0 + x_t'\gamma_0 - c \mid w)\mathbb{1}\{z_t'\beta_0 + x_t'\gamma_0 \geq c\} + \\
\mathbb{P}(Y_{it} = 0 \mid w)\mathbb{1}\{z_t'\beta_0 + x_t'\gamma_0 < c\},
\end{aligned}
$$

where the above condition holds since either $z_t'\beta_0 + x_t'\gamma_0 - c \geq 0$ so that there exists $y = z_t'\beta_0 + x_t'\gamma_0 - c \geq 0$ such that $\mathbb{P}(Y_{it} \leq y \mid w) = \mathbb{P}(v_{it} \geq c \mid w)$, or $\mathbb{P}(v_{it} \geq c \mid w) \leq \mathbb{P}(v_{it} \geq z_t'\beta_0 + x_t'\gamma_0 \mid w) = \mathbb{P}(Y_{it} = 0 \mid w)$ when $z_t'\beta_0 + x_t'\gamma_0 < c$.

Taking expectation over the endogenous covariate X_i yields the upper bound for the distribution $v_{it} \mid Z_i = z$:

$$\mathbb{P}(v_{it} \geq c \mid z) \leq \mathbb{P}(Y_{it} \leq z_t'\beta_0 + X_{it}'\gamma_0 - c,\ z_t'\beta_0 + X_{it}'\gamma_0 \geq c \mid z) + \mathbb{P}(Y_{it} = 0,\ z_t'\beta_0 + X_{it}'\gamma_0 < c \mid z).$$

Rearranging the formula, the above upper bound can be equivalently written as

$$\begin{aligned}
&\mathbb{P}(Y_{it} \leq z_t'\beta_0 + X_{it}'\gamma_0 - c,\ z_t'\beta_0 + X_{it}'\gamma_0 \geq c \mid z) + \mathbb{P}(Y_{it} = 0,\ z_t'\beta_0 + X_{it}'\gamma_0 < c \mid z) \\
=&\mathbb{P}(0 < Y_{it} \leq z_t'\beta + X_{it}'\gamma - c,\ z_t'\beta + X_{it}'\gamma \geq c \mid z) + \mathbb{P}(Y_{it} = 0 \mid z) \\
=&\mathbb{P}(0 < Y_{it} \leq z_t'\beta + X_{it}'\gamma - c \mid z) + \mathbb{P}(Y_{it} = 0 \mid z).
\end{aligned}$$

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

$$\mathbb{P}(v_{it} \geq c \mid w) \geq \mathbb{P}(Y_{it} \leq z_t'\beta_0 + x_t'\gamma_0 - c \mid w),$$

where the above condition holds since either $z_t'\beta_0 + x_t'\gamma_0 - c \geq 0$ so that there exists $y = z_t'\beta_0 + x_t'\gamma_0 - c \geq 0$ such that $\mathbb{P}(Y_{it} \leq y \mid w) = \mathbb{P}(v_{it} \geq c \mid w)$, or the lower bound is zero when $z_t'\beta_0 + x_t'\gamma_0 < c$.

Taking expectation over X_i leads to the following lower bound:

$$\mathbb{P}(v_{it} \geq c \mid z) \geq \mathbb{P}(Y_{it} \leq z_t'\beta_0 + X_{it}'\gamma_0 - c \mid z).$$

Given the bounds on the distribution $\mathbb{P}(v_{it} \geq c \mid z)$, the partial stationarity assumption implies the following identifying restriction for θ_0:

$$\max_t \mathbb{P}(Y_{it} \leq z_t'\beta_0 + X_{it}'\gamma_0 - c \mid z) \leq \max_s \{\mathbb{P}(0 < Y_{is} \leq z_s'\beta + X_{is}'\gamma - c \mid z) + \mathbb{P}(Y_{is} = 0 \mid z)\},$$

for any $c \in \mathcal{R}$ and any z. □

A.8 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 10 cannot be directly applied here. Due to this feature in the dynamic model, we need to adjust the results in

Proposition 10. 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 (10) to identify θ_0.

The lower bound in condition (10) can be bounded below as follows:

$$\begin{aligned}\mathbb{P}(Y_{it} \leq z_t'\beta + Y^*_{i,t-1}\gamma - c \mid z)\\ \geq \mathbb{P}(Y_{it} \leq z_t'\beta + Y_{i,t-1}\gamma - c, Y_{i,t-1} > 0 \mid z) := L_{t,cen}(c \mid z;\theta).\end{aligned}$$

Similarly, the upper bound in condition (10) can be further bounded above

$$\mathbb{P}(0 < Y_{is} \leq z_s'\beta + Y^*_{i,s-1}\gamma - c \mid z) + \mathbb{P}(Y_{is} = 0 \mid z) \leq U_{s,cen}(c \mid z;\theta),$$

where $U_{s,cen}(c \mid z;\theta)$ is defined as

$$\begin{aligned}U_{s,cen}(c \mid z;\theta) := \mathbb{P}(0 < Y_{is} \leq z_s'\beta + Y_{i,s-1}\gamma - c,\ Y_{i,s-1} > 0 \mid z)\\ + \mathbb{P}(Y_{is} > 0,\ Y_{i,s-1} = 0 \mid z) + \mathbb{P}(Y_{is} = 0 \mid z).\end{aligned}$$

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

Lemma 4. *Under Assumption 1 and $X_{it} = Y^*_{i,t-1}$, $\theta_0 \in \tilde{\Theta}_{I,cen}$, where the identified set $\tilde{\Theta}_{I,cen}$ consists of all $\theta = \left(\beta', \gamma'\right)' \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_x}$ such that*

$$\max_{t=1,\ldots,T} L_{t,cen}(c \mid z;\theta) \leq \min_{s=1,\ldots,T} U_{s,cen}(c \mid z;\theta).$$

for any $c \in \mathcal{R}$ and any realization $z = (z_1, ..., z_T)$ in the support of Z_i.