Identification of Dynamic Nonlinear Panel Models under Partial Stationarity∗

Wayne Yuan Gao† and Rui Wang‡

January 10, 2024

Abstract

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

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

∗We are grateful to Jason Blevins, Robert de Jong, Kyoo il Kim, Louise Laage, Eric Mbakop, Liyang Sun, and Jeffrey Wooldridge for helpful comments and suggestions.
†Department of Economics, University of Pennsylvania, 133 S 36th St, Philadelphia, PA 19104, USA. Email: waynegao@upenn.edu
‡Department of Economics, Ohio State University, 1945 N High St, Columbus, OH 43210, USA. Email: wang.16498@osu.edu
1 Introduction

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

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

\[ Y_{it} = 1\{Z_{it}'\beta_0 + X_{it}'\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\}, \]

where \( Y_{it} \in \{0, 1\} \) denotes a binary outcome variable for individual \( i = 1, 2, \ldots \) and time \( t = 1, \ldots, T \), 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 \( \epsilon_{it} \) denotes the unobserved time-varying error term for individual \( i \) at time \( t \). The objective is to identify the parameter \( \theta_0 := (\beta_0', \gamma_0')' \) using a panel of observed variables \((Z_{it}, X_{it}, Y_{it})_{it}\).

We focus on short panels, where the number of time periods \( T \geq 2 \) is fixed and finite.

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

\[ \epsilon_{is} \mid Z_{is}, Z_{it}, X_{is}, X_{it}, \alpha_i \overset{d}{\sim} \epsilon_{it} \mid Z_{is}, Z_{it}, X_{is}, X_{it}, \alpha_i \text{ for any } s, t \leq T. \]

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

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

\[ \text{For instance, suppose } \mathbb{E}[\epsilon_{it} \mid Z_{is}, Z_{it}, X_{is}, X_{it}] = X_{it}'\eta, \text{ then the conditional distributions of } \epsilon_{it} \text{ and } \epsilon_{is} \text{ cannot be the same as long as } X_{it}'\eta \neq X_{is}'\eta, \text{ so condition (2) fails in general.} \]
diately. For another example, if $X_{it}$ includes “price” or other variables that may be endogenously chosen by economic agents, then the exogeneity restriction imposed by condition (2) will again fail to hold.

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

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

(3)

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

We describe how to exploit the partial stationarity condition (3) to derive the identified set on the model parameters $\theta_0$ through conditional moment inequalities, which take the form of lower and upper bounds for the conditional distribution $\epsilon_{it} + \alpha_i \mid Z_{is}, Z_{it}$, solely as functions of observed variables and the model parameters $\theta_0$. We show that these bounds must have nonzero intersections over time under the partial stationarity assumption, thereby forming identifying restrictions for the parameter $\theta_0$.

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

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

$^2$Condition (3) also accommodates the standard stationarity assumption conditional on all covariates.
the coefficients $\beta_0$ on exogenous covariates (under scale normalization) as well as the signs of the coefficients on endogenous covariates $\gamma_0$.

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

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

**Literature Review**

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

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

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

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

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

which is exactly a “partial stationarity” condition in the specific context of (4), and derives
the sharp identified set for \( \theta_0 \) 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. \( Y_{i,t-1}, Y_{i,t-2}, \ldots \) ) 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. More differently, some other papers work with sequential exogeneity in various dynamic panel mod-

---

3Our identification strategy shares some conceptual similarity with the idea of generalized instrumental variable (GIV) in Chesher and Rosen (2017), who proposes a general approach for representing the identified set of structural models with endogeneity. Chesher and Rosen (2017), Chesher and Rosen (2020), and Chesher, Rosen, and Zhang (2023) demonstrate how the GIV framework can be applied to various settings, but focus mostly on the use of exclusion restrictions and/or full independence assumptions. In this paper, we neither impose exclusion restrictions nor independence assumptions but instead explore identification under a partial stationarity condition.
els and provide (non-)identification results under different model restrictions. For example, Shiu and Hu (2013) imposes a high-level invertibility condition along with a restriction that rules out the dependence of covariates on past dependent variables. More recently, Bonhomme, Dano, and Graham (2023) investigates panel binary choice models with a single binary predetermined covariate under sequential exogeneity, whose evolution may depend on the past history of outcome and covariate values. The sequential exogeneity condition considered in these papers and the partial stationarity condition in ours again do not nest each other as special cases: in particular, our current paper accommodates contemporarily endogenous covariates that violate sequential exogeneity. In summary, the key assumptions, identification strategy, and identification results of these studies are substantially different from and thus not directly comparable to those in our current paper.

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

The rest of the paper is organized as follows. Section 2 studies the sharp identification of binary choice models with endogenous covariates. We provide sufficient conditions to achieve point identification and explore two classic models: static and dynamic models. Sections 3-5 apply the identification approach to ordered response models, multinomial choice models, and models with censored dependent variables. Section 6 broadens the scope to general nonseparable models. Section 7 presents simulation results about the finite-sample
performances of our approach and Section 8 explores the empirical application of income categories using various ordered response models. We conclude with Section 9.

2 Binary Choice Model

2.1 Model

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

\[
Y_{it} = 1\{Z_{it}'\beta_0 + X_{it}'\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\},
\]

where \(Y_{it} \in \{0, 1\}\) denotes the binary dependent variable for individual \(i \in \{1, 2, \ldots\}\) and time \(t \leq T\), \(Z_{it} \in \mathcal{R}^{d_z}\) denotes exogenous covariates, \(X_{it} \in \mathcal{R}^{d_x}\) denotes potentially endogenous covariates, \(\alpha_i \in \mathcal{R}\) denotes the unobserved fixed effect for individual \(i\), and \(\epsilon_{it}\) denotes the unobserved time-varying error term for individual \(i\) at time \(t\). The objective is to identify the parameter \(\theta_0 := (\beta_0', \gamma_0')'\) using a short panel of observed variables \((Z_{it}, X_{it}, Y_{it})_{it}\), where the number of time periods \(T \geq 2\) is fixed and finite.

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

\[
\epsilon_{is} | Z_{is}, Z_{it}, X_{is}, X_{it}, \alpha_i \sim d \epsilon_{it} | Z_{is}, Z_{it}, X_{is}, X_{it}, \alpha_i \quad \text{for any } s, t \leq T.
\]

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

Our paper aims to study identification of \(\theta_0\) robust to endogeneity of \(X_{it}\), where \(X_{it}\) can be arbitrarily dependent with \((\alpha_i, \epsilon_{it})\). Before introducing our assumptions, we first provide several applications to illustrate this endogeneity.

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

\[
Y_{it} = 1\{Z_{it}'\beta_0 + Y_{it-1}\gamma_0 + \alpha_i + \epsilon_{it} \geq 0\},
\]

\(^4\)See Section 1 for more detailed discussions on related literature.
under the stationarity assumption only given exogenous covariates $Z_{it}$: $\epsilon_{is} \mid Z_{is}, Z_{it}, \alpha_i \overset{d}{\sim} \epsilon_{it} \mid Z_{is}, Z_{it}, \alpha_i$. In this model, the full stationarity condition may naturally fail due to its dynamic nature. For example, suppose $\epsilon_{it}$ is i.i.d. across time and independent of $\{Z_{it}\}_{t=1}^T$, $\alpha_i$. Given $Z_{is} = Z_{it} = z, \alpha_i = a$, it is clear that the conditional distribution of $\epsilon_{it} \mid Y_{it-1}, Y_{it}, z, a$ will be different from $\epsilon_{it+1} \mid Y_{it-1}, Y_{it}, z, a \overset{d}{\sim} \epsilon_{it+1}$, since $Y_{it}$ is informative for $\epsilon_{it}$.

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

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

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

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

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

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

Motivated by the above examples, this paper aims to establish identification of $\theta_0$, while being robust to endogeneity of $X_{it}$ with respect to unobserved factors $\{\alpha_i, \epsilon_{it}\}$. Below we propose an approach that accommodates various types of endogeneity, regardless of whether the endogenous covariates $X_{it}$ are discrete lagged outcome variables (as in Example 1) or continuously-valued covariates with contemporary endogeneity as in other examples. To achieve this, we introduce the following partial conditional stationarity assumption:

**Assumption 1** (Partial Stationarity). The conditional distribution of $\epsilon_{it} \mid Z_{is}, Z_{it}, \alpha_i$ is stationary over time: $\epsilon_{is} \mid Z_{is}, Z_{it}, \alpha_i \overset{d}{\sim} \epsilon_{it} \mid Z_{is}, Z_{it}, \alpha_i$ for any $s, t \leq T$. 

8
Assumption 1 only assumes the stationarity given the covariate \((Z_{is}, Z_{it})\), but does not impose any restrictions on the endogenous covariate \(X_{it}\). The standard full stationarity condition (2) is nested in Assumption 1, and it is in general stronger than Assumption 1. The main distinction is that Assumption 1 allows \(X_{it}\) to be arbitrarily dependent with both the fixed effect \(\alpha_i\) and the time-changing error \(\epsilon_{it}\). In addition, their correlation can be different across time. Assumption 1 also shares similar features with condition (2), in that it does not impose any distributional assumption on \(\epsilon_{it}\) and allows \(\epsilon_{it}\) to be correlated across time. The standard i.i.d. assumption of \(\epsilon_{it}\) across time or independence of \(\epsilon_{it}\) with \((\alpha_i, Z_{it}, X_{it})\) are also nested in Assumption 1 as special cases.

2.2 Identification Strategy

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

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

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

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

However, Manski (1987)’s method is not applicable to our setting with Assumption 1, given that the endogenous regressor \(X_{it}\) can be correlated with \(v_{it}\). When the covariate \(X_{it}\) changes, the conditional distribution of \(v_{it}\) might potentially change as well. Variation in the choice probability contains a mixture of changes in the covariate index and the conditional distribution of \(v_{it}\). Hence, it is not feasible to derive identifying restrictions on \(\theta_0\) merely by examining the intertemporal variation in the choice probability.

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

\[
v_{is} \mid Z_{ist} \overset{d}{\sim} v_{it} \mid Z_{ist}.
\]

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

\footnote{Condition 2 implies the stationarity of \(v_{it}: v_{is} \overset{d}{\sim} v_{it} \mid W_{ist}.\)
functions of observed variables and the parameter \( \theta_0 \). Since the distribution of \( v_{it} | Z_{ist} \) is the same over time, then the bounds of the distribution across time should have intersections. This restriction serves as the identifying restriction for \( \theta_0 \). Next, we show how to derive bounds on the distribution of \( v_{it} | Z_{ist} \).

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

**Lower bound.** The conditional choice probability given \( W_{ist} = w_{st} \) is given as

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

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

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

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

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

\[
F_{v_{ist}|z_{st}}(a | z_{st}) \geq \int x_{st} Pr(Y_{it} = 1 | W_{ist} = w_{st}) \mathbb{1}\{w'_t \theta_0 \leq a\} dF_{X_{ist}|z_{st}}(x_{st} | z_{st})
\]

\[
= Pr(Y_{it} = 1, z'_t \theta_0 + X'_{it} \gamma_0 \leq a | z_{st}).
\]

The above lower bound only depends on observed variables \( (W_{ist}, Y_{it}) \) and \( \theta_0 \), so it is identified up to the parameter \( \theta_0 \). This bound will help construct identifying restrictions on \( \theta_0 \).

**Upper bound.** The idea of constructing the upper bound is similar. For any \( a \in \mathcal{R} \), the conditional distribution \( F_{v_{ist}|w_{st}} \) can be bounded above:

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

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

\[
F_{v_{ist}|z_{st}}(a | z_{st}) \leq \int x_{st} \left\{ Pr(Y_{it} = 1 | W_{ist} = w_{st}) \mathbb{1}\{w'_t \theta_0 \geq a\} + \mathbb{1}\{w'_t \theta_0 < a\} \right\} dF_{X_{ist}|z_{st}}(x_{st} | z_{st})
\]

\[
= Pr(Y_{it} = 1, z'_t \theta_0 + X'_{it} \gamma_0 \geq a | z_{st}) + Pr(z'_t \beta_0 + X'_{it} \gamma_0 < a | z_{st})
\]

\[
= 1 - Pr(Y_{it} = 0, z'_t \beta_0 + X'_{it} \gamma_0 \geq a | z_{st}).
\]

10
Given the lower and upper bound for the distribution \( F_{v_t|Z_{st}}(a \mid z_{st}) \), we are ready to establish identifying conditions for \( \theta_0 \). Since Assumption 1 requires that \( v_t \mid Z_{ist} \overset{d}{\sim} v_s \mid Z_{ist} \), it implies the upper bound for the distribution of \( v_s \mid Z_{ist} \) must be larger than the lower bound for the distribution of \( v_t \mid Z_{ist} \) otherwise Assumption 1 cannot hold. This restriction characterizes an identified set for \( \theta_0 \), presented in the following proposition.

**Proposition 1.** Under Assumption 1, an identified set \( \Theta_{I,1} \) for \( \theta_0 \) is the set of parameters that satisfy the following conditions:

\[
1 - \Pr(Y_{is} = 0, z_s' \beta + X_{is}' \gamma \geq a \mid z_{st}) \geq \Pr(Y_{it} = 1, z_t' \beta + X_{it}' \gamma \leq a \mid z_{st}),
\]

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

Proposition 1 characterizes an identified set for \( \theta_0 \), using the conditional joint distribution of \( (Y_{it}, X_{it}) \) given \( Z_{ist} \) across any pair of periods \((s, t)\). The identification result allows for endogeneity in the covariate \( X_{ist} \) and does not rely on any excluded instruments.\(^6\)

The identifying condition in (7) can be reformulated as conditional moment inequalities, by substituting the probability \( \Pr(Y_{it} = 0, z_t' \beta_0 + X_{it}' \gamma_0 \geq a \mid z_{st}) \) with the conditional expectation \( \mathbb{E} \left[ \mathbb{I} \{Y_{it} = 0, z_t' \beta_0 + X_{it}' \gamma_0 \geq a \} \mid z_{st} \right] \). Consequently, the estimation and inference for \( \theta_0 \) can be conducted using established methods such as Chernozhukov, Hong, and Tamer (2007), Andrews and Shi (2013), and Chernozhukov, Lee, and Rosen (2013).

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

\[
w_s' \theta_0 \geq w_t' \theta_0 \iff 1 - \Pr(Y_{is} = 0 \mid w_{st}) = \Pr(Y_{is} = 1 \mid w_{st}) \geq \Pr(Y_{it} = 1 \mid w_{st}).
\]

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

\[
1 - \Pr(Y_{is} = 0, z_s' \beta_0 + Y_{is-1} \gamma_0 \geq a \mid z_{st}) \geq \Pr(Y_{it} = 1, z_t' \beta_0 + Y_{it-1} \gamma_0 \leq a \mid z_{st}),
\]

for any \( a \in \{z_s' \beta_0, z_s' \beta_0 + \gamma_0, z_s' \beta_0, z_s' \beta_0 + \gamma_0\} \). As shown in Appendix A.4, the above identifying condition on \( \theta_0 \) implies the identifying conditions in Khan, Ponomareva, and Tamer (2023, Theorem 1, p. 8-9), under different values of \((z_s' \beta_0, z_s' \beta_0, \gamma_0)\).

\(^6\)When a sub-vector \( Z_{ist}^{exc} \) of \( Z_{ist} \) is assumed to be excluded from the binary choice model, then the identifying restriction in (7) is adjusted by setting the coefficient of \( Z_{ist}^{exc} \) as zero: \( \beta_0^{exc} = 0 \).
Proposition 1 does not impose restrictions on the support of the endogenous covariate $X_{it}$, allowing it to be either discrete or continuous. When $X_{it}$ is continuous, then Proposition 1 requires condition (7) to hold for any $a \in R$ so it exploits the whole distribution information of $(X_{it}, Y_{it})$ at any value $a$. For the estimation and inference of $\theta_0$, we can follow the conventional inference method to discretize the space of $R$ and pick $K_n$ points for $a$. In the case of discrete $X_{it}$, the dimension of $a$ can be further reduced, and the identified set $\Theta_{I,1}$ can be characterized by finite number of restrictions, shown in the following corollary.

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

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

We show that $\Theta_{I,1}$ is an identified set for $\theta_0$, however, it is still unclear whether condition (7) exploits all the available information for $\theta_0$. The following theorem establishes the sharpness of the identified set $\Theta_{I,1}$ when $X_{it}$ takes finite number of values.

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

Proposition 1 provides identification results without restrictions on the support of $X_{it}$, and Theorem 1 shows that the results already exhausted all available information for $\theta_0$ when $X_{it}$ only takes finite number of values. This results implies that for dynamic models with any number of lagged dependent variables, the identified set $\Theta_{I,1}$ is the smallest set we can obtain.

### 2.3 Point Identification

Proposition 1 characterizes the sharp identified set for $\theta_0$ by only imposing Assumption 1. This section provides sufficient conditions to achieve point identification for $\beta_0$ (up to scale) and the sign of $\gamma_0$ under support conditions on the exogenous covariate $Z_{it}$. We focus on the scenario where the endogenous covariate $X_{it}$ is discrete $X_{it} \in \mathcal{X} = \{a_1, a_2, ..., a_K\}$ and there are only two periods $T = 2$.

To point identify $\beta_0$, 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:

$$Z_1 := \{(z_1, z_2) \mid \exists x \in \mathcal{X} \text{ s.t. } 1 - \Pr(Y_{i1} = 0, X_{i1} = x \mid z_{12}) < \Pr(Y_{i2} = 1, X_{i2} = x \mid z_{12})\},$$

$$Z_2 := \{(z_1, z_2) \mid \exists x \in \mathcal{X} \text{ s.t. } 1 - \Pr(Y_{i1} = 1, X_{i1} = x \mid z_{12}) < \Pr(Y_{i2} = 0, X_{i2} = x \mid z_{12})\}.$$

Let $Z := Z_1 \cup Z_2$. Let $\Delta Z_i = Z_{i2} - Z_{i1}$ and $\Delta Z$ be defined as

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

As shown in Appendix A.3, when $\Delta z$ satisfies $\Delta z \in \Delta Z$, the sign of $\Delta z'\beta_0$ is identified. In the definition of the two sets $Z_1, Z_2$, 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 $\beta_0$ can be identified up to scale under the standard large support condition on $\Delta z$.

Let $\Delta z^j$ denote the $j$th element of $\Delta z$. The following is the support condition on the covariate.

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

**Proposition 2.** Under Assumptions 1-2, the parameter $\beta_0$ is point identified up to scale.

We provide point identification for $\beta_0$ 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 $\beta_0$, 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 $\Delta z$ has bounded support.

The parameter $\gamma_0$ 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 $\gamma_0$ under certain choice probabilities. We derive the sufficient conditions to identify the sign of $\gamma_0$.

Let $x^j$ denote the $j$-th element of $x$ and $\gamma_0^j$ denote the $j$-th coefficient of $\gamma_0$. We define the following two sets:

$$Z_{3j} := \{(z_1, z_2) \mid \exists x_1, x_2 \in \mathcal{X} \text{ with } x^j_1 \neq x^j_2, x^m_1 = x^m_2 \forall m \neq j \text{ s.t. } 1 - \Pr(Y_{i1} = 0, X_{i1} = x_1 \mid z_{12}) < \Pr(Y_{i2} = 1, X_{i2} = x_2 \mid z_{12})\};$$

$$Z_{4j} := \{(z_1, z_2) \mid \exists x_1, x_2 \in \mathcal{X} \text{ with } x^j_1 \neq x^j_2, x^m_1 = x^m_2, \forall m \neq j \text{ s.t. } 1 - \Pr(Y_{i1} = 1, X_{i1} = x_1 \mid z_{12}) < \Pr(Y_{i2} = 0, X_{i2} = x_2 \mid z_{12})\}.$$
From the identifying results in Proposition 1, the value of \((x^j_1 - x^j_2)\gamma^j_0\) can be bounded when \((z_1, z_2)\) belong to the two sets:

\[
(z_1, z_2) \in Z^j_3 \implies (x^j_1 - x^j_2)\gamma^j_0 < \Delta z' \beta_0,
\]

\[
(z_1, z_2) \in Z^j_4 \implies (x^j_1 - x^j_2)\gamma^j_0 > \Delta z' \beta_0.
\]

Then the sign of \(\gamma^j_0\) is identified if either the sign of \(\Delta z' \beta_0\) is identified as negative when \((z_1, z_2) \in Z_2\) or as positive when \((z_1, z_2) \in Z_1\).

**Proposition 3.** Under Assumptions 1, and for any \(1 \leq j \leq d_x\), either \(Z^j_3 \cap Z_2 \neq \emptyset\) or \(Z^j_4 \cap Z_1 \neq \emptyset\), then the sign of \(\gamma_0\) 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 \(Z^j_3, Z^j_4\) can be simplified as there existing \(x_1 \neq x_2\) such that the corresponding choice probability is observed. Besides the sign of \(\gamma_0\), the identification results can also bound the value of \(\gamma_0\) 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 \(\gamma_0 = (\gamma^1_0, \gamma^2_0)\), then \(\gamma^1_0\) is identified when the required choice probability in the two sets \(Z^j_3, Z^j_4\) 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 \(\gamma_0\), which may be stronger than necessary and can be relaxed in certain scenarios. For example, when we know that \(\gamma^1_0 + \gamma^2_0 > 0\) while \(\gamma^1_0 < 0\), we can infer that \(\gamma^2_0 > 0\) without requiring additional assumptions on the two sets \(Z^3_3, Z^2_4\).

## 3 Ordered Response Model

### 3.1 Model and Identification

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

\[
Y^*_it = Z'^t_{it} \beta_0 + X'^t_{it} \gamma_0 + \alpha_i + \epsilon_{it},
\]

\[
Y_{it} = \sum_{j=1}^J y_j 1\{b_j < Y^*_it \leq b_{j+1}\},
\]

where \(Y^*_it\) denotes the latent dependent variable, \(b_1 = -\infty, b_{J+1} = +\infty\), and the remaining threshold parameters \(b_j\) (where \(b_j \leq b_{j+1}\)) can be either known or unknown for \(2 \leq j \leq J-1\). The binary choice model in (1) is nested with \(J = 2\) and \(b_2 = 0\). Similar to model (1), \(X_{it}\)
is the potentially endogenous covariate that can be arbitrarily dependent on \((\alpha_i, \epsilon_{it})\), such as the lagged dependent variable \(Y_{i,t-1}\), while \(Z_{it}\) is the exogenous covariate that satisfies Assumption 1:

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

We now show how our key identification strategy based on partial stationarity can also be exploit to partially identify \(\theta_0\). Similar to Section 2, we still seek to establish both lower and upper bounds for the distribution \(F_{v_t|Z_{st}}(a \mid z_{st})\), where \(v_{it} := \epsilon_{it} + \alpha_i\). The distinction is that in the ordered choice setting we can obtain \(J\) different (upper and lower) bounds, which can then be aggregated over to form tighter bounds as shown below.

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

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

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

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

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

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

An equivalent expression for this lower bound is given as:

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

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

2. Upper bound. Similarly, we can derive the upper bound for \(F_{v_t|W_{st}}(a \mid w_{st})\) using
the sum of conditional choice probabilities, given as follows:

\[
F_{v_t \mid w_{st}}(a \mid w_{st}) = 1 - \Pr(v_t > a \mid w_{st}) \\
\leq 1 - \sum_{j=1}^{J} \Pr(b_j - w_t^t \theta_0 < v_t \leq b_j + 1 - w_t^t \theta_0 \mid w_{st}) \mathbb{1}\{b_j - w_t^t \theta_0 \geq a\} \\
\leq 1 - \sum_{j=1}^{J} \Pr(Y_{it} = y_j \mid w_{st}) \mathbb{1}\{b_j - w_t^t \theta_0 \geq a\}.
\]

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

\[
F_{v_t \mid z_{st}}(a \mid z_{st}) \leq 1 - \sum_{j=1}^{J} \Pr(Y_{it} = y_j, b_j - z_{it}' \beta_0 - X_{it}' \gamma_0 \geq a \mid z_{st}).
\]

Given the established lower and bounds on the distribution of \(v_t \mid z_{st}\), an identified set for \(\theta_0\) is characterized in the following proposition.

**Proposition 4.** Under Assumption 1, the identified set \(\Theta_{I,2}\) for \(\theta_0\) is the set of parameters \(\theta\) that satisfy the following conditions:

\[
1 - \sum_{j=1}^{J} \Pr(Y_{is} = y_j, b_j - z_{is}' \beta_0 - X_{is}' \gamma_0 \geq a \mid z_{st}) \geq \sum_{j=1}^{J} \Pr(Y_{it} = y_j, b_{j+1} - z_{it}' \beta_0 - X_{it}' \gamma \leq a \mid z_{st}),
\]

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

Proposition 4 characterizes the identified set for \(\theta_0\) for the general ordered response model with potential endogeneity. This result allows for both static and dynamic models, as discussed in Section 3.2. In comparison to the binary choice model, Proposition 4 aggregates information from all \(J\) ordered responses across different time periods to identify \(\theta_0\). When there are only two choices \((J = 2)\), this result simplifies to the identifying condition (7) in Proposition 1, with \(y_1 = 0, y_2 = 1\) and changing \(a\) to \(-a\). Again, the result in Proposition 4 accommodates both discrete and continuous endogenous covariates. When the endogenous \(X_{it}\) only takes a finite number of values, such as lagged dependent variable, we can reduce the number of restrictions in Proposition 4 to a finite number, as shown in the next corollary.

**Corollary 2.** Under Assumption 1 and \(X_{it}\) only takes finite number of values \(X_{it} \in \{a_1, ..., a_K\}\), the identified set \(\Theta_{I,2}\) is the set of parameters \(\theta\) that satisfy condition (9) for any \(a \in \mathcal{Q}_2 := \{b_j - z_{i}' \beta_0 - a_{i}' \gamma_0, ..., b_j - z_{i}' \beta_0 - a_{i}' \gamma_0, b_j - z_{i}' \beta_0 - a_{i}' \gamma, ..., b_j - z_{i}' \beta_0 - a_{i}' \gamma\}_{j=2}^J\), any \(s, t \leq T\), and any \(z_{st} \in \mathcal{R}^{d_z} \times \mathcal{R}^{d_z}.

Similar to Corollary 1, Corollary 2 reduces the number of conditional moment inequalities to \(2K \times (J - 1)\) numbers for any pair of two periods. For the dynamic model with one lagged
dependent variable $X_{it} = Y_{i,t-1} \in \{y_1, ..., y_J\}$, then there will be $2J(J - 1)$ number of restrictions.

### 3.2 Applications

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

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

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

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

**Lemma 1.** Assume that $\epsilon_{is} \mid (W_{ist}, \alpha_i) \overset{d}{\sim} \epsilon_{it} \mid (W_{ist}, \alpha_i)$, the identified set $\Theta_{I,2}$ for $\theta_0$ is the set of parameters $\theta$ that satisfy the following conditions:

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

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

In the static model, variation in conditional choice probability only comes from changes in the covariate $W_{st}$. Thus, we can directly establish the relationship between the variation in choice probability and the changes in covariates over time, eliminating the parameter $a$. Different from the binary choice model, the results in Lemma 1 also exploit intertemporal variation in the sum of multiple choices rather than investigating a single choice. Moreover, we can utilize variations in the sum of different choices across various periods to identify $\theta_0$.

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

**Dynamic model**: consider the following dynamic ordered response model with one
lagged dependent variable:

\[ Y_{it}^* = Z_{it}' \beta_0 + Y_{it-1} \gamma_0 + \alpha_i + \epsilon_{it}, \]

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

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

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

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

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

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

### 3.3 Point Identification

This section provides sufficient conditions to achieve point identification for \( \beta_0 \) (up to scale) and the sign of \( \gamma_0 \). Similar to Section 2.3, we still focus on the scenario where the endogenous covariate \( X_{it} \) only takes finite number of values \( X_{it} \in \mathcal{X} = \{a_1, a_2, \ldots, a_K\} \), and we illustrate the concept using two periods \( (T = 2) \).

To identify \( \beta_0 \), we define the following two sets:

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

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

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

\[
1 - \sum_{j=1}^{k-1} \Pr(Y_{i1} = y_j, X_{i1} = x \mid z_{12}) < \sum_{j=k}^{J} \Pr(Y_{i2} = y_j, X_{i2} = x \mid z_{12}) \right\}. \]
Let \( Z_{\text{order}} := Z_{1, \text{order}} \cup Z_{2, \text{order}} \). Let \( \Delta Z_i = Z_{i2} - Z_{i1} \) and \( \Delta Z_{\text{order}} \) be defined as
\[
\Delta Z_{\text{order}} := \left\{ \Delta z := z_2 - z_1 \mid (z_1, z_2) \in Z_{\text{order}} \right\}.
\]

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

**Proposition 5.** Under Assumptions 1 & 3, the parameter \( \beta_0 \) is point identified up to scale.

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

To identify the sign of \( \gamma_0 \), we define the following two sets: for \( j \leq d_x \),
\[
Z_{3, \text{order}}^j := \left\{ (z_1, z_2) \mid \exists x_1, x_2 \in \mathcal{X} \text{ with } x_1^j \neq x_2^j, x_1^m = x_2^m \forall m \neq j, \exists 2 \leq k \leq J \text{ s.t.} \right. \\
1 - \sum_{j=1}^{J} \Pr(Y_{i1} = y_j, X_{i1} = x_1 \mid z_{12}) < \sum_{j=1}^{k-1} \Pr(Y_{i2} = y_j, X_{i2} = x_2 \mid z_{12}) \right\};
\]
\[
Z_{4, \text{order}}^j := \left\{ (z_1, z_2) \mid \exists x_1, x_2 \in \mathcal{X} \text{ with } x_1^j \neq x_2^j, x_1^m = x_2^m \forall m \neq j, \exists 2 \leq k \leq J \text{ s.t.} \right. \\
1 - \sum_{j=1}^{J} \Pr(Y_{i1} = y_j, X_{i1} = x_1 \mid z_{12}) < \sum_{j=1}^{k-1} \Pr(Y_{i2} = y_j, X_{i2} = x_2 \mid z_{12}) \right\}.
\]

**Proposition 6.** Under Assumptions 1, and for any \( 1 \leq j \leq d_x \), either \( Z_{3, \text{order}}^j \cap Z_{2, \text{order}} \neq \emptyset \) or \( Z_{4, \text{order}}^j \cap Z_{1, \text{order}} \neq \emptyset \), then the sign of \( \gamma_0 \) is identified.

For the dynamic model with one lagged dependent variable \( X_{it} = Y_{i,t-1} \), then the joint choice probabilities defined in the two sets \( Z_{3, \text{order}}^j, Z_{4, \text{order}}^j \) can be simplified as taking different values for the lagged dependent variable: \( Y_{i0} = y_{j1}, Y_{i1} = y_{j2} \) with \( y_{j1} \neq y_{j2} \). The identifying restriction in Proposition 4, in general, can bound the value of \((y_j - y_k)\gamma_0\) using the variation in \( Z_{it} \). Proposition 6 ensures that its bound excludes zero for some \((y_j, y_k)\), thereby identifying the sign of \( \gamma_0 \). When there are more than two periods \((T > 2)\), then point identification of \( \beta_0 \) and the sign of \( \gamma_0 \) is achieved when there exists two periods such that the assumptions are satisfied.
4 Multinomial Choice Model

4.1 Model and Identification

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

\[
\begin{align*}
    u_{ijt} &= Z'_{ijt} \beta_0 + X'_{ijt} \gamma_0 + \alpha_{ij} + \epsilon_{ijt}, \\
    Y_{it} &= \arg\max_{j \in J} u_{ijt},
\end{align*}
\]

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

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

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

The identification of \( \theta_0 \) has been studied under the standard full stationarity condition given all covariates \( W_{ist} \) in different models, including Pakes and Porter (2022), Shi, Shum, and Song (2018), Gao and Li (2020), and Wang (2022). The main focus of this paper is to derive identification for \( \theta_0 \), while allowing for the endogenous covariate \( X_{it} \). As discussed in Section 4.2, our approach also accommodates the standard stationarity assumption, and the results are consistent with those in Pakes and Porter (2022).

Let \( v_{ijt} = \alpha_{ij} + \epsilon_{ijt} \) and \( v_{it} = \{v_{ijt}\}_{j \in J} \). The identification strategy is still to establish bounds on the distribution of \( v_{it} \mid Z_{ist} \) and derive identifying restrictions on \( \theta_0 \) based on partial stationarity. Here, the distinction is that the error term \( v_{it} \) is multi-dimensional instead of one-dimensional in binary or ordered choice models. Moreover, consumers’ choice involves the comparison between different choices.

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

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

The above probability restricts the distribution \( \Pr(v_{ikt} - v_{ijt} \leq (w_{jt} - w_{kt})' \theta_0 \forall k \neq j \mid w_{st}) \). From observed choice probability, we can derive bounds for the distribution \( \Pr(v_{ikt} - v_{ijt} \leq \)
\(a_{jk} \forall k \neq j \mid z_{st})\) as a function of \(\theta_0\) and observed variable \((X_{it}, Z_{it}, Y_{it})\). Therefore, the identifying condition for \(\theta_0\) is characterized by the restriction that the bounds over different periods must intersect. More generally, we observe the conditional probability \(\Pr(Y_{it} \in K \mid w_{ist})\) for any set \(K \subset J\) and can bound the probability \(\Pr(v_{ikt} - v_{ijt} \leq a_{jk} \exists j \in K, \forall k \in K^c \mid z_{st})\), where \(K^c := J \setminus K\). In the case where \(K\) is a singleton \((K = \{j\})\), it then becomes the conditional probability of selecting a single choice \(j\). The following proposition characterizes the identification results for \(\theta_0\).

**Proposition 7.** Under Assumption 1, an identified set \(\Theta_{I,3}\) for \(\theta_0\) is the set of parameters \(\theta = (\beta, \gamma)\) that satisfy the following condition:

\[
1 - \Pr(Y_{is} \in K^c, (z_{js} - z_{ks})'\beta + (X_{ij}s - X_{iks})'\gamma \geq a_{jk} \forall j \in K, k \in K^c \mid z_{st}) \\
\geq \Pr(Y_{it} \in K, (z_{jt} - z_{kt})'\beta + (X_{ij}t - X_{ikt})'\gamma \leq a_{jk} \forall j \in K, k \in K^c \mid z_{st}),
\]

(10)

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

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

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

### 4.2 Applications

We examine the two applications of Proposition 7 for the panel multinomial choice model. The first one studies the static model, where all covariates \(W_{it}\) are exogenous \((\epsilon_{is} \mid W_{ist}, \alpha_i \sim d \epsilon_{it} \mid W_{ist}, \alpha_i)\), which is the same setting as Pakes and Porter (2022). In this scenario, we show that Proposition 7 boils down to the identifying results in Pakes and Porter (2022). Another one explores the dynamic multinomial choice model, incorporating one lagged dependent variable \(Y_{i,t-1}\).
**Static model:** Pakes and Porter (2022) characterizes the sharp identified set for $\theta_0$ under the stationarity assumption given all covariates: $\epsilon_{is} \mid W_{is}, W_{it}, \alpha_i \sim \mathcal{N}(0, \alpha_i)$. Now, we show that the identifying restriction in (10) is consistent with the results in Pakes and Porter (2022).

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

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

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

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

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

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

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

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

## 5 Censored Dependent Variable Model

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

Consider the following panel models with censored outcomes:

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

$$Y_{it} = \max\{Y_{it}^*, 0\},$$
where \( X_{it} \) represents the endogenous covariate, \( Z_{it} \) stands for the exogenous covariate that satisfies Assumption 1, \( Y_{it}^* \) denotes the latent outcome not observed in the data, and \( Y_{it} \) represents the observed outcome, which is censored at zero. The threshold for censoring can be replaced with other nonzero constants. The identification strategy is still to derive conditions for \( \theta_0 \) based on the stationarity assumption. Here the conditional distribution of \( \alpha_i + \epsilon_{it} \mid Z_{st} \) is not identified since we only observe the censoring outcome \( Y_{it} \). While we can still bound the conditional distribution \( \alpha_i + \epsilon_{it} \mid Z_{st} \) using variation in the joint distribution \( (Y_{it}, Z_{it}, X_{it}) \). The following proposition presents the identification results of \( \theta_0 \) with censoring outcomes.

**Proposition 8.** Under Assumption 1, an identified set for \( \theta_0 \) is the set of parameters \( \theta \) that satisfy the following condition:

\[
\Pr(0 < Y_{is} \leq z_i' \beta + X_{is}' \gamma - a \mid z_{st}) + \Pr(Y_{is} = 0 \mid z_{st}) \geq \Pr(Y_{it} \leq z_i' \beta + X_{it}' \gamma - a \mid z_{st}),
\]

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

Similar to discrete choice models studied in previous sections, Proposition 8 characterizes an identified set for \( \theta_0 \) by exploiting the variation in the joint distribution \( (Y_{it}, X_{it}) \mid Z_{ist} \) over time and the variation in the exogenous covariates \( Z_{ist} \). The bounds on the distribution \( \alpha_i + \epsilon_{it} \mid Z_{st} \) can be derived either from the probability \( \Pr(0 < Y_{it} \leq y \mid z_{st}) \) or \( \Pr(Y_{it} = 0 \mid z_{st}) \), depending on the value of the covariate index \( z_i' \beta + X_{it}' \gamma \). This result still accommodates both static and dynamic models with censored outcomes.

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

\[
\Pr(0 < Y_{is} \leq w_s' \theta - a \mid w_{st}) + \Pr(Y_{is} = 0 \mid w_{st}) \geq \Pr(Y_{it} \leq w_t' \theta - a \mid w_{st}). \tag{11}
\]

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

\[
\Pr(Y_{is} = 0 \mid w_{st}) \geq \Pr(Y_{it} \leq w_t' \theta - a \mid w_{st}).
\]

When \( a \) satisfies \( a \leq \min\{w_s' \theta, w_t' \theta\} \), condition (11) transforms into

\[
\Pr(0 < Y_{is} \leq w_s' \theta - a \mid w_{st}) + \Pr(Y_{is} = 0 \mid w_{st}) = \Pr(Y_{is} \leq w_s' \theta - a \mid w_{st}) \geq \Pr(Y_{it} \leq w_t' \theta - a \mid w_{st}).
\]

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

The following lemma derives the identified set for \( \theta_0 \) using both conditional moment
inequalities and equalities.

**Lemma 2.** Assuming that $\epsilon_{it} | \alpha_i, W_{ist} \sim \epsilon_{is} | \alpha_i, W_{ist}$, an identified set for $\theta_0$ is characterized by the set of parameters $\theta$ that satisfy the following conditions:

\[
\begin{align*}
\{ & \text{If } w_s' \theta \leq a \leq w_t' \theta \implies \Pr(Y_{is} = 0 \mid w_{st}) \geq \Pr(Y_{it} \leq w_t' \theta - a \mid w_{st}) , \\
& \text{If } a \leq \min\{w_s' \theta, w_t' \theta\} \implies \Pr(Y_{is} \leq w_t' \theta - a \mid w_{st}) = \Pr(Y_{it} \leq w_t' \theta - a \mid w_{st}) , \\
\end{align*}
\]

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

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

\[
\begin{align*}
Y_{it}^* &= Z_{it}' \beta_0 + Y_{i,t-1} \gamma_0 + \alpha_i + \epsilon_{it} , \\
Y_{it} &= \max\{Y_{it}^*, 0\} .
\end{align*}
\]

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

### 6 Generalization to Nonseparable Models

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

To illustrate, consider the following nonseparable semiparametric model:

\[
Y_{it} = G \left( W_{it}' \theta_0, \alpha_i, \epsilon_{it} \right) ,
\]

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

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

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

24
As before, we decompose $W_{it}$, and correspondingly $\theta_0$, into two components, $W_{it} = (Z_{it}', X_{it}')'$ and $\theta_0 = (\beta'_0, \gamma'_0)'$ so that

$$W_{it}'\theta_0 = Z_{it}'\beta_0 + X_{it}'\gamma_0,$$

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

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

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

$$1\left\{ G(a, \alpha_i, \epsilon_{it}) \geq y \right\} \leq 1\left\{ w_{i}'\theta_0 \leq a \right\} 1\left\{ G(a, \alpha_i, \epsilon_{it}) \geq y \right\} \leq 1\left\{ w_{i}'\theta_0 \leq a \right\} 1\left\{ G(w_{i}'\theta_0, \alpha_i, \epsilon_{it}) \geq y \right\},$$

and above as

$$1\left\{ G(a, \alpha_i, \epsilon_{it}) \geq y \right\} \leq 1\left\{ w_{i}'\theta_0 > a \right\} 1\left\{ G(a, \alpha_i, \epsilon_{it}) \geq y \right\} + 1\left\{ w_{i}'\theta_0 \leq a \right\} \leq 1\left\{ w_{i}'\theta_0 > a \right\} 1\left\{ G(w_{i}'\theta_0, \alpha_i, \epsilon_{it}) \geq y \right\} + 1\left\{ w_{i}'\theta_0 \leq a \right\} = 1 - 1\left\{ w_{i}'\theta_0 \leq a \right\} 1\left\{ G(w_{i}'\theta_0, \alpha_i, \epsilon_{it}) < y \right\}.$$

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

$$\mathbb{P}(Y_{it} \geq y \mid w_{st}) 1\left\{ w_{i}'\theta_0 \leq a \right\} \leq \mathbb{P}(G(a, \alpha_i, \epsilon_{it}) \geq y \mid w_{st}) \leq 1 - \mathbb{P}(Y_{it} < y \mid w_{st}) 1\left\{ w_{i}'\theta_0 > a \right\}.$$

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

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

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

We summarize the identifying restriction above by the following proposition:
Proposition 9. Under Assumptions 1 and 4, the following condition holds:

\[ 1 - P(Y_{it} < y, z_{it}' \beta_0 + X_{it}' \gamma_0 > a | z_{st}) \geq P(Y_{is} \geq y, z_{is}' \beta_0 + X_{is}' \gamma_0 \leq a | z_{st}), \]  

(12)

for any \( a \in \mathcal{R}, y \in \mathcal{Y}, s \neq t \leq T, \) and \( z_{st} \in \mathcal{R}_d^z \times \mathcal{R}_d^z \).

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

7 Simulation

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

7.1 Static Ordered Choice Model

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

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

where the covariate \( Z_{it}^k \) satisfies \( Z_{it}^k \sim \mathcal{N}(0, \sigma_z) \) for \( k \in \{1, 2\} \); the fixed effects \( \alpha_i \) 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 \( (\epsilon_{it1}, \epsilon_{it2}) \) follows the normal distribution \( \mathcal{N}(\mu, \Sigma) \) with \( \mu = (0, 0) \) and \( \Sigma = (1 \rho; \rho 1) \).

The true parameter is \( \beta_0 := (\beta_{01}, \beta_{02})' = (1, 1)' \), 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 \( b_3 = 1 \).
We consider the covariates \( Z_{it} \) to be exogenous and Lemma 1 characterizes the identified set for \( \beta_0 \) using the following conditional moment inequalities:

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

where

\[
g(Z_{ist}, Y_{ist}; \beta_0) = \begin{cases} 
1\{b_2 - Z_{ist}^\prime \beta \geq b_2 - Z_{it}^\prime \beta_0\}(1\{Y_{is} = 1\} - 1\{Y_{it} = 1\}); \\
1\{b_2 - Z_{ist}^\prime \beta \geq b_3 - Z_{it}^\prime \beta_0\}(1\{Y_{is} = 1\} - 1\{Y_{it} \in \{1, 2\}\}); \\
1\{b_3 - Z_{ist}^\prime \beta \geq b_2 - Z_{it}^\prime \beta_0\}(1\{Y_{is} \in \{1, 2\}\} - 1\{Y_{it} = 1\}); \\
1\{b_3 - Z_{ist}^\prime \beta \geq b_3 - Z_{it}^\prime \beta_0\}(1\{Y_{is} \in \{1, 2\}\} - 1\{Y_{it} \in \{1, 2\}\}).
\end{cases}
\]

The first element \( \beta_{01} \) of the parameter \( \beta_0 \) is normalized to one, and we are interested in conducting inference for the parameter \( \beta_{02} \) using the CLR approach. Tables 1 and 2 report the average confidence interval (CI) for \( \beta_{02} \), the coverage probability (CP), the average length of the CI (length), the power of the test at zero (power), and the mean absolute deviation of the lower bound (\( l_{MAD} \)) and upper bound (\( u_{MAD} \)) of the CI.

Table 1: Performance of \( \beta_{02} \) under different values of \( \sigma_z \) (\( \rho = 0.25 \))

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

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

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>CI</th>
<th>CP length</th>
<th>Power</th>
<th>$l_{MAD}$</th>
<th>$u_{MAD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N = 2000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0$</td>
<td>[0.537, 1.755]</td>
<td>0.895</td>
<td>1.218</td>
<td>1.000</td>
<td>0.476</td>
</tr>
<tr>
<td>$\rho = 0.25$</td>
<td>[0.537, 1.760]</td>
<td>0.876</td>
<td>1.222</td>
<td>1.000</td>
<td>0.476</td>
</tr>
<tr>
<td>$\rho = 0.5$</td>
<td>[0.511, 1.765]</td>
<td>0.909</td>
<td>1.254</td>
<td>1.000</td>
<td>0.497</td>
</tr>
<tr>
<td></td>
<td>$N = 8000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0$</td>
<td>[0.584, 1.553]</td>
<td>0.933</td>
<td>0.969</td>
<td>1.000</td>
<td>0.436</td>
</tr>
<tr>
<td>$\rho = 0.25$</td>
<td>[0.570, 1.532]</td>
<td>0.939</td>
<td>0.962</td>
<td>1.000</td>
<td>0.439</td>
</tr>
<tr>
<td>$\rho = 0.5$</td>
<td>[0.573, 1.526]</td>
<td>0.934</td>
<td>0.954</td>
<td>1.000</td>
<td>0.442</td>
</tr>
</tbody>
</table>

7.2 Dynamic Ordered Choice Model

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

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

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

The observed dependent variable $Y_{it}$ is given as

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

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

In this dynamic model, the covariates $Z_i := (Z_{it})_{t=1}^T$ and the initial value $Y_{i0}$ are exogenous, while the lagged variable $Y_{i,t-1}$ is endogenous. Proposition 4 characterizes the identified set for $\theta_0$ with the following conditional moment inequalities:
(1) When \( s \in \{2, 3\} \),

\[
1 - \sum_{j=2}^{3} \Pr(Y_{i1} = y_j \mid z, y_0) \ast \mathbb{1}\{b_j - z_1'\beta - y_0\gamma \geq a\}
\]

\[
\geq \sum_{j=1}^{2} \Pr(Y_{is} = y_j, b_{j+1} - z_{s}'\beta - Y_{is-1}\gamma \leq a \mid z, y_0),
\]

\[
1 - \sum_{j=2}^{3} \Pr(Y_{is} = y_j, b_j - z_{s}'\beta - Y_{is-1}\gamma \geq a \mid z, y_0)
\]

\[
\geq \sum_{j=1}^{2} \Pr(Y_{is} = y_j, b_j - z_{s}'\beta - y_0\gamma \leq a \mid z, y_0) \ast \mathbb{1}\{b_j - z_1'\beta - y_0\gamma \geq a\},
\]

for any \( a \in \{b_j - z_1'\beta - y_0\gamma, b_j - z_2'\beta - \gamma, b_j - z_3'\beta - 2\gamma, b_j - z_3'\beta - 3\gamma\} \).

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

\[
1 - \sum_{j=2}^{3} \Pr(Y_{is} = y_j, b_j - z_{s}'\beta - Y_{is-1}\gamma \geq a \mid z, y_0)
\]

\[
\geq \sum_{j=1}^{2} \Pr(Y_{it} = y_j, b_{j+1} - z_{t}'\beta - Y_{it-1}\gamma \leq a \mid z, y_0),
\]

for any \( a \in \{b_j - z_1'\beta - \gamma, b_j - z_2'\beta - 2\gamma, b_j - z_3'\beta - 3\gamma, b_j - z_3'\beta - 2\gamma, b_j - z_3'\beta - 3\gamma\} \).

We normalize the first parameter \( \beta_0 \) to one, and report the performance of the coefficient \( \gamma_0 \) 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 \( \sigma_z \) 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 \( \gamma_0 \) depends on the variation in \( \Delta z_\cdot'\beta_0 \). A larger variation in \( \Delta z_\cdot'\beta_0 \) 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 $\gamma_0$ under different values of $\sigma_z$ ($\rho = 0.25$)

<table>
<thead>
<tr>
<th>$\sigma_z$</th>
<th>CI</th>
<th>CP</th>
<th>length</th>
<th>Power</th>
<th>$l_{MAD}$</th>
<th>$u_{MAD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_z = 1$</td>
<td>[0.446, 1.606]</td>
<td>0.935</td>
<td>1.160</td>
<td>1.000</td>
<td>0.565</td>
<td>0.625</td>
</tr>
<tr>
<td>$\sigma_z = 1.5$</td>
<td>[0.375, 1.673]</td>
<td>0.959</td>
<td>1.298</td>
<td>1.000</td>
<td>0.629</td>
<td>0.693</td>
</tr>
<tr>
<td>$\sigma_z = 2$</td>
<td>[0.311, 1.730]</td>
<td>0.960</td>
<td>1.418</td>
<td>1.000</td>
<td>0.700</td>
<td>0.739</td>
</tr>
</tbody>
</table>

| $\sigma_z = 1$ | [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 $\gamma_0$ under different values of $\rho$ ($\sigma_z = 1$)

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>CI</th>
<th>CP</th>
<th>length</th>
<th>Power</th>
<th>$l_{MAD}$</th>
<th>$u_{MAD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho = 0$</td>
<td>[0.472, 1.593]</td>
<td>0.932</td>
<td>1.121</td>
<td>1.000</td>
<td>0.550</td>
<td>0.607</td>
</tr>
<tr>
<td>$\rho = 0.25$</td>
<td>[0.446, 1.606]</td>
<td>0.935</td>
<td>1.160</td>
<td>1.000</td>
<td>0.565</td>
<td>0.625</td>
</tr>
<tr>
<td>$\rho = 0.5$</td>
<td>[0.457, 1.631]</td>
<td>0.943</td>
<td>1.173</td>
<td>1.000</td>
<td>0.548</td>
<td>0.648</td>
</tr>
</tbody>
</table>

| $\rho = 0$ | [0.528, 1.472] | 0.958 | 0.945  | 1.000 | 0.475 | 0.487 |
| $\rho = 0.25$ | [0.529, 1.495] | 0.969 | 0.966  | 1.000 | 0.473 | 0.504 |
| $\rho = 0.5$ | [0.535, 1.515] | 0.975 | 0.980  | 1.000 | 0.467 | 0.519 |
8 Empirical Application

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

<table>
<thead>
<tr>
<th></th>
<th>income category</th>
<th>residence</th>
<th>tenure /100</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>1.990</td>
<td>0.799</td>
<td>0.825</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.810</td>
<td>0.401</td>
<td>0.738</td>
</tr>
<tr>
<td>25% quantile</td>
<td>1.000</td>
<td>1.000</td>
<td>0.220</td>
</tr>
<tr>
<td>median</td>
<td>2.000</td>
<td>1.000</td>
<td>0.605</td>
</tr>
<tr>
<td>75% quantile</td>
<td>3.000</td>
<td>1.000</td>
<td>1.280</td>
</tr>
<tr>
<td>minimum</td>
<td>1.000</td>
<td>1.000</td>
<td>0.010</td>
</tr>
<tr>
<td>maximum</td>
<td>3.000</td>
<td>1.000</td>
<td>4.850</td>
</tr>
</tbody>
</table>

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

\(^7\)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.
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>
<thead>
<tr>
<th></th>
<th>$\beta_{0,1}$ (residence)</th>
<th>$\beta_{0,2}$ (tenure)</th>
<th>$\gamma_0$ (lag)</th>
</tr>
</thead>
<tbody>
<tr>
<td>exogenous static model</td>
<td>1</td>
<td>[0.612, 0.939]</td>
<td>-</td>
</tr>
<tr>
<td>endogenous static model</td>
<td>1</td>
<td>[0.041, 0.939]</td>
<td>-</td>
</tr>
<tr>
<td>dynamic model</td>
<td>1</td>
<td>[0.531, 0.694]</td>
<td>[0.286, 0.612]</td>
</tr>
</tbody>
</table>

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

9 Conclusion

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

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

---

8Botosaru and Muris (2022) proposes an approach to obtain bounds on counterfactual CCPs in semi-parametric dynamic panel data models, assuming that the index parameters are (partially) identified.
(parametric index in our case) by an arbitrary constant so as to obtain an object free of endogeneity issues may have broader applicability beyond the models studied in this work, and it remains to see whether our key identification strategy can be further adapted to other structures.

\section{Appendix}

\subsection{A.1 Proof of Lemma 1}

\textit{Proof.} The identifying condition (7) for the parameter \( \theta \) in Proposition 1 is given as

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

(13)

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

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

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

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

\hfill \Box

\subsection{A.2 Proof of Theorem 1}

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

Let \( P_{st}(j, k \mid w_{st}) := \Pr(Y_{is} = j, Y_{it} = k \mid W_{ist} = w_{st}) \) denote the observed joint probability of selecting \( j \) at period \( s \) and selecting \( k \) at period \( t \). The first requirement for sharpness
is that the constructed distribution of \((v_{is}, v_{it}) \mid W_{ist}\) matches the observed joint choice probability according to the model: given \(W_{ist} = w_{st}\),
\[
\begin{align*}
P_{st}(1, 1 \mid w_{st}) &= \Pr(v_{is} \leq w_{s,t}^\theta, v_{it} \leq w_{t,t}^\theta \mid w_{st}), \\
P_{st}(1, 0 \mid w_{st}) &= \Pr(v_{is} \leq w_{s,t}^\theta \mid w_{st}) - \Pr(v_{is} \leq w_{s,t}^\theta, v_{it} \leq w_{t,t}^\theta \mid w_{st}), \\
P_{st}(0, 1 \mid w_{st}) &= \Pr(v_{it} \leq w_{t,t}^\theta \mid w_{st}) - \Pr(v_{is} \leq w_{s,t}^\theta, v_{it} \leq w_{t,t}^\theta \mid w_{st}),
\end{align*}
\]
and \(P_{st}(0,0 \mid w_{st})\) can be matched automatically when the above conditions hold.

Another condition for sharpness requires the marginal distribution of \(v_{it} \mid W_{ist}\) to satisfy the following stationarity condition in Assumption 1:
\[
F_{v_{it}|z_{st}}(a \mid z_{st}) = F_{v_{is}|z_{st}}(a \mid z_{st}) \quad \forall a \in \mathcal{R}.
\]

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

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

Now, for any \(\theta \in \Theta_{I,1}\), our objective is to construct the distribution of \(v_{it} \mid W_{st}\) and \(v_{is} \mid W_{st}\) such that conditions (15) and (16) hold. Consider that the endogenous covariate \(X_{it}\) only takes \(K\) values: \(X_{it} \in \mathcal{X} = \{a_1, ..., a_K\}\) for any \(t \leq T\). \(^9\) Given \(Z_{it} = z_t\), the covariate index \(z_t^\beta + x_t^\gamma\) belongs to the \(K\) points \(\{z_1^\beta + a_1^\gamma, ..., z_K^\beta + a_K^\gamma\}\). Without loss of generalization, we can rank the \(K\) points as \(a_1^\gamma \leq a_2^\gamma \leq \ldots \leq a_K^\gamma\) since we can reindex the \(K\) values of \(\{a_1, ..., a_K\}\) such that this ranking is satisfied. Let \(m_k := z_k^\beta + a_k^\gamma\) denote the \(K\) points for which \(w_{t,t}^\theta\) can take, and \(n_k := z_k^\beta + a_k^\gamma\) for \(1 \leq k \leq K\) denote the \(K\) points \(w_{s,t}^\theta\) can take. Let \(\bar{m} = \max\{z_1^\beta, z_k^\beta\} + a_k^\gamma + \epsilon\) for some \(\epsilon > 0\).

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

\(^9\) The covariate \(X_{it}\) is allowed to have different support across time \(t\), then \(\mathcal{X}\) is the union of the support of \(X_{it}\) over all periods.
distribution $F_{v_t|Z_{st}}(a | z_{st})$, given as follows:

$$F_{v_t|Z_{st}}(a | z_{st}) \leq \sum_{x_{st} \in \mathcal{X}} \left\{ F_{v_t|W_{st}}(w'_t \theta | w_{st}) \mathbb{1}\{w'_t \theta \geq a\} + \mathbb{1}\{w'_t \theta < a\} \right\} \Pr(X_{st} = x_{st} | z_{st})$$

$$:= U_t(a | z_{st}),$$

$$F_{v_t|Z_{st}}(a | z_{st}) \geq \sum_{x_{st} \in \mathcal{X}} F_{v_t|W_{st}}(w'_t \theta | w_{st}) \mathbb{1}\{w'_t \theta \leq a\} \Pr(X_{st} = x_{st} | z_{st}) := L_t(a | z_{st}).$$

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

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

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

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

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

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

**Construction**

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

**Step 1:**

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

\[
F_{v_{st} \mid W_{st}}^{U}(a \mid w_{st}) = \begin{cases} 
0 & \text{if } a < m_1, \\
P_t(w_{st}) & \text{if } m_1 \leq a < w_t'\theta + \epsilon, \\
1 & \text{if } w_t'\theta + \epsilon \leq a,
\end{cases}
\]

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

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

\[
F_{v_{st} \mid W_{st}}^{L}(a \mid w_{st}) = \begin{cases} 
0 & \text{if } a < w_t'\theta, \\
P_t(w_{st}) & \text{if } w_t'\theta \leq a < \bar{m}, \\
1 & \text{if } \bar{m} \leq a
\end{cases}
\]

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

**Step 2:** this step establishes the construction of the distribution \( F_{v_{st} \mid W_{st}}^{*}(a \mid w_{st}) \) at the points \( m_k \). The observed marginal distribution given in condition (16) only restricts the distribution of \( v_{st} \mid W_{st} \) at each point \( m_k \), given \( w_{st} = (z_{st}, x_{s}, a_{k}) \),

\[ P_t(w_{st}) = \Pr(v_{st} \leq m_k \mid z_{st}, x_{s}, a_k). \]

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

\[
\bar{L}(m_K \mid z_{st}) = \begin{cases} 
U_t(m_1 \mid z_{st}) & \text{when } j_1 = 1, \\
U_t(m_K-1 \mid z_{st}) & \text{when } j_1 = 2, \\
U_t(m_K \mid z_{st}) & \text{when } j_1 = K.
\end{cases}
\]

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

\[
\bar{L}(m_K \mid z_{st}) = \alpha_1 U_t(m_{j_1-1} \mid z_{st}) + (1 - \alpha_1) U_t(m_{j_1} \mid z_{st}).
\]

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

\[
F_{v_{st} \mid W_{st}}^{*}(m_K \mid w_{st}) = \alpha_1 F_{t}^{U}(m_{j_1-1} \mid w_{st}) + (1 - \alpha_1) F_{t}^{U}(m_{j_1} \mid w_{st}).
\]
The above construction generates the marginal distribution $\tilde{L}(m_K \mid z_{st})$. Moreover, it matches the marginal choice probability when $x_t = a_K$ in condition (16) since $F^U_t(m_j \mid w_{st}) = P_t(w_{st})$ if $x_t = a_K$ for any $1 \leq j \leq K$ from the definition of $F^U_t(m_j \mid w_{st})$.

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

\[
\begin{array}{c}
L_t(m_{K-1} \mid z_{st}) & U_t(m_1 \mid z_{st}) = L_t(m_K \mid z_{st}) & U_t(m_{K-1} \mid z_{st}) U_t(m_K \mid z_{st})
\end{array}
\]

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

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

Accordingly, the distribution $F^*_t|w_{st}(m_{K-1} \mid w_{st})$ at $m_{K-1}$ is constructed as

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

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

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

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

**Step 3:** we show the construction at the points $a = n_k$ for $1 \leq k \leq K$. For points $a = n_k$, we only need to construct a proper distribution to be consistent with Equation (17). We still start with the largest point $n_K$. The distribution $\tilde{L}(n_K \mid z_{st})$ can be expressed as
follows in different scenarios: for some $2 \leq j_3 \leq K$

$$
L(n_K \mid z_{st}) = \begin{cases} 
\alpha_3 \tilde{L}(m_K \mid z_{st}) + (1 - \alpha_3) \cdot 1 & \text{if } n_K > m_K, \\
\alpha_4 \tilde{L}(m_{j_{3-1}} \mid z_{st}) + (1 - \alpha_4) \tilde{L}(m_{j_3} \mid z_{st}) & \text{if } m_{j_{3-1}} < n_K < m_{j_3}, \\
\alpha_5 \cdot 0 + (1 - \alpha_5) \tilde{L}(m_1 \mid z_{st}) & \text{if } n_K < m_1,
\end{cases}
$$

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

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

$$
F^*_{v_t | W_{st}}(n_K \mid w_{st}) = \begin{cases} 
\alpha_3 F^*_{v_t | W_{st}}(m_K \mid w_{st}) + (1 - \alpha_3) \cdot 1 & \text{if } n_K > m_K, \\
\alpha_4 F^*_{v_t | W_{st}}(m_{j_{3-1}} \mid w_{st}) + (1 - \alpha_4) F^*_{v_t | W_{st}}(m_{j_3} \mid w_{st}) & \text{if } m_{j_{3-1}} < n_K < m_{j_3}, \\
\alpha_5 \cdot 0 + (1 - \alpha_5) F^*_{v_t | W_{st}}(m_1 \mid w_{st}) & \text{if } n_K < m_1.
\end{cases}
$$

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

$$
\tilde{L}(n_{K-1} \mid z_{st}) = \begin{cases} 
\alpha_6 \tilde{L}(m_K \mid z_{st}) + (1 - \alpha_6) \tilde{L}(n_K \mid z_{st}) & \text{if } m_K < n_{K-1} < n_K, \\
\alpha_7 \tilde{L}(m_{j_{3-1}} \mid z_{st}) + (1 - \alpha_7) \tilde{L}(n_K \mid z_{st}) & \text{if } m_{j_{3-1}} < n_{K-1} < n_K < m_{j_3}, \\
\alpha_8 \tilde{L}(m_{j_{4-1}} \mid z_{st}) + (1 - \alpha_8) \tilde{L}(m_{j_4} \mid z_{st}) & \text{if } m_{j_{4-1}} < n_{K-1} < m_{j_4} \leq m_{j_3-1}, \\
\alpha_9 \cdot 0 + (1 - \alpha_9) \tilde{L}(n_K \mid z_{st}) & \text{if } n_{K-1} < n_K < m_1,
\end{cases}
$$

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

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

$$
F^*_{v_t | W_{st}}(n_K \mid w_{st}) = \begin{cases} 
\alpha_6 F^*_{v_t | W_{st}}(m_K \mid w_{st}) + (1 - \alpha_6) F^*_{v_t | W_{st}}(n_K \mid w_{st}) & \text{if } m_K < n_{K-1} < n_K, \\
\alpha_7 F^*_{v_t | W_{st}}(m_{j_{3-1}} \mid w_{st}) + (1 - \alpha_7) F^*_{v_t | W_{st}}(n_K \mid w_{st}) & \text{if } m_{j_{3-1}} < n_{K-1} < n_K < m_{j_3}, \\
\alpha_8 F^*_{v_t | W_{st}}(m_{j_{4-1}} \mid w_{st}) + (1 - \alpha_8) F^*_{v_t | W_{st}}(m_{j_4} \mid w_{st}) & \text{if } m_{j_{4-1}} < n_{K-1} < m_{j_4} \leq m_{j_3-1}, \\
(1 - \alpha_9) F^*_{v_t | W_{st}}(n_K \mid w_{st}) & \text{if } n_{K-1} < n_K < m_1.
\end{cases}
$$

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

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

38
A.3 Proof of Proposition 2 and 3

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

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

$$1 - \Pr(Y_{i1} = 0, z_1' \beta_0 + X_{i1}' \gamma_0 \geq a \mid z_{12}) \geq \Pr(Y_{i2} = 1, z_2' \beta_0 + X_{i2}' \gamma_0 \leq a \mid z_{12}),$$

for $a \in \{z_1' \beta_0 + a_k' \gamma_0, \ldots, z_1' \beta_0 + a_K' \gamma_0, z_2' \beta_0 + a_1' \gamma_0, \ldots, z_2' \beta_0 + a_K' \gamma_0\}$, and another identifying condition switches the order of period 1 and 2.

Let $a = z_1' \beta_0 + a_k' \gamma_0$, then the above upper bound can be further bounded as

$$1 - \Pr(Y_{i1} = 0, z_1' \beta_0 + X_{i1}' \gamma_0 \geq z_1' \beta_0 + a_k' \gamma_0 \mid z_{12}) \leq 1 - \Pr(Y_{i1} = 0, X_{i1} = a_k \mid z_{12}).$$

When $z_1' \beta_0 - z_2' \beta_0 \geq 0$ which implies $z_1' \beta_0 + a_k' \gamma_0 \geq z_2' \beta_0 + a_k' \gamma_0$, then the lower bound can be bounded below as

$$\Pr(Y_{i2} = 1, z_2' \beta_0 + X_{i2}' \gamma_0 \leq z_1' \beta_0 + a_k' \gamma_0 \mid z_{12}) \leq \Pr(Y_{i2} = 1, X_{i2} = a_k \mid z_{12}).$$

Combining the above results leads to

If $z_1' \beta_0 - z_2' \beta_0 \geq 0 \implies 1 - \Pr(Y_{i1} = 0, X_{i1} = a_k \mid z_{12}) \geq \Pr(Y_{i2} = 1, X_{i2} = a_k \mid z_{12}).$

The contraposition of the above inequality yields

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

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

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

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

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

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

---

10 The value of $a = z_2' \beta_0 + a_k' \gamma_0$ leads to the same identifying condition.
discuss three cases: $b^* < 0$, $b^* = 0$, and $b^* > 0$.

Case 1: $b^* < 0$. When the covariate $\Delta z^*$ takes a large positive value $\Delta z^* \to +\infty$ 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: $b^* = 0$. For any value $\Delta z$, the value of $\Delta z'b$ is either positive or nonpositive. When $\Delta z'b > 0$ is positive, then let $\Delta z^*$ take a large negative value $\Delta z^* \to -\infty$ 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 $\Delta z^* \to +\infty$ such that $\Delta z'\beta_0 > 0$.

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

For the identification of $\gamma_0$, under the similar analysis for $\beta_0$, we have

$$(z_1, z_2) \in Z_3^j \implies (x_1^j - x_2^j)\gamma_0^j < \Delta z'\beta_0,$$

$$(z_1, z_2) \in Z_4^j \implies (x_1^j - x_2^j)\gamma_0^j > \Delta z'\beta_0.$$

As previously shown, when $(z_1, z_2) \in Z_2$, it implies that $\Delta z'\beta_0 < 0$. Therefore, when $(z_1, z_2) \in Z_2 \cap Z_3^j$, we have $(x_1^j - x_2^j)\gamma_0^j < \Delta z'\beta_0 < 0$ and the sign of $\gamma_0^j$ is identified given $x_1^j \neq x_2^j$. Similarly, when $(z_1, z_2) \in Z_1 \cap Z_4^j$, the sign of $\gamma_0^j$ is also identified given $(x_1^j - x_2^j)\gamma_0^j > \Delta z'\beta_0 > 0$. Proposition 3 requires that for any $j \leq d_x$, either $Z_2 \cap Z_3^j \neq \emptyset$ or $Z_1 \cap Z_4^j \neq \emptyset$ so that the sign of $\gamma_0^j$ is identified for any $j$. 

\[ \square \]
A.4 Relationship to Khan, Ponomareva, and Tamer (2023)

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

(i) $P(y_{it} = 1 \mid z_{st}) > P(y_{is} = 1 \mid z_{st}) \implies (z_t - z_s)\beta_0 + \gamma_0 > 0$,

(ii) $P(y_{it} = 1 \mid z_{st}) > 1 - P(y_{is-1} = 1, y_{is} = 0 \mid z_{st}) \implies (z_t - z_s)\beta_0 - \min\{0, \gamma_0\} > 0$,

(iii) $P(y_{it} = 1 \mid z_{st}) > 1 - P(y_{is-1} = 0, y_{is} = 0 \mid z_{st}) \implies (z_t - z_s)\beta_0 + \max\{0, \gamma_0\} > 0$,

(iv) $P(y_{it-1} = 1, y_{it} = 1 \mid z_{st}) > P(y_{is} = 1 \mid z_{st}) \implies (z_t - z_s)\beta_0 + \max\{0, \gamma_0\} > 0$,

(v) $P(y_{it-1} = 1, y_{it} = 1 \mid z_{st}) > 1 - P(y_{is-1} = 1, y_{is} = 0 \mid z_{st}) \implies (z_t - z_s)\beta_0 > 0$,

(vi) $P(y_{it-1} = 1, y_{it} = 1 \mid z_{st}) > 1 - P(y_{is-1} = 0, y_{is} = 0 \mid z_{st}) \implies (z_t - z_s)\beta_0 + \gamma_0 > 0$,

(vii) $P(y_{it-1} = 0, y_{it} = 1 \mid z_{st}) > P(y_{is} = 1 \mid z_{st}) \implies (z_t - z_s)\beta_0 - \min\{0, \gamma_0\} > 0$,

(viii) $P(y_{it-1} = 0, y_{it} = 1 \mid z_{st}) > 1 - P(y_{is-1} = 1, y_{is} = 0 \mid z_{st}) \implies (z_t - z_s)\beta_0 - \gamma_0 > 0$,

(ix) $P(y_{it-1} = 0, y_{it} = 1 \mid z_{st}) > 1 - P(y_{is-1} = 0, y_{is} = 0 \mid z_{st}) \implies (z_t - z_s)\beta_0 > 0$.

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

$$U_s(a) := 1 - \Pr(Y_{is} = 0, z_s'\beta_0 + Y_{i,s-1}\gamma_0 \geq a \mid z_{st})$$

$$\geq \Pr(Y_{it} = 1, z_t'\beta_0 + Y_{i,t-1}\gamma_0 \leq a \mid z_{st}) := L_t(a).$$

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

$$U_s(a) = 1 - \Pr(Y_{is} = 0, Y_{i,s-1} = 1 \mid z_{st}) \mathbb{1}\{z_s'\beta_0 + \gamma_0 \geq a\} - \Pr(Y_{is} = 0, Y_{i,s-1} = 0 \mid z_{st}) \mathbb{1}\{z_s'\beta_0 \geq a\}. $$

Similarly, the lower bound is given as

$$L_t(a) = \Pr(Y_{it} = 1, Y_{i,t-1} = 1 \mid z_{st}) \mathbb{1}\{z_t'\beta_0 + \gamma_0 \leq a\} + \Pr(Y_{it} = 1, Y_{i,t-1} = 0 \mid z_{st}) \mathbb{1}\{z_t'\beta_0 \leq a\}. $$

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

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

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

$$L_t(a) = \Pr(Y_{it} = 1, Y_{i,t-1} = 1 \mid z_{st}) + \Pr(Y_{it} = 1, Y_{i,t-1} = 0 \mid z_{st}) = \Pr(Y_{it} = 1 \mid z_{st}).$$
By exploiting the condition \( \max\{0, \gamma_0\} = \min\{0, \gamma_0\} = |\gamma_0|\), the identifying condition is given as

\[
(z_i' \beta_0 - z_s' \beta_0) + |\gamma_0| \leq 0 \implies U_s(a) = \Pr(Y_{is} = 1 | z_{st}) \geq L_t(a) = \Pr(Y_{it} = 1 | z_{st}).
\]

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

\[
\Pr(Y_{it} = 1 | z_{st}) > \Pr(Y_{is} = 1 | z_{st}) \implies (z_i' \beta_0 - z_s' \beta_0) + |\gamma_0| > 0.
\]

**Case 2:** consider that \( z_i' \beta_0 + \max\{0, \gamma_0\} \leq z_s' \beta_0 + \gamma_0 \). Let \( a = z_s' \beta_0 + \gamma_0 \), the bounds are simplified as

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

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

\[
z_i' \beta_0 - z_s' \beta_0 - \min\{0, \gamma_0\} \leq 0 \implies 1 - \Pr(Y_{is} = 0, Y_{i,s-1} = 1 | z_{st}) \geq \Pr(Y_{it} = 1 | z_{st}).
\]

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

\[
\Pr(Y_{it} = 1 | z_{st}) > 1 - \Pr(Y_{is} = 0, Y_{i,s-1} = 1 | z_{st}) \implies z_i' \beta_0 - z_s' \beta_0 - \min\{0, \gamma_0\} > 0.
\]

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

\( \square \)

### A.5 Proof of Proposition 2

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

\[
U_{t,\text{order}}(a | z_{st}) := 1 - \sum_{j=1}^{J} \Pr(Y_{is} = y_j, b_j - z_s' \beta - X_{is}' \gamma \geq a | z_{st}) \geq \sum_{j=1}^{J} \Pr(Y_{it} = y_j, b_{j+1} - z_i' \beta - X_{it}' \gamma \leq a | z_{st}) := L_{t,\text{order}}(a | z_{st}).
\] (18)

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

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

When \( a < q_1 \), then \( \Pr(Y_{it} = y_j, b_{j+1} - z_i' \beta - X_{it}' \gamma \leq a | z_{st}) = 0 \) for any \( 2 \leq j \leq K \) so that the lower bound is zero \( L_{t,\text{order}}(a | z_{st}) = 0 \) and it is not informative for \( \theta \).

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

\[\]

A.6 Proof of Lemma 1

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

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

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

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

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

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

\[
\text{if } b_{j_1} - w'_s\theta_0 \geq b_{j_2+1} - w'_t\theta_0 \implies \sum_{j=1}^{j_1-1} \Pr(Y_{is} = y_j \mid w_{st}) = 1 - \sum_{j=j_1}^{J} \Pr(Y_{is} = y_j \mid w_{st}) \\
\geq \sum_{j=1}^{j_2} \Pr(Y_{it} = y_j \mid w_{st}).
\]

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

\[
\]

A.7 Proof of Proposition 5 and Proposition 6

**Proof.** We first prove that \(\beta_0\) is point identified up to scale by showing that for any \(\Delta z \in \Delta \mathcal{Z}_{\text{order}},\) the sign of \(\Delta z'\beta_0\) is identified. The identifying condition for \(\theta_0\) in Proposition 4 is
given as

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

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

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

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

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

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

Combining the results yields the following conditions:

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

The contraposition of the above condition generates

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

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

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

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

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

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

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

Similarly, the contraposition of the above condition gives

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

Therefore, when \( \Delta z \in \mathcal{Z}_{3,\text{order}} \cap \mathcal{Z}_{2,\text{order}} \), it implies that

\[
0 < \Delta z' \beta_0 < (x_1 - x_2)' \gamma_0.
\]

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

\[\square\]

A.8 Proof of Proposition 7

Proof. For any set \( K \subset J \), the conditional probability of selecting a choice \( j \in K \) given \( W_{i,t} = w_{i,t} \) is:

\[
\Pr(Y_{i,t} \in K \mid w_{i,t}) = \Pr(\exists j \in K \text{ s.t. } w'_{ij,t} \theta_0 + v_{ij,t} \geq w'_{ikt} \theta_0 + v_{ikt} \forall k \in K^c \mid w_{i,t}).
\]

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

\[
Q_t(a_{jk} \mid w_{i,t}) := \Pr(\exists j \in K \text{ s.t. } v_{ikt} - v_{ij,t} \leq a_{jk} \forall k \in K^c \mid w_{i,t}).
\]

When \( a_{jk} \) satisfies \( a_{jk} \geq (w_{ij,t} - w_{ikt})' \theta_0 \) for any \((j,k)\), then the above probability can be bounded below as

\[
Q_t(a_{jk} \mid w_{i,t}) \geq \Pr(\exists j \in K \text{ s.t. } v_{ikt} - v_{ij,t} \leq (w_{ij,t} - w_{ikt})' \theta_0 \forall k \in K^c \mid w_{i,t})
\]

\[
= \Pr(Y_{i,t} \in K \mid w_{i,t}).
\]

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

\[
Q_t(a_{jk} \mid w_{i,t}) \geq \Pr(Y_{i,t} \in K \mid w_{i,t}) \mathbb{1}\{a_{jk} \geq (w_{ij,t} - w_{ikt})' \theta_0 \forall j \in K, k \in K^c\}.
\]

The above inequality holds since either \( a_{jk} \geq (w_{ij,t} - w_{ikt})' \theta_0 \) or the lower bound is zero.
Since Assumption 1 is only conditioning on the exogenous covariate \( z_{st} \), we can bound the conditional distribution \( Q_t(a_{jk} \mid z_{st}) \) by taking expectation of \( X_{it} \) given \( z_{st} \):

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

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

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

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

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

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

Rearranging the above formula yields

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

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

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

\( \square \)

### A.9 Proof of Proposition 8

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

\[
\Pr(Y_{it} = 0 \mid w_{st}) = \Pr(Y_{it}^* \leq 0 \mid w_{st}) = \Pr(v_{it} \geq z_t'\beta_0 + x_t'\gamma_0 \mid w_{st}).
\]

46
When \( y > 0 \), the conditional distribution is given as
\[
\Pr(Y_{it} \leq y \mid w_{it}) = \Pr(Y_{it}^* \leq 0, Y_{it} \leq y \mid w_{it}) + \Pr(0 < Y_{it}^*, Y_{it} \leq y \mid w_{it}) \\
= \Pr(Y_{it}^* \leq 0 \mid w_{it}) + \Pr(0 < Y_{it}^* \leq y \mid w_{it}) \\
= \Pr(Y_{it}^* \leq y \mid w_{it}) \\
= \Pr(v_{it} \geq z_t^* \beta_0 + x_t^* \gamma_0 - y \mid w_{it}).
\]

Combining the two scenarios, the full conditional distributional of \( Y_{it} \mid W_{ist} \) is characterized as follows:
\[
\Pr(Y_{it} \leq y \mid w_{it}) = \begin{cases}
\Pr(v_{it} \geq z_t^* \beta_0 + x_t^* \gamma_0 - y \mid w_{it}) & \text{if } y \geq 0, \\
0 & \text{if } y < 0.
\end{cases}
\]

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

\[
\Pr(v_{it} \geq a \mid w_{st}) \leq \Pr(Y_{it} \leq z_t^* \beta_0 + x_t^* \gamma_0 - a) \mathbb{1}\{z_t^* \beta_0 + x_t^* \gamma_0 \geq a\} + \Pr(Y_{it} = 0 \mid z_{st}) \mathbb{1}\{z_t^* \beta_0 + x_t^* \gamma_0 < a\},
\]

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

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

\[
\Pr(v_{it} \geq a \mid z_{st}) \leq \Pr(Y_{it} \leq z_t^* \beta_0 + X_t^* \gamma_0 - a, z_t^* \beta_0 + X_t^* \gamma_0 \geq a \mid z_{st}) + \Pr(Y_{it} = 0, z_t^* \beta_0 + X_t^* \gamma_0 < a \mid z_{st}).
\]

The above upper bound can be also also expressed as

\[
\Pr(Y_{it} \leq z_t^* \beta_0 + X_t^* \gamma_0 - a, z_t^* \beta_0 + X_t^* \gamma_0 \geq a \mid z_{st}) + \Pr(Y_{it} = 0, z_t^* \beta_0 + X_t^* \gamma_0 < a \mid z_{st}) \\
= \Pr(0 < Y_{it} \leq z_t^* \beta + X_t^* \gamma - a, z_t^* \beta + X_t^* \gamma \geq a \mid z_{st}) + \Pr(Y_{it} = 0 \mid z_{st}) \\
= \Pr(0 < Y_{it} \leq z_t^* \beta + X_t^* \gamma - a \mid z_{st}) + \Pr(Y_{it} = 0 \mid z_{st}).
\]

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

\[
\Pr(v_{it} \geq a \mid w_{st}) \geq \Pr(Y_{it} \leq z_t^* \beta_0 + x_t^* \gamma_0 - a),
\]

where the above condition holds since either \( z_t^* \beta_0 + x_t^* \gamma_0 - a \geq 0 \) so that there exists \( y = z_t^* \beta_0 + x_t^* \gamma_0 - a \geq 0 \) such that \( \Pr(Y_{it} \leq y \mid w_{st}) = \Pr(v_{it} \geq a \mid w_{st}) \), or the lower bound is zero when \( z_t^* \beta_0 + x_t^* \gamma_0 < a \).
Taking expectation over $X_{ist}$ leads to the following lower bound:
\[
Pr(v_{it} \geq a \mid z_{st}) \geq Pr(Y_{it} \leq z'_i\beta_0 + X'_{it}\gamma_0 - a \mid z_{st}).
\]

The conditional stationarity condition requires $v_{is} \mid Z_{ist} \overset{d}{\sim} v_{it} \mid Z_{ist}$, which implies that the bounds for $v_{it} \mid Z_{ist}$ must have intersections over any pair of periods $(s, t)$. This restriction generates the following identifying condition for $\theta_0$:
\[
Pr(0 < Y_{is} \leq z'_s\beta + X'_{is}\gamma - a \mid z_{st}) + Pr(Y_{is} = 0 \mid z_{st}) \geq Pr(Y_{it} \leq z'_i\beta_0 + X'_{it}\gamma_0 - a \mid z_{st})
\]
for any $a \in \mathcal{R}$, any $s, t$, and any $z_{st}$.

\[\square\]

A.10 Dynamic Censored Models with Latent Dependent Variables

Consider the following dynamic censored models with the latent lagged outcome $Y_{i,t-1}^*$:
\[
Y_{it}^* = Z'_{it}\beta_0 + Y_{i,t-1}^*\gamma_0 + \alpha_i + \epsilon_{it},
\]
\[
Y_{it} = \max\{Y_{it}^*, 0\}.
\]

In this model, the endogenous variable $X_{it}$ is the lagged outcome: $X_{it} = Y_{i,t-1}^*$. However, the variable $Y_{i,t-1}^*$ is not observed in data, so the results in Proposition 8 cannot be directly applied here. Due to this feature in the dynamic model, we need to adjust the results in Proposition 8. Given that $Y_{i,t-1}^* = Y_{i,t-1}$ when $Y_{i,t-1} > 0$, we can further relax the lower and upper bounds in (8) to identify $\theta_0$.

The lower bound in condition (8) can be bounded below as follows:
\[
Pr(Y_{it} \leq z'_i\beta + Y_{i,t-1}^*\gamma - a \mid z_{st}) \\
\geq Pr(Y_{it} \leq z'_i\beta + Y_{i,t-1}^*\gamma - a, Y_{i,t-1} > 0 \mid z_{st}) := L_{t,cen}(a \mid z_{st}; \theta). \\
\]

Similarly, the upper bound in condition (8) can be further bounded above
\[
Pr(0 < Y_{is} \leq z'_s\beta + Y_{i,s-1}^*\gamma - a \mid z_{st}) + Pr(Y_{is} = 0 \mid z_{st}) \leq U_{s,cen}(a \mid z_{st}; \theta),
\]
where $U_{s,cen}(a \mid z_{st}; \theta)$ is defined as
\[
U_{s,cen}(a \mid z_{st}; \theta) := Pr(0 < Y_{is} \leq z'_s\beta + Y_{i,s-1}^*\gamma - a, Y_{i,s-1} > 0 \mid z_{st}) \\
+ Pr(Y_{is} > 0, Y_{i,s-1} = 0 \mid z_{st}) + Pr(Y_{is} = 0 \mid z_{st}).
\]

For the dynamic model, an identified set for $\theta_0$ is characterized by the following lemma:

**Lemma 4.** Under Assumption 1 and $X_{it} = Y_{i,t-1}^*$, the identified set for $\theta_0$ is characterized...
by the set of parameters $\theta$ that satisfy the following conditions:

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

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

References


49


51