Aristodemou, Eleni

Working Paper

Semiparametric Identification in Panel Data Discrete Response Models

Tinbergen Institute Discussion Paper, No. TI 2018-065/III

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Aristodemou, Eleni (2018) : Semiparametric Identification in Panel Data Discrete Response Models, Tinbergen Institute Discussion Paper, No. TI 2018-065/III, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/185584

TI 2018-065/III
Tinbergen Institute Discussion Paper

Semiparametric Identification in Panel Data Discrete Response Models

Eleni (E.) Aristodemou[1]

[1] University of Amsterdam, The Netherlands

Semiparametric Identification in Panel Data Discrete Response Models*

Eleni Aristodemou†

University of Amsterdam

August 9, 2018

Abstract

This paper studies semiparametric identification in linear index discrete response panel data models with fixed effects. Departing from the classic binary response static panel data model, this paper examines identification in the binary response dynamic panel data model and the ordered response static panel data model. It is shown that under mild distributional assumptions on the fixed effect and the time-varying unobservables, point-identification fails but informative bounds on the regression coefficients can still be derived. Partial identification is achieved by eliminating the fixed effect and discovering features of the distribution of the unobservable time-varying components that do not depend on the unobserved heterogeneity. Numerical analyses illustrate how the identified set changes as the support of the explanatory variables varies.

Keywords: Static and Dynamic Panel Data, Binary Response Models, Ordered Response Models, Semiparametric Identification, Partial Identification

JEL classification Numbers: C01, C33, C35.

*I would like to thank Andrew Chesher, Geert Dhaene, Bo Honoré, Frank Kleibergen, Lars Nesheim, Adam Rosen, Martin Weidner and Frank Windmeijer for helpful discussions and comments. Any errors are my own.

†Address: Quantitative Economics Section, Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1001 NJ, Amsterdam. Email: e.aristodemou@uva.nl

1 Introduction

This paper provides new results on semiparametric identification in fixed effects panel data discrete response models, and in particular linear index binary response dynamic panel data models and linear index ordered response static panel data models with additively separable fixed effects. It is shown that under mild distributional assumptions on the fixed effect and the time-varying unobservables, point-identification fails but informative identification bounds for the regression coefficients can still be derived. In the dynamic binary response setting, partial identification of the regression coefficients is achieved by observing individuals who switch in two consecutive time periods, conditional on their initial condition. In the static ordered response setting on the other hand, in addition to the individuals who switch from one period to the next, it is shown that individuals who choose the "in-between" category in two consecutive periods are also a useful source of identification.

As pointed out by Heckman (1981a) intertemporal correlation in the decisions of individuals in panel data models comes in general through the presence of time-invariant unobservables and lagged dependent variables in the underlying functional form specification. Ignoring this dynamic behaviour can result in inconsistent estimates of the regression coefficients and other quantities of interest, while distinguishing between the causes of autocorrelation may have important policy implications.

Linear panel data model settings where the dependent outcome is continuous, can be seen as solving an omitted variables problem, arising from the presence of this additively separable fixed effect. Even when this fixed effect is not restricted to be independent of the explanatory variables, point-identification of the regression parameters can be achieved by differencing out the fixed effect.

In non-linear panel data models with additively separable fixed effects, the differencing approach can not be directly implemented. Identification and estimation in these models rely heavily on the assumptions econometricians place on the individual specific heterogeneity. The challenges these models pose have been well documented in the literature. Choosing between random and fixed effects, how to deal with initial conditions and lagged dependent variables,

as well as the incidental parameters problem and the calculation of marginal effects, have all been extensively studied by a number of authors. A detailed summary of developments can be found in Arellano and Honoré (2001) and more recently in Arellano and Bonhomme (2011).

This paper studies semiparametric identification in linear index discrete response panel data models with fixed effects in two different settings. The first setting corresponds to the binary response dynamic panel data model, where the individuals' choice set consists of a binary outcome, for example the choice of whether or not to purchase a specific product in a given period. The main focus is on eliminating the unrestricted fixed effect, which is allowed to be correlated with the explanatory variables and the time-varying unobservables, without imposing distributional assumptions on the time-varying unobservables. Identification is achieved by finding features of the distribution that do not depend on the fixed effect. Rosen and Weidner (2013,WP), thereafter RW2013, took such an approach for deriving bounds in the static binary outcome setting. In contrast, this paper analyses dynamic binary response models, where last period's choice directly enters the current period's decision rule. Conditioning on the initial condition, identification relies on individuals who switch options from one period to the next. It is shown that the joint probability of the choices these individuals make in two consecutive periods is bounded by features of the distribution invariant to the fixed effect. Under an exogeneity condition for the time-varying unobservables, this allows for the derivation bounds for the coefficients of the contemporaneous explanatory variables and the coefficient of the lagged dependent variable.

Point-identification of the regression parameters in binary response panel data models relies on strong and restrictive assumptions which might be untestable and difficult to satisfy in many applications. Several papers including Chamberlain (1984, 2010), Honoré (2002) and Honoré and Kyriazidou (2000), have shown that in linear index panel data models with binary outcomes, parametric point-identification of the regression parameters when regressors have bounded support can only be achieved under the assumption of independently and identically logistically distributed time-varying unobservables. Manski (1987), using a conditional version of the Maximum Score Estimator, shows that in the static panel data binary response model inference is possible under a time-stationarity condition, when the strictly exogenous

explanatory variables vary enough over time with at least one component having unbounded support. Honoré and Lewbel (2002) show point-identification in binary panel data models with predetermined regressors, if there exists a special regressor that is independent of the fixed effect, conditional on the rest of the regressors and the instruments .

In an attempt to generalize these findings, several papers study semiparametric identification in a general class of panel data models that results in partial identification of parameters and quantities of interest. For example, Chernozhukov, Hahn, and Newey (2005) focus on nonparametric bound analysis in multinomial panel data models with correlated random effects, while Chernozhukov, Fernández-Val, Hahn, and Newey (2013) provide sharp identification bounds for the average and quantile treatment effects in fully parametric and semiparametric nonseparable panel data models. Honoré and Tamer (2006) study bounds on parameters in dynamic discrete choice models, mainly focusing on the initial condition problem. In linear panel data settings Rosen (2012) studies the identifying power of conditional quantile restrictions in short panels with fixed effects.

Panel data binary response models are suitable for explaining individual choices when the choice set includes two alternatives. However, when the choice set includes more than two alternatives the binary choice model may fail to take into account all the information. The second setting examined in this paper is the static ordered response setting, where the choice set consists of more than two ordered alternatives. The shape restrictions imposed by the ordered response model allows for the characterization of the identified set, without imposing distributional assumptions on the unobserved time-varying components or the fixed effect. The bounds are achieved by relying on observable implications in which the fixed effect does not appear. In contrast to the binary case where information on the parameters of interest only comes through individuals who switch, in the ordered model it is shown that individuals who choose the "in-between" categories also provide a useful source of information. The information provided by the individuals who stay with the same option might be useful in comparing the behaviour of switchers to non-switchers. Furthermore, the greater number of choice-pairs that can be used in the ordered model in comparison to the binary model might help in achieving tighter bounds for the regression parameters.

Several papers examined identification in multinomial response panel data models where the choice set includes a variety of unordered alternatives. In a recent working paper, Pakes and Porter (2014) provide set identification results in multinomial models with additively separable fixed effects, where the key assumption is a group homogeneity condition on the disturbances conditional on the contemporaneous explanatory variables and the fixed effects. Shi, Shum, and Song (2017) develop a semiparametric identification and estimation approach to panel data multinomial choice models based on cyclic monotonicity, which point-identifies the model parameters. Although, these papers provide clear identification results, they usually require the comparison of each option against every other alternative, which might be intractable and computationally heavy in practice. This paper departs from these models and imposes some additional shape restrictions on the functional form, thus reducing the number of between alternatives comparisons needed to determine the optimal choice.

Identification in panel data ordered response models has not been extensively studied in the literature. Following the work by Honoré (1992) that shows how to consistently estimate the parameters in the truncated/censored panel data model, this paper focuses on the "in-between" case of ordered outcomes. Since every ordered response model can be expressed as a dichotomous/binary response model, parametric point-identification can be achieved under the assumptions of logistically distributed time-varying unobservables as in Chamberlain (1984, 2010). As discussed in Baetschmann, Staub, and Winkelmann (2015) the literature has estimated the fixed effects ordered logit model either with a single dichotomization with constant or individual-specific thresholds, or by combining all possible dichotomizations by various estimation methods. In a recent paper, Muris (2017) introduces a new estimator for the fixed effects ordered logit model which allows for estimation of the differences in the cut points, in addition to a more efficient estimation of the regression coefficients. He achieves this by using the fact that the ordered logit with J outcomes and T time periods has $(J-1)^T$ ways to be converted into a binary choice logit model. This paper departs from these approaches, by relaxing the logistic distribution assumption and using the complete structure of the ordered choice model, thus utilizing more information.

As already discussed, allowing for the fixed effect to be correlated with the explanatory

variables, creates an endogeneity problem. Chesher (2010) and Chesher and Smolinski (2012) derive sharp identification bounds in nonparametric cross-sectional ordered response models in the presence of endogenous variables, that can shrink at a relatively fast rate as the relevance of the instruments increases and as the number of ordered outcomes becomes larger. They show that when the ordered model is expressed as a binary response model, the identified sets are not sharp. This paper departs from the instrumental variables approach and tackles this problem by directly eliminating this unobserved heterogeneity.

The rest of the paper is structured as follows. Section 2 examines identification in the dynamic binary response models. Section 3 extends the static binary response panel data model to a static ordered response panel data model and examines identification under weak distributional conditions. Section 4 gives some numerical results for the models discussed. Section 5 concludes with some final remarks. All the proofs are provided in the Appendix.

2 The Dynamic Binary Response Panel Data Model

Binary response panel data models are widely used to model situations where individuals are observed over time making choices from a set that includes two alternatives, for example the choice of seeking employment or not or the choice of travelling by train or by car in a specific period. The leading example in the literature has been the static binary response model, where individuals' choices are correlated across different periods only through the presence of a time-varying unobserved heterogeneity, see for instance Chamberlain (1984, 2010) and Manski (1987). In the simplest form of this model, each individual in the population is observed for two time periods, $t = 1$ and $t = 2$, and in each time period he can choose one option from the set $\mathcal{Y}_t \equiv \{0, 1\}$. Therefore, each individual is characterized by a set of observables (Y, X) such that $Y = (Y_1, Y_2)$, $X = (X_1, X_2)$, and a set of unobservables (V, α), where $V = (V_1, V_2)$ and $\alpha \in \mathbb{R}$. The utility an individual, with covariates x_t and unobservables v_t, α, receives from choosing a specific outcome y in period t, y_t, is given by

$$U_t = X_t\beta + \alpha + V_t \tag{1}$$

where, X_t are observed individual characteristics, α is the unobserved to the econometrician time-invariant individual fixed effect and V_t is the unobserved to the econometrician time-varying component. Normalizing the utility of the outside option, $Y_t = 0$, in each period to be zero and under the assumption that in each period t individuals choose the outcome to maximize (1), the static panel data binary response model can be shown to be equivalent to,

$$y_t = 1(X_t\beta + \alpha + V_t > 0)$$

where

$$1(X_t\beta + \alpha + V_t > 0) = \begin{cases} 1 & if\ X_t\beta + \alpha + V_t > 0 \\ 0 & otherwise \end{cases} \tag{2}$$

RW2013 provide partial identification results in this kind of models and find features of the distribution that do not depend on α, by considering less restrictive conditions on the distribution of the time-invariant unobervables than the ones discussed in Section 1.

This section extends the linear index binary response static model to the linear index binary response dynamic model. This is of practical relevance because in panel data settings with repeated observations it is evident and natural to assume that individuals' past choices directly affect current and future decisions. For example, an individual's decision to seek employment in the current period is likely to be affected by his employment status last period. This allows for correlation in choices to come through two sources, the fixed effect and the lagged dependent variable. The dynamic binary response panel data model that includes the lagged dependent variable as an additional explanatory variable can be expressed as,

$$y_t = 1(X_t\beta + 1(Y_{t-1} = 1)\gamma + \alpha + V_t > 0) \tag{3}$$

where

$$1(X_t\beta + 1(Y_{t-1} = 1)\gamma + \alpha + V_t > 0) = \begin{cases} 1 & if\ X_t\beta + 1(Y_{t-1} = 1)\gamma + \alpha + V_t > 0 \\ 0 & otherwise \end{cases}$$

and

$$1(Y_{t-1} = 1) = \begin{cases} 1 & if\ Y_{t-1} = 1 \\ 0 & otherwise \end{cases}$$

In this model each individual is observed for three periods, $t = \{0, 1, 2\}$, and is characterized by a set of observables $Y = (Y_0, Y_1, Y_2)$, $X = (X_1, X_2)$, and a set of unobservables (V, α), where $V = (V_1, V_2)$ and $\alpha \in \mathbb{R}$. The setting is equivalent to the standard specification of dynamic binary panel data models, where the lagged dependent variable enters as an additional regressor in the form of Y_{t-1} and the parameter γ measures the "impact" of choosing option $Y = 1$ in period $t-1$. Last period's choice directly affects the decision so the choice in period $t-1$ needs to be taken into account. This creates an initial condition problem in modeling the choice in period $t = 1$, since the choice in period $t = 1$ depends on the choice in period $t = 0$. To deal with this issue, similar to Wooldridge (2005), it is assumed that the outcome in period $t = 0$, $Y_0 = y_0$, is observed, however no assumptions about its generation or its relation with the fixed effect are imposed, such that the set of conditioning covariates consists of $(x, y_0) \in \mathcal{X} \times \mathcal{Y}_0$. Section 2.1 formalizes the assumptions.

2.1 Model Assumptions

Assumption 1. *The observed data comprise a random sample of N individuals from the population. For each individual (Y, Y_0, X, V, α) are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ contains the Borel Sets. The support of (Y_0, X, V, α) is $(\mathcal{Y}_0 \times \mathcal{X} \times \mathcal{V} \times \mathcal{A})$ where $\mathcal{V} \subseteq \mathbb{R}^2$ and $\mathcal{A} \subseteq \mathbb{R}$.*

Assumption 2. *For each value of $x \in \mathcal{X}$ and $y_0 \in \mathcal{Y}_0$ there is a proper conditional distribution of (Y_1, Y_2) given $X = x$ and $Y_0 = y_0$,*

$$P^0(y_1, y_2|x, y_0) \equiv \mathbb{P}(Y_1 = y_1 \wedge Y_2 = y_2|X = x, Y_0 = y_0)$$

and the conditional probability of each pair (y_1, y_2), is point-identified over the support of (Y_1, Y_2) for almost every $x \in \mathcal{X}$ and $y_0 \in \mathcal{Y}_0$.

Assumption 3. *The conditional distribution of V given $(X = x, Y_0 = y_0)$, $F_{V|X,Y_0}$ is absolutely continuous with respect to the Lebesgue measure with an everywhere positive density and the marginal distribution of $\Delta V|X = x, Y_0 = y_0$ is given by $F_{\Delta V|X,Y_0}$, where $\Delta V = V_2 - V_1$.*

Assumption 4. *The conditional distribution of α given $X = x, Y_0 = y_0$ is absolutely continuous with respect to the Lebesgue measure with an everywhere positive density on $\mathbb{R}$ and marginal distribution $F_{\alpha|X,Y_0}$.*

Assumption 5. *X and V are stochastically independent conditional on Y_0, i.e. $V \perp X|Y_0$. α is allowed to be correlated with both V and X in an arbitrary way. The joint distribution of (V, α) conditional on $(X = x, Y_0 = y_0)$ is given by $F_{(V,\alpha)|X,Y_0}$.*

Assumption 1 defines the underlying probability space and notation for the support of the random variables (Y, Y_0, X, V, α). Assumption 2 stipulates that the conditional distribution of (Y_1, Y_2) given covariates x and the initial condition y_0 is point-identified for almost every $x \in \mathcal{X}$ and $y_0 \in \mathcal{Y}_0$, as would be the case for example under random sampling. Assumption 3 and Assumption 4 require the time-varying unobservable V and the unobserved heterogeneity α to be absolutely continuously distributed conditional on X and Y_0 with full support in the Euclidean space. Assumption 5 imposes only conditional independence between X and V conditional on Y_0, which is less restrictive than the assumptions imposed in the literature, such as $V \perp (X, Y_0)$ or specifying the conditional distribution $V|X, Y_0$. This assumption allows for example for correlation between V_0 and (V_1, V_2).

Assumption 4 imposes no restrictions on the time-invariant unobservable and allows correlation with the explanatory variables. Therefore, the presence of α creates an endogeneity problem, that needs to be addressed for identification and consistent estimation of the parameters of interest. In linear panel data models with continuous outcomes, differencing out the fixed effect is sufficient to guarantee point-identification of the regression parameters. This paper examines identification in linear index discrete panel data models and mimics the approach used in linear panel data models with continuous outcomes to solve the problem of the fixed effect.

Assumptions 3-5 provide restrictions on the conditional distributions of the unobservables given the observed covariates. These conditional distributions are elements of the generic collection of conditional distributions defined as,

- $\{F_{V|X,Y_0}(\cdot|x,y_0) : x \in \mathcal{X}, y_0 \in \mathcal{Y}_0\}$ is an element of $\mathcal{F}_{V|X,Y_0}$
- $\{F_{\Delta V|X,Y_0}(\cdot|x,y_0) : x \in \mathcal{X}, y_0 \in \mathcal{Y}_0\}$ is an element of $\mathcal{F}_{\Delta V|X,Y_0}$
- $\{F_{\alpha|X,Y_0}(\cdot|x,y_0) : x \in \mathcal{X}, y_0 \in \mathcal{Y}_0\}$ is an element of $\mathcal{F}_{\alpha|X,Y_0}$
- $\{F_{(V,\alpha)|X,Y_0}(\cdot|x,y_0) : x \in \mathcal{X}, y_0 \in \mathcal{Y}_0\}$ is an element of $\mathcal{F}_{(V,\alpha)|X,Y_0}$

Finally, the set of admissible structures $S \equiv (\beta, \gamma, F_{(V,\alpha)|X,Y_0})$ is defined in Assumption 6.

Assumption 6. *$S \equiv (\beta, \gamma, F_{(V,\alpha)|X,Y_0}) \in \mathcal{S}$ is a specified collection of parameters β and γ, and joint distributions of the time-varying unobservables and the unobserved heterogeneity, $F_{(V,\alpha)|X,Y_0}$. Such a S is called a structure.*

Under Assumptions 1-6 the identified set of admissible structures, denoted by S^0, is characterized. Define by $u(\tilde{y}_1, \tilde{y}_2, x, y_0, \alpha, v_1, v_2; \beta, \gamma)$, the utility an individual with covariates x, y_0 and unobservables α, v_1, v_2 receives from choosing any pair $(\tilde{y}_1, \tilde{y}_2) \in (\mathcal{Y}_1, \mathcal{Y}_2)$ and by $\mathcal{R}_{(y_1,y_2)}(x, y_0; \beta, \gamma)$ the region of unobservables (V, α) associated with (y_1, y_2) maximizing $u(\tilde{y}_1, \tilde{y}_2, x, y_0, \alpha, v_1, v_2; \beta, \gamma)$, such that

$$\mathcal{R}_{(y_1,y_2)}(x, y_0; \beta, \gamma) \equiv \left\{ (V, \alpha) \in (\mathcal{V}, \mathcal{A}) : (y_1, y_2) = arg \max_{(\tilde{y}_1,\tilde{y}_2) \in (\mathcal{Y}_1 \times \mathcal{Y}_2)} u(\tilde{y}_1, \tilde{y}_2, x, y_0, \alpha, v_1, v_2; \beta, \gamma) \right\}$$

Then S^0 is characterized by,

$$S^0 = \left\{ \begin{array}{c} (\beta, \gamma, F_{(V,\alpha)|X,Y_0}) \in \mathcal{S} : \forall (y_1, y_2) \in (\mathcal{Y}_1 \times \mathcal{Y}_2), \\ F_{(V,\alpha)|X,Y_0}\left(\mathcal{R}_{(y_1,y_2)}(x, y_0; \beta, \gamma)\right) = P^0(y_1, y_2|x, y_0) \\ a.e.\ x \in \mathcal{X} \text{ and } y_0 \in \mathcal{Y}_0 \end{array} \right\} \tag{4}$$

and the identified set for the model parameters (β, γ) is then characterized by,

$$\Theta^0 = \left\{ \begin{array}{c} (\beta, \gamma) \in \Theta : \forall (y_1, y_2) \in (\mathcal{Y}_1 \times \mathcal{Y}_2), \exists F_{(V,\alpha)|X,Y_0} \in \mathcal{F}_{(V,\alpha)|X,Y_0} \\ F_{(V,\alpha)|X,Y_0}\left(\mathcal{R}_{(y_1,y_2)}(x, y_0; \beta, \gamma)\right) = P^0(y_1, y_2|x, y_0) \\ a.e.\ x \in \mathcal{X} \text{ and } y_0 \in \mathcal{Y}_0 \end{array} \right\} \tag{5}$$

Point-identification of the regression coefficients in the dynamic binary response model under the logistic distribution assumption comes by observing individuals for (at least) four

time periods who change behaviour from period $t = 2$ and $t = 3$, as shown in Honoré and Kyriazidou (2000). This behaviour gives rise to features of the distribution that do not depend on the unobserved heterogeneity. In the semiparametric approach of this paper, that relaxes the distributional assumption on the time-varying unobservables, finding features of the distribution that do not depend on the unobserved heterogeneity leads to partial identification of the regression parameters.

2.2 Identified Set: Binary Response Dynamic Panel Data Model

The model in (3) is equivalent to standard linear index dynamic binary response panel data models, like the one discussed in Honoré and Kyriazidou (2000), where the lagged binary dependent variable enters as an additional regressor. Identification of the model parameters (β, γ) comes through features of the distribution that are invariant to changes in α, by considering the joint probability of the choices individuals make in periods $t = 1$ and $t = 2$, conditional on the choice in period $t = 0$. The regions $\mathcal{R}^{DB}_{(y_1,y_2)}(x, y_0; \beta, \gamma)$ that partition the $supp(V, \alpha)$ such that for all $(V, \alpha) \in (\mathcal{V}, \mathcal{A})$, $(Y_1, Y_2) = (y_1, y_2)$ when $X = x$ and $Y_0 = y_0$ are:

$$
\begin{aligned}
\mathcal{R}^{DB}_{(0,0)}(x,0;\beta,\gamma) &= \{(V,\alpha) \in (\mathcal{V},\mathcal{A}) : x_1\beta + \alpha + V_1 \leq 0 \ \& \ x_2\beta + \alpha + V_2 \leq 0\} \\
\mathcal{R}^{DB}_{(0,1)}(x,0;\beta,\gamma) &= \{(V,\alpha) \in (\mathcal{V},\mathcal{A}) : x_2\beta + \alpha + V_2 > 0 \geq x_1\beta + \alpha + V_1\} \\
\mathcal{R}^{DB}_{(1,0)}(x,0;\beta,\gamma) &= \{(V,\alpha) \in (\mathcal{V},\mathcal{A}) : x_2\beta + \gamma + \alpha + V_2 \leq 0 < x_1\beta + \alpha + V_1\} \\
\mathcal{R}^{DB}_{(1,1)}(x,0;\beta,\gamma) &= \{(V,\alpha) \in (\mathcal{V},\mathcal{A}) : x_1\beta + \alpha + V_1 > 0 \ \& \ x_2\beta + \gamma + \alpha + V_2 > 0\} \\
\mathcal{R}^{DB}_{(0,0)}(x,1;\beta,\gamma) &= \{(V,\alpha) \in (\mathcal{V},\mathcal{A}) : x_1\beta + \gamma + \alpha + V_1 \leq 0 \ \& \ x_2\beta + \alpha + V_2 \leq 0\} \\
\mathcal{R}^{DB}_{(0,1)}(x,1;\beta,\gamma) &= \{(V,\alpha) \in (\mathcal{V},\mathcal{A}) : x_2\beta + \alpha + V_2 > 0 \geq x_1\beta + \gamma + \alpha + V_1\} \\
\mathcal{R}^{DB}_{(1,0)}(x,1;\beta,\gamma) &= \{(V,\alpha) \in (\mathcal{V},\mathcal{A}) : x_2\beta + \gamma + \alpha + V_2 \leq 0 < x_1\beta + \gamma + \alpha + V_1\} \\
\mathcal{R}^{DB}_{(1,1)}(x,1;\beta,\gamma) &= \{(V,\alpha) \in (\mathcal{V},\mathcal{A}) : x_1\beta + \gamma + \alpha + V_1 > 0 \ \& \ x_2\beta + \gamma + \alpha + V_2 > 0\} \qquad (6)
\end{aligned}
$$

The conditional joint probabilities of $F_{(V,\alpha)|X,Y_0}$ for any given pair $(y_1, y_2) \in (\mathcal{Y}_1 \times \mathcal{Y}_2)$ for almost every $x \in \mathcal{X}$ and $y_0 \in \mathcal{Y}_0$ are thus defined by,

$$
P(y_1, y_2|x, y_0) = F_{(V,\alpha)|X,Y_0}\left(\mathcal{R}^{DB}_{(y_1,y_2)}(x, y_0; \beta, \gamma)|X = x, Y_0 = y_0\right) \qquad (7)
$$

where $P(y_1, y_2|x, y_0) = P(Y_1 = y_1 \wedge Y_2 = y_2|X = x, Y_0 = y_0)$.

From the regions in (6) and Figure 1(a), it can be seen that the model in (3) is complete in the sense that conditioning on any value of the explanatory variables and the initial condition, there is a unique solution to the individual's problem with probability one and identification sets of the form of (4) and (5) can be derived. Since the outcome in period $t = 0$ appears in the generation of the outcome in period $t = 1$, the outcomes in all three periods $t = \{0, 1, 2\}$ are used. Assumption 4 imposes no restrictions on the fixed effect and identification of the parameters β and γ through the elimination of α comes only by observing individuals who switch in periods $t = 1$ and $t = 2$ for each of the values of y_0. This implies that consideration of observations of the following types are required,

$$\begin{aligned} A &= \{Y_0 = 0 \wedge Y_1 = 0 \wedge Y_2 = 1\} \\ B &= \{Y_0 = 0 \wedge Y_1 = 1 \wedge Y_2 = 0\} \\ C &= \{Y_0 = 1 \wedge Y_1 = 0 \wedge Y_2 = 1\} \\ D &= \{Y_0 = 1 \wedge Y_1 = 1 \wedge Y_2 = 0\} \end{aligned} \tag{8}$$

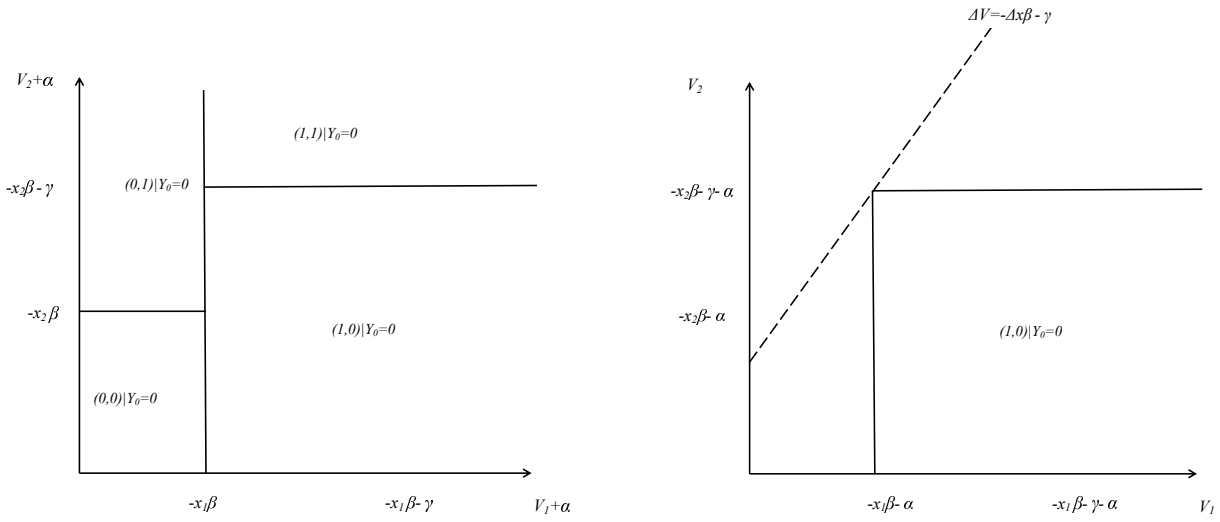
Theorem 1. *Let $S^{DB} = (\beta^{DB}, \gamma^{DB}, F^{DB}_{(V,\alpha)|X,Y_0})$ be a structure admitted by model (3) that is observationally equivalent to S^0, then under Assumptions 1-6, $(\beta^{DB}, \gamma^{DB})$ satisfy the following inequalities*

$$\begin{aligned} 1 - P(0,1|x,0) &\geq P_{\Delta V|Y_0}[\Delta V < -\Delta x\beta^{DB}|Y_0 = 0] \\ P(1,0|x,0) &\leq P_{\Delta V|Y_0}[\Delta V < -\Delta x\beta^{DB} - \gamma^{DB}|Y_0 = 0] \\ 1 - P(0,1|x,1) &\geq P_{\Delta V|Y_0}[\Delta V < -\Delta x\beta^{DB} + \gamma^{DB}|Y_0 = 1] \\ P(1,0|x,1) &\leq P_{\Delta V|Y_0}[\Delta V < -\Delta x\beta^{DB}|Y_0 = 1] \end{aligned}$$

where $P(y_1, y_2|x, y_0) = P(Y_1 = y_1 \wedge Y_2 = y_2|X = x, Y_0 = y_0)$, $\Delta X = X_2 - X_1$ with $(X_1, X_2) \in \mathcal{X} \times \mathcal{X}$ and $\Delta V = V_2 - V_1$ with $(V_1, V_2) \in \mathcal{V} \times \mathcal{V}$.

Proof. The proof is provided in Appendix A.1. □

The above relations provide restrictions on the distribution of $\Delta V|X, Y_0$ for any realization of $x \in \mathcal{X}$ and $y_0 \in \mathcal{Y}_0$ that do not depend on the fixed effect α. The events $\{Y_1 = 0 \wedge Y_2 = 0\}$ and $\{Y_1 = 1 \wedge Y_2 = 1\}$ provide no restrictions on ΔV and can not be used to eliminate the fixed effect α. Similarly to the binary logit fixed effects model and as discussed in Honoré (2002), the individuals who do not switch can not be used to identify the regression parameters, since for any value of β the choices these individuals make can be rationalized by extremely large or extremely small values of the fixed effect. In other words, these events provide no restrictions on the values the fixed effect can take for a given value of the regression parameters. The distribution of $\Delta V|X, Y_0 \sim F_{\Delta V|X,Y_0}$ is equivalent to $\Delta V|Y_0 \sim F_{\Delta V|Y_0}$ by Assumption 5. Notice that in order for the bounds in Theorem 1 to be informative, there should exists $x \in \mathcal{X}$ such that $x_1 \neq x_2$ with positive probability.



(a) Regions defined in equations (6)

(b) Upper Bound for $(1,0)|Y_0 = 0$

Figure 1: Regions of unobservables for each (Y_1, Y_2) choice when $\gamma < 0$ and $Y_0 = 0$

Figures 1(a) and 1(b) plot the regions of unobservables conditional on $Y_0 = 0$ and $\gamma < 0$ given in equations (6) and provide an outline of the main idea. Using the fact that the model in (3) is complete, it can be shown that the probability of any switching event is bounded by the probability of an event that is independent of the fixed effect. Figure 1(b) illustrates this result for the event $(Y_1, Y_2) = (1, 0)$ conditional on $Y_0 = 0$ and $\gamma < 0$. Changing α

moves the region of $(1,0)|Y_0 = 0$ up and down the line $\Delta V = -\Delta X\beta - \gamma$. Therefore, it is clear that $(V^*, \alpha^*) \in R^{DB}_{(1,0)}(x, 0; \beta, \gamma)$ implies $(V^*, \alpha^*) \in \{(V, \alpha) : \Delta V < -\Delta x\beta - \gamma\}$ and $P(1, 0|x, 0) \leq P_{\Delta V|Y_0}[\Delta V < -\Delta x\beta - \gamma|Y_0 = 0]$.

Theorem 2. *Following Theorem 1, under Assumptions 1-6 the bounds for β, γ are given by the set:*

$$\Theta^{DB} = \left\{ \begin{array}{c} (\beta, \gamma) \in \Theta : \ \forall \omega \in \mathbb{R}, \\ \sup_{x:-\Delta x\beta-\gamma\leq\omega} P(1,0|x,0) \leq \inf_{x:-\Delta x\beta\geq\omega} 1 - P(0,1|x,0) \\ and \\ \sup_{x:-\Delta x\beta\leq\omega} P(1,0|x,1) \leq \inf_{x:-\Delta x\beta+\gamma\geq\omega} 1 - P(0,1|x,1) \\ a.e.\ x \in \mathcal{X} \end{array} \right\}$$

Proof. The proof is provided in Appendix A.2. □

In dynamic models this period's choice depends on last periods's choice. This implies that the choice in period $t = 1$ depends on the choice in period $t = 0$, which is the first period observed in the sample. Unless this period coincides with the first period of the process, it will depend on previous (not observed) periods, the exogenous variables in period $t = 0$ and the joint distribution of the outcome in the first period and the unobserved heterogeneity. This joint distribution is (in general) different from the joint distribution of future outcomes and the unobserved heterogeneity. Therefore, since $V \not\perp \alpha$, Assumption 5 of $V \perp X|Y_0$ does not imply (in general) $V \perp (X, Y_0)$. Notice that unlike the Honoré and Kyriazidou (2000) result, where point-identification in the dynamic binary panel data model is achieved iff the errors are logistically distributed and 4 time periods are observed, the identified set in Theorem 2 only requires 3 time periods.

Finally, by replacing Assumption 5 of $X \perp V|Y_0$, with the unconditional independence assumption, $X \perp V$, the identified set for β and γ can be expressed in terms of the unconditional probabilities, given in Theorem 3.

Theorem 3. *Let Assumptions 1-4, $X \perp V$ and 6 hold. Then the (unconditional) identified set for (β, γ) is given by,*

$$\Theta_U^{DB} = \left\{ \begin{array}{c} (\beta,\gamma) \in \Theta : \ \forall \omega \in \mathbb{R}, \\ \sup_{x \in \mathcal{X}} \{\underline{G}(\omega|x,0)P_0(x) + \underline{G}(\omega|x,1)P_1(x)\} \leq \inf_{x \in \mathcal{X}} \{\overline{G}(\omega|x,0)P_0(x) + \overline{G}(\omega|x,1)P_1(x)\} \end{array} \right\}$$

where

$$\begin{aligned} \underline{G}(\omega|x,0) &= P[(Y_1,Y_2) = (1,0) \wedge -\Delta X\beta - \gamma \leq \omega | X = x, Y_0 = 0] \\ \underline{G}(\omega|x,1) &= P[(Y_1,Y_2) = (1,0) \wedge -\Delta X\beta \leq \omega | X = x, Y_0 = 1] \\ \overline{G}(\omega|x,0) &= 1 - P[(Y_1,Y_2) = (0,1) \wedge -\Delta X\beta \geq \omega | X = x, Y_0 = 0] \\ \overline{G}(\omega|x,1) &= 1 - P[(Y_1,Y_2) = (0,1) \wedge -\Delta X\beta + \gamma \geq \omega | X = x, Y_0 = 1] \end{aligned}$$

and

$$\begin{aligned} P(Y_0 = 0|X = x) &= P_0(x) \\ P(Y_0 = 1|X = x) &= P_1(x) \end{aligned}$$

Proof. The proof is provided in Appendix A.3. □

3 The Static Ordered Response Panel Data Model

Section 2 studies identification in binary response panel data models. As discussed in Section 1 several papers, such as Chintagunta, Kyriazidou, and Perktold (2001) and Pakes and Porter (2014), have extended the binary response panel data model to multinomial response models, where individuals choose from a set of unordered alternatives. This paper extends the binary response panel data model to one where the choice set consists of alternatives that can be ordered, such as the choice between unemployment, part-time employment or full-time employment and the choice of flying first, business or economy class. This approach might be beneficial in reducing the dimension of search for the identified set, since the shape restriction imposed by the ordering specification, reduces the between alternatives comparisons needed to determine the option chosen by the individual.

This section extends the model in (2) to a model of three ordered outcomes, where in every period $t = 1$ and $t = 2$ each individual chooses one option from the set $\mathcal{Y}_t = \{0, 1, 2\}$. Such a model could be used, for example, in describing consumers' choices when faced with vertically differentiated alternatives such that if all the options were offered at the same price everyone would choose option $Y = 2$, i.e. $(Y_t = 1) \preceq (Y_t = 2)$, where $A \preceq B$ denotes that B is weakly preferred to A. $Y_t = 0$ denotes the outside option and corresponds to not choosing any of the available alternatives. The panel data ordered response model for each individual can be expressed as,

$$Y_t = \begin{cases} 0 \; if \;\; X_t\beta + \alpha + V_t < c_{1t} \\ 1 \; if \; c_{1t} < X_t\beta + \alpha + V_t \leq c_{2t} \\ 2 \; if \; c_{2t} < X_t\beta + \alpha + V_t \end{cases} \tag{9}$$

where X_t are observed individual characteristics, α is the unobserved to the econometrician time-invariant individual heterogeneity, V_t is the time-varying component unobserved to the econometrician and $c = \{c_{11}, c_{12}, c_{21}, c_{22}\} \in \mathcal{C}$ are the threshold parameters in the ordered model, such that $C \subseteq \mathbb{R}^4$ and $c_{2t} > c_{1t}$, $\forall t \in \{1, 2\}$. For the rest of the section c is assumed to be observed to reduce the dimension of the identified set[1].

As already discussed in Section 1 the ordered response panel data model has not been extensively studied in the literature and the work has mainly focused in redefining the ordered response model as a set of binary response models and imposing logistically distributed unobservables. This paper departs from this approach and directly uses the ordered structure of the model to characterize the identified set, without imposing distributional assumption on the unobserved time-varying components or the fixed effect. Such an approach utilizes more information than the binary response representation and provides informative identification bounds for the regression parameters. Section 3.1 formalizes the assumptions imposed on model (9).

[1] Allowing c to be part of the identified set does not have an effect on the identification strategy.

3.1 Model Assumptions

Assumption 7. *The observed data comprise a random sample of N individuals from the population. For each individual (Y, X, V, α) are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ contains the Borel Sets. The support of (X, V, α) is $(\mathcal{X} \times \mathcal{V} \times \mathcal{A})$ where $\mathcal{V} \subseteq \mathbb{R}^2$ and $\mathcal{A} \subseteq \mathbb{R}$.*

Assumption 8. *For each value of $x \in \mathcal{X}$ there is a proper conditional distribution of (Y_1, Y_2) given $X = x$,*

$$P^0(y_1, y_2|x, c) \equiv \mathbb{P}(Y_1 = y_1 \wedge Y_2 = y_2|X = x, c)$$

and the conditional probability of each pair (y_1, y_2), is point-identified over the support of (Y_1, Y_2) for almost every $x \in \mathcal{X}$, for any fixed c.

Assumption 9. *The conditional distribution of V given $(X = x)$, $F_{V|X}$ is absolutely continuous with respect to the Lebesgue measure with an everywhere positive density and the marginal distribution of $\Delta V|X = x$ is given by $F_{\Delta V|X}$, where $\Delta V = V_2 - V_1$.*

Assumption 10. *The conditional distribution of α given $(X = x)$ is absolutely continuous with respect to the Lebesgue measure with an everywhere positive density on $\mathbb{R}$ and marginal distribution $F_{\alpha|X}$.*

Assumption 11. *X and V are stochastically independent. α is allowed to be correlated with both V and X in an arbitrary way. The joint distribution of (V, α) conditional on $(X = x)$ is given by $F_{(V,\alpha)|X}$.*

Assumption 7 defines the underlying probability space and notation for the support of the random variables (Y, X, V, α). Assumption 8 stipulates that the conditional distribution of (Y_1, Y_2) given covariates x is point-identified for almost every $x \in \mathcal{X}$, as would be the case for example under random sampling. Assumption 9 and Assumption 10 require the time-varying unobservable V and the unobserved heterogeneity α to be absolutely continuously distributed conditional on X with full support in the Euclidean space. Assumption 11 imposes independence of X and V, but allows α and V to be arbitrary correlated with a joint distribution $F_{(V,\alpha)|X}$.

Assumptions 9-11 provide restrictions on the conditional distributions of the unobservables given the observed covariates. These conditional distributions are elements of the generic collection of conditional distributions defined as,

- $\{F_{V|X}(\cdot|x) : x \in \mathcal{X}\}$ is an element of $\mathcal{F}_{V|X}$
- $\{F_{\Delta V|X}(\cdot|x) : x \in \mathcal{X}\}$ is an element of $\mathcal{F}_{\Delta V|X}$
- $\{F_{\alpha|X}(\cdot|x) : x \in \mathcal{X}\}$ is an element of $\mathcal{F}_{\alpha|X}$
- $\{F_{(V,\alpha)|X}(\cdot|x) : x \in \mathcal{X}\}$ is an element of $\mathcal{F}_{(V,\alpha)|X}$

Finally the set of structures $S \equiv (\beta, F_{(V,\alpha)|X})$ admitted by model (9) is defined through Assumption 12.

Assumption 12. *$S \equiv (\beta, F_{(V,\alpha)|X}) \in \mathcal{S}$ is a specified collection of parameters β and joint distributions of the time-varying unobservable and the unobserved heterogeneity, $F_{(V,\alpha)|X}$. Such a S is called a structure.*

Under Assumptions 7-12 the identified set of admissible structures, denoted by S^0, is characterized. Define by $u(\tilde{y}_1, \tilde{y}_2, x, c, \alpha, v_1, v_2; \beta)$, the utility an individual receives from choosing any pair $(\tilde{y}_1, \tilde{y}_2) \in (\mathcal{Y}_1, \mathcal{Y}_2)$ given the observables x, fixed values of c and the unobservables α, v_1, v_2 and by $\mathcal{R}_{(y_1,y_2)}(x, c; \beta)$ the region of unobservables (V, α) associated with (y_1, y_2) maximizing $u(\tilde{y}_1, \tilde{y}_2, x, c, \alpha, v_1, v_2; \beta)$, such that

$$\mathcal{R}_{(y_1,y_2)}(x, c; \beta) \equiv \{(V, \alpha) \in (\mathcal{V}, \mathcal{A}) : (y_1, y_2) = arg \max_{(\tilde{y}_1,\tilde{y}_2)\in(\mathcal{Y}_1\times\mathcal{Y}_2)} u(\tilde{y}_1, \tilde{y}_2, x, c, \alpha, v_1, v_2; \beta)\}.$$

Then S^0 is characterized by:

$$S^0 = \left\{ \begin{array}{c} (\beta, F_{(V,\alpha)|X}) \in S : \forall (y_1, y_2) \in (\mathcal{Y}_1 \times \mathcal{Y}_2) \\ F_{(V,\alpha)|X}(\mathcal{R}_{(y_1,y_2)}(x, c; \beta)) = P^0(y_1, y_2|x, c) \\ a.e.\ x \in \mathcal{X} \text{ and } c \in \mathcal{C} \end{array} \right\}$$

and the identified set for the model parameters β is given by:

$$\Theta^0 = \left\{ \begin{array}{c} \beta \in \Theta : \ \forall (y_1, y_2) \in (\mathcal{Y}_1, \mathcal{Y}_2), \exists F_{(V,\alpha)|X} \in \mathcal{F}_{(V,\alpha)|X} \\ F_{(V,\alpha)|X,c}(\mathcal{R}_{(y_1,y_2)}(x, c; \beta)) = P^0(y_1, y_2|x, c) \\ a.e.\ x \in \mathcal{X} \text{ and } c \in \mathcal{C} \end{array} \right\}$$

3.2 Identified Set: Ordered Response Static Panel Data Model

When no assumptions are imposed on the fixed effect, identification bounds for the regression parameters, β, are derived by finding feature of the distribution that do not depend on α. The regions $\mathcal{R}^{SO}_{(y_1,y_2)}(x,c;\beta)$ that partition the support of (V,α) such that $(Y_1,Y_2)=(y_1,y_2)$ when $X=x$ and for any fixed c are provided in Appendix A.4 and plotted in Figure 2. For any fixed vector c the model in (9) is complete in the sense that conditional on any value of the explanatory variables $x \in \mathcal{X}$, there is a unique solution to the individual decision problem with probability one. The conditional joint probabilities of $(V,\alpha)|X,c$, $F_{(V,\alpha)|X,c}$, for any given pair $(y_1,y_2) \in (\mathcal{Y}_1 \times \mathcal{Y}_2)$, conditional on $x \in \mathcal{X}$ and for any fixed $c \in \mathcal{C}$ are given by

$$P(y_1,y_2|x,c) = F_{(V,\alpha)|X}\left(\mathcal{R}^{SO}_{(y_1,y_2)}(x,c;\beta)\,\middle|X=x\right)$$

where $P(y_1,y_2|x,c) = P(Y_1=y_1 \wedge Y_2=y_2|X=x,c)$.

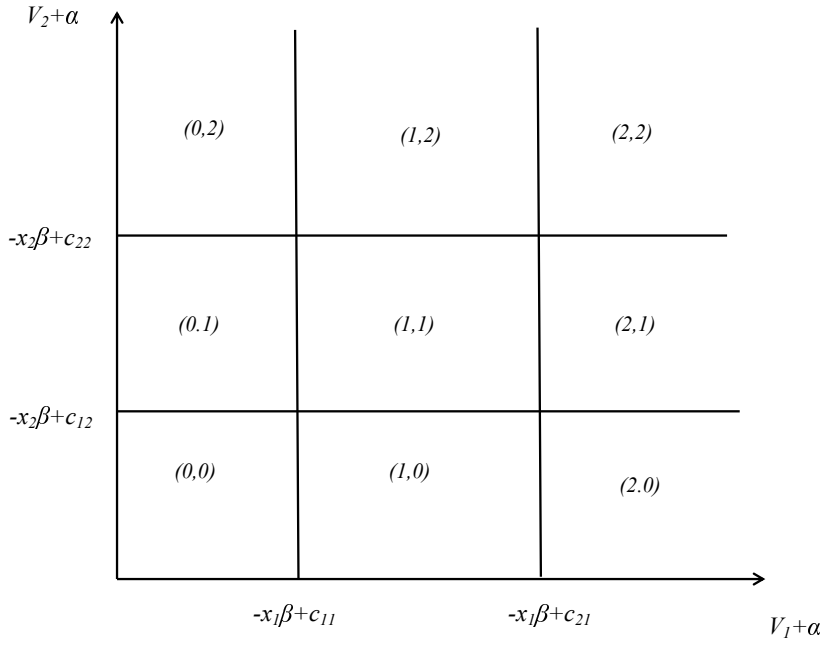


Figure 2: Regions of unobservables for each (Y_1,Y_2) choice for a fixed vector c

Similarly to the binary panel data model, individuals who switch from period $t=1$ to $t=2$, can be used for identification of the parameters β, without imposing any assumptions on the fixed effect. The transitions that are informative are therefore,

$$
\begin{aligned}
&\{Y_1 = 0 \ \wedge \ Y_2 = 1\} \\
&\{Y_1 = 1 \ \wedge \ Y_2 = 0\} \\
&\{Y_1 = 0 \ \wedge \ Y_2 = 2\} \\
&\{Y_1 = 2 \ \wedge \ Y_2 = 0\} \\
&\{Y_1 = 1 \ \wedge \ Y_2 = 2\} \\
&\{Y_1 = 2 \ \wedge \ Y_2 = 1\}
\end{aligned} \tag{10}
$$

In addition to the individuals who change from period $t = 1$ to $t = 2$ it can be shown that information that is independent of α is also provided by considering individuals who choose the same option, $Y = 1$, in periods $t = 1$ and $t = 2$ such that,

$$
\{Y_1 = 1 \wedge Y_2 = 1\} \tag{11}
$$

This might prove helpful when comparing the behaviour of switchers to non-switchers.

Theorem 4. *Let $S^{SO} = \left(\beta^{SO}, F^{SO}_{(V,\alpha)|X}\right)$ be a structure admitted by model (9). Under Assumptions 7-12, for any fixed parameter vector c, if S^{SO} is an observationally equivalent structure to S^0 then, for any $x \in \mathcal{X}$, β^{SO} satisfies,*

$$
\begin{aligned}
P(1,0|x,c) &\leq F_{\Delta V}[-\Delta x\beta^{SO} - c_{11} + c_{12}] \\
&\quad F_{\Delta V}[-\Delta x\beta^{SO} - c_{11} + c_{12}] \leq 1 - P(0,1|x,c) \\
P(2,0|x,c) &\leq F_{\Delta V}[-\Delta x\beta^{SO} + c_{12} - c_{21}] \\
&\quad F_{\Delta V}[-\Delta x\beta^{SO} - c_{11} + c_{22}] \leq 1 - P(0,2|x,c) \\
P(2,1|x,c) &\leq F_{\Delta V}[-\Delta x\beta^{SO} + c_{22} - c_{21}] \\
&\quad F_{\Delta V}[-\Delta x\beta^{SO} + c_{22} - c_{21}] \leq 1 - P(1,2|x,c) \\
P(1,1|x,c) &\leq P_{\Delta V}[-\Delta x\beta^{SO} + c_{22} - c_{11} > \Delta V > -\Delta x\beta^{SO} + c_{12} - c_{21}]
\end{aligned}
$$

where $\Delta X = X_2 - X_1$ and $\Delta V = V_2 - V_1$.

Proof. The proof is provided in Appendix A.5. □

Notice that in addition to the probabilities of the switching events, the conditional probability of the "in-between" event $(1,1)$ is also bounded by a conditional probability invariant to the fixed effect. The rationale behind this result is that, as discussed in Section 2, in the dynamic binary model the events $(Y_1, Y_2) = (0,0)$ *and* $(1,1)$ give no information on β since the behaviour for these cases can be matched by extremely small or extremely large values of α regardless of the value of β. This is also true for the ordered model for the events $(0,0)$ and $(2,2)$. However, as it can also be seen from Figure 2 in the ordered response model where $\mathcal{Y}_t = \{0,1,2\}$, considered in this section, the non-switchers who choose the "in-between" option $(Y_1, Y_2) = (1,1)$, provide restrictions on the possible values the fixed effect can take for each value of the parameter β, and hence can be used to identify the regression coefficients. Theorem 5 formalizes the result.

Theorem 5. *Let Assumptions 7-12 hold, for any fixed parameter vector c. Using the definitions in* (12)*, an outer region for β is given by the set:*

$$\Theta^{SO} = \left\{ \begin{array}{c} \beta \in \Theta : \forall \omega, \omega', \omega'' \in \mathbb{R}, \\ \max[s_{(1,0)}(\omega), s_{(2,0)}(\omega), s_{(2,1)}(\omega), s_{(1,1)}(\omega)] \leq \min[i_{(0,1)}(\omega), i_{(0,2)}(\omega), i_{(1,2)}(\omega), i_{(1,1)}(\omega)] \\ a.e.\ x \in \mathcal{X} \text{ and } c \in \mathcal{C} \end{array} \right\}$$

where,

$$\begin{aligned}
s_{(1,0)}(\omega) &= \sup_{x: -\Delta x\beta - c_{11} + c_{12} \leq \omega} P(1,0|x,c) \\
s_{(2,0)}(\omega) &= \sup_{x: -\Delta x\beta + c_{12} - c_{21} \leq \omega} P(2,0|x,c) \\
s_{(2,1)}(\omega) &= \sup_{x: -\Delta x\beta + c_{22} - c_{21} \leq \omega} P(2,1|x,c) \\
s_{(1,1)}(\omega) &= \sup_{x \in X^{**}} P(1,1|x,c) \\
i_{(0,1)}(\omega) &= \inf_{x: -\Delta x\beta - c_{11} + c_{12} \geq \omega} [1 - P(0,1|x,c)] \\
i_{(0,2)}(\omega) &= \inf_{x: -\Delta x\beta - c_{11} + c_{22} \geq \omega} [1 - P(0,2|x,c)] \\
i_{(1,2)}(\omega) &= \inf_{x: -\Delta x\beta + c_{22} - c_{21} \geq \omega} [1 - P(1,2|x,c)] \\
i_{(1,1)}(\omega) &= 1 - \inf_{x \in X*} P(1,1|x,c)
\end{aligned} \tag{12}$$

and for any fixed $\omega, \omega', \omega'' \in \mathbb{R}$

$$\begin{aligned} X^* &= \{x : \omega' \geq -\Delta x\beta + c_{22} - c_{11} \wedge -\Delta x\beta + c_{12} - c_{21} \geq \omega\} \\ X^{**} &= \{x : \omega \geq -\Delta x\beta + c_{22} - c_{11} \wedge -\Delta x\beta + c_{12} - c_{21} \geq \omega''\} \end{aligned} \tag{13}$$

Proof. The proof is provided in Appendix A.6. □

4 Numerical Examples

This section provides numerical illustrations of the identified sets derived in Sections 2 and 3 for the dynamic binary response model and the static ordered response model under different support conditions and Probability Generating Processes (PGP). For expositional purposes this section starts with some numerical illustrations of the identified sets for the binary response static panel data model derived in RW2013. All the models examined in this section have discrete support for the explanatory variables, X. Even though point-identification fails, the numerical examples illustrate that informative bounds can be achieved as the support of the discrete explanatory variables increases.

4.1 Static Binary Response Panel Data Model

4.1.1 Example 1

Consider the two time period static binary panel data model,

$$Y_t = 1(X_t\beta + \alpha + V_t > 0) \tag{14}$$

where $X_t = (X_{1t}, X_{2t})$, $\beta = (\beta_1, \beta_2)$, $\alpha|X \sim N(\overline{X}\delta, 1)$ with $\overline{X} = \frac{1}{2}(X_1 + X_2)$ and $\delta = (1, -1)$ and $V_t|X, \alpha \overset{iid}{\sim} f()$. The true parameter of $\beta_2 = 1$ after the normalization of $\beta_1 = 1$[2].

Tables 1 and 2 give the identified sets for β_2 under different specification for the distribution of the time-varying unobservables and as the support of the discrete explanatory

[2] This baseline PGP is similar to the one used in RW2013. The Normal distribution for the fixed effect was approximated on a grid with 100 evenly spaced support points on $[\overline{X}\delta - 4, \overline{X}\delta + 4]$, and the approximation error should be small.

variables (X_{1t}, X_{2t}) changes. The first Probit specification in Table 1 with $V_t|X, \alpha \overset{iid}{\sim} N(0,1)$ is the same as in RW2013. Table 1 also provides identified sets under the probit specification with $V_t|X, \alpha \overset{iid}{\sim} N\left(0, \frac{\pi^2}{3}\right)$ and the standard logit specification $V_t \perp (X, \alpha)$ with iid logistic distribution.

Table 1: Identified sets for β_2 under Probit and Logit specifications with symmetric support for (X_{1t}, X_{2t}) around zero

	Support of (X_{1t}, X_{2t})		
	$\{-1,0,1\}$	$\{-2,-1,0,1,2\}$	$\{-3,-2,-1,0,1,2,3\}$
$V \overset{iid}{\sim} N(0,1)$	$(0,\infty)$	$(0.5,2)$	$(0.667,1.5)$
$V \overset{iid}{\sim} N\left(0, \frac{\pi^2}{3}\right)$	$(-0.5,\infty)$	$(0.286,2.667)$	$(0.545,1.714)$
Logistic	$(-0.5,\infty)$	$(0.286,2.667)$	$(0.545,1.714)$

Table 2: Identified sets for β_2 under Probit and Logit specifications with asymmetric support for (X_{1t}, X_{2t}) around zero

	Support of (X_{1t}, X_{2t})		
	$\{-1,0,1,2,3\}$	$\{-2,-1,0,1,2,3,4\}$	$\{-4,-3,-2,-1,0,1,2\}$
$V \overset{iid}{\sim} N(0,1)$	$(0.5,6)$	$(0.667,2)$	$(0.667,2)$
$V \overset{iid}{\sim} N\left(0, \frac{\pi^2}{3}\right)$	$(0.286,\infty)$	$(0.545,2.5)$	$(0.545,2.5)$
Logistic	$(0.286,\infty)$	$(0.545,2.5)$	$(0.545,2.5)$

From Tables 1 and 2 two main conclusions can be drawn. The first one is that as the support of the explanatory variables increases the identified sets become narrower. This suggests that even thought the model only partially identifies the regression parameters, those sets shrink around the true value as the support of the explanatory variables increases.

Secondly, it is evident that the model with $V_t \overset{iid}{\sim} N\left(0, \frac{\pi^2}{3}\right)$ errors and the standard logit model with variance of the unobservables $var(V_t) = \frac{\pi^2}{3}$ give similar identified sets. This might

have implications on the identification and estimation of such models under mispecification[3].

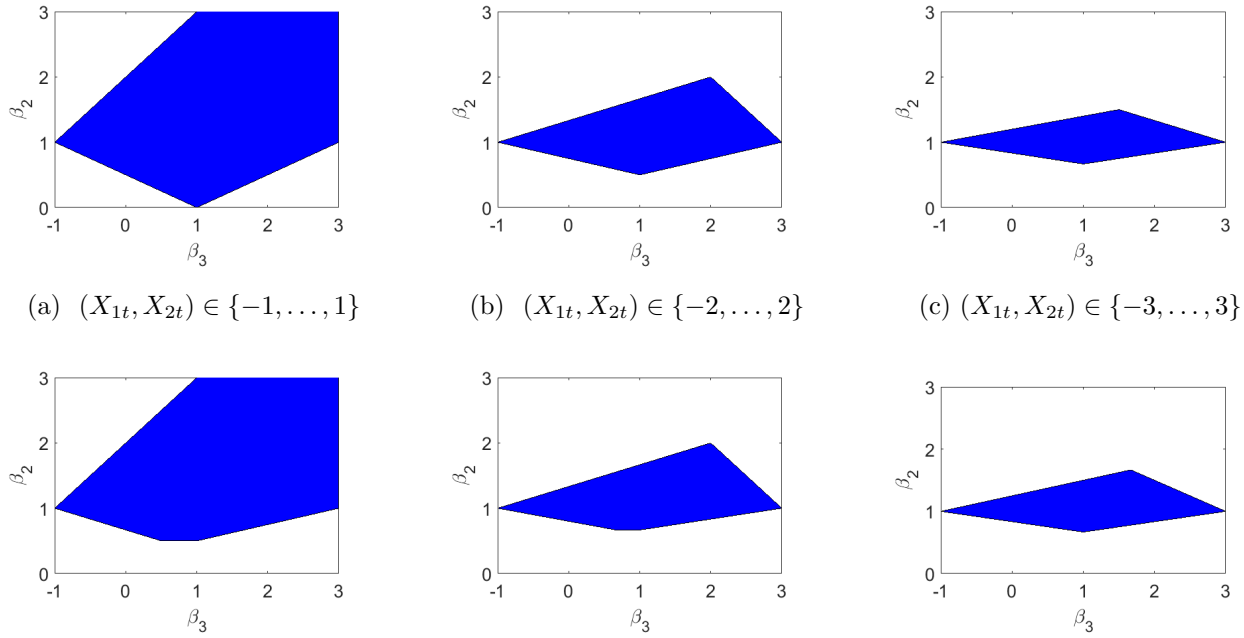
4.1.2 Example 2

Consider the two period static binary response panel data model as in (14) ,

$$Y_t = 1(X_t\beta + \alpha + V_t > 0)$$

where $X_t = (X_{1t}, X_{2t}, X_{3t})$ with $X_{3t} \in \{0, 1\}$, $\beta = (\beta_1, \beta_2, \beta_3)$, $\alpha|X \sim N(\overline{X}\delta, 1)$ with $\overline{X} = \frac{1}{2}(X_1 + X_2)$ and $\delta = (1, -1, 0)$ and $V_t|X, \alpha \overset{iid}{\sim} N(0, 1)$. The true parameter of $\beta_2 = 1$ and $\beta_3 = 1$ after the normalization of $\beta_1 = 1$.

Figure 3 provides identified set for (β_2, β_3) for different values of β_3 in the grid $[-1, 3]$. Similarly to Example 1, the sets shrink as the support of the discrete explanatory variables (X_{1t}, X_{2t}) changes, even if the support of X_3 remains fixed.



(a) $(X_{1t}, X_{2t}) \in \{-1, \ldots, 1\}$ (b) $(X_{1t}, X_{2t}) \in \{-2, \ldots, 2\}$ (c) $(X_{1t}, X_{2t}) \in \{-3, \ldots, 3\}$

(d) $(X_{1t}, X_{2t}) \in \{-1, \ldots, 3\}$ (e) $(X_{1t}, X_{2t}) \in \{-2, \ldots, 4\}$ (f) $(X_{1t}, X_{2t}) \in \{-4, \ldots, 2\}$

Figure 3: Joint identified sets for (β_2, β_3) when $X_{3t} \in \{0, 1\}$

[3]This might be related to the discussion in Ruud (1983).

4.2 Dynamic Binary Response Panel Data Model

4.2.1 Example 3

Consider the three period dynamic binary response panel data model as the one described in Section 2,

$$Y_t = 1(X_t\beta + 1(Y_{t-1} = 1)\gamma + \alpha + V_t > 0) \tag{15}$$

where $X_t = (X_{1t}, X_{2t})$, $\beta = (\beta_1, \beta_2)$, $\alpha|X \sim N(\overline{X}\delta, 1)$ with $\overline{X} = \frac{1}{2}(X_1 + X_2)$ and $\delta = (1, -1)$ and $V_t|X, Y_0, \alpha \overset{iid}{\sim} f()$. The true parameter of $\beta_2 = 1$ and $\gamma = 0.5$ after normalizing $\beta_1 = 1$.

Table 3 and Table 4 provide the identified sets for β_2 as described in Theorem 2, when $\gamma = 0.5$ is fixed, under different specification for the distribution of the time-varying unobservables and as the support of the discrete explanatory variables (X_{1t}, X_{2t}) changes.

It can be concluded that in the dynamic binary response panel data model with fixed γ, the identified set for β_2 shrinks as the support of the explanatory variables increases. Furthermore, the identified sets increase as the variance of the time-varying unobservables increases. Finally, similarly to the binary response static panel data model, the standard logit model and the model with $V_t|X, Y_0, \alpha \overset{iid}{\sim} N\left(0, \frac{\pi^2}{3}\right)$ errors give similar identified sets.

Table 3: Outer regions for β_2 under Probit and Logit specifications with symmetric support for (X_{1t}, X_{2t}) around zero when $\gamma = 0.5$

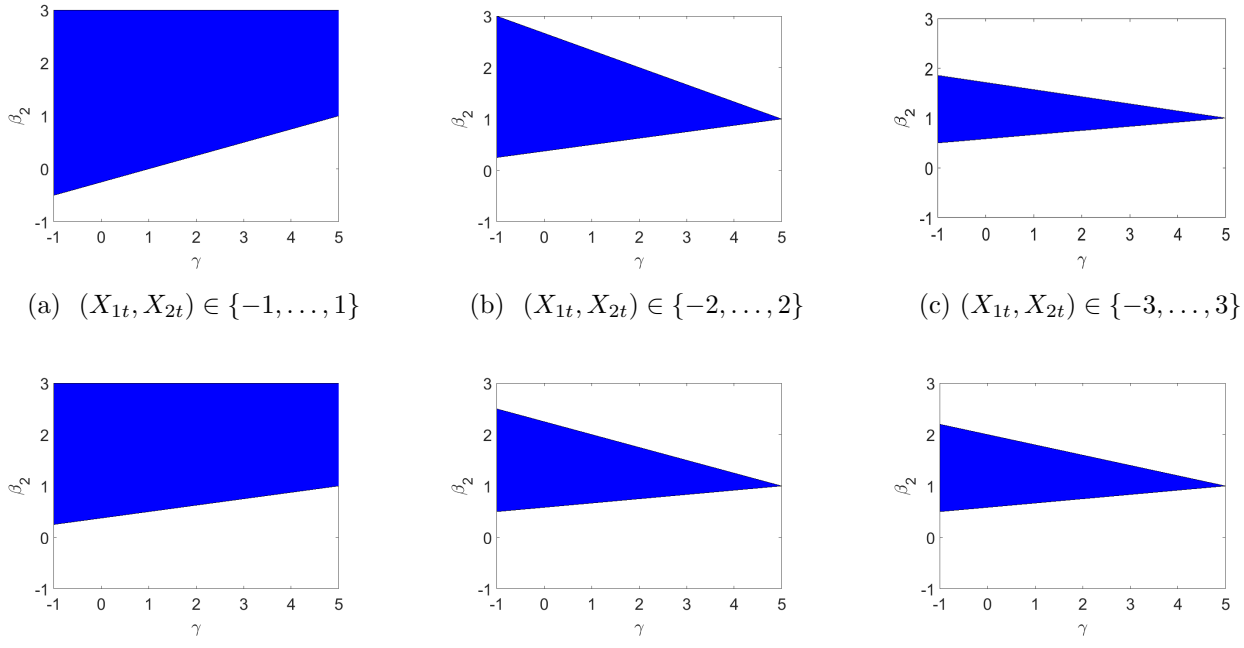
	Support of (X_{1t}, X_{2t})		
	$\{-1, 0, 1\}$	$\{-2, -1, 0, 1, 2\}$	$\{-3, -2, -1, 0, 1, 2, 3\}$
$V \overset{iid}{\sim} N(0, 1)$	$(-0.125, \infty)$	$(0.438, 2.5)$	$(0.625, 1.643)$
$V \overset{iid}{\sim} N\left(0, \frac{\pi^2}{3}\right)$	$(-0.375, \infty)$	$(0.313, 3.75)$	$(0.542, 1.917)$
Logistic	$(-0.375, \infty)$	$(0.313, 3.75)$	$(0.542, 1.917)$

Table 4: Outer regions for β_2 under Probit and Logit specifications with asymmetric support for (X_{1t}, X_{2t}) around zero when $\gamma = 0.5$

	Support of (X_{1t}, X_{2t})		
	$\{-1,0,1,2,3\}$	$\{-2,-1,0,1,2,3,4\}$	$\{-4,-3,-2,-1,0,1,2\}$
$V \overset{iid}{\sim} N(0,1)$	$(0.438, \infty)$	$(0.625, 2.125)$	$(0.625, 1.9)$
$V \overset{iid}{\sim} N\left(0, \frac{\pi^2}{3}\right)$	$(0.313, \infty)$	$(0.542, 2.833)$	$(0.542, 2.375)$
Logistic	$(0.313, \infty)$	$(0.542, 2.833)$	$(0.542, 2.375)$

4.2.2 Example 4

Consider the binary response dynamic panel data model in (15) with $V_t|X, Y_0, \alpha \overset{iid}{\sim} N(0,1)$, but assume no knowledge of β_2 or γ. Figure 4 provides identified sets for (β_2, γ) as described in Theorem 2, for different values of γ in the grid $[-1, 6]$, as the support of (X_{1t}, X_{2t}) changes, with true values $\beta_2 = 1$ and $\gamma = 0.5$.



(a) $(X_{1t}, X_{2t}) \in \{-1, \ldots, 1\}$ (b) $(X_{1t}, X_{2t}) \in \{-2, \ldots, 2\}$ (c) $(X_{1t}, X_{2t}) \in \{-3, \ldots, 3\}$

(d) $(X_{1t}, X_{2t}) \in \{-1, \ldots, 3\}$ (e) $(X_{1t}, X_{2t}) \in \{-2, \ldots, 4\}$ (f) $(X_{1t}, X_{2t}) \in \{-4, \ldots, 2\}$

Figure 4: Joint identified sets for (β_2, γ)

It is clear that for the specific range of values for the grid of γ chosen the identified set for β_2 shrinks as the support of the discrete explanatory variables (X_{1t}, X_{2t}) increases, however with three periods the identified set for γ is only bounded from above. Recall that as shown in Honoré and Kyriazidou (2000) the parameters in the dynamic binary response model with one lagged dependent variable and logistically distributed unobservables are point-identified with at least four time periods.

4.3 Static Ordered Response Panel Data Model

4.3.1 Example 5

Consider the two period static ordered response panel data model as discussed in Section 3,

$$Y_t = \begin{cases} 0 \; if \quad X_t\beta + \alpha + V_t < c_{1t} \\ 1 \; if \; c_{1t} < X_t\beta + \alpha + V_t \leq c_{2t} \\ 2 \; if \; c_{2t} < X_t\beta + \alpha + V_t \end{cases} \tag{16}$$

where $X_t = (X_{1t}, X_{2t})$, $\beta = (\beta_1, \beta_2)$, $\alpha|X \sim N(\overline{X}\delta, 1)$ with $\overline{X} = \frac{1}{2}(X_1 + X_2)$ and $\delta = (1, -1)$ and $V_t|X, \alpha \overset{iid}{\sim} N(0, 1)$. The true parameter of $\beta_2 = 1$ after the normalization of $\beta_1 = 1$. The threshold parameters are assumed to be time-invariant such that $c_{11} = c_{12}$ and $c_{21} = c_{22}$[4], and $c_{11} = c_{12}$ is normalized to 0. Table 5 and Table 6 provide the outer sets for β_2 as described in Theorem 5, when $c = c_{21} = c_{22}$ is fixed, for different values of c, and as the support of the discrete explanatory variables (X_{1t}, X_{2t}) changes[5]. For any given c it is evident that as the support of X_t increases the identified sets decrease. Furthermore, it can be also clearly seen that the choice of the threshold c affects the size of the identified sets. This indicates that in the ordered choice model the threshold plays a crucial role in the identifying power of the model.

[4]In a recent working paper Botosaru and Muris (2017) consider estimation of the regression parameters and the thresholds in a fixed effects ordered logit model with a time-varying link function.

[5]The outer sets are calculated by setting $\omega' \to \infty$ and $\omega'' \to -\infty$.

Table 5: Identified sets for β_2 under Probit and Logit specifications with symmetric support for (X_{1t}, X_{2t}) around zero

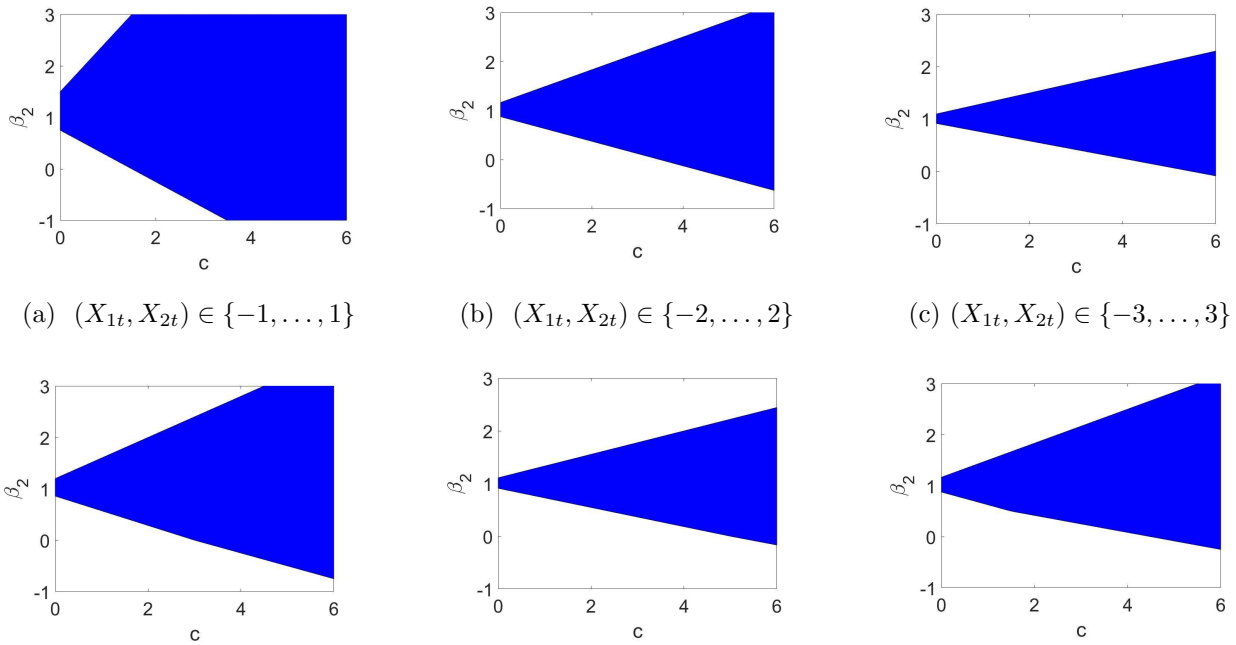
	Support of (X_{1t}, X_{2t})		
	$\{-1, 0, 1\}$	$\{-2, -1, 0, 1, 2\}$	$\{-3, -2, -1, 0, 1, 2, 3\}$
$c = 0.5$	$(0, 5)$	$(0.5, 1.8)$	$(0.667, 1.444)$
$c = 1.5$	$(0, 3)$	$(0.5, 1.667)$	$(0.667, 1.4)$
$c = 2.5$	(-0.5,2.333)	$(0.286, 1.571)$	$(0.556, 1.364)$

Table 6: Identified sets for β_2 under Probit and Logit specifications with asymmetric support for (X_{1t}, X_{2t}) around zero

	Support of (X_{1t}, X_{2t})		
	$\{-1, 0, 1, 2, 3\}$	$\{-2, -1, 0, 1, 2, 3, 4\}$	$\{-4, -3, -2, -1, 0, 1, 2\}$
$c = 0.5$	$(0.5, 3)$	$(0.667, 1.667)$	$(0.636, 1.8)$
$c = 1.5$	$(0.429, 1.8)$	$(0.636, 1.444)$	(0.5,1.667)
$c = 2.5$	$(0.375, 1.571)$	$(0.6, 1.364)$	$(0.333, 1.571)$

4.3.2 Example 6

Consider the ordered response static panel data model in (16), but assume no knowledge of β_2 or c. Figure 5 provides identified sets for (β_2, c) for different values of c in the grid [0,6], when the true value of $\beta_2 = 1$ and $c = 1.5$, as the support of (X_{1t}, X_{2t}) increases. Similarly to the previous examples, the identified sets decrease as the support of the explanatory variables increases.



(a) $(X_{1t}, X_{2t}) \in \{-1, \ldots, 1\}$ (b) $(X_{1t}, X_{2t}) \in \{-2, \ldots, 2\}$ (c) $(X_{1t}, X_{2t}) \in \{-3, \ldots, 3\}$

(d) $(X_{1t}, X_{2t}) \in \{-1, \ldots, 3\}$ (e) $(X_{1t}, X_{2t}) \in \{-2, \ldots, 4\}$ (f) $(X_{1t}, X_{2t}) \in \{-4, \ldots, 2\}$

Figure 5: Joint identified sets for (β_2, c)

5 Concluding Remarks

This paper studies identification in discrete response panel data models with fixed effects. Under fairly mild conditions, identified sets for the regression parameters in the dynamic binary and static ordered response models are derived, without assuming distributional assumptions on the time-varying unobservables or the fixed effect. The bounds are achieved by relying on observable implications in which the fixed effect does not appear. Similar to the binary static panel data model, the dynamic binary and the static ordered panel data models are complete and informative identification bounds for the parameters of interest can be derived.

As discussed in Section 1 when the time-varying unobservables are independent and identically distributed with a logistic distribution, then the regression parameters in the linear index binary and ordered response panel data models can be point-identified. The feature of the distribution these papers use that does not depend on the unobservable α is the conditional

probability of the outcome variable in a specific period taking a specific value, conditional on the event that individuals change at some period in the past. The feature of the distribution that does not depend on the unobserved heterogeneity in this paper is the joint probability of two outcome variables that take different values in two periods. If Assumptions 3 and 9 are strengthened such that the time-varying unobservables follow an independent and identically distributed logistic distribution, the bounds provided in this paper might still fail to be singletons. Chamberlain (1984, 2010) and Honoré and Kyriazidou (2000) prove point-identification of the regression parameters, when the regressors have bounded support, under the assumptions that the time-varying unobservables follow a logistic distribution and are independent of both the explanatory variables (X_1, X_2) and α. Under Assumptions 5 and 11, V is allowed to be correlated with α. Imposing the assumption that $V \perp \alpha$ may or may not lead to point-identification and would require an additional assumption which might not be credible or testable.

In conclusion, even though the identification bounds in this paper might not be singleton sets, they provide information on the regression parameters under fairly weak conditions. Since the sets do not depend on any distributional assumption on the unobservables, they can provide information for a general class of linear index static and dynamic discrete response panel data models with fixed effects. Furthermore, they are relatively simple to construct and therefore might be easy to use for computation and inference.

A Appendix

A.1 Proof of Theorem 1

This section provides the proof of Theorem 1. It can be shown that the event probabilities described in Theorem 1 do not depend on α. Consider the sequences of events in (8). The conditional probability for the event $(Y_1, Y_2) = (0, 1)$ conditional on $Y_0 = 0$ is,

$$\begin{aligned} P(0,1|x,0) &= P_{(V,\alpha)|X,Y_0}[\{X_1\beta + \alpha + V_1 \leq 0\} \wedge \{X_2\beta + \alpha + V_2 > 0\}|X = x, Y_0 = 0] \\ &\leq P_{V|X,Y_0}[(X_2 - X_1)\beta + V_2 - V_1 > 0|X = x, Y_0 = 0] \\ &= P_{\Delta V|X,Y_0}[\Delta V > -\Delta X\beta|X = x, Y_0 = 0] \end{aligned}$$

The conditional probability for the event $(Y_1, Y_2) = (1, 0)$ conditional on $Y_0 = 0$ is given by,

$$\begin{aligned} P(1,0|x,0) &= P_{(V,\alpha)|X,Y_0}[\{X_1\beta + \alpha + V_1 \geq 0\} \wedge \{X_2\beta + \gamma + \alpha + V_2 < 0\}|X = x, Y_0 = 0] \\ &\leq P_{V|X,Y_0}[(X_2 - X_1)\beta + \gamma + V_2 - V_1 < 0|X = x, Y_0 = 0] \\ &= P_{\Delta V|X,Y_0}[\Delta V < -\Delta X\beta - \gamma|X = x, Y_0 = 0] \end{aligned}$$

The conditional probability for the event $(Y_1, Y_2) = (0, 1)$ conditional on $Y_0 = 1$ is,

$$\begin{aligned} P(0,1|x,1) &= P_{(V,\alpha)|X,Y_0}[\{X_1\beta + \gamma + \alpha + V_1 \leq 0\} \wedge \{X_2\beta + \alpha + V_2 > 0\}|X = x, Y_0 = 1] \\ &\leq P_{V|X,Y_0}[(X_2 - X_1)\beta - \gamma + V_2 - V_1 > 0|X = x, Y_0 = 1] \\ &= P_{\Delta V|X,Y_0}[\Delta V > -\Delta X\beta + \gamma|X = x, Y_0 = 1] \end{aligned}$$

The conditional probability for the event $(Y_1, Y_2) = (1, 0)$ conditional on $Y_0 = 1$ is equivalent to,

$$\begin{aligned} P(1,0|x,1) &= P_{(V,\alpha)|X,Y_0}[\{X_1\beta + \gamma + \alpha + V_1 \geq 0\} \wedge \{X_2\beta + \gamma + \alpha + V_2 < 0\}|X = x, Y_0 = 1] \\ &\leq P_{V|X,Y_0}[(X_2 - X_1)\beta + V_2 - V_1 < 0|X = x, Y_0 = 1] \\ &= P_{\Delta V|X,Y_0}[\Delta V < -\Delta X\beta|X = x, Y_0 = 1] \end{aligned}$$

For any fixed $X = x$ and by applying Assumption 5 , the above inequalities imply that

$$\begin{aligned} 1 - P(0,1|x,0) &\geq P_{\Delta V|Y_0}[\Delta V < -\Delta x\beta | Y_0 = 0] \\ P(1,0|x,0) &\leq P_{\Delta V|Y_0}[\Delta V < -\Delta x\beta - \gamma | Y_0 = 0] \\ 1 - P(0,1|x,1) &\geq P_{\Delta V|Y_0}[\Delta V < -\Delta x\beta + \gamma | Y_0 = 1] \\ P(1,0|x,1) &\leq P_{\Delta V|Y_0}[\Delta V < -\Delta x\beta | Y_0 = 1] \end{aligned}$$

which completes the proof.

A.2 Proof of Theorem 2

This section proves Theorem 2. Consider any constant $\omega \in \mathbb{R}$, conditioning on $Y_0 = 0$, then

$$\begin{aligned} (Y_1, Y_2) = (0,1) \wedge -\Delta X\beta \geq \omega \Rightarrow \Delta V > \omega \\ (Y_1, Y_2) = (1,0) \wedge -\Delta X\beta - \gamma \leq \omega \Rightarrow \Delta V < \omega \end{aligned} \tag{17}$$

and conditioning on $Y_0 = 1$, then

$$\begin{aligned} (Y_1, Y_2) = (0,1) \wedge -\Delta X\beta + \gamma \geq \omega \Rightarrow \Delta V > \omega \\ (Y_1, Y_2) = (1,0) \wedge -\Delta X\beta \leq \omega \Rightarrow \Delta V < \omega \end{aligned} \tag{18}$$

The relations in (17) and (18) imply that, $\forall \omega \in \mathbb{R}$:

$$\begin{aligned} 1 - P[(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta \geq \omega | X = x, Y_0 = 0] &\geq P[\Delta V < \omega | X = x, Y_0 = 0] \\ P[(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta - \gamma \leq \omega | X = x, Y_0 = 0] &\leq P[\Delta V < \omega | X = x, Y_0 = 0] \\ 1 - P[(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta + \gamma \geq \omega | X = x, Y_0 = 1] &\geq P[\Delta V < \omega | X = x, Y_0 = 1] \\ P[(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta \leq \omega | X = x, Y_0 = 1] &\leq P[\Delta V < \omega | X = x, Y_0 = 1] \end{aligned}$$

$$\Longleftrightarrow$$

When $-\Delta X\beta \geq \omega$

$$P[(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta \geq \omega | X = x, Y_0 = 0] = P(0,1|x,0)$$

otherwise

$$P[(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta \geq \omega | X = x, Y_0 = 0] = 0$$

When $-\Delta X\beta - \gamma \leq \omega$

$$P[(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta - \gamma \leq \omega | X = x, Y_0 = 0] = P(1,0|x,0)$$

otherwise

$$P[(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta - \gamma \leq \omega | X = x, Y_0 = 0] = 0$$

When $-\Delta X\beta + \gamma \geq \omega$

$$P[(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta + \gamma \geq \omega | X = x, Y_0 = 1] = P(0,1|x,1)$$

otherwise

$$P[(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta + \gamma \geq \omega | X = x, Y_0 = 1] = 0$$

When $-\Delta X\beta \leq \omega$

$$P[(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta \leq \omega | X = x, Y_0 = 1] = P(1,0|x,1)$$

otherwise

$$P[(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta \leq \omega | X = x, Y_0 = 1] = 0$$

These equalities in combination with (17), (18) and by applying Assumption 5 imply that,

$$\begin{aligned}
\inf_{x:-\Delta x\beta \geq \omega} 1 - P(0,1|x,0) &\geq F_{\Delta V|Y_0}(\omega|0) \\
\sup_{x:-\Delta x\beta - \gamma \leq \omega} P(1,0|x,0) &\leq F_{\Delta V|Y_0}(\omega|0) \\
\inf_{x:-\Delta x\beta + \gamma \geq \omega} 1 - P(0,1|x,1) &\geq F_{\Delta V|Y_0}(\omega|1) \\
\sup_{x:-\Delta x\beta \leq \omega} P(1,0|x,1) &\leq F_{\Delta V|Y_0}(\omega|1)
\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
\sup_{x:-\Delta x\beta - \gamma \leq \omega} P(1,0|x,0) &\leq F_{\Delta V|Y_0}(\omega|0) \leq \inf_{x:-\Delta x\beta \geq \omega} 1 - P(0,1|x,0) \\
\sup_{x:-\Delta x\beta \leq \omega} P(1,0|x,1) &\leq F_{\Delta V|Y_0}(\omega|1) \leq \inf_{x:-\Delta x\beta + \gamma \geq \omega} 1 - P(0,1|x,1)
\end{aligned} \tag{19}$$

which completes the proof.

A.3 Proof of Theorem 3

To prove the unconditional identified set in Theorem 3, first notice that

$$F_{\Delta V|X}(\omega|x) = F_{\Delta V|X,Y_0}(\omega|x,0)P(Y_0=0|X=x) + F_{\Delta V|X,Y_0}(\omega|x,1)P(Y_0=1|X=x).$$

Define $P(Y_0=0|X=x) = P_0(x)$ and $P(Y_0=1|X=x) = P_1(x)$, which are fully observed. The relations (17) and (18), $\forall \omega \in \mathbb{R}$ imply:

$$\begin{aligned}
1 - P[(Y_1,Y_2)=(0,1) \wedge -\Delta X\beta \geq \omega | X=x, Y_0=0] &\geq P[\Delta V < \omega | X=x, Y_0=0] \\
P[(Y_1,Y_2)=(1,0) \wedge -\Delta X\beta - \gamma \leq \omega | X=x, Y_0=0] &\leq P[\Delta V < \omega | X=x, Y_0=0] \\
1 - P[(Y_1,Y_2)=(0,1) \wedge -\Delta X\beta + \gamma \geq \omega | X=x, Y_0=1] &\geq P[\Delta V < \omega | X=x, Y_0=1] \\
P[(Y_1,Y_2)=(1,0) \wedge -\Delta X\beta \leq \omega | X=x, Y_0=1] &\leq P[\Delta V < \omega | X=x, Y_0=1].
\end{aligned} \tag{20}$$

Define by

$$\begin{aligned}
\underline{G}(\omega|x,0) &= P[(Y_1,Y_2)=(1,0) \wedge -\Delta X\beta - \gamma \leq \omega | X=x, Y_0=0] \\
\underline{G}(\omega|x,1) &= P[(Y_1,Y_2)=(1,0) \wedge -\Delta X\beta \leq \omega | X=x, Y_0=1] \\
\overline{G}(\omega|x,0) &= 1 - P[(Y_1,Y_2)=(0,1) \wedge -\Delta X\beta \geq \omega | X=x, Y_0=0] \\
\overline{G}(\omega|x,1) &= 1 - P[(Y_1,Y_2)=(0,1) \wedge -\Delta X\beta + \gamma \geq \omega | X=x, Y_0=1].
\end{aligned}$$

Then multiplying by $P_0(x)$ and $P_1(x)$ such that,

$$\begin{aligned}
1 - P[(Y_1,Y_2)=(0,1) \wedge -\Delta X\beta \geq \omega | X=x, Y_0=0]P_0(x) &\geq P[\Delta V < \omega | X=x, Y_0=0]P_0(x) \\
P[(Y_1,Y_2)=(1,0) \wedge -\Delta X\beta - \gamma \leq \omega | X=x, Y_0=0]P_0(x) &\leq P[\Delta V < \omega | X=x, Y_0=0]P_0(x) \\
1 - P[(Y_1,Y_2)=(0,1) \wedge -\Delta X\beta + \gamma \geq \omega | X=x, Y_0=1]P_1(x) &\geq P[\Delta V < \omega | X=x, Y_0=1]P_1(x) \\
P[(Y_1,Y_2)=(1,0) \wedge -\Delta X\beta \leq \omega | X=x, Y_0=1]P_1(x) &\leq P[\Delta V < \omega | X=x, Y_0=1]P_1(x),
\end{aligned}$$

implies that the relations in (20) can be expressed as

$$\begin{aligned}
\underline{G}(\omega|x,0)P_0(x) &\leq F_{\Delta V|X,Y_0}(\omega|x,0)P_0(x) \leq \overline{G}(\omega|x,0)P_0(x) \\
\underline{G}(\omega|x,1)P_1(x) &\leq F_{\Delta V|X,Y_0}(\omega|x,1)P_1(x) \leq \overline{G}(\omega|x,1)P_1(x)
\end{aligned}$$

which implies,

$$\underline{G}(\omega|x,0)P_0(x)+\underline{G}(\omega|x,1)P_1(x)\leq F_{\Delta V|X}(\omega|x)\leq \overline{G}(\omega|x,0)P_0(x)+\overline{G}(\omega|x,1)P_1(x)$$

$$\Longleftrightarrow$$

$$\underline{G}(\omega|x,0)P_0(x)+\underline{G}(\omega|x,1)P_1(x)\leq F_{\Delta V}(\omega)\leq \overline{G}(\omega|x,0)P_0(x)+\overline{G}(\omega|x,1)P_1(x)$$

where the last result follows from Assumption $V \perp X$. The last relation implies that,

$$\sup_{x\in\mathcal{X}}\{\underline{G}(\omega|x,0)P_0(x)+\underline{G}(\omega|x,1)P_1(x)\}\leq \inf_{x\in\mathcal{X}}\{\overline{G}(\omega|x,0)P_0(x)+\overline{G}(\omega|x,1)P_1(x)\}.$$

This completes the proof.

A.4 Regions of unobservables for the static ordered model

The regions $\mathcal{R}^{SO}_{(y_1,y_2)}(x,c;\beta)$ that partition the support of (V,α) such that $(Y_1,Y_2)=(y_1,y_2)$ when $X=x$ and for any fixed c and plotted in Figure 2, are given by the following set of inequalities,

$$\begin{aligned}
\mathcal{R}^{SO}_{(0,0)}(x,c;\beta) &= \left\{(V,\alpha)\in(\mathcal{V},\mathcal{A}): \begin{array}{c} x_1\beta+\alpha+V_1-c_{11}\leq 0 \\ \& \\ x_2\beta+\alpha+V_2-c_{12}\leq 0 \end{array}\right\} \\
\mathcal{R}^{SO}_{(0,1)}(x,c;\beta) &= \left\{(V,\alpha)\in(\mathcal{V},\mathcal{A}): \begin{array}{c} x_1\beta+\alpha+V_1-c_{11}\leq 0 \\ \& \\ -x_2\beta+c_{12}<\alpha+V_2\leq -x_2\beta+c_{22} \end{array}\right\} \\
\mathcal{R}^{SO}_{(0,2)}(x,c;\beta) &= \left\{(V,\alpha)\in(\mathcal{V},\mathcal{A}): \begin{array}{c} x_1\beta+\alpha+V_1-c_{11}\leq 0 \\ \& \\ 0<x_2\beta+\alpha+V_2-c_{22} \end{array}\right\}
\end{aligned}$$

$$\mathcal{R}^{SO}_{(1,0)}(x,c;\beta) = \left\{ (V,\alpha) \in (\mathcal{V},\mathcal{A}) : \begin{array}{c} -x_1\beta + c_{11} < \alpha + V_1 \leq -x_1\beta + c_{21} \\ \& \\ x_2\beta + \alpha + V_2 - c_{12} \leq 0 \end{array} \right\}$$

$$\mathcal{R}^{SO}_{(1,1)}(x,c;\beta) = \left\{ (V,\alpha) \in (\mathcal{V},\mathcal{A}) : \begin{array}{c} -x_1\beta + c_{11} < \alpha + V_1 \leq -x_1\beta + c_{21} \\ \& \\ -x_2\beta + c_{12} < \alpha + V_2 \leq -x_2\beta + c_{22} \end{array} \right\}$$

$$\mathcal{R}^{SO}_{(1,2)}(x,c;\beta) = \left\{ (V,\alpha) \in (\mathcal{V},\mathcal{A}) : \begin{array}{c} -x_1\beta + c_{11} < \alpha + V_1 \leq -x_1\beta + c_{21} \\ \& \\ 0 < x_2\beta + \alpha + V_2 - c_{22} \end{array} \right\}$$

$$\mathcal{R}^{SO}_{(2,0)}(x,c;\beta) = \left\{ (V,\alpha) \in (\mathcal{V},\mathcal{A}) : \begin{array}{c} 0 < x_1\beta + \alpha + V_1 - c_{21} \\ \& \\ x_2\beta + \alpha + V_2 - c_{12} \leq 0 \end{array} \right\}$$

$$\mathcal{R}^{SO}_{(2,1)}(x,c;\beta) = \left\{ (V,\alpha) \in (\mathcal{V},\mathcal{A}) : \begin{array}{c} 0 < x_1\beta + \alpha + V_1 - c_{21} \\ \& \\ -x_2\beta + c_{12} < \alpha + V_2 \leq -x_2\beta + c_{22} \end{array} \right\}$$

$$\mathcal{R}^{SO}_{(2,2)}(x,c;\beta) = \left\{ (V,\alpha) \in (\mathcal{V},\mathcal{A}) : \begin{array}{c} 0 < x_1\beta + \alpha + V_1 - c_{21} \\ \& \\ 0 < x_2\beta + \alpha + V_2 - c_{22} \end{array} \right\}$$

A.5 Proof of Theorem 4

To prove Theorem 4 first notice that the conditional probabilities $P(Y_1 = y_1 \wedge Y_2 = y_2|X = x, c) = P(y_1, y_2|x, c)$ of the events given in (10) are given by:

$$\begin{aligned}
P(0,1|x,c)) &= P_{(V,\alpha)|X,c}[\{0 \geq X_1\beta + \alpha + V_1 - c_{11}\} \\
&\wedge \{X_2\beta + \alpha + V_2 - c_{12} > 0 \wedge 0 \geq X_2\beta + \alpha + V_2 - c_{22}\}|X = x, c] \\
P(1,0|x,c) &= P_{(V,\alpha)|X,c}[\{X_1\beta + \alpha + V_1 - c_{11} > 0 \wedge 0 \geq X_1\beta + \alpha + V_1 - c_{21}\} \\
&\wedge \{0 \geq X_2\beta + \alpha + V_2 - c_{12}\}|X = x, c] \\
P(0,2|x,c) &= P_{(V,\alpha)|X,c}[\{0 \geq X_1\beta + \alpha + V_1 - c_{11}\} \wedge \{X_2\beta + \alpha + V_2 - c_{22} > 0\}|X = x, c] \\
P(2,0|x,c) &= P_{(V,\alpha)|X,c}[\{X_1\beta + \alpha + V_1 - c_{21} > 0\} \wedge \{0 \geq X_2\beta + \alpha + V_2 - c_{12}\}|X = x, c] \\
P(1,2|x,c) &= P_{(V,\alpha)|X,c}[\{X_1\beta + \alpha + V_1 - c_{11} > 0 \wedge 0 \geq X_1\beta + \alpha + V_1 - c_{21}\} \\
&\wedge \{X_2\beta + \alpha + V_2 - c_{22} > 0\}|X = x, c] \\
P(2,1|x,c) &= P_{(V,\alpha)|X,c}[\{X_1\beta + \alpha + V_1 - c_{21} > 0\} \\
&\wedge \{X_2\beta + \alpha + V_2 - c_{12} > 0 \wedge 0 \geq X_2\beta + \alpha + V_2 - c_{22}\}|X = x, c] \qquad (21)
\end{aligned}$$

From (21) it can be shown that:

$$\begin{aligned}
P(0,1|x,c) &\leq P_{(V,\alpha)|X,c}(\{0 \geq X_1\beta + \alpha + V_1 - c_{11}\} \wedge \{X_2\beta + \alpha + V_2 - c_{12} > 0\}|X = x, c) \\
&\leq P_{(V,\alpha)|X,c}(0 > (X_1 - X_2)\beta + (V_1 - V_2) - c_{11} + c_{12}|X = x, c) \\
P(1,0|x,c) &\leq P_{(V,\alpha)|X,c}[\{X_1\beta + \alpha + V_1 - c_{11} > 0\} \wedge \{0 \geq X_2\beta + \alpha + V_2 - c_{12}\}|X = x, c] \\
&\leq P_{(V,\alpha)|X,c}(0 > (X_2 - X_1)\beta + (V_2 - V_1) - c_{12} + c_{11}|X = x, c) \\
P(0,2|x,c) &\leq P_{(V,\alpha)|X,c}(0 > (X_1 - X_2)\beta + (V_1 - V_2) - c_{11} + c_{22}|X = x, c) \\
P(2,0|x,c) &\leq P_{(V,\alpha)|X,c}(0 > (X_2 - X_1)\beta + (V_2 - V_1) - c_{12} + c_{21}|X = x, c) \\
P(1,2|x,c) &\leq P_{(V,\alpha)|X,c}(\{0 \geq X_1\beta + \alpha + V_1 - c_{21}\} \wedge \{X_2\beta + \alpha + V_2 - c_{22} > 0\}|X = x, c) \\
&\leq P_{(V,\alpha)|X,c}(0 > (X_1 - X_2)\beta + V_1 - V_2 + c_{22} - c_{21}|X = x, c) \\
P(2,1|x,c) &\leq P_{(V,\alpha)|X,c}[\{X_1\beta + \alpha + V_1 - c_{21} > 0\} \wedge \{0 \geq X_2\beta + \alpha + V_2 - c_{22}\}|X = x, c] \\
&\leq P_{(V,\alpha)|X,c}(0 > (X_2 - X_1)\beta + V_2 - V_1 + c_{21} - c_{22}|X = x, c)
\end{aligned} \qquad (22)$$

Then using Assumptions 11 for any fixed $c \in \mathcal{C}$, the relations in (22) can be expressed as:

$$\begin{aligned}
P(0,1|x,c) &\leq P_{V|X,c}(V_2 - V_1 > -\Delta X\beta - c_{11} + c_{12}|X = x, c) \\
&= 1 - F_{\Delta V}[-\Delta x\beta - c_{11} + c_{12}] \\
P(1,0|x,c) &\leq P_{V|X,c}(V_2 - V_1 < -\Delta X\beta + c_{12} - c_{11}|X = x, c) \\
&= F_{\Delta V}[-\Delta x\beta - c_{11} + c_{12}] \\
P(0,2|x,c) &\leq P_{V|X,c}(-\Delta X\beta - c_{11} + c_{22} < V_2 - V_1|X = x, c) \\
&= 1 - F_{\Delta V}[-\Delta x\beta - c_{11} + c_{22}] \\
P(2,0|x,c) &\leq P_{V|X,c}(V_2 - V_1 < -\Delta X\beta + c_{12} - c_{21}|X = x, c) \\
&= F_{\Delta V}[-\Delta x\beta + c_{12} - c_{21}] \\
P(1,2|x,c) &\leq P_{V|X,c}(-\Delta X\beta + c_{22} - c_{21} < V_2 - V_1|X = x, c) \\
&= 1 - F_{\Delta V}[-\Delta x\beta + c_{22} - c_{21}] \\
P(2,1|x,c) &\leq P_{V|X,c}(V_2 - V_1 < -\Delta X\beta + c_{22} - c_{21}|X = x, c) \\
&= F_{\Delta V}[-\Delta x\beta + c_{22} - c_{21}] \qquad (23)
\end{aligned}$$

The inequalities in (23) lead to bounds for β, such that for any given $X = x$,

$$\begin{aligned}
P(1,0|x,c) &\leq F_{\Delta V}[-\Delta x\beta - c_{11} + c_{12}] \leq 1 - P(0,1|x,c) \\
P(2,0|x,c) &\leq F_{\Delta V}[-\Delta x\beta + c_{12} - c_{21}] \\
&\quad F_{\Delta V}[-\Delta x\beta - c_{11} + c_{22}] \leq 1 - P(0,2|x,c) \\
P(2,1|x,c) &\leq F_{\Delta V}[-\Delta x\beta + c_{22} - c_{21}] \leq 1 - P(1,2|x,c)
\end{aligned} \qquad (24)$$

The above relations show that changing choices from period $t = 1$ to $t = 2$ provide restrictions on the distributions of ΔV that do not depend on the fixed effect, α. In the binary case discussed in Section 2 the events $(Y_1, Y_2) = (0,0)$ *and* $(1,1)$ gave no information on β since the behaviour for these cases can be matched by extremely small or extremely large values of α. This is also true for the static ordered model for the events $(0,0)$ and $(2,2)$. However, in the ordered response model where $\mathcal{Y}_t = \{0,1,2\}$, considered in this section, there is an

"in-between" event, $(Y_1, Y_2) = (1, 1)$, that provides information for β without involving the fixed effect. To see that consider the joint probability of choosing the event (11):

$$\begin{aligned} P(1,1|x,c) &= P_{(V,\alpha)|X,c}[\{X_1\beta + \alpha + V_1 - c_{11} > 0 \wedge 0 \geq X_1\beta + \alpha + V_1 - c_{21}\} \\ &\wedge \{X_2\beta + \alpha + V_2 - c_{12} > 0 \wedge 0 \geq X_2\beta + \alpha + V_2 - c_{22}\}|X = x, c] \\ &= P_{(V,\alpha)|X,c}[X_1\beta + \alpha + V_1 - c_{11} > 0 \wedge 0 \geq X_1\beta + \alpha + V_1 - c_{21} \\ &\wedge X_2\beta + \alpha + V_2 - c_{12} > 0 \wedge 0 \geq X_2\beta + \alpha + V_2 - c_{22}|X = x, c] \end{aligned} \tag{25}$$

By combining

$$X_1\beta + \alpha + V_1 - c_{11} > 0 \; and \; 0 \geq X_2\beta + \alpha + V_2 - c_{22}$$

and

$$0 \geq X_1\beta + \alpha + V_1 - c_{21} \; and \; X_2\beta + \alpha + V_2 - c_{12} > 0$$

it can be shown that:

$$\begin{aligned} P(1,1|x,c) &\leq P_V[0 > (X_2 - X_1)\beta + (V_2 - V_1) - c_{22} + c_{11} \\ &\wedge (X_2 - X_1)\beta + (V_2 - V_1) - c_{12} + c_{21} > 0|X = x, c] \\ P(1,1|x,c) &\leq P_{\Delta V}[-\Delta X\beta + c_{22} - c_{11} > \Delta V > -\Delta X\beta + c_{12} - c_{21}|X = x, c] \end{aligned} \tag{26}$$

which does not depend on α. This completes the proof.

A.6 Proof of Theorem 5

This section proves the identified set of the static ordered model given in Theorem 5. As discussed in Section 3 the events $\{(0,1),(1,0),(0,2),(2,0),(1,2),(2,1),(1,1)\}$ provide restrictions on the distribution of ΔV that do not depend on the unobserved heterogeneity, α. From the inequalities in (24) and (26) it can be concluded that,

$$
\begin{aligned}
(Y_1, Y_2) = (0,1) \quad &\Rightarrow \quad \Delta V > -\Delta X\beta - c_{11} + c_{12} \\
(Y_1, Y_2) = (1,0) \quad &\Rightarrow \quad \Delta V < -\Delta X\beta - c_{11} + c_{12} \\
(Y_1, Y_2) = (0,2) \quad &\Rightarrow \quad \Delta V > -\Delta X\beta - c_{11} + c_{22} \\
(Y_1, Y_2) = (2,0) \quad &\Rightarrow \quad \Delta V < \Delta X\beta + c_{12} - c_{21} \\
(Y_1, Y_2) = (1,2) \quad &\Rightarrow \quad \Delta V > -\Delta X\beta + c_{22} - c_{21} \\
(Y_1, Y_2) = (2,1) \quad &\Rightarrow \quad \Delta V < -\Delta X\beta + c_{22} - c_{21} \\
(Y_1, Y_2) = (1,1) \quad &\Rightarrow \quad -\Delta X\beta + c_{22} - c_{11} \geq \Delta V \geq -\Delta X\beta + c_{12} - c_{21}
\end{aligned}
$$

For a given set of arbitrary constants $\omega, \omega', \omega'' \in \mathbb{R}$ the above relations imply that,

$$
\begin{aligned}
(Y_1, Y_2) = (0,1) \quad &\wedge \quad -\Delta X\beta - c_{11} + c_{12} \ \geq \omega \Rightarrow \Delta V > \omega \\
(Y_1, Y_2) = (1,0) \quad &\wedge \quad -\Delta X\beta - c_{11} + c_{12} \ \leq \omega \Rightarrow \Delta V < \omega \\
(Y_1, Y_2) = (0,2) \quad &\wedge \quad -\Delta X\beta - c_{11} + c_{22} \ \geq \omega \Rightarrow \Delta V > \omega \\
(Y_1, Y_2) = (2,0) \quad &\wedge \quad -\Delta X\beta + c_{12} - c_{21} \ \leq \omega \Rightarrow \Delta V < \omega \\
(Y_1, Y_2) = (1,2) \quad &\wedge \quad -\Delta X\beta + c_{22} - c_{21} \ \geq \omega \Rightarrow \Delta V > \omega \\
(Y_1, Y_2) = (2,1) \quad &\wedge \quad -\Delta X\beta + c_{22} - c_{21} \ \leq \omega \Rightarrow \Delta V < \omega
\end{aligned}
$$

$$
\begin{aligned}
(Y_1, Y_2) = (1,1) \quad &\wedge \quad \omega' \geq -\Delta X\beta + c_{22} - c_{11} \wedge -\Delta X\beta + c_{12} - c_{21} \geq \omega \ \Rightarrow \omega' > \Delta V > \omega \\
(Y_1, Y_2) = (1,1) \quad &\wedge \quad \omega \geq -\Delta X\beta + c_{22} - c_{11} \wedge -\Delta X\beta + c_{12} - c_{21} \geq \omega'' \ \Rightarrow \omega > \Delta V > \omega''
\end{aligned}
\tag{27}
$$

The relations in (27) and Assumption 11 imply that for each fixed $c \in \mathcal{C}$, $\forall \omega, \omega', \omega'' \in \mathbb{R}$:

$$\begin{aligned}
1 - P[(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta - c_{11} + c_{12} \geq \omega | X = x, c] &\geq P[\Delta V < \omega] \\
P[(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta - c_{11} + c_{12} \leq \omega | X = x, c] &\leq P[\Delta V < \omega] \\
1 - P[(Y_1, Y_2) = (0,2) \wedge -\Delta X\beta - c_{11} + c_{22} \geq \omega | X = x, c] &\geq P[\Delta V < \omega] \\
P[(Y_1, Y_2) = (2,0) \wedge -\Delta X\beta + c_{12} - c_{21} \leq \omega | X = x, c] &\leq P[\Delta V < \omega] \\
1 - P[(Y_1, Y_2) = (1,2) \wedge -\Delta X\beta + c_{22} - c_{21} \geq \omega | X = x, c] &\geq P[\Delta V < \omega] \\
P[(Y_1, Y_2) = (2,1) \wedge -\Delta X\beta + c_{22} - c_{21} \leq \omega | X = x, c] &\leq P[\Delta V < \omega] \\
P[(Y_1, Y_2) = (1,1) \wedge \{\omega' \geq -\Delta X\beta + c_{22} - c_{11} \wedge -\Delta X\beta + c_{12} - c_{21} \geq \omega\} | X = x, c] & \\
\leq P[\omega < \Delta V < \omega'] & \\
P[(Y_1, Y_2) = (1,1) \wedge \{\omega \geq -\Delta X\beta + c_{22} - c_{11} \wedge -\Delta X\beta + c_{12} - c_{21} \geq \omega''\} | X = x, c] & \\
\leq P[\omega'' < \Delta V < \omega] & \qquad (28)
\end{aligned}$$

Following the same arguments as in proving Theorem 2 it can be shown that for any given $\omega, \omega', \omega'' \in \mathbb{R}$,

When $-\Delta X\beta - c_{11} + c_{12} \geq \omega$

$$1 - P[(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta - c_{11} + c_{12} \geq \omega | X = x, c] = 1 - P(0,1|x,c)$$

otherwise

$$1 - P[(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta - c_{11} + c_{12} \geq \omega | X = x, c] = 1$$

When $-\Delta X\beta - c_{11} + c_{12} \leq \omega$

$$P[(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta - c_{11} + c_{12} \leq \omega | X = x, c] = P(1,0|x,c)$$

otherwise

$$P[(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta - c_{11} + c_{12} \leq \omega | X = x, c] = 0$$

When $-\Delta X\beta - c_{11} + c_{22} \geq \omega$

$$1 - P[(Y_1, Y_2) = (0,2) \wedge -\Delta X\beta - c_{11} + c_{22} \geq \omega | X = x, c] = 1 - P(0,2|x,c)$$

otherwise

$$1 - P[(Y_1, Y_2) = (0,2) \wedge -\Delta X\beta - c_{11} + c_{22} \geq \omega | X = x, c] = 1$$

When $-\Delta X\beta + c_{12} - c_{21} \leq \omega$

$$P[(Y_1, Y_2) = (2,0) \wedge -\Delta X\beta + c_{12} - c_{21} \leq \omega | X = x, c] = P(2,0|x,c)$$

otherwise

$$P[(Y_1, Y_2) = (2,0) \wedge -\Delta X\beta + c_{12} - c_{21} \leq \omega | X = x, c] = 0$$

When $-\Delta X\beta + c_{22} - c_{21} \geq \omega$

$$1 - P[((Y_1, Y_2) = (1,2) \wedge -\Delta X\beta + c_{22} - c_{21} \geq \omega | X = x, c] = 1 - P(1,2|x,c)$$

otherwise

$$1 - P[(Y_1, Y_2) = (1,2) \wedge -\Delta X\beta + c_{22} - c_{21} \geq \omega | X = x, c] = 1$$

When $-\Delta X\beta + c_{22} - c_{21} \leq \omega$

$$P[(Y_1, Y_2) = (2,1) \wedge -\Delta X\beta + c_{22} - c_{21} \leq \omega | X = x, c] = P(2,1|x,c)$$

otherwise

$$P[(Y_1, Y_2) = (2,1) \wedge -\Delta X\beta + c_{22} - c_{21} \leq \omega | X = x, c] = 0$$

When $\omega' \geq -\Delta X\beta + c_{22} - c_{11} \wedge -\Delta X\beta + c_{12} - c_{21} \geq \omega$,

$$P[(Y_1, Y_2) = (1,1) \wedge \{\omega' > -\Delta X\beta + c_{22} - c_{11} \wedge -\Delta X\beta + c_{12} - c_{21} \geq \omega\} | X = x, c] = P(1,1|x,c)$$

otherwise

$$P[(Y_1, Y_2) = (1,1) \wedge \{\omega' > -\Delta X\beta + c_{22} - c_{11} \wedge -\Delta X\beta + c_{12} - c_{21} \geq \omega\} | X = x, c] = 0$$

When $\omega \geq -\Delta X\beta + c_{22} - c_{11} \wedge -\Delta X\beta + c_{12} - c_{21} \geq \omega''$

$$P[(Y_1, Y_2) = (1,1) \wedge \{\omega \geq -\Delta X\beta + c_{22} - c_{11} \wedge -\Delta X\beta + c_{12} - c_{21} > \omega''\} | X = x, c] = P(1,1|x,c)$$

otherwise

$$P[(Y_1, Y_2) = (1,1) \wedge \{\omega \geq -\Delta X\beta + c_{22} - c_{11} \wedge -\Delta X\beta + c_{12} - c_{21} > \omega''\} | X = x, c] = 0.$$

The above relations show that depending on the different values of $x \in \mathcal{X}$ the lower and upper bounds of the inequalities in (28) change. Since for any value of β in the identified set

the relations in (28) should hold simultaneously for all $\omega, \omega', \omega'' \in \mathbb{R}$, a.e. $x \in \mathcal{X}$, combining (24), (26), (28) and the above relations for the different values of $x \in \mathcal{X}$, the distribution of ΔV is shown to be bounded by,

$$\begin{aligned}
\sup_{x:-\Delta x\beta - c_{11} + c_{12} \leq \omega} P(1,0|x,c) &\leq F_{\Delta V}(\omega) \\
F_{\Delta V}(\omega) &\leq \inf_{x:-\Delta x\beta - c_{11} + c_{12} \geq \omega} [1 - P(0,1|x,c)] \\
\sup_{x:-\Delta x\beta + c_{12} - c_{21} \leq \omega} P(2,0|x,c) &\leq F_{\Delta V}(\omega) \\
F_{\Delta V}(\omega) &\leq \inf_{x:-\Delta x\beta - c_{11} + c_{22} \geq \omega} [1 - P(0,2|x,c)] \\
\sup_{x:-\Delta x\beta + c_{22} - c_{21} \leq \omega} P(2,1|x,c) &\leq F_{\Delta V}(\omega) \\
F_{\Delta V}(\omega) &\leq \inf_{x:-\Delta x\beta + c_{22} - c_{21} \geq \omega} [1 - P(1,2|x,c)] \\
\sup_{x \in X^*} P(1,1|x,c) &\leq F_{\Delta V}(\omega') - F_{\Delta V}(\omega) \\
\sup_{x \in X^{**}} P(1,1|x,c) &\leq F_{\Delta V}(\omega) - F_{\Delta V}(\omega'')
\end{aligned}$$

where

$$\begin{aligned}
X^* &= \{x : \omega' \geq -\Delta x\beta + c_{22} - c_{11} \wedge -\Delta x\beta + c_{12} - c_{21} \geq \omega\} \\
X^{**} &= \{x : \omega \geq -\Delta x\beta + c_{22} - c_{11} \wedge -\Delta x\beta + c_{12} - c_{21} \geq \omega''\}
\end{aligned}$$

Furthermore, notice that at the limit $\omega' \to \infty$ and $\omega'' \to -\infty$:

$$\begin{aligned}
P(1,1|x,c) &\leq 1 - F_{\Delta V}(\omega) \Leftrightarrow F_{\Delta V}(\omega) \leq 1 - P(1,1|x,c), \ \textit{when } x \in X^* \\
P(1,1|x,c) &\leq F_{\Delta V}(\omega), \ \textit{when } x \in X^{**}
\end{aligned}$$

and

$$\begin{aligned}
F_{\Delta V}(\omega) &\leq \inf_{x \in X^*} 1 - P(1,1|x,c) \\
\sup_{x \in X^{**}} P(1,1|x,c) &\leq F_{\Delta V}(\omega)
\end{aligned} \tag{29}$$

which completes the proof.

References

Arellano, M., and S. Bonhomme (2011): "Nonlinear panel data analysis," *Annual Review of Economics*, 3, 395–424.

Arellano, M., and B. E. Honoré (2001): "Panel data models: some recent developments," *Handbook of econometrics*, 5, 3229–3296.

Baetschmann, G., K. E. Staub, and R. Winkelmann (2015): "Consistent estimation of the fixed effects ordered logit model," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(3), 685–703.

Botosaru, I., and C. Muris (2017): "Binarization for panel models with fixed effects," Discussion paper, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Chamberlain, G. (1984): "Panel data," *Handbook of econometrics*, 2, 1247–1318.

——— (2010): "Binary response models for panel data: Identification and information," *Econometrica*, 78(1), 159–168.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81(2), 535–580.

Chernozhukov, V., J. Hahn, and W. Newey (2005): "Bound analysis in panel models with correlated random effects," Discussion paper, Unpublished manuscript, MIT.

Chesher, A. (2010): "Instrumental variable models for discrete outcomes," *Econometrica*, 78(2), 575–601.

Chesher, A., and K. Smolinski (2012): "IV models of ordered choice," *Journal of Econometrics*, 166(1), 33–48.

Chintagunta, P., E. Kyriazidou, and J. Perktold (2001): "Panel data analysis of household brand choices," *Journal of Econometrics*, 103(1), 111–153.

Heckman, J. J. (1981a): "Heterogeneity and state dependence," in *Studies in labor markets*. University of Chicago Press.

Honoré, B. E. (1992): "Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects," *Econometrica*, 60(3), 533–65.

——— (2002): "Nonlinear models with panel data," *Portuguese Economic Journal*, 1(2), 163–179.

Honoré, B. E., and E. Kyriazidou (2000): "Panel data discrete choice models with lagged dependent variables," *Econometrica*, 68(4), 839–874.

Honoré, B. E., and A. Lewbel (2002): "Semiparametric binary choice panel data models without strictly exogeneous regressors," *Econometrica*, 70(5), 2053–2063.

Honoré, B. E., and E. Tamer (2006): "Bounds on parameters in panel dynamic discrete choice models," *Econometrica*, 74(3), 611–629.

Manski, C. (1987): "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica*, 55(2), 357–362.

Muris, C. (2017): "Estimation in the fixed-effects ordered logit model," *Review of Economics and Statistics*, 99(3), 465–477.

Pakes, A., and J. Porter (2014): "Moment Inequalities for Semiparametric Multinomial Choice with Fixed Effects," Discussion paper.

Rosen, A. M. (2012): "Set identification via quantile restrictions in short panels," *Journal of Econometrics*, 166(1), 127–137.

Rosen, A. M., and M. Weidner (2013,WP): "Bounds for Semiparametric Binary Panel Data Models," *in preparation, slides available at https://goo.gl/dVUW3V*.

Ruud, P. (1983): "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecifications of Distribution in Multinomial Discrete Choice Models," *Econometrica*, 51(1), 225–28.

Shi, X., M. Shum, and W. Song (2017): "Estimating semi-parametric panel multinomial choice models using cyclic monotonicity," *Forthcoming in Econometrica*.

Wooldridge, J. M. (2005): "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," *Journal of Applied Econometrics*, 20(1), 39–54.