Semiparametric identification in panel data discrete response models

Aristodemou, E.

DOI
10.1016/j.jeconom.2020.04.002

Publication date
2021

Document Version
Final published version

Published in
Journal of Econometrics

Journal of Econometrics 220 (2021) 253–271

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Semiparametric identification in panel data discrete response models

Eleni Aristodemou *

University of Amsterdam and Tinbergen Institute, The Netherlands

A R T I C L E I N F O

Article history:
Available online 5 May 2020

JEL classification:
C01
C33
C35

Keywords:
Static and dynamic panel data
Binary response models
Ordered response models
Semiparametric identification
Partial identification

A B S T R A C T

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 identification bounds change as the support of the explanatory variables varies.

1. Introduction

This paper provides new results on semiparametric identification in fixed effects linear index binary response dynamic panel data models and ordered response static panel data models. Under mild distributional assumptions on the fixed effect and the time-varying unobservables point-identification fails, but informative identification bounds on 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, 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 behavior 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 models with continuous dependent variables, 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 allowed to be correlated with the explanatory variables, point-identification of the regression parameters can be achieved by differencing out this fixed effect.

* Correspondence to: Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB, Amsterdam, The Netherlands.
E-mail address: e.aristodemou@uva.nl.

https://doi.org/10.1016/j.jeconom.2020.04.002

In non-linear panel data models with additively separable fixed effects, the differencing approach cannot be directly implemented. Identification and estimation in these models rely heavily on the assumptions placed 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. 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 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. The main focus is on eliminating the unrestricted fixed effect, which is allowed to be correlated with the explanatory variables, without imposing distributional assumptions on the time-varying unobservables. Partial identification is achieved by finding features of the distribution that are independent of the fixed effect. Rosen and Weidner (2013,WP), thereafter RW2013, took such an approach for deriving bounds in the static binary outcome model. This paper analyzes dynamic binary response models, where last period's choice directly enters current period's decision rule. Conditioning on the initial condition, identification relies on individuals who switch choices 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 partial identification of the coefficients of the contemporaneous explanatory variables and 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. This paper provides partial identification results without imposing the logistic distributional assumption, or relying on the existence of special regressors or regressors with unbounded support.

This paper falls within the general category of papers studying semiparametric and partial identification in panel data models. For example, Chernozhukov et al. (2005) focus on nonparametric bound analysis in multinomial panel data models with correlated random effects, while Chernozhukov et al. (2013) provide sharp identification sets 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.

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 allow for partial identification of the parameters of interest, 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 behavior 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 on 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 et al. (2018) 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 et al. (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. This paper departs from these approaches, by relaxing the logistic distribution assumption and using the complete structure of the ordered choice model.

The rest of the paper is structured as follows. Section 2 examines identification in the dynamic binary response model. 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 includes some general discussion and Section 6 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 traveling 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-invariant unobserved heterogeneity, as in 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 the individual can choose one option from the set $\mathcal{Y}_t = \{0, 1\}$. 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 $\alpha \in \mathbb{R}$. Define by $1(\cdot)$ the indicator function which equals to 1 if $(\cdot)$ is true and 0 otherwise, then the static panel data binary response model is given by,

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

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-varying 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 in addition to other factors, such as potential income and years of education. 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,

$$Y_t = 1(X_t\beta + Y_{t-1}\gamma + \alpha + V_t > 0) \tag{2}$$

In this model each individual is observed for three periods, $t = \{0, 1, 2\}$, and is characterized by a set of observables $Y_0, Y = (Y_1, Y_2), X = (X_1, X_2)$, and a set of unobservables (V, α), where $V = (V_1, V_2)$ and $\alpha \in \mathbb{R}$. Like Honoré and Kyriazidou (2000), the parameter γ measures the "impact" of choosing option $Y = 1$ in period $t - 1$, or the true state dependence parameter, for example the effect of being employed in period $t - 1$, and β can be interpreted as the effect of other personal characteristics on the employment decision. 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 $t = 0$.[1] 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 (*Random Sampling*). 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 $\mathcal{A} \subseteq \mathbb{R}$.

[1] Period $t = 0$ denotes 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.

Assumption 2 (*Conditional Independence*)**.** X and V are stochastically independent conditional on Y_0, i.e. $V \perp X|Y_0$.

Assumption 1 defines the underlying probability space and notation for the support of the random variables (Y, Y_0, X, V, α). Under this assumption the data comprise a random sample and therefore the conditional distribution $P(y_1, y_2|x, y_0) \equiv P(Y_1 = y_1 \wedge Y_2 = y_2|X = x, Y_0 = y_0)$ 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 2 imposes only conditional independence between X and V conditional on Y_0,[2] 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). α is allowed to be correlated with both V and X in an arbitrary way. The set of all possible (conditional) distributions of V, α given X, Y_0 is $\mathcal{F}_{(V,\alpha)|X,Y_0}$ and the set of all possible (conditional) distributions of ΔV given X, Y_0, where $\Delta V = V_2 - V_1$, is $\mathcal{F}_{\Delta V|X,Y_0}$.

The potential correlation of the time-invariant unobservable with the explanatory variables, creates an additional endogeneity problem, that needs to be addressed for identification and consistent estimation of the parameters of interest.[3] In linear panel data models with continuous outcomes differencing out the fixed effect guarantees point-identification of the regression parameters, under some sufficient conditions. This paper mimics the approach used in linear panel data models with continuous outcomes to solve the problem of the fixed effect.

Finally, define a structure $S \equiv (\beta, \gamma, F_{(V,\alpha)|X,Y_0})$ as a specified collection of parameters β and γ, and joint distributions of the time-varying unobservables and the unobserved heterogeneity, $F_{(V,\alpha)|X,Y_0}$. The set of admissible structures is thus defined in Assumption 3.

Assumption 3 (*Admissible Structures*)**.** The structure S admitted by the model belongs to a collection $\mathcal{S}$ of parameters β and γ belonging to a parameter space Θ and joint distributions of the time-varying unobservables and the unobserved heterogeneity, $F_{(V,\alpha)|X,Y_0} \in \mathcal{F}_{(V,\alpha)|X,Y_0}$.

Following Assumptions 1–3, the identified set of admissible structures, denoted by $\mathcal{S}^0$, is characterized by,

$$\mathcal{S}^0 = \left\{ \begin{array}{c} (\beta, \gamma, F_{(V,\alpha)|X,Y_0}) \in \mathcal{S} : \forall (y_1, y_2) \in (\mathcal{Y}_1 \times \mathcal{Y}_2), \\ F_{(V,\alpha)|X,Y_0}\left(\mathcal{R}^{DB}_{(y_1,y_2)}(x, y_0; \beta, \gamma)\right) = P(y_1, y_2|x, y_0) \\ a.e.\ x \in \mathcal{X} \text{ and } y_0 \in \mathcal{Y}_0 \end{array} \right\} \tag{3}$$

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

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

where following the definition of the dynamic binary response model (*DB*) in (2), $\mathcal{R}^{DB}_{(y_1,y_2)}(x, y_0; \beta, \gamma)$ are defined as the regions of unobservables (V, α) that partition the *supp*(V, α) such that for all $(V, \alpha) \in (\mathcal{V}, \mathcal{A})$, $(Y_1, Y_2) = (y_1, y_2)$ when $X = x$ and $Y_0 = y_0$,

$$\mathcal{R}^{DB}_{(y_1,y_2)}(x, y_0; \beta, \gamma) = \{(V, \alpha) \in (\mathcal{V}, \mathcal{A}) : y_t = 1(x_t\beta + y_{t-1}\gamma + \alpha + V_t > 0), t = 1, 2\} \tag{5}$$

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 their choice from period $t = 1$ and $t = 2$, as shown in Honoré and Kyriazidou (2000). This 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 logistic distributional assumption, finding features of the distribution that do not depend on the unobserved heterogeneity leads to partial identification of the regression parameters.

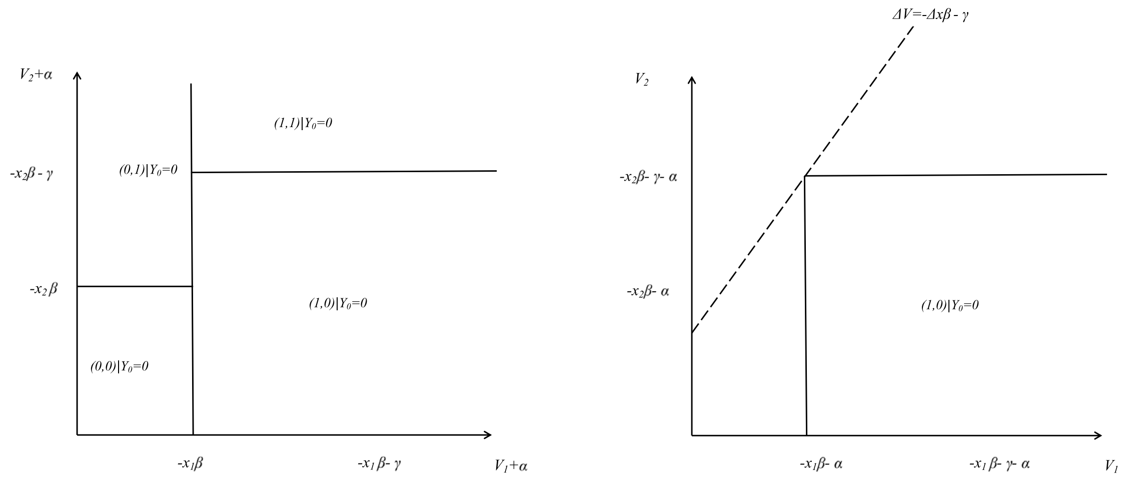
2.2. Identification bounds: Binary response dynamic panel data model

Identification of the parameters of interest (β, γ) in model (2) 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 $t = 0$.

From the regions in (5) and Fig. 1(a), it can be seen that the model in (2) is complete and coherent in the sense that conditioning on any value of the explanatory variables and the initial condition, for every (V, α) the model predicts a unique (y_1, y_2) outcome with probability one and identification bounds of the form of (3) and (4) can be derived. Since

[2] For example, suppose that Y is an employment indicator and (X_1, X_2) is potential income in periods 1 and 2 and (V_1, V_2) contains the (unobserved) family size in periods 1 and 2. The employment status in period $t = 0$, Y_0, affects both potential income and family size. Therefore, potential income and family size will be correlated through Y_0, but uncorrelated conditional on Y_0.

[3] In dynamic panel data models there is an endogeneity problem by construction since α is correlated with the lagged dependent variable, y_{t-1}.



(a) Regions defined in equations (5) (b) Upper Bound for $(1,0)|Y_0=0$

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

the outcome in period $t = 0$ directly affects the outcome in period $t = 1$, all three periods $t = \{0, 1, 2\}$ are used. No restriction is imposed 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 namely,

$$\begin{aligned} A &= \{Y_0 = 0 \wedge Y_1 = 0 \wedge Y_2 = 1\} \\ B &= \{Y_0 = 0 \wedge Y_1 = 1 \wedge Y_2 = 0\} \\ C &= \{Y_0 = 1 \wedge Y_1 = 0 \wedge Y_2 = 1\} \\ D &= \{Y_0 = 1 \wedge Y_1 = 1 \wedge Y_2 = 0\} \end{aligned} \tag{6}$$

Theorem 1. *Let $S^{DB} = (\beta^{DB}, \gamma^{DB}, F^{DB}_{(V,\alpha)|X,Y_0})$ be a structure admitted by model (2) such that $S^{DB} \in \mathcal{S}^0$, then under Assumptions 1–3, $(\beta^{DB}, \gamma^{DB})$ 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 $\Delta X = X_2 - X_1$ and $\Delta V = V_2 - V_1$.

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.[4] The events $\{Y_1 = 0 \wedge Y_2 = 0\}$ and $\{Y_1 = 1 \wedge Y_2 = 1\}$ provide no restrictions on ΔV and cannot 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 cannot be used to identify the regression parameters, since for any value of (β, γ) 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 any given value of the regression parameters. Notice that in order for the bounds in Theorem 1 to be informative, there should exist $x \in \mathcal{X}$ such that $x_1 \neq x_2$ with positive probability.

Fig. 1(a) and (b) plot the regions of unobservables conditional on $Y_0 = 0$ and $\gamma < 0$ given in equations (5) and provide an outline of the main idea. 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. Fig. 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 $P(1, 0|x, 0) \leq P_{\Delta V|Y_0}[\Delta V < -\Delta x\beta - \gamma | Y_0 = 0]$.

[4] 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 2.

Theorem 2. *Following Theorem 1, under Assumptions 1–3 the bounds on β, γ are given by,*

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

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

Notice that unlike the Honoré and Kyriazidou (2000) result, where point-identification in the dynamic binary panel data model is achieved if the errors are logistically distributed and 4 time periods are observed, the identification bounds in Theorem 2 only require 3 time periods. Furthermore, the bounds in Theorem 2 may not be sharp. This is in contrast to the static case in RW2013 in which sharpness was proved. The presence of the lagged dependent variable complicates the analysis, therefore the focus of the current paper is only on deriving identification bounds on (β, γ).

Finally, by replacing Assumption 2 of $X \perp V|Y_0$, with the unconditional independence assumption, $X \perp V$, the identification bounds on β and γ can be expressed in terms of the unconditional probabilities, given in Theorem 3.

Theorem 3. *Let Assumptions 1, $X \perp V$ and 3 hold. Then the (unconditional) identification bounds on (β, γ) are 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. □

2.3. Dynamic panel data model with more periods

In Section 2, informative bounds on the parameters (β, γ), in the situation where each individual is observed for three periods, $t = \{0, 1, 2\}$, were derived. The results can be extended to more periods, where each individual is observed for T periods and is characterized by the set of observables $Y_0, \tilde{Y} = (Y_1, \ldots, Y_T), \tilde{X} = (X_1, \ldots, X_T)$, and a set of unobservables $(\tilde{V}, \alpha)$, where $\tilde{V} = (V_1, \ldots, V_T)$ and $\alpha \in \mathbb{R}$. Define by $V = (V_t, V_{t+1})$, $\Delta V = V_{t+1} - V_t$ and by $X = (X_t, X_{t+1})$. Under the random sampling imposed by Assumption 1 $P(y_t, y_{t+1}|x, y_{t-1}) \equiv P(Y_t = y_t \wedge Y_{t+1} = y_{t+1}|X = x, Y_{t-1} = y_{t-1})$ is point-identified over the support of (Y_t, Y_{t+1}) for almost every $x \in \mathcal{X}$ and $y_{t-1} \in \mathcal{Y}_{t-1}$. By extending Assumption 2 of $(X_1, X_2) \perp (V_1, V_2)|Y_0$ to $V \perp X|Y_{t-1}$ for all $t = \{1, \ldots, T\}$ the identification bounds on (β, γ) can be characterized as in Theorem 4.

Theorem 4. *Let Assumptions 1, 3 and the extended version of Assumption 2 hold, then the bounds on β, γ are given by the set $\Theta^{DB} = \bigcap_{1 \leq t \leq T-1} \Theta_t^{DB}$, where:*

$$\Theta_t^{DB} = \left\{ \begin{array}{c} \sup_{x: -\Delta x\beta - \gamma \leq \omega} P(1,0|x,Y_{t-1} = 0) \leq \inf_{x: -\Delta x\beta \geq \omega} 1 - P(0,1|x,Y_{t-1} = 0) \\ \text{and} \\ \sup_{x: -\Delta x\beta \leq \omega} P(1,0|x,Y_{t-1} = 1) \leq \inf_{x: -\Delta x\beta + \gamma \geq \omega} 1 - P(0,1|x,Y_{t-1} = 1) \end{array} \right\}$$

Proof. The proof is an extension of the proofs of Theorems 1 and 2 and is omitted. □

Increasing the number of time periods can lead to tighter identification bounds. However, if the conditional distribution of $(Y_t, Y_{t+1}|X_t, X_{t+1}, Y_{t-1})$ is stationary, i.e. invariant with respect to the choice of $(t, t+1)$, this will not result in tighter bounds.[5]

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, including for example Chintagunta et al. (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 identification bounds, since the shape restrictions imposed by the ordering specification, reduce the between alternatives comparisons needed to determine the optimal choice.

This section extends the model in (1) 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 $Y = 2$ and $Y = 0$ denotes the outside option, that corresponds to not choosing any of the available alternatives. The static panel data ordered response model for each individual can be expressed as,

$$Y_t = \begin{cases} 0 \text{ if } \; X_t\beta + \alpha + V_t \leq c_1 \\ 1 \text{ if } \; c_1 < X_t\beta + \alpha + V_t \leq c_2 \\ 2 \text{ if } \; c_2 < X_t\beta + \alpha + V_t \end{cases} \tag{7}$$

where X_t are observed individual characteristics, α is the unobserved time-invariant individual heterogeneity, V_t is the time-varying unobserved component and $c = (c_1, c_2) \in \mathcal{C}$ are the (unknown) threshold parameters in the ordered response model, such that $\mathcal{C} \subseteq \mathbb{R}^2$ and $c_2 > c_1$.[6] Furthermore, it is clear that the model in (7) is observationally equivalent to the model with, $\tilde{c}_1 = 0$, $\tilde{\alpha} = \alpha - c_1$ and $\tilde{c}_2 = c_2 - c_1$, therefore c_1 is normalized to zero.

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 with logistically distributed unobservables. This paper departs from this approach and uses the ordered structure of the model to characterize the identification bounds, without imposing distributional assumptions 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 on the regression parameters. Section 3.1 formalizes the assumptions imposed on model (7).

3.1. Model assumptions

Assumption 4 (*Random Sampling*)**.** 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 5 (*Independence*)**.** X and V are stochastically independent.

Assumption 4 defines the underlying probability space and notation for the support of the random variables (Y, X, V, α). Under this assumption the data comprise a random sample and therefore the conditional distribution $P(y_1, y_2|x) \equiv P(Y_1 = y_1 \wedge Y_2 = y_2|X = x)$ is point-identified over the support of (Y_1, Y_2) for almost every $x \in \mathcal{X}$. Assumption 5 imposes independence of X and V,[7] but allows α to be arbitrary correlated with both V and X. The set of all possible (conditional) distributions of V, α given X is $\mathcal{F}_{(V,\alpha)|X}$ and the set of all possible (conditional) distributions of ΔV given X, where $\Delta V = V_2 - V_1$, is $\mathcal{F}_{\Delta V|X}$.

Finally define a structure $S \equiv (\beta, c_2, F_{(V,\alpha)|X})$ as a specified collection of parameters β and c_2, and joint distributions of the time-varying unobservables and the unobserved heterogeneity, $F_{(V,\alpha)|X}$. The set of admissible structures is thus defined in Assumption 6.

---

5 I would like to thank an anonymous referee for pointing this out.

6 The threshold parameters are assumed to be constant over time as in many applications of panel data ordered response models. 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.

7 This is the strict exogeneity assumption imposed in many panel data settings. The model can be applied to settings of vertically differentiated alternatives where higher values of Y correspond to better quality products. In this setting the regressors might be consumer characteristics which are uncorrelated with demand shocks.

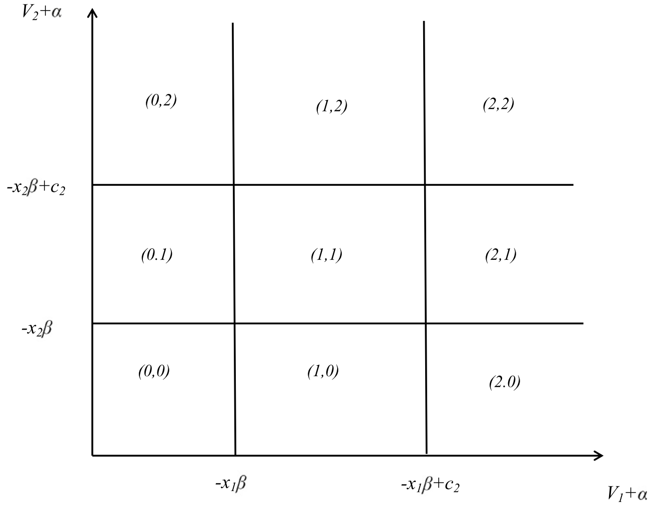


Fig. 2. Regions of unobservables for each (Y_1, Y_2) choice.

Assumption 6 (*Admissible Structures*)**.** The structure S admitted by the model belongs to a collection $\mathcal{S}$ of parameters β and c_2 belonging to a parameter space Θ and joint distributions of the time-varying unobservables and the unobserved heterogeneity, $F_{(V,\alpha)|X} \in \mathcal{F}_{(V,\alpha)|X}$.

Under Assumptions 4–6 the identified set of admissible structures, denoted by $\mathcal{S}^0$, is characterized by,

$$\mathcal{S}^0 = \left\{ \begin{array}{c} (\beta, c_2, F_{(V,\alpha)|X}) \in S : \forall (y_1, y_2) \in (\mathcal{Y}_1 \times \mathcal{Y}_2) \\ F_{(V,\alpha)|X}(\mathcal{R}^{SO}_{(y_1,y_2)}(x; \beta, c_2)) = P(y_1, y_2|x) \\ a.e.\ x \in \mathcal{X} \end{array} \right\}$$

and the identified set for the model parameters (β, c_2) is given by,

$$\Theta^0 = \left\{ \begin{array}{c} (\beta, c_2) \in \Theta : \exists F_{(V,\alpha)|X} \in \mathcal{F}_{(V,\alpha)|X}, \forall (y_1, y_2) \in (\mathcal{Y}_1, \mathcal{Y}_2), \\ F_{(V,\alpha)|X}(\mathcal{R}^{SO}_{(y_1,y_2)}(x; \beta, c_2)) = P(y_1, y_2|x) \\ a.e.\ x \in \mathcal{X} \end{array} \right\}$$

where following the definition of the static ordered response model (SO) in (7), $\mathcal{R}^{SO}_{(y_1,y_2)}(x; \beta, c_2)$ are defined as the regions of unobservables (V, α) that partition the support of (V, α) such that $(Y_1, Y_2) = (y_1, y_2)$ when $X = x$.

3.2. Identification bounds: Ordered response static panel data model

When no assumptions are imposed on the fixed effect, identification bounds on the regression parameters, (β, c_2), are derived by finding features of the distribution that do not depend on α. Fig. 2 plots the regions $\mathcal{R}^{SO}_{(y_1,y_2)}(x; \beta, c_2)$ for any fixed x. The model in (7) is complete and coherent in the sense that conditional on any value of the explanatory variables $x \in \mathcal{X}$, for every (V, α) the model predicts a unique (y_1, y_2) outcome with probability one.

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, c_2), without imposing any assumptions on the fixed effect. The transitions that are informative are therefore,

$$\{Y_1 = i \wedge Y_2 = j\}, \forall i, j = 0, 1, 2 \text{ and } i \neq j. \tag{8}$$

In addition, 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,

$$\{Y_1 = 1 \wedge Y_2 = 1\}. \tag{9}$$

Theorem 5. *Let $S^{SO} = \left(\beta^{SO}, c_2^{SO}, F^{SO}_{(V,\alpha)|X}\right)$ be a structure admitted by model (7). Under Assumptions 4–6, if $S^{SO} \in \mathcal{S}^0$ then, for any $x \in \mathcal{X}$, (β^{SO}, c_2^{SO}) satisfy,*

$$P(1, 0|x) \leq F_{\Delta V}[-\Delta x \beta^{SO}]$$
$$F_{\Delta V}[-\Delta x \beta^{SO}] \leq 1 - P(0, 1|x)$$

$$
\begin{aligned}
P(2,0|x) &\leq F_{\Delta V}[-\Delta x\beta^{SO} - c_2^{SO}] \\
&\qquad F_{\Delta V}[-\Delta x\beta^{SO} + c_2^{SO}] \leq 1 - P(0,2|x) \\
P(2,1|x) &\leq F_{\Delta V}[-\Delta x\beta^{SO}] \\
&\qquad F_{\Delta V}[-\Delta x\beta^{SO}] \leq 1 - P(1,2|x) \\
P(1,1|x) &\leq P_{\Delta V}[-\Delta x\beta^{SO} + c_2^{SO} > \Delta V > -\Delta x\beta^{SO} - c_2^{SO}]
\end{aligned}
$$

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

Proof. The proof is provided in Appendix A.4. □

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. As it can also be seen from Fig. 2, the choice $(Y_1, Y_2) = (1,1)$ provides restrictions on the possible values the fixed effect can take for each value of (β, c_2), and hence can be used to identify the regression coefficients. Similar to the dynamic binary model in Section 2, in the ordered model the events $(0,0)$ and $(2,2)$ give no information on β and c_2 since these cases can be matched by extremely small or extremely large values of α regardless of the value of (β, c_2).

Theorem 5 has two important implications. It is known that each ordered response model can be expressed as a binary response model. Consider the situation where the model in (7) is re-expressed as a binary response model where the choices $(Y_t = 1, Y_t = 2)$ were merged together. Then, following the same identification strategy, Theorem 5 would only include $P(1,0|x) \leq F_{\Delta V}[-\Delta x\beta] < 1 - P(0,1|x)$. It is thus evident that the set of inequalities defining the identification bounds in the ordered response contains the set of inequalities defining the identification bounds in the binary response representation. Furthermore, in the binary response model identifying restrictions involve only individuals who switch. In the ordered response model non-switchers also provide information which might prove helpful when comparing the behavior of switchers to non-switchers. Theorem 6 formalizes the identification bounds on (β, c_2).

Theorem 6. *Let Assumptions 4–6 hold. Using the definitions in (10), the identification bounds on* (β, c_2) *are given by the set:*

$$
\Theta^{SO} = \left\{ \begin{array}{c} (\beta, c_2) \in \Theta : \forall \omega \in \mathbb{R}, \\ \max[s_{(1,0)}(\omega), s_{(2,0)}(\omega), s_{(2,1)}(\omega), s_{(1,1)}(\omega)] \leq \min[i_{(0,1)}(\omega), i_{(0,2)}(\omega), i_{(1,2)}(\omega), i_{(1,1)}(\omega)] \end{array} \right\}
$$

where,

$$
\begin{aligned}
s_{(1,0)}(\omega) &= \sup_{x:-\Delta x\beta \leq \omega} P(1,0|x) \\
s_{(2,0)}(\omega) &= \sup_{x:-\Delta x\beta - c_2 \leq \omega} P(2,0|x) \\
s_{(2,1)}(\omega) &= \sup_{x:-\Delta x\beta \leq \omega} P(2,1|x) \\
s_{(1,1)}(\omega) &= \sup_{x:-\Delta x\beta + c_2 \leq \omega} P(1,1|x) \\
i_{(0,1)}(\omega) &= \inf_{x:-\Delta x\beta \geq \omega} [1 - P(0,1|x)] \\
i_{(0,2)}(\omega) &= \inf_{x:-\Delta x\beta + c_2 \geq \omega} [1 - P(0,2|x)] \\
i_{(1,2)}(\omega) &= \inf_{x:-\Delta x\beta \geq \omega} [1 - P(1,2|x)] \\
i_{(1,1)}(\omega) &= \inf_{x:-\Delta x\beta - c_2 \geq \omega} [1 - P(1,1|x)]
\end{aligned}
\tag{10}
$$

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

Similarly to Theorem 2, the identification bounds on (β, c_2) in Theorem 6 might not be sharp.

3.3. Static ordered response panel data model with more choices

Consider now the extension of model (7) where $\mathcal{Y}_t = \{0, \ldots, K\}$ and $c = (c_1, c_2, \ldots, c_K)$, with $c_1 < c_2 < \cdots < c_K$ and c_1 normalized to zero, such that,

$$
Y_t = \begin{cases} 0 \text{ if } \; X_t\beta + \alpha + V_t \leq 0 \\ 1 \text{ if } \; 0 < X_t\beta + \alpha + V_t \leq c_2 \\ \ldots \\ K \text{ if } \; c_K < X_t\beta + \alpha + V_t \end{cases} \tag{11}
$$

In this case any combination (Y_t, Y_{t+1}) except $(Y_t, Y_{t+1}) = (0,0)$ and $(Y_t, Y_{t+1}) = (K,K)$, provides information on $(\beta, c_2, \ldots, c_K)$. The increased number of inequalities might result in tighter bounds.

If model (7) or model (11) is extended to more periods $t \in \{1, \ldots, T\}$, then the bounds on (β, c) might also become tighter than in the case of two periods, unless the conditional distribution of $(Y_t, Y_{t+1}|X_t, X_{t+1})$ is stationary.

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})		
	$\{-1, 0, 1\}$	$\{-2, -1, 0, 1, 2\}$	$\{-3, -2, -1, 0, 1, 2, 3\}$
$V_t \overset{iid}{\sim} N(0, 1)$	$(0, \infty)$	$(0.5, 2)$	$(0.667, 1.5)$
$V_t \overset{iid}{\sim} N\left(0, \frac{\pi^2}{3}\right)$	$(-0.5, \infty)$	$(0.286, 2.667)$	$(0.545, 1.714)$
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})		
	$\{-1, 0, 1, 2, 3\}$	$\{-2, -1, 0, 1, 2, 3, 4\}$	$\{-4, -3, -2, -1, 0, 1, 2\}$
$V_t \overset{iid}{\sim} N(0, 1)$	$(0.5, 6)$	$(0.667, 2)$	$(0.667, 2)$
$V_t \overset{iid}{\sim} N\left(0, \frac{\pi^2}{3}\right)$	$(0.286, \infty)$	$(0.545, 2.5)$	$(0.545, 2.5)$
Logistic	$(0.286, \infty)$	$(0.545, 2.5)$	$(0.545, 2.5)$

4. Numerical examples

This section provides numerical illustrations of the identification bounds 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 response panel data model,

$$Y_t = 1(X_t\beta + \alpha + V_t > 0) \tag{12}$$

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 value of $\beta_2 = 1$ after the normalization of $\beta_1 = 1$.[8]

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

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 though 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 give similar identified sets. As discussed in Amemiya (1981) and Maddala (1983) the two distributions are very close to each other except at the tails. Since the observable implications used for identification provide restrictions on the realization of ΔV, extreme values of ΔV are unlikely, making the identified sets indistinguishable.

4.1.2. Example 2

Consider the two period static binary response panel data model as in (12),

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

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 values of $\beta_2 = 1$ and $\beta_3 = 1$ after normalizing $\beta_1 = 1$. Fig. 3 provides the joint identified sets for (β_2, β_3) as the support of (X_{1t}, X_{2t}) changes.[9] Similarly to Example 1, the sets shrink as the support of the discrete explanatory variables (X_{1t}, X_{2t}) increases, even if the support of X_3 remains fixed.

[8] 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.

[9] The identified sets were constructed for values of β_3 in the grid [−1,3].

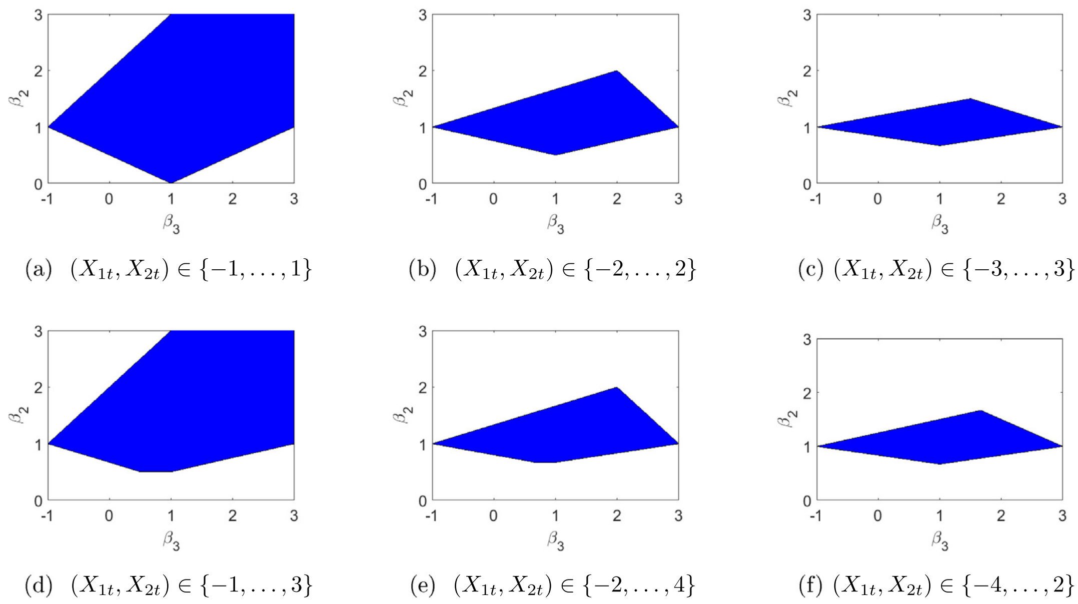


Fig. 3. Joint identified sets for (β_2, β_3) when $X_{3t} \in \{0, 1\}$.

Table 3
Identification bounds on β_2 under Probit and Logit specifications with symmetric support for (X_{1t}, X_{2t}) around zero and known $\gamma = 0.5$.

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

Table 4
Identification bounds on β_2 under Probit and Logit specifications with asymmetric support for (X_{1t}, X_{2t}) around zero and known $\gamma = 0.5$.

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

Table 5
Identification bounds on β_2 with symmetric support and asymmetric support for (X_{1t}, X_{2t}) around zero and known $\beta_3 = 0.5$ when $X_{3t} \in \{0, 1\}$.

	Support of (X_{1t}, X_{2t})		
	$\{-1, 0, 1\}$	$\{-2, -1, 0, 1, 2\}$	$\{-3, -2, -1, 0, 1, 2, 3\}$
$V_t \overset{iid}{\sim} N(0, 1)$	$(0, 5)$	$(0.5, 1.8)$	$(0.667, 1.444)$
	Support of (X_{1t}, X_{2t})		
	$\{-1, 0, 1, 2, 3\}$	$\{-2, -1, 0, 1, 2, 3, 4\}$	$\{-4, -3, -2, -1, 0, 1, 2\}$
$V_t \overset{iid}{\sim} N(0, 1)$	$(0.5, 5)$	$(0.667, 1.8)$	$(0.667, 1.667)$

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,

$$Y_t = 1(X_t\beta + Y_{t-1}\gamma + \alpha + V_t > 0) \tag{13}$$

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 values of $\beta_2 = 1$ and $\gamma = 0.5$ after normalizing $\beta_1 = 1$.

Tables 3 and 4 provide the identification bounds on β_2 as described in Theorem 2, when $\gamma = 0.5$ is known, under different specifications 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 known γ, the identification bounds on β_2 shrink as the support of the explanatory variables increases. Furthermore, the identification bounds 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 identification bounds.

Consider now the static model discussed in Section 4.1. An interesting comparison would be to compare the identification bounds for the dynamic binary model with $(X_{1t}, X_{2t}, Y_{t-1})$ with the static model with strictly exogenous regressors (X_{1t}, X_{2t}, X_{3t}).[10] Table 5 provides the identification bounds on β_2, when $X_3 = \{0, 1\}$ and $\beta_3 = 0.5$ and $V_t|X, \alpha \overset{iid}{\sim} N(0, 1)$. It is clear that in the case of binary exogenous X_3 and $\beta_3 = 0.5$ the identification bounds on β_2 are smaller in comparison to the dynamic model with a binary Y_{t-1} and $\gamma = 0.5$ given in Tables 3 and 4. This is a consequence of the stronger assumptions in the static binary model, where X_3 is restricted to be exogenous.

4.2.2. Example 4

Consider the binary response dynamic panel data model in (13) with $V_t|X, Y_0, \alpha \overset{iid}{\sim} N(0, 1)$, but assume no knowledge of β_2 or γ. Fig. 4 provides the joint identification bounds on (β_2, γ) as described in Theorem 2, when the true values of $\beta_2 = 1$ and $\gamma = 0.5$, as the support of (X_{1t}, X_{2t}) changes.[11]

It is clear that for the specific range of values for the grid of γ chosen, the bounds on β_2 shrink as the support of the discrete explanatory variables (X_{1t}, X_{2t}) increases, however with three periods γ is only bounded from above. In an earlier paper, Honoré and Kyriazidou (2000) showed that in the dynamic binary response model with one lagged dependent variable and logistically distributed unobservables the parameters are point-identified with at least four time periods. This paper avoids making any parametric distributional assumptions, such as the logistic distribution, and draws the attention to the three period model, which results in partial identification.

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,

$$Y_t = \begin{cases} 0 \text{ if } \; X_t\beta + \alpha + V_t \le 0 \\ 1 \text{ if } \; 0 < X_t\beta + \alpha + V_t \le c_2 \\ 2 \text{ if } \; c_2 < X_t\beta + \alpha + V_t \end{cases} \tag{14}$$

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 $V_t|X, \alpha \overset{iid}{\sim} N(0, 1)$. The true value of $\beta_2 = 1$ after normalizing $\beta_1 = 1$. Tables 6 and 7 provide the identification bounds on β_2 described in Theorem 6 for different known values of c_2, as the support of the discrete explanatory variables (X_{1t}, X_{2t}) changes.[12] For any given c_2 it is evident that as the support of X_t increases the bounds shrink. Furthermore, it is evident that the value of the threshold parameter c_2 affects the size of the identification bounds. This indicates that in the ordered response model the threshold plays a crucial role in the identifying power of the model. In addition, the designs in this section are one-to-one compatible with the designs in Section 4.1 where $V_t \overset{iid}{\sim} N(0, 1)$, and it is clear that the bounds in the ordered response model are contained within the binary response model bounds.

4.3.2. Example 6

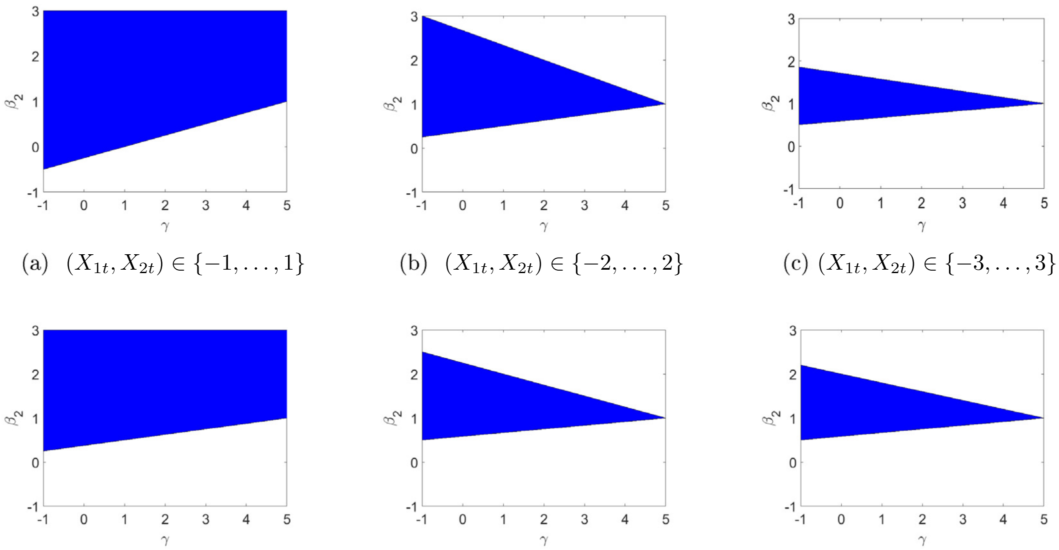
Consider the ordered response model in (14), with unknown β_2 and c_2. Fig. 5 provides the joint identification bounds on (β_2, c_2) when the true values of $\beta_2 = 1$ and $c_2 = 1.5$, as the support of (X_{1t}, X_{2t}) changes.[13] Similarly to the previous examples, the bounds shrink as the support of the explanatory variables increases.

10 I would like to thank an anonymous referee for this suggestion.

11 The identification bounds were constructed for values of γ in the grid $[-1, 6]$.

12 The identification bounds were constructed by setting 50 grid points for β_2 in the grid $[-1, 4]$.

13 The bounds were constructed by setting 50 points for β_2 in the grid [-1,4] and 20 points for c_2 in the grid [0,3].



(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) $(X_{1t}, X_{2t}) \in \{-4, \ldots, 2\}$

Fig. 4. Joint identification bounds on (β_2, γ).

Table 6
Identification bounds on β_2 under the Probit specification with symmetric support for (X_{1t}, X_{2t}) around zero and known values of c_2.

	Support of (X_{1t}, X_{2t})		
	$\{-1, 0, 1\}$	$\{-2, -1, 0, 1, 2\}$	$\{-3, -2, -1, 0, 1, 2, 3\}$
$c_2 = 0.5$	$(0.225, \infty)$[a]	(0.633, 1.551)	(0.735, 1.245)
$c_2 = 1.5$	(0.327, 2.674)	(0.633, 1.551)	(0.837, 1.245)
$c_2 = 2.5$	(0.327, 2.265)	(0.633, 1.551)	(0.837, 1.347)

[a]The upper bound for β_2 was recorded as ∞ whenever it reached the upper bound of the grid for β_2.

Table 7
Identification bounds on β_2 under the Probit specification with asymmetric support for (X_{1t}, X_{2t}) around zero and known values of c_2.

	Support of (X_{1t}, X_{2t})		
	$\{-1, 0, 1, 2, 3\}$	$\{-2, -1, 0, 1, 2, 3, 4\}$	$\{-4, -3, -2, -1, 0, 1, 2\}$
$c_2 = 0.5$	(0.633, 2.98)	(0.735, 1.551)	(0.735, 1.551)
$c_2 = 1.5$	(0.633, 1.653)	(0.837, 1.347)	(0.735, 1.551)
$c_2 = 2.5$	(0.633, 1.551)	(0.837, 1.245)	(0.735, 1.551)

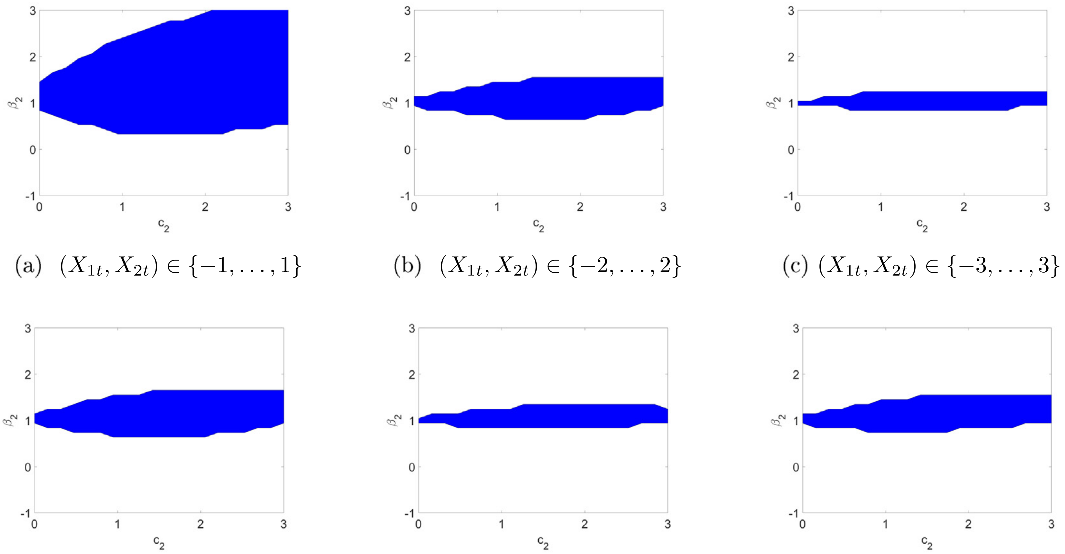
5. Discussion

Section 4 provides numerical illustrations of the models in Sections 2 and 3. An interesting discussion involves the elementary theoretical properties of the identification bounds, such as whether the identification regions are convex and bounded. These properties are important for estimation and inference.[14] First, consider the identification bounds defined in Theorem 2 for the dynamic binary response model. The identification bounds can be re-written as,[15]

$$\Theta^{DB} = \left\{ \begin{array}{c} (\beta, \gamma) \in \Theta : \\ \forall (x_i, x_j), -(\Delta x_i - \Delta x_j)\beta \leq \gamma \Rightarrow P(1, 0|x_i, 0) + P(0, 1|x_j, 0) \leq 1 \\ and \\ \forall (\tilde{x}_i, \tilde{x}_j), -(\Delta \tilde{x}_i - \Delta \tilde{x}_j)\beta \leq \gamma \Rightarrow P(1, 0|\tilde{x}_i, 1) + P(0, 1|\tilde{x}_j, 1) \leq 1 \end{array} \right\} \tag{15}$$

14 See for example Beresteanu and Molinari (2008). Ho and Rosen (2017) provide a discussion of how inference can be performed in more general settings.

15 This representation also appears in RW2013.



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

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

Fig. 5. Joint identification bounds on (β_2, c_2).

The computation of the identification bounds was carried out by noticing that the identification bounds in equation (15) imply the following,

$$\Theta^{DB} = \left\{ \begin{array}{c} (\beta, \gamma) \in \Theta : \\ \forall (x_i, x_j), P(1, 0|x_i, 0) + P(0, 1|x_j, 0) \geq 1 \Rightarrow -(\Delta x_i - \Delta x_j)\beta \geq \gamma \\ and \\ \forall (\tilde{x}_i, \tilde{x}_j), P(1, 0|\tilde{x}_i, 1) + P(0, 1|\tilde{x}_j, 1) \geq 1 \Rightarrow -(\Delta \tilde{x}_i - \Delta \tilde{x}_j)\beta \geq \gamma \end{array} \right\} \tag{16}$$

Since the last inequality takes the same form for both constraints, the computation of the identification bounds was carried out by choosing all the (x_i, x_j) such that either $P(1, 0|x_i, 0) + P(0, 1|x_j, 0) \geq 1$ or $P(1, 0|x_i, 1) + P(0, 1|x_j, 1) \geq 1$ was satisfied, and then finding the values of β on the grid values of γ that satisfied the constraint $-(\Delta x_i - \Delta x_j)\beta \geq \gamma$. Notice that regardless of the dimension of X, (16) implies

$$\Delta\Delta X\beta + \gamma \leq 0 \tag{17}$$

where $\Delta\Delta X = \Delta x_i - \Delta x_j$ and (x_i, x_j) satisfies $P(1, 0|x_i, 0) + P(0, 1|x_j, 0) \geq 1$ or $P(1, 0|x_i, 1) + P(0, 1|x_j, 1) \geq 1$. For every $\Delta\Delta X$, (17) defines a half-space in (β, γ)-space, so $(\beta^{DB}, \gamma^{DB})$ must lie in the intersection of these half-spaces, and the identifications bounds are convex.[16]

Whether the identification bounds are finite or not, subject to the normalization $\beta_1 = 1$, depends on the matrix formed by $\Delta\Delta X$ for the various (x_i, x_j). A necessary but not sufficient condition for boundedness is that the number of constraints in (17) is at least $dim(\beta) + 1$. This can be determined by linear programming.[16] For example, consider the case of a single β_2 after the normalization of $\beta_1 = 1$. The constraints in (17) can be shown to satisfy,

$$\begin{aligned} &when\ \Delta\Delta X_2 > 0,\ \frac{-\Delta\Delta X_1 - \gamma}{\Delta\Delta X_2} \geq \beta_2 \\ &when\ \Delta\Delta X_2 < 0,\ \frac{-\Delta\Delta X_1 - \gamma}{\Delta\Delta X_2} \leq \beta_2 \end{aligned} \tag{18}$$

where $\Delta\Delta X_1 = \Delta x_{1i} - \Delta x_{1j}$ and $\Delta\Delta X_2 = \Delta x_{2i} - \Delta x_{2j}$. If (X_i, X_j) varies enough to guarantee that $P(\Delta\Delta X_2 > 0) > 0$ and $P(\Delta\Delta X_2 < 0) > 0$ then for any fixed value of γ, β_2 is bounded.

Turning to the ordered response model, convexity and boundedness of the identification regions, derived in Theorem 6, cannot be shown by standard arguments. This is due to the presence of many inequalities, which make the representation of the identification bounds in the form of (16) not straight forward, as well as the presence of the *sup* and the *inf* functions which imply different sets of $\mathcal{X}(\beta, c_2)$ for different values of (β, c_2).

[16] I would like to thank an anonymous referee for pointing these simplifications and generalizations out.

6. Conclusion

This paper studies identification in discrete response panel data models with fixed effects. Under fairly mild conditions, informative identification bounds on 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. How tight and informative these bounds are, depends on the observed probability distribution as well as on the variation on the explanatory variables, with more observed implications and higher variation, leading (in general) to smaller bounds. Nevertheless, as also shown in the numerical examples, these identification bounds can be presumably wide or even infinite and the information content might be quite low especially when the support of the regressors is small. This is a direct effect of the weak assumptions imposed that cannot be necessarily strengthened in all circumstances. When the assumptions can be strengthened it might be possible to achieve tighter identification bounds, however this might come to an expense of the credibility of the conclusions.[17]

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. Identification in this paper relies on weaker conditions than in the point-identified cases. 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 the current paper is the joint probability of two outcome variables taking different values in two periods. These observable implications use less restrictive assumptions and in particular no information on the distribution of the time-varying unobservables. Therefore, even under the logistic distribution assumption, the bounds provided in this paper might still fail to be singletons. Furthermore, the papers proving point-identification under the assumption that the time-varying unobservables follow a logistic distribution, also impose independence of V and α. Assumptions 2 and 5 are less restrictive since V is allowed to be correlated with α.

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 bounds 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.

Acknowledgments

This paper is a revised version of Chapters 5 and 6 of my UCL PhD dissertation (2016) and subsequently Tinbergen Institute Discussion Paper TI 2018-065/III. I would like to thank the guest editors and two anonymous referees for their insightful suggestions and comments. I would also like to thank Andrew Chesher, Geert Dhaene, Bo Honoré, Frank Kleibergen, Lars Nesheim, Adam Rosen, Martin Weidner, Frank Windmeijer and several conference and seminar audiences for helpful discussions and comments. I gratefully acknowledge financial support from ESRC and UCL. Any errors are my own.

Appendix

A.1. Proof of Theorem 1

Consider event $A = \{Y_0 = 0 \wedge Y_1 = 0 \wedge Y_2 = 1\}$. The conditional probability for the event $(Y_1, Y_2) = (0, 1)$ conditional on $Y_0 = 0$ is,

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

Applying the same argument for the rest of the sequences of events in (6) implies that,

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

[17] As first introduced by Manski (2003) *The Law of Decreasing Credibility* states that "the credibility of inference decreases with the strength of the assumptions maintained".

For any fixed $X = x$ and by applying Assumption 2, 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

Consider any constant $\omega \in \mathbb{R}$, conditioning on $Y_0 = 0$ implies that,

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

and conditioning on $Y_0 = 1$ implies that,

$$\begin{aligned} &(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta + \gamma \geq \omega \Rightarrow \Delta V > \omega \\ &(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta \leq \omega \Rightarrow \Delta V < \omega \end{aligned} \tag{A.2}$$

The relations in (A.1) and (A.2) imply that, $\forall \omega \in \mathbb{R}$:

$$\begin{aligned} 1 - P[(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta \geq \omega | X = x, Y_0 = 0] &\geq P[\Delta V < \omega | X = x, Y_0 = 0] \\ P[(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta - \gamma \leq \omega | X = x, Y_0 = 0] &\leq P[\Delta V < \omega | X = x, Y_0 = 0] \\ 1 - P[(Y_1, Y_2) = (0,1) \wedge -\Delta X\beta + \gamma \geq \omega | X = x, Y_0 = 1] &\geq P[\Delta V < \omega | X = x, Y_0 = 1] \\ P[(Y_1, Y_2) = (1,0) \wedge -\Delta X\beta \leq \omega | X = x, Y_0 = 1] &\leq P[\Delta V < \omega | X = x, Y_0 = 1] \end{aligned} \tag{A.3}$$

Consider the event $A = \{Y_0 = 0 \wedge Y_1 = 0 \wedge Y_2 = 1\}$. When $-\Delta X\beta \geq \omega$,

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

and when $-\Delta X\beta < \omega$,

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

Using the same argument, lower and upper bounds on the rest of the inequalities in (A.3) can be derived. These bounds in combination with (A.1), (A.2) and Assumption 2 imply that,

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

$$\Longleftrightarrow$$

$$\begin{aligned} \sup_{x:-\Delta x\beta - \gamma \leq \omega} P(1,0|x,0) \leq F_{\Delta V|Y_0}(\omega|0) \leq \inf_{x:-\Delta x\beta \geq \omega} 1 - P(0,1|x,0) \\ \sup_{x:-\Delta x\beta \leq \omega} P(1,0|x,1) \leq F_{\Delta V|Y_0}(\omega|1) \leq \inf_{x:-\Delta x\beta + \gamma \geq \omega} 1 - P(0,1|x,1) \end{aligned} \tag{A.4}$$

which completes the proof.

A.3. Proof of Theorem 3

To prove the unconditional identification bounds 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, and define the relations in (A.3) as,

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

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

Then multiplying by $P_0(x)$ the probabilities conditional on $Y_0 = 0$ and by $P_1(x)$ the probabilities conditional on $Y_0 = 1$ imply that the relations in (A.3) can be expressed as

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

which imply,

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

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. Proof of Theorem 5

The proof of Theorem 5 follows similar arguments as the proof of Theorem 1. For example consider the event $\{Y_1 = 0 \wedge Y_2 = 1\}$. It can be shown that,

$$\begin{aligned}
P(0,1|x) &= P_{(V,\alpha)|X}[\{0 \geq X_1\beta + \alpha + V_1\} \\
&\quad \wedge \{X_2\beta + \alpha + V_2 > 0 \wedge 0 \geq X_2\beta + \alpha + V_2 - c_2\}|X=x] \\
&\leq P_{(V,\alpha)|X}(\{0 \geq X_1\beta + \alpha + V_1\} \wedge \{X_2\beta + \alpha + V_2 > 0\}|X=x) \\
&\leq P_{V|X}(0 > (X_1 - X_2)\beta + (V_1 - V_2)|X=x) \\
&= 1 - F_{\Delta V}[-\Delta x\beta]
\end{aligned}$$

where the last inequality follows from Assumption 5. Applying the same arguments for the rest of the relations in (8) it can be shown that for any given $X = x$,

$$\begin{aligned}
P(1,0|x) &\leq F_{\Delta V}[-\Delta x\beta] \\
&\quad F_{\Delta V}[-\Delta x\beta] \leq 1 - P(0,1|x) \\
P(2,0|x) &\leq F_{\Delta V}[-\Delta x\beta - c_2] \\
&\quad F_{\Delta V}[-\Delta x\beta + c_2] \leq 1 - P(0,2|x) \\
P(2,1|x) &\leq F_{\Delta V}[-\Delta x\beta] \\
&\quad F_{\Delta V}[-\Delta x\beta] \leq 1 - P(1,2|x)
\end{aligned} \tag{A.5}$$

The above relations show that changing choices from period $t = 1$ to $t = 2$ provide restrictions on the distribution of ΔV that do not depend on the fixed effect, α. In addition it can be shown that the "in-between" event, $(Y_1, Y_2) = (1, 1)$, also provides information on (β, c_2) without involving the fixed effect. To see that consider the joint probability of choosing the event (9):

$$\begin{aligned}
P(1,1|x) &= P_{(V,\alpha)|X}[\{X_1\beta + \alpha + V_1 > 0 \wedge 0 \geq X_1\beta + \alpha + V_1 - c_2\} \\
&\quad \wedge \{X_2\beta + \alpha + V_2 > 0 \wedge 0 \geq X_2\beta + \alpha + V_2 - c_2\}|X=x]
\end{aligned}$$

By combining

$$X_1\beta + \alpha + V_1 > 0 \textit{ and } 0 \geq X_2\beta + \alpha + V_2 - c_2$$

and

$$0 \geq X_1\beta + \alpha + V_1 - c_2 \textit{ and } X_2\beta + \alpha + V_2 > 0$$

it can be shown that,

$$\begin{aligned}
P(1,1|x) &\leq P_V[0 > (X_2 - X_1)\beta + (V_2 - V_1) - c_2 \\
&\quad \wedge (X_2 - X_1)\beta + (V_2 - V_1) + c_2 > 0|X=x] \\
P(1,1|x) &\leq P_{\Delta V}[-\Delta X\beta + c_2 > \Delta V > -\Delta X\beta - c_2|X=x]
\end{aligned} \tag{A.6}$$

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

A.5. Proof of Theorem 6

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 (A.5) and (A.6) and for a given set of arbitrary constants $\omega, \omega', \omega'' \in \mathbb{R}$, such that $\omega'' < \omega < \omega'$ it follows that,

$$
\begin{aligned}
&(Y_1, Y_2) = (0, 1) \wedge -\Delta X\beta \geq \omega \Rightarrow \Delta V > \omega \\
&(Y_1, Y_2) = (1, 0) \wedge -\Delta X\beta \leq \omega \Rightarrow \Delta V < \omega \\
&(Y_1, Y_2) = (0, 2) \wedge -\Delta X\beta + c_2 \geq \omega \Rightarrow \Delta V > \omega \\
&(Y_1, Y_2) = (2, 0) \wedge -\Delta X\beta - c_2 \leq \omega \Rightarrow \Delta V < \omega \\
&(Y_1, Y_2) = (1, 2) \wedge -\Delta X\beta \geq \omega \Rightarrow \Delta V > \omega \\
&(Y_1, Y_2) = (2, 1) \wedge -\Delta X\beta \leq \omega \Rightarrow \Delta V < \omega \\
&(Y_1, Y_2) = (1, 1) \wedge \omega' \geq -\Delta X\beta + c_2 \wedge -\Delta X\beta - c_2 \geq \omega \Rightarrow \omega' > \Delta V > \omega \\
&(Y_1, Y_2) = (1, 1) \wedge \omega \geq -\Delta X\beta + c_2 \wedge -\Delta X\beta - c_2 \geq \omega'' \Rightarrow \omega > \Delta V > \omega''
\end{aligned} \tag{A.7}
$$

The relations in (A.7) and Assumption 5 imply that,

$$
\begin{aligned}
1 - P[(Y_1, Y_2) &= (0, 1) \wedge -\Delta X\beta \geq \omega | X = x] \geq P[\Delta V < \omega] \\
P[(Y_1, Y_2) &= (1, 0) \wedge -\Delta X\beta \leq \omega | X = x] \leq P[\Delta V < \omega] \\
1 - P[(Y_1, Y_2) &= (0, 2) \wedge -\Delta X\beta + c_2 \geq \omega | X = x] \geq P[\Delta V < \omega] \\
P[(Y_1, Y_2) &= (2, 0) \wedge -\Delta X\beta - c_2 \leq \omega | X = x] \leq P[\Delta V < \omega] \\
1 - P[(Y_1, Y_2) &= (1, 2) \wedge -\Delta X\beta \geq \omega | X = x] \geq P[\Delta V < \omega] \\
P[(Y_1, Y_2) &= (2, 1) \wedge -\Delta X\beta \leq \omega | X = x] \leq P[\Delta V < \omega] \\
P[(Y_1, Y_2) &= (1, 1) \wedge \{\omega' \geq -\Delta X\beta + c_2 \wedge -\Delta X\beta - c_2 \geq \omega\} | X = x] \\
&\leq P[\omega < \Delta V < \omega'] \\
P[(Y_1, Y_2) &= (1, 1) \wedge \{\omega \geq -\Delta X\beta + c_2 \wedge -\Delta X\beta - c_2 \geq \omega''\} | X = x] \\
&\leq P[\omega'' < \Delta V < \omega]
\end{aligned} \tag{A.8}
$$

Following the same arguments as in proving Theorem 2 it can be shown that for any given $\omega'' < \omega < \omega' \in \mathbb{R}$, depending on the different values of $x \in \mathcal{X}$ the lower and upper bounds of the inequalities in (A.8) change. Since for any value of (β, c_2) in the identification bounds the relations in (A.8) should hold simultaneously the distribution of ΔV is shown to be bounded by,

$$
\begin{aligned}
\sup_{x: -\Delta x\beta \leq \omega} P(1, 0|x) &\leq F_{\Delta V}(\omega) \\
F_{\Delta V}(\omega) &\leq \inf_{x: -\Delta x\beta \geq \omega} [1 - P(0, 1|x)] \\
\sup_{x: -\Delta x\beta - c_2 \leq \omega} P(2, 0|x) &\leq F_{\Delta V}(\omega) \\
F_{\Delta V}(\omega) &\leq \inf_{x: -\Delta x\beta + c_2 \geq \omega} [1 - P(0, 2|x)] \\
\sup_{x: -\Delta x\beta \leq \omega} P(2, 1|x) &\leq F_{\Delta V}(\omega) \\
F_{\Delta V}(\omega) &\leq \inf_{x: -\Delta x\beta \geq \omega} [1 - P(1, 2|x)] \\
\sup_{x \in X^*} P(1, 1|x) &\leq F_{\Delta V}(\omega') - F_{\Delta V}(\omega) \\
\sup_{x \in X^{**}} P(1, 1|x) &\leq F_{\Delta V}(\omega) - F_{\Delta V}(\omega'')
\end{aligned}
$$

where

$$
\begin{aligned}
X^* &= \{x : \omega' \geq -\Delta x\beta + c_2 \wedge -\Delta x\beta - c_2 \geq \omega\} \\
X^{**} &= \{x : \omega \geq -\Delta x\beta + c_2 \wedge -\Delta x\beta - c_2 \geq \omega''\}.
\end{aligned}
$$

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

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

and

$$F_{\Delta V}(\omega) \leq \inf_{x:-\Delta x\beta - c_2 \geq \omega} 1 - P(1, 1|x)$$

$$\sup_{x:-\Delta x\beta + c_2 \leq \omega} P(1, 1|x) \leq F_{\Delta V}(\omega) \tag{A.9}$$

which completes the proof.

References

Amemiya, T., 1981. Qualitative response models: A survey. J. Econ. Lit. 19 (4), 1483–1536.
Arellano, M., Bonhomme, S., 2011. Nonlinear panel data analysis. Annu. Rev. Econ. 3, 395–424.
Arellano, M., Honoré, B.E., 2001. Panel data models: some recent developments. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, Vol. 5. North-Holland, Amsterdam, pp. 3229–3296.
Baetschmann, G., Staub, K.E., Winkelmann, R., 2015. Consistent estimation of the fixed effects ordered logit model. J. Roy. Statist. Soc. Ser. A 178 (3), 685–703.
Beresteanu, A., Molinari, F., 2008. Asymptotic properties for a class of partially identified models. Econometrica 76 (4), 763–814.
Botosaru, I., Muris, C., 2017. Binarization for Panel Models with Fixed Effects. Cemmap working paper CWP31/17.
Chamberlain, G., 1984. Panel data. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, Vol. 2. North-Holland, Amsterdam, pp. 1247–1318.
Chamberlain, G., 2010. Binary response models for panel data: Identification and information. Econometrica 78 (1), 159–168.
Chernozhukov, V., Fernández-Val, I., Hahn, J., Newey, W., 2013. Average and quantile effects in nonseparable panel models. Econometrica 81 (2), 535–580.
Chernozhukov, V., Hahn, J., Newey, W., 2005. Bound Analysis in Panel Models with Correlated Random Effects. Unpublished manuscript, MIT.
Chintagunta, P., Kyriazidou, E., Perktold, J., 2001. Panel data analysis of household brand choices. J. Econometrics 103 (1-2), 111–153.
Heckman, J.J., 1981a. Heterogeneity and state dependence. In: Rosen, S. (Ed.), Studies in Labor Markets. National Bureau of Economic Research, University of Chicago Press, Chicago, pp. 91–139.
Ho, K., Rosen, A.M., 2017. Partial identification in applied research: Benefits and challenges. In: Honoré, B., Pakes, A., Piazzesi, M., Samuelson, L. (Eds.), Advances in Economics and Econometrics: Eleventh World Congress of the Econometric Society (Econometric Society Monographs), Vol. 2. Cambridge University Press, Cambridge, UK, pp. 307–359.
Honoré, B.E., 1992. Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. Econometrica 60 (3), 533–565.
Honoré, B.E., 2002. Nonlinear models with panel data. Port. Econ. J. 1 (2), 163–179.
Honoré, B.E., Kyriazidou, E., 2000. Panel data discrete choice models with lagged dependent variables. Econometrica 68 (4), 839–874.
Honoré, B.E., Lewbel, A., 2002. Semiparametric binary choice panel data models without strictly exogeneous regressors. Econometrica 70 (5), 2053–2063.
Honoré, B.E., Tamer, E., 2006. Bounds on parameters in panel dynamic discrete choice models. Econometrica 74 (3), 611–629.
Maddala, G.S., 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge, UK.
Manski, C.F., 1987. Semiparametric analysis of random effects linear models from binary panel data. Econometrica 55 (2), 357–362.
Manski, C.F., 2003. Partial Identification of Probability Distributions. In: Springer Series in Statistics, Springer, New York.
Muris, C., 2017. Estimation in the fixed-effects ordered logit model. Rev. Econ. Stat. 99 (3), 465–477.
Pakes, A., Porter, J., 2014. Moment Inequalities for Multinomial Choice with Fixed Effects. Working paper, Harvard University.
Rosen, A.M., 2012. Set identification via quantile restrictions in short panels. J. Econometrics 166 (1), 127–137.
Rosen, A.M., Weidner, M., 2013,WP. Bounds for binary panel data models with fixed effects. in preparation, slides available at https://goo.gl/dVUW3V.
Shi, X., Shum, M., Song, W., 2018. Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. Econometrica 86 (2), 737–761.
Wooldridge, J.M., 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. J. Appl. Econometrics 20 (1), 39–54.