Identification of Dynamic Panel Binary Response Models*

Shakeeb Khan† Maria Ponomareva‡ Elie Tamer §

March 2019

Abstract

We analyze identification in dynamic binary response models with fixed effects under general conditions. This class of models has been used to distinguish between state dependence and heterogeneity. More recently, there has been renewed interest in these models, especially in empirical IO and health economics, to explain switching costs, inertia or stickiness. This paper revisits the identification in these models under minimal assumptions. We first characterize the sharp set for latent utility parameters under *conditional stationarity* on the errors allowing for dynamics in the presence of fixed effects. The identified set can be characterized by a union of convex polyhedrons that can be easily numerically constructed. We conduct the same exercise under the stronger assumption of *conditional exchangeability*, and establish its incremental identifying power. We extend our identification approach to study models with more time periods as well. This shows how longer horizons help in shrinking the identified set. Throughout, we also provide sufficient conditions for point identification. For inference in cases with discrete regressors, we also provide an approach to constructing confidence sets for the identified sets using a linear program. Simulation based evidence on the size and shape of the identified sets in varying designs is also given. These illustrate the informational content of different assumptions.

Keywords: Binary Choice, Dynamic Panel Data, Identification.

*We are grateful for helpful comments from S. Chen, A. Chesher, H.Ichimura, T. Magnac, as well as conference and seminar participants at various institutions for comments. X. Lin and P. Sarkis provided excellent research assistance.

†Department of Economics, Boston College Chestnut Hill, MA, `shakeeb.khan@bc.edu`

‡Department of Economics Northern Illinois University, DeKalb, IL, `mponomareva@niu.edu`

§Department of Economics Harvard University, Cambridge, MA, `elietamer@fas.harvard.edu`

1 Introduction

There has been a renewed interests in empirical economics recently in estimating models of discrete choice over time. This interest is partly motivated by empirical regularities: certain individuals are more likely to stick with a choice if they have experienced that choice in the past. This choice "stickiness" has been attributed variably in this literature to *inertia* or *switching costs*. For example, Handel (2013) estimates a model of health insurance choice in a large firm documenting *inertia* in choices overtime. Dubé, Hitsch, and Rossi (2010) empirically find that this "inertia" in packaged goods markets is likely caused by brand loyalty. Polyakova (2016) studies the important question of quantifying the effect of switching costs in Medicare Part D markets and its relation to adversely selected plans. See also Ketcham, Lucarelli, and Powers (2015) for work on switching costs in the presence of many choices in Medicare. Other recent papers on switching costs and inertia include Raval and Rosenbaum (2018) on hospital delivery choice and Illanes (2016) on switching costs in pension plan choices. So, this recent availability of these panel data in such important markets on the one hand and the central role that the dynamic discrete choice literature played in econometric theory on the other provide the main motivation for this paper.

The dynamic discrete choice model figures prominently in econometrics. In fundamental work, Heckman (1981) discusses two different explanations for the empirical regularity that an individual is more likely to experience a state after having experienced it in the past. The first explanation, termed *state dependence*, is a genuine behavioral response to occupying the state in the past in the sense that a similar individual who did not experience the state in the past is less likely to experience it now. The current literature relabels state dependence as switching costs, inertia or stickiness and can be thought of as a *causal effect* of past occupancy of the state[1]. The second explanation advanced by Heckman is *heterogeneity*, whereby individuals are different in unobservables ways and if these unobservables are correlated over time, this will lead to said regularity. This serial correlation in the unobservables (or *heterogeneity*) is a competing explanation to state dependence and each of these lead to a different policy prescription. Hence, the econometrics literature since then has focused on models under which we are able to identify state dependence while allowing for serial correlation. These models differ in the kinds of assumptions used. We revisit this question

[1]See the recent work in Torgovitsky (2016) that provides characterization of this causal effect under minimal assumptions.

of identification of finite dimensional parameters in these models and analyze identification under minimal assumptions.

Concretely, our benchmark model is the semiparametric binary dynamic panel data model of Honoré and Kyriazidou (2000). This model relates a binary outcome in period t, y_t (we abstract from subscripting also by i) to its lagged value y_{t-1} in the following way

$$y_t = I\{u_t \leq x_t'\beta + \gamma y_{t-1} + \alpha\} \quad t = 1, 2, ...T$$

Here, γ is meant to measure the effect of state dependence (also switching costs or inertia). This parameter γ is treated as a fixed constant while the unobservables take the standard form $u_t - \alpha$ where α is individual specific and time independent and is meant to capture the systematic correlation of the unobservables over time[2]. A fixed effect treats α as possibly arbitrarily correlated with the regressor vector $x = (x_1', \ldots, x_T')'$. The challenge here is to identify (β, γ) under general assumptions on the conditional distribution of $u = (u_1, \ldots, u_T)'$ given x. For important work on inference on (β, γ) , see Heckman (1981), Chamberlain (1984), and Honoré and Kyriazidou (2000)[3].

Our paper's main focus is the question of identification of $\theta = (\gamma, \beta')'$ under minimal assumptions. This includes characterizing the sharp identification region[4]. We derive a formulation for the (sharp) identified set for θ under different sets of restrictions on the (joint) conditional distribution of u. We generalize the results in the literature in many directions. For example, we maintain stationarity restrictions on the distribution of u conditional on x and α and derive the identified set for θ when $T = 2$ (and also when $T = 3$ and larger). We then strengthen the stationarity assumption on u (which allows for serial correlation) to conditional exchangeability and derive the identified set under this assumption. These identified sets do not require any restrictions on the distribution of α, or impose any restrictions on the support of the regressor vector x over time. For example, this allows for time trends, time dummies and/or only discrete regressors. Throughout, these restrictions do not condition on the initial condition y_0 and so are internally consistent as we add more time periods. We also provide sufficient point identification conditions in terms of variation in

[2]Though in this paper we treat γ as a fixed parameter to be estimated, it is possible to extend the approaches here to cases where γ can be modeled as some function of regressors.

[3]For other work on dynamic models, see Altonji and Matzkin (2005), Honoré and Tamer (2006), Honoré and Lewbel (2002), Honoré (1993), Hu (2002), and Kyriazidou (2001).

[4]Other recent work that established sharp identification regions for structural parameters in nonlinear models includes Khan, Ponomareva, and Tamer (2011), Khan, Ponomareva, and Tamer (2016).

the support of x that provide new point identification results. Furthermore, we provide a novel linear programming based inference procedure that provides a confidence set for the identified set that is simple to compute. We also provide extensive Monte Carlo evidence on the size of the identified set and how sensitive it is to varying assumptions.

More recently, there has been work on the econometrics question of dynamics in discrete choice models. For example, Pakes and Porter (2014) provide novel methods for inference in multinomial choice models with individual fixed effects. Shi, Shum, and Song (2018) study also a multinomial choice model with fixed effects (but no dynamics) under cyclic monotonicity requirements. Aguirregabiria, Gu, and Luo (2018) study a version of the dynamic discrete choice model with logit errors by deriving sufficient statistics for the unobserved fixed effect. Also, Honoré and Tamer (2006) provide bounds on θ in a parametric random effects model without assumptions on the initial condition distribution. Finally, there is also a complementary literature that is interested in inference on average effects in panel data models. See for example Chernozhukov, Fernández-Val, Hahn, and Newey (2013) and Torgovitsky (2016). The latter paper constructs identified sets for average causal effect of lagged outcomes in binary response models under minimal assumptions.

The rest of our paper is organized as follows. The next section introduces the main model we wish to consider the identification of and previews the conditions we will be assuming. Section 3 begins our analysis by focusing on the setting of a panel data with two periods. Section 3.1 addresses identification of the structural parameters in this setting under a *conditional stationarity* assumption, as was introduced in Manski (1987) for the static binary response panel data model. Section 3.2 explores the identified region for the same model but under stricter conditions un the unobserved components. Specifically we strengthen our restrictions from *stationarity* to *exchangeability* and then *serial independence*. These models are proven to have informational content in the sense that they result in smaller identified regions than the model in Section 3.1, yet still more general than the models introduced in Chamberlain (1985) and Honoré and Kyriazidou (2000). Section 4 considers extensions of the model to allow for a panel data with a longer time series. Specifically, in 4.1 we add additional time periods to the panel, exploring the time component's informational content by showing how the identified region shrinks when more periods are available. Section 5 compliments our identification results in the previous sections by proposing computationally attractive methods to conduct inference on the structural parameters. This will enable testing, for example, if there is indeed *persistence* in the binary variable of interest. Section 6 reports results from simulation studies which explore how the identified region varies across the different models considered. Section 7 concludes with discussions on areas for future research, such as the effect of introducing more choices available to the agent, by studying a dynamic multinomial choice model with individual and choice effects, as first introduced in Chamberlain (1984) and more recently in Pakes and Porter (2014), Ouyang, Khan, and Tamer (2019).

2 Dynamic Panel Binary Choice Model

In this section, we consider the following model:

$$y_t = I\{u_t \leq x_t'\beta + \gamma y_{t-1} + \alpha\} \tag{2.1}$$

where u_t is an unobserved scalar random variable, x_t is an observed $k-$dimensional vector of covariates, β denotes an unknown k dimensional vector of regression coefficients, α denotes the unobserved scalar individual specific effect. The observed binary variable y_t takes the value 1 if the argument inside the indicator function $I\{\cdot\}$ is true, and 0 otherwise. Finally, we let the unknown scalar parameter γ denote the measure of persistence in the model.

In what follows we will will explore the identifiability of the unknown parameters β, γ, when making one of the following assumptions about the distribution of $u_1, u_2, \ldots, u_T$:

(STAT) Stationarity (identical distribution):

$$u_t \sim u_1 | \alpha, x, \text{ for all } t = 2, \ldots, T$$

(CEX) Exchangeability: for any $k \leq T$ and any permutation $(s_1, \ldots, s_k)$ of $(1, \ldots, k)$,

$$(u_1, \ldots, u_k) \sim (u_{s_0}, \ldots, u_{s_k}) | \alpha, x, y_0$$

(CIID) Conditional i.i.d.:

$$u_1, \ldots, u_T \sim i.i.d. | \alpha, x$$

(IND) Full independence, semiparametric:

$$u_1, \ldots, u_T \sim F(\cdot)|\alpha, x$$

where $F(\cdot)$ does not depend on α, x.

These differing assumptions relate to each other in terms of their level of generality. Specifically, (STAT) is the weakest assumption, assumption (CEX) implies stationarity assumption (STAT), (CIID) implies (CEX), and finally, (IND) implies (CIID). As wee will show the converses, however, are not true for any of these relationships. The (CIID) and (CEX) assumptions can be equivalent under a particular form of conditional exchangeability ($T = \infty$ in (CEX) assumption). Note also that (STAT) and (CEX) do not involve conditioning on the initial condition y_0.

We note Assumption (CIID) can be further strengthened to $u_1, \ldots, u_T$ being mutually independent and identically distributed *and also* independent from x and α (IND assumption). This latter version is the one used for identification in semiparametric model in Honoré and Kyriazidou (2000). As we show in this paper, this assumption can potentially provide additional restrictions on the parameters of interest.

We also require the conditions below.

A1. $u = (u_1, \ldots, u_T)$ is absolutely continuous conditional on α, x.

A2. $\Delta u_{ts} = u_t - u_s$ is also absolutely continuous for any $t, s = 1, \ldots, T$ and $t \neq s$.

A3. The support of $u_t|x, \alpha$ is $\mathbb{R}$.

A4. We observe $i = 1, \ldots, n$ iid draws from (2.1): $\{y_i = (y_{i0}, \ldots, y_{iT}), x_i = (x_{i1}, \ldots, x_{iT}), i = 1, \ldots, n\}$. Assume n is large relative to T, so any notions of asymptotics are derived under the assumption that $n \to \infty$, while T is fixed.

These are common conditions. The first three allow us to use strict monotonicity of the distribution functions of various objects. The last condition above is a requirement that data are obtained on observations starting with y_0.

The next section analyzes in details the sharp identification of the model when $T = 2$. We provide characterization of the identified set without making any assumptions on the support of the regressors. The Section contains the main results in the paper.

3 Identification with $T = 2$

We first analyze what can be learned with $T = 2$ time periods. This assumes that we have access to the initial period $t = 0$ and then two more periods $t = 1$ and $t = 2$. Throughout this section and the next, we will be using the following notation when, say, $T = 3$: for $d_1, d_2, d_3 \in \{0, 1\}$

$$p_1(d_1|x, y_0) \equiv P(y_1 = d_1|x, y_0)$$
$$p_2(d_1, d_2|x, y_0) \equiv P(y_1 = d_1, y_2 = d_2|x, y_0)$$
$$p_3(d_1, d_2, d_3|x, y_0) \equiv P(y_1 = d_1, y_2 = d_2, y_3 = d_3|x, y_0)$$

3.1 Stationarity with $T = 2$

We start by analyzing identifying power of stationarity assumption (STAT). Specifically, we assume the econometrician observes a random sample for the random variables y_0, y_1, y_2, x_1, x_2 and maintain the assumption:

Assumption 3.1. *(STAT(T=2))*

$$u_2 \sim u_1|\alpha, x$$

Note that we require that stationarity holds conditional on the fixed effect and the vector of all lead and lags of the regressor x. On the other hand, no assumptions are made on the support of x (this can only be a time trend or time dummy for example), and we do not condition on the initial condition y_0.

The result below gives us the set of all parameters that are observationally equivalent to the true parameter under this stationarity assumption (so that this set is the *sharp* identified set under (STAT) assumption).

Theorem 3.1. *Let $\Theta_{I,stat}^{\{1,2\}}$ be the set of $\theta = (\beta', \gamma)'$ that satisfy the restrictions below: if for some x,*

(1) $P(y_2 = 1|x) \geq P(y_1 = 1|x) \Rightarrow (x_2 - x_1)'\beta + |\gamma| \geq 0$;

(2) $P(y_1 = 1|x) \geq P(y_2 = 1|x) \Rightarrow (x_2 - x_1)'\beta - |\gamma| \leq 0$*;*

(3) $P(y_1 = 0, y_2 = 1|x) \geq P(y_1 = 1|x)$ *or* $P(y_0 = 1, y_1 = 0|x) \geq P(y_2 = 0|x) \Rightarrow (x_2 - x_1)'\beta - \min\{0, \gamma\} \geq 0$*;*

(4) $P(y_1 = 1, y_2 = 0|x) \geq P(y_1 = 0|x)$ *or* $P(y_0 = 0, y_1 = 1|x) \geq P(y_2 = 1|x) \Rightarrow (x_2 - x_1)'\beta + \min\{0, \gamma\} \leq 0$*;*

(5) $P(y_0 = 0, y_1 = 0|x) \geq P(y_2 = 0|x) \Rightarrow (x_2 - x_1)'\beta + \max\{0, \gamma\} \geq 0$*;*

(6) $P(y_0 = 1, y_1 = 1|x) \geq P(y_2 = 1|x) \Rightarrow (x_2 - x_1)'\beta - \max\{0, \gamma\} \leq 0$*;*

(7) $P(y_0 = 0, y_1 = 0|x) + P(y_1 = 0, y_2 = 1|x) \geq 1 \Rightarrow (x_2 - x_1)'\beta \geq 0$*;*

(8) $P(y_0 = 1, y_1 = 1|x) + P(y_1 = 1, y_2 = 0|x) \geq 1 \Rightarrow (x_2 - x_1)'\beta \leq 0$*;*

(9) $P(y_0 = 1, y_1 = 0|x) + P(y_1 = 0, y_2 = 1|x) \geq 1 \Rightarrow (x_2 - x_1)'\beta - \gamma \geq 0$*;*

(10) $P(y_0 = 0, y_1 = 1|x) + P(y_1 = 1, y_2 = 0|x) \geq 1 \Rightarrow (x_2 - x_1)'\beta + \gamma \leq 0$

for all $\theta = (\beta', \gamma)' \in \Theta_{I,stat}^{\{1,2\}}$. *Then* $\Theta_{I,stat}^{\{1,2\}}$ *is the sharp identified set for* θ *under stationarity assumption (STAT) and* $T = 2$.

Proof: Let $v_1 = u_1 - \alpha$ and $v_2 = u_2 - \alpha$. Note that u_1 and u_2 are identically distributed conditional on x and α if and only if v_1 and v_2 are also identically distributed conditional on x and α. This implies that v_1 and v_2 must be identically distributed conditional on x, so let $F(\cdot|x)$ be the marginal distribution of v_t for $t = 1, 2$, conditional on x. The following are sharp restrictions on $F(\cdot|x)$: for $v_1 = u_1 - \alpha$ we have

$$\begin{aligned}
F(x_1'\beta + \min\{0, \gamma\}) &\leq P(y_1 = 1|x) \\
P(y_1 = 1|x) &\leq F(x_1'\beta + \max\{0, \gamma\}) \\
P(y_0 = 1, y_1 = 1|x) &\leq F(x_1'\beta + \gamma|x) \\
F(x_1'\beta + \gamma|x) &\leq 1 - P(y_0 = 1, y_1 = 0|x) \\
P(y_0 = 0, y_1 = 1|x) &\leq F(x_1'\beta|x) \\
F(x_1'\beta|x) &\leq 1 - P(y_0 = 0, y_1 = 0|x)
\end{aligned}$$

and for $v_2 = u_2 - \alpha$ we have:

$$
\begin{aligned}
F(x_2'\beta + \min\{0,\gamma\}) &\leq P(y_2 = 1|x) \\
P(y_2 = 1|x) \leq& F(x_2'\beta + \max\{0,\gamma\}) \\
P(y_1 = 1, y_2 = 1|x) \leq& F(x_2'\beta + \gamma|x) \\
F(x_2'\beta + \gamma|x) &\leq 1 - P(y_1 = 1, y_2 = 0|x) \\
P(y_1 = 0, y_2 = 1|x) \leq& F(x_2'\beta|x) \\
F(x_2'\beta|x) &\leq 1 - P(y_1 = 0, y_2 = 0|x)
\end{aligned}
$$

Under conditional stationarity, the model provides no other restrictions on the shape of $F(\cdot|x)$. The rest of the proof closely follows the proof of sharpness of $\Theta_{I,cex}^{\{1,2\}}(1)$ in Lemma 3.2 below with the following copula equations that match conditional probabilities of observed outcomes for $j = 0, 1$:

$$
\begin{aligned}
\tilde{C}(P(y_0 = j|x), \tilde{q}_{1j}(x)|x) =& P(y_0 = j, y_1 = 1|x) \\
\tilde{C}(P(y_0 = 0|x), \tilde{q}_{1j}(x), \tilde{q}_{21}(x)|x) =& P(y_0 = j, y_1 = 1, y_2 = 1|x) \\
\tilde{C}(P(y_0 = 0|x), \tilde{q}_{1j}(x), \tilde{q}_{20}(x)|x) =& \tilde{C}(P(y_0 = j|x), 1, \tilde{q}_{20}(x)|x) \\
& - P(y_0 = j, y_1 = 0, y_2 = 1|x)
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{q}_{1j}(x) = P(\tilde{v}_1 \leq x_1\tilde{\beta} + \tilde{\gamma}j|x) \equiv \tilde{F}(x_1\tilde{\beta} + \tilde{\gamma}j|x) \\
\tilde{q}_{2j}(x) = P(\tilde{v}_2 \leq x_2\tilde{\beta} + \tilde{\gamma}j|x) \equiv \tilde{F}(x_2\tilde{\beta} + \tilde{\gamma}j|x)
\end{aligned}
$$

and $\tilde{F}(\cdot|x)$ is the conditional distribution of some $\tilde{v}_t$. Let $\tilde{\alpha}$ be some scalar random variable (or even a constant), and define $\tilde{u}_t = \tilde{v}_t + \tilde{\alpha}$. Then $\tilde{u}_1$ and $\tilde{u}_2$ are identically distributed conditional on x and $\tilde{\alpha}$, and with $\tilde{y}_0 = y_0$, the joint distribution of $\{x, \tilde{y}_0, \tilde{y}_t = 1\{\tilde{u}_1 \leq \tilde{\beta}x_t + \tilde{\gamma}\tilde{y}_{t-1} + \tilde{\alpha}, t = 1, 2\}$ matches the distribution of observables. That is, for any $\tilde{\theta} \in \Theta_{I,stat}^{\{1,2\}}$ we are able to find some distribution G of $(\tilde{u}_1, \tilde{u}_2, \tilde{\alpha})$ such that $p(d, x|\theta, F) = p(d, x|\tilde{\theta}, G)$ and where $\tilde{u}_1 - \tilde{\alpha}$ and $\tilde{u}_2 - \tilde{\alpha}$ are identically distributed conditional on y_0, x with marginal distribution $\tilde{F}$. That is, $\theta \overset{o.e.}{\sim}_{\mathcal{F}} \tilde{\theta}$ under stationarity assumption. ■.

A key part in the above proof is the use of the results from the next subsection on exchangeability. There, a series of Lemmas contains the critical steps needed for the above argument to hold.

Stationarity (or identical distribution of error terms) is the key identifying assumption in Manski (1987) for the static binary response model with fixed effects. The above characterization extends this result to dynamic models without any restrictions on the support of the covariate distribution. In addition, the above is also *constructive*. By this we mean we can use the above inequalities to construct set estimation and inference procedures for the parameters β, γ.

For example, we note that the left hand side of the above inequalities are all conditional choice probabilities that can be be estimated from the data. Consequently, from the identification perspective, we can treat these conditional probabilities as *known* functions. The right hand side of the above inequalities involve indexes which are based on both observed variables and unknown parameters. Consequently our identified set and estimator thereof will be the set of all the values of $\tilde{\beta}, \tilde{\gamma}$ where the above inequalities hold true. As we show here, this set will generally not reduce to a unique point, at least not without other conditions on the disturbances and the regressors. Our main point which can be concluded from the theorem is the set attained based on the above inequalities, is the smallest set attainable based on the stated assumptions in the model.

3.2 Exchangeability with $T = 2$

We strengthen the above model under stationarity and impose *exchangeability.*[5] This assumption adapts (CEX) above to the $T = 2$ case. Note here that exchangeability of (u_0, u_1, u_2) conditional on (α, x) implies exchangeability of (u_1, u_2) conditional on (α, x, y_0). This subsection contains the critical results in the paper.

Definition 1. *We define exchangeability and independence as follows.*

- *(CEX(T=2)) We have (1)* $u_1 \sim u_2|\alpha, x, y_0$. *Also, let* F_{12} (F_{21}) *be the distribution of* $(u_1, u_2|\alpha, x, y_0)$ $((u_2, u_1|\alpha, x, y_0))$ *respectively. Then, let (2)* $F_{12} = F_{21}$ *hold.*
- *(CIID(T=2)) We have* $u_1, u_2 \sim i.i.d|\alpha, x, y_0$.

[5] This assumption has been used in the literature for different models to identify parameters of interest. See, for example Altonji and Matzkin (2005).

The exchangeability assumption (CEX) involves both a *marginal* part which requires that u_1 and u_2 have the same distribution, but it also requires that *jointly*, (u_1, u_2) has the same distribution of the random variable to permutes the two indices. Building the identified set under exchangeability involves two steps. The first step provides the identified set under part (2) of CEX(T=2) 1 above. Then, the following Theorem provides the identified set when we have both (1) and (2). As for conditional independence, it turns out that under a particular form of exchangeability, conditional independence and exchangeability are equivalent and so we formally show that the identified set is the same whether we assume CIID/T=2 or CEX/T=2. We start with the result that provides the identifying power of the second part of the exchangeability condition above.

Proposition 3.1. *Let CEX(T=2) of Definition 1 above hold and let $y_0 \in \{0, 1\}$. Then parameters β and γ must satisfy the following conditions for every (x, y_0) in the support:*

(i) If $(x_2 - x_1)'\beta - \gamma y_0 \leq 0$ and $(x_2 - x_1)'\beta - \gamma y_0 + \gamma \leq 0$, then $P(y_1 = 1, y_2 = 0|x, y_0) \geq P(y_1 = 0, y_2 = 1|x, y_0)$;

(ii) If $(x_2 - x_1)'\beta - \gamma y_0 \geq 0$ and $(x_2 - x_1)'\beta - \gamma y_0 + \gamma \geq 0$, then $P(y_1 = 1, y_2 = 0|x, y_0) \leq P(y_1 = 0, y_2 = 1|x, y_0)$

with at least one strict inequality on the left-hand side implying strict inequality on the right-hand side.

Proof: Conditional exchangeability assumption implies that

$$\begin{aligned}
\frac{P(y_1 = 1, y_2 = 0|\alpha, x, y_0)}{P(y_1 = 0, y_2 = 1|\alpha, x, y_0)} &= \frac{P(u_1 \leq \alpha + x_1'\beta + \gamma y_0, u_2 > \alpha + x_2'\beta + \gamma|\alpha, x, y_0)}{P(u_1 > \alpha + x_1'\beta + \gamma y_0, u_2 \leq \alpha + x_2'\beta|\alpha, x, y_0)} \\
&= \frac{P(u_1 \leq \alpha + x_1'\beta + \gamma y_0, u_2 > \alpha + x_2'\beta + \gamma|\alpha, x, y_0)}{P(u_1 \leq \alpha + x_2'\beta, u_2 > \alpha + x_1'\beta + \gamma y_0|\alpha, x, y_0)}
\end{aligned}$$

And since $P(u_1 \leq a_1, u_2 > a_2|\alpha, x, y_0)$ is strictly increasing in a_1 and strictly decreasing in a_2, we have the following:

(1) If $x_1'\beta + \gamma y_0 \geq x_2'\beta$ and $x_1'\beta + \gamma y_0 \geq x_2'\beta + \gamma$, then $P(y_1 = 1, y_2 = 0|\alpha, x, y_0) \geq P(y_1 = 0, y_2 = 1|\alpha, x, y_0)$

(2) If $x_1'\beta + \gamma y_0 \leq x_2'\beta$ and $x_1'\beta + \gamma y_0 \leq x_2'\beta + \gamma$, then $P(y_1 = 1, y_2 = 0|\alpha, x, y_0) \leq P(y_1 = 0, y_2 = 1|\alpha, x, y_0)$

Integrating α out gives us conditions (1) and (2). ■

The conditional exchangeability assumption of Proposition 3.1 is a stronger assumption than stationarity of u_t, but it is weaker than the conditional iid assumption, and even weaker than the full iid assumption in Honoré and Kyriazidou (2000).

Note here that if $\gamma = 0$, conditions (1) and (2) in Proposition 3.1 reduce to a variant of Manski's identification conditions for the static panel data binary choice model, for outcomes such that $y_1 + y_2 = 0$:

1. $P(0, 1|x, y_0) \geq P(1, 0|x, y_0)$ implies that $(x_2 - x_1)'\beta \geq 0$
2. $P(1, 0|x, y_0) \geq P(0, 1|x, y_0)$ implies that $(x_2 - x_1)'\beta \leq 0$
3. $P(1, 0|x, y_0) = P(0, 1|x, y_0)$ if and only if $(x_2 - x_1)'\beta = 0$.

Manski (1987) provided sufficient conditions on the support of the regressor vector x that leads to point identification of β (essentially requiring full support for the regression index). This is interesting since the result in Proposition 3.1 provides a characterization of the identified set in the Manski model without any conditions on the support of x.

The next Theorem, whose proof follows is the main result in this section.

Theorem 3.2. *Let $\Theta^{\{1,2\}}_{I,cex}(2)$ be the set of parameters that satisfy conditions (i) and (ii) of Proposition 3.1. Also let $\Theta^{\{1,2\}}_{I,cex}(1)$ satisfiy the following restriction: if for some $z = (x, y_0)$*

(1) $P(y_1 = 1|z) \geq P(y_2 = 1|z) \Rightarrow (x_2 - x_1)\tilde{\beta} + \min\{0, \gamma\} - \gamma y_0 \leq 0$*;*

(2) $P(y_1 = 1|z) \leq P(y_2 = 1|z) \Rightarrow (x_2 - x_1)\tilde{\beta} + \max\{0, \gamma\} - \gamma y_0 \geq 0$*;*

(3) $P(y_1 = 0, y_2 = 1|z) \geq P(y_1 = 1|z) \Rightarrow (x_2 - x_1)\tilde{\beta} - \tilde{\gamma} y_0 \geq 0$*;*

(4) $P(y_1 = 1, y_2 = 0|z) \geq P(y_1 = 0|z) \Rightarrow (x_2 - x_1)\tilde{\beta} + \tilde{\gamma}(1 - y_0) \leq 0$*;*

for all $\tilde{\theta} \in \Theta^{\{1,2\}}_{I,cex}(1)$. Then $\Theta^{\{1,2\}}_{I,cex} = \Theta^{\{1,2\}}_{I,cex}(2) \cap \Theta^{\{1,2\}}_{I,cex}(1)$ is the sharp identified set for θ under either conditional exchangeability (CEX/T=2) assumption or conditional independence (CIID/T=2) of Definition 1 above.

In this result, $\Theta^{\{1,2\}}_{I,cex}(2)$ is the sharp identified set for θ under the assumption that pairs (u_1, u_2) and (u_2, u_1) are identically distributed conditional on y_0, x and α. Conditional exchangeability assumption also implies that u_1 and u_2 are identically distributed, too, so $\Theta^{\{1,2\}}_{I,cex}(1)$ is the sharp identified set for θ that respect the condition where u_1 and u_2 are identically distributed conditional on y_0, x, and α.

3.2.1 Proof of Theorem 3.2

The proof of Theorem 3.2 above consists of three Lemmas. First, Lemma 3.1 shows that infinite conditional exchangeability and conditional independence models are observationally equivalent. This result is interesting by itself but here it allows us to connect conditional exchangeability and CIID in a natural way. Lemma 3.2 is a fundamental result in this paper that is used for proving sharpness in the stationary model also. This result uses interesting marginal probability vs joint probability comparisons that we see in the characterization of the identified sets. Finally, Lemma 3.3 connects both pieces (1) and (2) of exchageability in Assumption 1 above.

Lemma 3.1. *Let $u_1, \ldots, u_T, \ldots$ be conditionally (on α and $z = (y_0, x)$) exchangeable for any $T \geq 2$, and continuously distributed. We observe $\{(y_1 = 1\{u_t \leq \alpha + a_1\}, \ldots, y_T = 1\{u_T \leq \alpha + a_T\}), y_0, x\}$, where $a_t = x_t'\beta + \gamma y_{t-1}$ for $t = 1, \ldots, T$. Then there exist $\tilde{u}_1, \ldots, \tilde{u}_T, \tilde{\alpha}$ such that $\tilde{u}_1, \ldots, \tilde{u}_T$ are iid conditional on $\tilde{\alpha}$ and $z = (y_0, x)$; and for $\tilde{y}_1 = 1\{\tilde{u}_t \leq \tilde{\alpha} + a_1\}, \ldots, \tilde{y}_T = 1\{\tilde{u}_T \leq \tilde{\alpha} + a_T\})$ we have the following:*

$$P(\tilde{y}_1 = 1, \ldots, \tilde{y}_T = 1 | y_0, x) = P(y_1 = 1, \ldots, y_T = 1 | y_0, x)$$

Proof: Although Theorem 3.1 requires us to look at only some sequences $(a_1, \ldots, a_T)$ and to match only the distribution of indicator variables $y_1, \ldots, y_T$, infinite exchangeability of $u_1, \ldots, u_T, \ldots$ allows us to get a much stronger result: conditional on z, we can match the whole distribution of $(u_1-\alpha,\dots,u_T-\alpha)$ to a conditional iid model $(\tilde{u}_1-\tilde{\alpha},\dots,\tilde{u}_T-\tilde{\alpha})$. So below we prove a general result for an arbitrary sequence $(a_1,\dots,a_T)$. First, note that if $u_1,\dots,u_T$ are exchangeable conditional on α and z, then $u_1-\alpha,\dots,u_T-\alpha$ are exchangeable conditional on z.

Next, conditional on z, for infinitely exchangeable sequences, Theorem 3 in Olshen (1973) implies that there exists a scalar random variable $\tilde{\alpha}$ such that conditional on $\tilde{\alpha}$, $u_1-\alpha, u_2-\alpha,\dots,u_T-\alpha$ are conditionally iid:

$$P(u_1-\alpha\le a_1, u_1-\alpha\le a_2,\dots,u_T-\alpha\le a_T|z,\tilde{\alpha})=\prod_{t=1}^{T}P(u_t-\alpha\le a_t|z,\tilde{\alpha})$$

Let $\tilde{u}_t=u_t-\alpha+\tilde{\alpha}$. Then we have the following:

$$P(\tilde{u}_1\le a_1+\tilde{\alpha},\dots,\tilde{u}_T\le u_T+\tilde{\alpha}|z,\tilde{\alpha})=\prod_{t=1}^{T}P(\tilde{u}_t\le a_t+\tilde{\alpha}|z,\tilde{\alpha})$$

That is, for an exchangeable sequence $(u_1-\alpha,\dots,u_T-\alpha)$ we were able to construct a model $(\tilde{u}_1,\dots,\tilde{u}_T,\tilde{\alpha})$ such that

(i) both models are observationally equivalent:

$$\begin{aligned}P(u_1\le a_1+\alpha,\dots,u_T\le a_T+\alpha|z)=&\int P(u_1-\alpha\le a_1,\dots,u_T-\alpha\le a_T|z,\tilde{\alpha})dF(\tilde{\alpha}|z)\\=&\int P(\tilde{u}_1\le a_1+\tilde{\alpha},\dots,\tilde{u}_T\le a_T+\tilde{\alpha}|z,\tilde{\alpha})dF(\tilde{\alpha}|z)\\=&P(\tilde{u}_1\le a_1+\tilde{\alpha},\dots,\tilde{u}_T\le a_T+\tilde{\alpha}|z)\end{aligned}$$

(ii) $\tilde{u}_1,\dots,\tilde{u}_T$ are iid conditional on z and $\tilde{\alpha}$.

This implies that infinite conditional exchangeability and conditional iid assumptions produce observationally equivalent models. ■

Again, when $T=2$, the exchangeability assumption is equivalent to the following two conditions:

1. u_1 and u_2 are identically distributed conditional on y_0, x, and α (this is the stationarity

of the distribution of u_t);

2. (u_1, u_2) and (u_2, u_1) are identically distributed conditional on y_0, x, and α.

The stationarity of u_t places certain restrictions on θ which are given by the set $\Theta_{I,cex}^{\{1,2\}}(1)$ defined as all parameters $\theta \in \Theta$ that satisfy the following conditions for all $z = (y_0, x)$:

(i) $P(y_1 = 1|z) \geq P(y_2 = 1|z)$, then $(x_2 - x_1)\tilde{\beta} + \min\{0, \gamma\} - \gamma y_0 \leq 0$;

(ii) $P(y_1 = 1|z) \leq P(y_2 = 1|z)$, then $(x_2 - x_1)\tilde{\beta} + \max\{0, \gamma\} - \gamma y_0 \geq 0$;

(iii) $P(y_1 = 0, y_2 = 1|z) \geq P(y_1 = 1|z)$, then $(x_2 - x_1)\tilde{\beta} - \tilde{\gamma} y_0 \geq 0$;

(iv) $P(y_1 = 1, y_2 = 0|z) \geq P(y_1 = 0|z)$, then $(x_2 - x_1)\tilde{\beta} + \tilde{\gamma}(1 - y_0) \leq 0$;

The next lemma shows that $\Theta_{I,cex}^{\{1,2\}}(1)$ is the sharp identified set for θ under the assumption that u_t is stationary conditional on α, x, y_0. This is a key result in the paper. It shows that for any parameter $\theta \in \Theta_{I,cex}^{\{1,2\}}(1)$, we are able to construct a distribution for the unobservables that obeys the exchangeability assumptions and matches the distribution of the data.

Lemma 3.2. *Let $\mathcal{F}_{cex,1}$ be the set of all distributions $F_{u,\alpha|z}$ such that $F_{u|\alpha,z}$ satisfies stationarity part of the exchangeability assumption for all α and $z = (x, y_0)$. Then for any $\tilde{\theta} \in \Theta_{I,cex}^{\{1,2\}}(1)$, $\theta \overset{o.e.}{\sim}_{\mathcal{F}} \tilde{\theta}$ under $\mathcal{F}_{cex,1}$.*

Proof: For each z, we define

$$
\begin{aligned}
\tilde{q}_1(z) &= P(\tilde{v}_1 \leq x_1\tilde{\beta} + \tilde{\gamma} y_0|z) = \tilde{F}(x_1\tilde{\beta} + \tilde{\gamma} y_0|z) \\
\tilde{q}_{20}(z) &= P(\tilde{v}_2 \leq x_2\tilde{\beta}|z) = \tilde{F}(x_2\tilde{\beta}|z) \\
\tilde{q}_{21}(z) &= P(\tilde{v}_2 \leq x_2\tilde{\beta} + \tilde{\gamma}|z) = \tilde{F}(x_2\tilde{\beta} + \tilde{\gamma}|z)
\end{aligned}
\tag{3.1}
$$

where $\tilde{F}$ is the conditional (on z) distribution of $\tilde{v}_t = \tilde{\alpha} + \tilde{u}_t$.

An arbitrary $\tilde{\theta} = (\tilde{\beta}, \tilde{\gamma})$ will induce an arbitrary order between $x_1\tilde{\beta} + \tilde{\gamma} y_0$, $x_2\tilde{\beta} + \tilde{\gamma}$ and $x_2\tilde{\beta}$, which in turn induces the same ordering between $\tilde{q}_1(z), \tilde{q}_{20}(z), \tilde{q}_{21}(z)$ so that we can always find a distribution that passes through these three points.

The parameter $\tilde{\theta}$ is observationally equivalent to the true parameter θ if the following holds: for an ordering between $\tilde{q}_1(z), \tilde{q}_{20}(z), \tilde{q}_{21}(z)$ that is induced by this $\tilde{\theta}$, there exists a bivariate copula $\tilde{C}(\cdot, \cdot|z)$ such that

$$\begin{aligned} \tilde{q}_1(z) &= P(y_1 = 1|z) \\ \tilde{C}(\tilde{q}_1(z), \tilde{q}_{21}(z)|z) &= P(y_1 = 1, y_2 = 1|z) \\ \tilde{C}(\tilde{q}_1(z), \tilde{q}_{20}(z)|z) &= \tilde{q}_{20}(z) - P(y_1 = 0, y_2 = 1|z) \end{aligned} \tag{3.2}$$

The first two equations in (3.2) match $P(y_1 = 1, y_2 = 0|z)$ and $P(y_1 = 1, y_2 = 1|z)$, while the last equations matches $P(y_1 = 0, y_2 = 1|z)$.

Copulas are non-negative, non-decreasing functions with $\tilde{C}(\tilde{q}_1(z), 0|z) = 0$ and $\tilde{C}(\tilde{q}_2(z), 1|z) = \tilde{q}_1(z)$. Next, for any q_2, $\tilde{C}(\tilde{p}_1(z), p_2|z)$ is bounded by Fréchet-Hoeffding bounds:

$$\max\{\tilde{q}_1(z) + q_2 - 1, 0\} \leq \tilde{C}(\tilde{q}_1(z), q_2|z) \leq \min\{\tilde{q}_1(z), q_2\}$$

To simplify the notation, let us denote $P_j^t = P(y_t = j|z)$ and $P_{jk} = P(y_1 = j, y_2 = k|z)$ for $t = 1, 2$ and $j, k = 0, 1$. Then in Figure 1 below, the solution to (3.2) will be represented by the intersection of a non-decreasing in q_2 function $\tilde{C}(P_1^1, q_2)$ that lies within the Fréchet-Hoeffding bounds, with the red line ($\tilde{q}_{21}(z)$ in the second equation in (3.2)), and the blue line ($\tilde{q}_{20}(z)$ in the third equation in (3.2)). If that function crosses the red line to the left of the blue line, then $\tilde{q}_{21}(z) < \tilde{q}_{20}(z)$; if it crosses the red line to the right of the blue line, then $\tilde{q}_{21}(z) < \tilde{q}_{20}(z)$.

Since copulas are non-negative, we must have $\tilde{q}_{20}(z) \geq P_{01}$. So if $P(y_1 = 0, y_2 = 1|z) \geq P(y_1 = 1|z)$, we must have $\tilde{q}_{20}(z) > \tilde{q}_1(z)$ and so we need to place the following restrictions on $\tilde{\beta}, \tilde{\gamma}$:

$$(x_2 - x_1)\tilde{\beta} - \tilde{\gamma} y_0 \geq 0$$

A copula function (dashed or dash-dotted line in Figure 1 can cross the red line to the right of P_1^1 (so that $\tilde{q}_{21}(z) > \tilde{q}_1(z)$) only if the lower copula bound evaluated at $q_2 = P_1^1$ is smaller than $P_{11} = P(y_1 = 1, y_2 = 1|z)$. If this is not true, that is if

$$\max\{2P_1^1 - 1, 0\} > P_{11}$$

then $\tilde{q}_{21}(z) \leq \tilde{q}_1(z)$. That is, if $P(y_1 = 0|z) < P(y_1 = 1|z) - P(y_1 = 1, y_2 = 1|z) \equiv P(y_1 =$

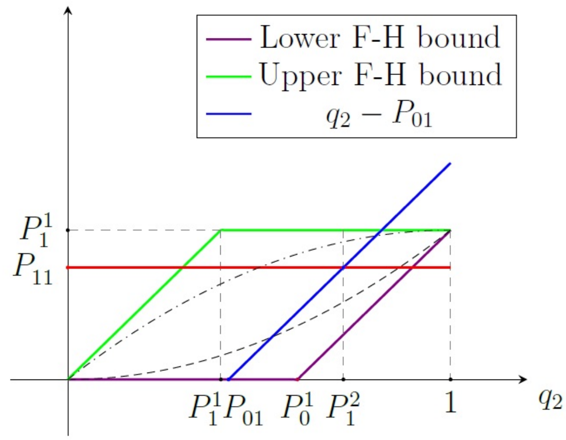

Figure 1: $P_j^t = P(y_t = j|z)$ and $P_{jk} = P(y_1 = j, y_2 = k|z)$. The dashed curves represent potential copula function $\tilde{C}(P_1^1, q_2)$ as a function of q_2.

$1, y_2 = 0|z)$, then we have to place the following restrictions on $\tilde{\beta}, \tilde{\gamma}$:

$$(x_2 - x_1)\tilde{\beta} + \tilde{\gamma}(1 - y_0) \leq 0$$

Finally, since copulas are non-decreasing, one solution (either $\tilde{q}_{20}(z)$ or $\tilde{q}_{21}(z)$) to second and third equations in (3.2) will have to be to the left of the intersection of blue and red line, and another solution has to be to the right of that intersection. Which means that if $P_1^2 \leq P_1^1$, then

$$\min\{\tilde{q}_{20}(z), \tilde{q}_{21}(z)\} \leq P_1^1 = \tilde{q}_1(z)$$

and similarly, if $P_1^2 \geq P_1^1$, then

$$\max\{\tilde{q}_{20}(z), \tilde{q}_{21}(z)\} \geq P_1^1 = \tilde{q}_1(z)$$

That is, if $P(y_1 = 1|z) \leq P(y_2 = 1|z)$, we have to have

$$x_2'\tilde{\beta} + \min\{0, \tilde{\gamma}\} \leq x_1'\tilde{\beta} + \tilde{\gamma}y_0$$

and if $P(y_1 = 1|z) \geq P(y_2 = 1|z)$, we have to have

$$x_2'\tilde{\beta} + \max\{0, \tilde{\gamma}\} \geq x_1'\tilde{\beta} + \tilde{\gamma}y_0$$

Now, let $\tilde{\theta}$ be an arbitrary parameter that belongs to, and let G be the joint distribution of $(\tilde{v}_1, \tilde{v}_2)$ defined as

$$G(\cdot,\cdot|z) = \tilde{C}(\tilde{F}(\cdot), \tilde{F}(\cdot)|z)$$

where $\tilde{F}$ is a distribution that satisfies (3.1). By construction, $p(z;\theta,F) = p(z;\tilde{\theta},G)$, and G has identical marginals. That is, $\tilde{\theta}$ is observationally equivalent to the true θ under the identical distribution assumption. ∎

Next, we show that with finite 2-exchangeability, the identified set is characterized by exactly the same inequalities as with conditional iid assumptions, although the two models are not necessarily observationally equivalent. First, we show that for any $\tilde{\theta}$ in the "stationarity" set $\Theta_{I,cex}^{\{1,2\}}(1)$ such that it does not belong to $\Theta_{I,cex}^{\{1,2\}}(2)$, there exists no $G \in \mathcal{F}_{cex}$ such that $p(y,x|\tilde{\theta},G) = p(y,x|\theta,F)$ for almost all x.

Let us define

$$A(\theta, z) = \begin{pmatrix} (x_1 - x_2)'\beta + \gamma y_0 \\ (x_1 - x_2)'\beta + \gamma(y_0 - 1) \end{pmatrix}$$

and for any $\tilde{\theta} \neq \theta$, define

$$Q(\tilde{\theta}) = \{z \in \mathcal{Z} : 1\{A(\tilde{\theta}, z) > 0\} = 1 = 1\{A(\theta, z) \leq 0\} \text{ or } 1\{A(\tilde{\theta}, z) < 0\} = 1 = 1\{A(\theta, z) \geq 0\}\}$$

That is, $Q(\tilde{\theta})$ is a set of covariates z such that inequalities in Proposition 3.1 hold with different signs for θ and $\tilde{\theta}$.

Lemma 3.3. *Let $\tilde{\theta} \in \Theta_{I,cex}^{\{1,2\}}(1)$ and suppose that $P\{Z \in Q(\tilde{\theta})\} > 0$. Then θ is identified relative to $\tilde{\theta}$ under both $\mathcal{F}_{cex}$ and $\mathcal{F}_{ciid}$. That is, there exists no $G \in \mathcal{F}_{cex}$ or $G \in \mathcal{F}_{ciid}$ such that $p(y,x|\theta,F) = p(y,x|\tilde{\theta},G)$ for almost all x.*

Proof: Without loss of generality, let $z \in Q(\tilde{\theta})$ be such that $1\{A(\tilde{\theta}, z) > 0\} = 1 = \{A(\theta, z) \leq 0\}$ and let $G \in \mathcal{F}_{cex}$. Since G belongs to the exchangeability (iid) class, $A(\tilde{\theta}, z) > 0$ implies that

$$p(y_0 = 1, y_2 = 0, z|\tilde{\theta}, G) > p(y_0 = 1, y_2 = 1, z|\tilde{\theta}, G)$$

At the same time, we have $A(\theta, z) \leq 0$ and since F belongs to the exchangeability class, it must be the case that

$$p(y_0 = 1, y_2 = 0, z|\theta, F) \leq p(y_0 = 0, y_2 = 1, z|\theta, F)$$

That is, for this z there exists no $G \in \mathcal{F}_{ex}$ such that $p(y, x|\tilde{\theta}, G) = p(y, x|\theta, F)$ for almost all x. Similar arguments can be used to show that if $1\{A(\tilde{\theta}, z) < 0\} = 1 = 1\{A(\theta, z) \geq 0\}$, then again there exists no $G \in \mathcal{F}_{cex}$ such that $p(y, x|\tilde{\theta}, G) = p(y, x|\theta, F)$ for almost all x.

Finally, since $\mathcal{F}_{ciid} \subset \mathcal{F}_{cex}$, the above argument equally applies to the identification under serial independence and stationarity. ■

Next, we show that any $\tilde{\theta} \in \Theta^{\{1,2\}}_{I,cex}(1) \bigcap \Theta^{\{1,2\}}_{I,cex}(2)$ is observationally equivalent to θ.

Lemma 3.4. *Let $\tilde{\theta} \in \Theta^{\{1,2\}}_{I,cex}(1)$ and suppose that $P\{Z \in Q(\tilde{\theta})\} = 0$. Then $\theta \overset{o.e.}{\sim}_{\mathcal{F}} \tilde{\theta}$ under both $\mathcal{F}_{cex}$ and $\mathcal{F}_{ciid}$.*

Proof: We start with $\mathcal{F}_{cex}$: we want to show that if $P\{Z \in Q(\tilde{\theta})\} = 0$, then there exists $G \in \mathcal{F}_{cex}$ such that $p(y, x|\tilde{\theta}, G) = p(y, x|\theta, F)$ where $F \in \mathcal{F}_{cex}$ is the true distribution that generated the data. We first fix α and $z = (y_0, x)$, and construct $G_{u|\alpha,z}(u_1, u_2; \theta, F)$ in such a way that for any α and z:

$$\begin{aligned}
&P(u_1 > x_1'\tilde{\beta} + \tilde{\gamma}y_0 + \alpha, u_2 > x_2'\tilde{\beta} + \alpha|\alpha, z; G) = P(u_1 > x_1'\beta + \gamma y_0 + \alpha, u_2 > x_2'\beta + \alpha|\alpha, z; F)\\
&P(u_1 > x_1'\tilde{\beta} + \tilde{\gamma}y_0 + \alpha, u_2 \leq x_2'\tilde{\beta} + \alpha|\alpha, z; G) = P(u_1 > x_1'\beta + \gamma y_0 + \alpha, u_2 \leq x_2'\beta + \alpha|\alpha, z; F)\\
&P(u_1 \leq x_1'\tilde{\beta} + \tilde{\gamma}y_0 + \alpha, u_2 > x_2'\tilde{\beta} + \tilde{\gamma} + \alpha|\alpha, z; G) = P(u_1 \leq x_1'\beta + \gamma y_0 + \alpha, u_2 > x_2'\beta + \gamma + \alpha|\alpha, z; F)
\end{aligned} \tag{3.3}$$

Since for almost all z, $z \notin Q(\tilde{\theta})$, we will be able to find $G_{u|\alpha,z}(u_1, u_2; \theta, F)$ that satisfies (3.3) and exchangeability condition. Finally, let $G_{\alpha|z} = F_{\alpha|z}$. Then $p(y, x|\theta, F) = p(y, x|\tilde{\theta}, G)$ by construction.

Now, suppose that $F \in \mathcal{F}_{ciid}$. For any fixed α and z, we define $G_{u_1|\alpha,z}(\cdot; \theta, F)$ in such a way that

$$G_{u_1|\alpha,z}(x_1'\tilde{\beta} + \tilde{\gamma}y_0 + \alpha; \theta, F) = F_{u_1|\alpha,z}(x_1'\beta + \gamma y_0 + \alpha)$$

similarly, we define $G_{u_2|\alpha,z}$ so that the following holds:

$$G_{u_2|\alpha,z}(x_2'\tilde{\beta}+\tilde{\gamma}d+\alpha;\theta,F)=F_{u_2|\alpha,z}(x_2'\beta+\gamma d+\alpha) \text{ for } d=0,1$$

We have only the following three restrictions on $G_{u_t|\alpha,z}$:

- $0 \leq G_{u_t|\alpha,z}(\cdot;\theta,F) \leq 1$;
- $G_{u_t|\alpha,z}(a;\theta,F)$ is strictly increasing in a;
- $G_{u_1|\alpha,z}(\cdot;\theta,F)=G_{u_2|\alpha,z}(\cdot;\theta,F)$.

The first restriction obviously holds. Also, since $F_{u_1|\alpha,z}=F_{u_2|\alpha,z}$ and $z \notin Q(\tilde{\theta})$, then one can find $G_{u_t|\alpha,z}$ that satisfies the last restriction. By construction, $G_{u_1|\alpha,z}(a;\theta,F)$ is pinned down only at $a=x_1'\tilde{\beta}+\tilde{\gamma}y_0$, while $G_{u_2|\alpha,z}(a;\theta,F)$ is pinned down either at $a=x_2'\tilde{\beta}$ or at $a=x_2'\tilde{\beta}+\tilde{\gamma}$. Since $\tilde{\theta}\in\Theta_{I,cex}^{T=2}$ and $P\{Z\in Q(\tilde{\theta})\}=0$, we'll be able to find $G_{u_t|\alpha,z}$ that satisfies the monotonicity restriction. Finally, we define $G_{u_1,u_2|\alpha,z}(a_1,b_1)=G_{u_1|\alpha,z}(a_1)G_{u_2|\alpha,z}(b_1)$. That is, by construction this distribution satisfies serial independence. If we choose $G_{\alpha|z}=F_{\alpha|z}$, then by construction $p(y,x|\theta,F)=p(y,x|\tilde{\theta},G)$ for almost all x. That is, $\tilde{\theta}$ is observationally equivalent to the true parameter θ. ■

Finally, it is easy to see that Lemmas 3.2, 3.3 and 3.4 immediately imply Theorem 3.2.

3.3 Independence with $T=2$: a Non-Stationary Model

Next, we strengthen the exchangeability assumption (CEX) to full independence assumption (IND). This model of independence, unlike the stationary and exchangeability models previously, does not restrict the correlation between u_1 and u_2. So, we have a non-stationary model with independence.

Proposition 3.2. *Assume that $u=(u_1,u_2)$ is independent from (x,y_0). Then for pairs of (x,y_0) and $(\tilde{x},\tilde{y}_0)$, the parameters β and γ satisfy the following restrictions:*

(i) If $p_2(1,0|x,y_0)\geq 1-p_2(0,1|\tilde{x},\tilde{y}_0)$, then $((x_1-x_2)-(\tilde{x}_1-\tilde{x}_2))'\beta+\gamma(y_0-\tilde{y}_0-1)\geq 0$.

(ii) If $p_2(1,0|x,y_0)> 1-p_2(0,1|\tilde{x},\tilde{y}_0)$, then $((x_1-x_2)-(\tilde{x}_1-\tilde{x}_2))'\beta+\gamma(y_0-\tilde{y}_0-1)> 0$.

Proof: Event $(u_1 \leq \alpha + x_1'\beta + \gamma y_0, u_2 > \alpha + x_2'\beta + \gamma)$ implies that $u_1 - u_2 \leq (x_1 - x_2)'\beta + \gamma(y_0 - 1)$. That is,

$$P(1,0|\alpha,x,y_0) \leq P(u_1 - u_2 \leq (x_1 - x_2)'\beta + \gamma(y_0 - 1)|\alpha,x,y_0)$$

Integrating α out, we get

$$p_2(1,0|x,y_0) \leq P(u_1 - u_2 \leq (x_1 - x_2)'\beta + \gamma(y_0 - 1)|x,y_0) = F_{12}((x_1 - x_2)'\beta + \gamma(y_0 - 1))$$

where $F_{12}(\cdot)$ is the cdf of $u_1 - u_2$ and the last equality follows from the fact that $u_1 - u_2$ is independent from (x, y_0). So, the following inequality that must hold for any x and y_0 in the support:

$$p_2(1,0|x,y_0) \leq F_{12}((x_1 - x_2)'\beta + \gamma(y_0 - 1))$$

Similarly, we can show that

$$P(0,1|\alpha,x,y_0) \leq P(u_1 - u_2 > (x_1 - x_2)'\beta + \gamma y_0|\alpha,x,y_0)$$

so that we have the following:

$$F_{12}((x_1 - x_2)'\beta + \gamma y_0) \leq 1 - p_2(0,1|x,y_0)$$

To sum up, for all x, y_0 in the support we have the following:

$$\begin{aligned} p_2(1,0|x,y_0) &\leq F_{12}((x_1 - x_2)'\beta + \gamma(y_0 - 1)) \\ F_{12}((x_1 - x_2)'\beta + \gamma y_0) &\leq 1 - p_2(0,1|x,y_0) \end{aligned}$$

Conditions (i) and (ii) immediately follow from the fact that $F_{12}(\cdot)$ is a strictly increasing function. $\square$

Remark 3.1. *Note that the independence condition of Proposition 3.2 keeps the independence assumption form Honoré and Kyriazidou (2000) but relaxes stationarity.*

Given that $y_0, \tilde{y}_0$ are binary, by looking at (i) and (ii) in the Proposition above, we that only the sign of γ will be identified, but we may get some meaningful identification for β. We can also potentially add more restrictions to shrink the identified set even further, For example, if we assume that $Med(u_1 - u_2|x, y_0) = 0$, then the following must hold for any x, y_0 in the support:

(i) If $p_2(1, 0|x, y_0) \geq 0.5$, then $(x_1 - x_2)'\beta + \gamma(y_0 - 1) \geq 0$.

(ii) If $p_2(0, 1|x, y_0) \geq 0.5$, then $(x_1 - x_2)'\beta + \gamma y_0 \leq 0$.

Now we are ready to discuss identifying power of the semiparametric model in Honoré and Kyriazidou (2000), where error terms u_1, u_2 are assumed to be iid and independent from α, x, y_0.

In their semiparametric model setup, Honoré and Kyriazidou (2000) make two assumptions about unobservable components u_1, u_2:

(IND) u_1, u_2 are independent from α, x, y_0;

(TIID) u_1 and u_2 are mutually independent (independent over time) and identically distributed (stationarity).

So far, we showed that condition (TIID) does not have any additional identifying power over the conditional exchangeability. However, independence condition (IND) can help to shrink the identified set in Theorem 3.2.

3.4 Point Identification: T=2 case

We explore the question of whether and under what conditions do the sharp sets characterized in Theorems 3.1 and 3.2 shrink to a point. Naturally, we expect sufficient point identification conditions to rely on enough variation in the regressor distribution. As will be shown, it turns out that with $T = 2$, under stationarity, β can be point identified and γ can generally not be point identified. This is in contrast to the $T = 2$ with exchangeability where both β and γ can be point identified, though as will be shown, the latter only can when it is nonnegative.

3.4.1 Point Identification under Stat (T=2)

There are interesting implications of the conclusion of Theorem 3.1. Most notable, we establish here that under support conditions on x the parameter vector β can be *point* identified, but γ *cannot*, though its *sign* can be. We will formally state these results here as a theorem, whose proof is left to the appendix. Before doing so, we note that the parameters γ, β are only identified up to scale which is a feature in binary response models. Consequently, we impose scale normalizations here. As is often followed in the literature (see, e.g. Horowitz (1992), Honoré and Kyriazidou (2000)) we decompose the vector x_t into $x_{t1}, \tilde{x}_t$ where x_{t1} denotes the first component of x_t and $\tilde{x}_t$ denotes its remaining components. We will normalize the coefficient of x_{t1} to 1, and denote the remaining components of β by θ. Therefore, we can write $x_t'\beta = x_{t1} + \tilde{x}_t'\tilde{\theta}$. Before stating the results on point identification, we make the following assumptions regarding $x \equiv (x_1, x_2)$, whose support we denote by $\mathcal{X}$

Assumption 3.2. *Define the following subsets*[6] *of* $\mathcal{X}$*:*

1. $\mathcal{X}_1 = \{x \in \mathcal{X} : P(y_2 = 1|x) \geq P(y_1 = 1|x)\}$
2. $\mathcal{X}_2 = \{x \in \mathcal{X} : P(y_1 = 1|x) \geq P(y_2 = 1|x)\}$
3. $\mathcal{X}_{3a} = \{x \in \mathcal{X} : P(y_1 = 0, y_2 = 1|x) \geq P(y_1 = 1|x)\}$
4. $\mathcal{X}_{3b} = \{x \in \mathcal{X} : P(y_0 = 1, y_1 = 0|x) \geq P(y_2 = 0|x)\}$
5. $\mathcal{X}_{4a} = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 0|x) \geq P(y_1 = 0|x)\}$
6. $\mathcal{X}_{4b} = \{x \in \mathcal{X} : P(y_0 = 0, y_1 = 1|x) \geq P(y_2 = 1|x)\}$
7. $\mathcal{X}_5 = \{x \in \mathcal{X} : P(y_0 = 0, y_1 = 0|x) \geq P(y_2 = 0|x)\}$
8. $\mathcal{X}_6 = \{x \in \mathcal{X} : P(y_0 = 1, y_1 = 1|x) \geq P(y_2 = 1|x)\}$
9. $\mathcal{X}_7 = \{x \in \mathcal{X} : P(y_0 = 0, y_1 = 0|x) + P(y_1 = 0, y_2 = 1|x) \geq 1\}$
10. $\mathcal{X}_8 = \{x \in \mathcal{X} : P(y_0 = 1, y_1 = 1|x) + P(y_1 = 1, y_2 = 0|x) \geq 1\}$
11. $\mathcal{X}_9 = \{x \in \mathcal{X} : P(y_0 = 1, y_1 = 0|x) + P(y_1 = 0, y_2 = 1|x) \geq 1\}$
12. $\mathcal{X}_{10} = \{x \in \mathcal{X} : P(y_0 = 0, y_1 = 1|x) + P(y_1 = 1, y_2 = 0|x) \geq 1\}$

PID1 *Assume each of* $\mathcal{X}_j$*,* $j = 1 : 10$ *have positive measure, and furthermore, assume* $\mathcal{X}_{7,8} \equiv \{x : x \in \mathcal{X}_7 \cap \mathcal{X}_8)\}$ *has positive measure.*

[6] We note the subsets here correspond directly to the implications in Theorem 3.1.

PID2 *For each* $x \in \mathcal{X}_j$ $j = 1 : 10, \mathcal{X}_{7,8}$, *conditional on* α, x_{11}, x_{21}, $\tilde{x}_2 - \tilde{x}_1$ *has full column rank.*

PID3 *For each* $x \in \mathcal{X}_j$ $j = 1 : 10, \mathcal{X}_{7,8}$, *conditional on* $\tilde{x}_1, \tilde{x}_2, \alpha$, $x_{21} - x_{11}$ *has support on the real line.*

Under the above conditions we can attain point identification for β, though not γ. This is stated in the following theorem, whose proof is delegated to the Appendix.

Theorem 3.3. *Suppose Assumptions* $A1 - A4$, *and Assumption 3.2 hold. Then* β *is point identified;* γ *is not point identified but its sign can be.*

Remark 3.2. *The point identification in the Theorem above relies on large support for* $x_{21} - x_{11}$ *after conditioning on subsets (assumed to be nonempty) of the regressor space* $\mathcal{X}$. *This type of condition is analogous to that assumed in Manski (1987) for establishing point identification in the static binary choice panel data model.*

Remark 3.3. *As also discussed in the proof in the Appendix, it may be possible to attain point identification of* γ *under additional conditions. Specifically, these would involve further "intersection" subsets, such as, for example* $\mathcal{X}_{10} \cap \mathcal{X}_1$ *having positive measure. However, as explained in Appendix such a condition can generally only be true if the corresponding probabilities in Theorem 3.1 attain their limiting values of 0 and 1. As such, we are not comfortable concluding point identification for* γ *under the our stationarity condition when* $T = 2$. *As we will see in the next section, point identification of* γ *will be more easily attainable when we either strengthen the stationarity condition to exchangeability, or increase the panel width from* $T = 2$ *to* $T = 3$.

Next, we examine point identification under exchangeability.

3.4.2 Point Identification under Exchangeabililty (T=2)

As was the case under the STAT assumption, there are interesting special cases of the conclusion of Theorem 3.2. Specifically, like there, we can attain point identification results with additional conditions on observed regressors. However, in contrast to our results with stationarity, under CEX(T=2) of Definition 1 we can now attain point identification of both $\tilde{\beta}$ and $\tilde{\gamma}$. This powerful result illustrates the informational content of CEX, and is based on the following conditions stated as an Assumption.

Assumption 3.3. *Let the following hold.*

P2ID1 *Define the following subsets of $\mathcal{X}$:*

1. $\mathcal{X}2_1 = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 0|y_0, x) \geq P(y_1 = 0, y_2 = 1|y_0, x)\}$
2. $\mathcal{X}2_2 = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 0|y_0, x) \leq P(y_1 = 0, y_2 = 1|y_0, x)\}$
3. $\mathcal{X}2_3 = \{x \in \mathcal{X} : P(y_1 = 1|y_0, x) \geq P(y_2 = 1|y_0, x)\}$
4. $\mathcal{X}2_4 = \{x \in \mathcal{X} : P(y_1 = 1|y_0, x) \leq P(y_2 = 1|y_0, x)\}$
5. $\mathcal{X}2_5 = \{x \in \mathcal{X} : P(y_1 = 0, y_2 = 1|y_0, x) \leq P(y_1 = 1|y_0, x)\}$
6. $\mathcal{X}2_6 = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 0|y_0, x) \leq P(y_1 = 0|y_0, x)\}$

Assume each of $\mathcal{X}2_j,\ j = 1 : 6$, has positive measure.

P2ID2 *For $x \in \mathcal{X}2_j$ $j = 1 : 6$, conditional on α, x_{11}, x_{12}, $\tilde{x}_2 - \tilde{x}_1$ has full column rank.*

P2ID3 *For $x \in \mathcal{X}2_j$ $j = 1 : 6$, conditional on $\tilde{x}_1, \tilde{x}_2, \alpha$, $x_{21} - x_{11}$ has support on the real line.*

We state the point identification result in the next Theorem.

Theorem 3.4. *Suppose Assumptions $A1 - A4$, Definition 1, and 3.3 hold. Then both β and the sign of γ are point identified. Furthermore, if $\gamma \geq 0$, then γ is point identified as well.*

3.4.3 Comparison to Honore and Kyriazidou

Our result in Theorem 3.4 and the conditions it is based on are worth comparing to Honoré and Kyriazidou (2000) (HK hereafter) , who also consider identification and estimation of dynamic binary choice models with fixed effects. First, with $T = 2$, citing the important result in Chamberlain (1993), HK conclude that the parameters $\tilde{\beta}, \tilde{\gamma}$ are *not* point identified, even under the *parametric* assumption that the disturbance terms are conditionally iid logistic, which is a stronger than the assumption in Definition 1 imposed here. The lack of point identification motivated HK to pursue identification results for the $T = 3$ case. Our explanation on the difference in identification conclusions would primarily be based on the support conditions on the strictly exogenous variables x_{it}. For example, in HK's conditional logit model, x_{it} can all be discrete with bounded support, which is ruled out by our assumption P2ID3. But given the distribution-free nature of our assumption, such an Assumption P2ID3 is necessary to attain point identification, even for *static* binary choice models, as shown in Chamberlain (2010). In fact, HK impose a regressor support condition like our Assumption P2ID3 when they relax the conditional Logit assumption in their model with $T = 3$.

Second, HK attain point identification for the dynamic binary choice model but only with $T \geq 3$, which we do not require in Theorem 3.4. Furthermore, the identification result and estimation procedure in HK requires *each* component of *all* exogenous variables x_{it} satisfy a time varying overlap support condition in 2 of the periods. This is very restrictive as it rules out, for example, time trends. Such a strong condition is not imposed in Assumptions P2ID1-P2ID3, so here we can attain point identification even when components of x_{it} have support that vary with t. So, for example one component of x_{it} could be t itself in our Theorem 3.4, which is in contrast to the results in HK. In fact, from a partial identification perspective, their proposed procedure will not even provide a bound for the coefficients if there is no overlap in the support of x_{it} across time.

4 Additional Time Periods

In the previous section we explored identified regions for dynamic binary choice models with fixed effects under a varying assumptions, generally showing that these regions shrink as we strengthened our conditions. In this section we extend the model in the time dimension. Specifically we will explore the identifying power of time, by considering models where $T \geq 3$.. As we will show, there is identifying power in observing agents across more time periods. Our most striking finding is that in the dynamic binary choice model with $T \geq 3$ the regression coefficients can be point identified, even under conditions analogous to Assumption 3.1. Note this is in contrast to what we found for the $T = 2$ case.

4.1 Stationarity and Exchangeability with $T = 3$

In this section we explore the informational content of observing data over additional time periods, so, for example $T = 3$. It is clear that having access to longer horizons for individuals would have more identifying power. For example, in their model, Honoré and Kyriazidou (2000) demonstrate how β, γ could be point identified in the $T = 3$ case but not the $T = 2$ setting. Here we will derive the identified region for β, γ for $T \geq 3$ under our stationarity condition and exchangeability condition from before. We start with the the identified set for $\theta \equiv (\beta, \gamma)$ under the (STAT) assumption which is given in the result below.

Theorem 4.1. *Let $\Theta_{I,stat}^{\{2,3\}}$ be the set of $\theta \in \Theta$ that is constructed in a same way as $\Theta_{I,stat}^{\{1,2\}}$ by shifting the time subscript by +1 (for details see the proof below). Additionally, let $\Theta_{I,stat}^{\{1,3\}}$ be the set of parameters that satisfy the following restrictions: if for some x,*

(1) $P(y_3 = 1|x) \geq P(y_1 = 1|x) \Rightarrow (x_3 - x_1)'\beta + |\gamma| \geq 0$;

(2) $P(y_1 = 1|x) \geq P(y_3 = 1|x) \Rightarrow (x_3 - x_1)'\beta - |\gamma| \leq 0$;

(3) $P(y_0 = 0, y_1 = 0|x) \geq P(y_3 = 0|x)$ *or* $P(y_2 = 1, y_3 = 1|x) \geq P(y_1 = 1|x) \Rightarrow (x_3 - x_1)'\beta + \max\{0, \gamma\} \geq 0$;

(4) $P(y_0 = 1, y_1 = 1|x) \geq P(y_3 = 1|x)$ *or* $P(y_2 = 0, y_3 = 0|x) \geq P(y_1 = 0|x) \Rightarrow (x_3 - x_1)'\beta - \max\{0, \gamma\} \leq 0$;

(5) $P(y_0 = 1, y_1 = 0|x) \geq P(y_3 = 0|x)$ *or* $P(y_2 = 0, y_3 = 1|x) \geq P(y_1 = 1|x) \Rightarrow (x_3 - x_1)'\beta - \min\{0, \gamma\} \geq 0$;

(6) $P(y_0 = 0, y_1 = 1|x) \geq P(y_3 = 1|x)$ *or* $P(y_2 = 1, y_3 = 0|x) \geq P(y_1 = 0|x) \Rightarrow (x_3 - x_1)'\beta + \min\{0, \gamma\} \leq 0$;

(7) $P(y_2 = 1, y_3 = 1|x) + P(y_0 = 1, y_1 = 0|x) \geq 1 \Rightarrow (x_3 - x_1)'\beta \geq 0$;

(8) $P(y_2 = 1, y_3 = 0|x) + P(y_0 = 1, y_1 = 1|x) \geq 1 \Rightarrow (x_3 - x_1)'\beta \leq 0$;

(9) $P(y_2 = 1, y_3 = 1|x) + P(y_0 = 0, y_1 = 0|x) \geq 1 \Rightarrow (x_3 - x_1)'\beta + \gamma \geq 0$;

(10) $P(y_2 = 1, y_3 = 0|x) + P(y_0 = 0, y_1 = 1|x) \geq 1 \Rightarrow (x_3 - x_1)'\beta + \gamma \leq 0$;

(11) $P(y_2 = 0, y_3 = 1|x) + P(y_0 = 1, y_1 = 0|x) \geq 1 \Rightarrow (x_3 - x_1)'\beta - \gamma \geq 0$;

(12) $P(y_2 = 0, y_3 = 0|x) + P(y_0 = 1, y_1 = 1|x) \geq 1 \Rightarrow (x_3 - x_1)'\beta - \gamma \leq 0$.

Then $\Theta_{I,stat}^{T=3} = \Theta_{I,stat}^{\{1,2\}} \cap \Theta_{I,stat}^{\{2,3\}} \cap \Theta_{I,stat}^{\{1,3\}}$ is the sharp identified set for θ under the assumption that u_1, u_2, u_3 are identically distributed conditional on x and α.

Proof: the proof closely follows the proof of Theorem 3.1, with the addition of the following sharp restrictions on the marginal distribution of $v_3 = u_3 - \alpha$ (conditional on x and α):

$$\begin{aligned}
&F(x_3'\beta + \min\{0,\gamma\}) \le P(y_3 = 1|x) \\
P(y_3 = 1|x) \le &F(x_3'\beta + \max\{0,\gamma\}) \\
P(y_2 = 1, y_3 = 1|x) \le &F(x_3'\beta + \gamma|x) \\
&F(x_3'\beta + \gamma|x) \le 1 - P(y_2 = 1, y_3 = 0|x) \\
P(y_2 = 0, y_3 = 1|x) \le &F(x_3'\beta|x) \\
&F(x_3'\beta|x) \le 1 - P(y_2 = 0, y_3 = 0|x)
\end{aligned}$$

Combining these restrictions with restrictions for v_1 and the conditional stationarity assumption gives us $\Theta^{\{1,3\}})I, stat$; and $\Theta^{\{2,3\}}_{I,stat}$ is obtained by combining these restrictions with the restrictions for v_2. In particular, $\Theta^{\{2,3\}}_{I,stat}$ is given by the restrictions: if for some x,

(1) $P(y_3 = 1|x) \ge P(y_2 = 1|x) \Rightarrow (x_3 - x_2)'\beta + |\gamma| \ge 0$;

(2) $P(y_2 = 1|x) \ge P(y_3 = 1|x) \Rightarrow (x_3 - x_2)'\beta - |\gamma| \le 0$;

(3) $P(y_2 = 0, y_3 = 1|x) \ge P(y_2 = 1|x)$ or $P(y_1 = 1, y_2 = 0|x) \ge P(y_3 = 0|x) \Rightarrow (x_3 - x_2)'\beta - \min\{0,\gamma\} \ge 0$;

(4) $P(y_2 = 1, y_3 = 0|x) \ge P(y_2 = 0|x)$ or $P(y_1 = 0, y_2 = 1|x) \ge P(y_3 = 1|x) \Rightarrow (x_3 - x_2)'\beta + \min\{0,\gamma\} \le 0$;

(5) $P(y_1 = 0, y_2 = 0|x) \ge P(y_3 = 0|x) \Rightarrow (x_3 - x_2)'\beta + \max\{0,\gamma\} \ge 0$;

(6) $P(y_1 = 1, y_2 = 1|x) \ge P(y_3 = 1|x) \Rightarrow (x_3 - x_2)'\beta - \max\{0,\gamma\} \le 0$;

(7) $P(y_1 = 0, y_2 = 0|x) + P(y_2 = 0, y_3 = 1|x) \ge 1 \Rightarrow (x_3 - x_2)'\beta \ge 0$;

(8) $P(y_1 = 1, y_2 = 1|x) + P(y_2 = 1, y_3 = 0|x) \ge 1 \Rightarrow (x_3 - x_2)'\beta \le 0$;

(9) $P(y_1 = 1, y_2 = 0|x) + P(y_2 = 0, y_3 = 1|x) \ge 1 \Rightarrow (x_3 - x_2)'\beta - \gamma \ge 0$;

(10) $P(y_1 = 0, y_2 = 1|x) + P(y_2 = 1, y_3 = 0|x) \ge 1 \Rightarrow (x_3 - x_2)'\beta + \gamma \le 0$

■

Again, the above adds to the previous stationarity Theorem with $T = 2$ a set of moment inequalities that shrink the set. It is also intuitive that as we add more time periods, similar type inequalities can be added.

Next, we provide the identified set in the exchangeability case with T=3. Similar to the two-period case, exchangeability assumption will help to further shrink the identified set. Indeed, when $T = 3$, conditional exchangeability assumption now involves 12 addition moment inequalities of the "if,then" restrictions that are summarized in Table 1.

Proposition 4.1. *Suppose that $u = (u_1, u_2, u_3)$ is exchangeable conditional on α, x, y_0. Then parameters β and γ must satisfy the restrictions in Table 1 for every (x, y_0) in the support.*

Note that the 3-exchangeability result in Proposition 4.1 also implies the 2-exchangeability result from Proposition 3.1. Additionally, we can obtain Honoré and Kyriazidou (2000) point identifying restrictions as a particular implication of Proposition 4.1, as summarized below.

Remark 4.1. *When $x_2 = x_3$, restrictions for event pairs $\{(0,1,0),(1,0,0)\}$ and $\{(0,1,1),(1,0,1)\}$ in Table 1 reduce to the following set:*

(i) If $(x_1 - x_2)'\beta + \gamma(y_0 - 1) \geq 0$, then $p_3(1,0,1|\alpha, x, y_0) \geq p_3(0,1,1|\alpha, x, y_0)$

(ii) If $(x_1 - x_2)'\beta + \gamma(y_0 - 1) \leq 0$, then $p_3(1,0,1|\alpha, x, y_0) \leq p_3(0,1,1|\alpha, x, y_0)$

(iii) If $(x_1 - x_2)'\beta + \gamma y_0 \geq 0$, then $p_3(1,0,0|\alpha, x, y_0) \geq p_3(0,0,1|\alpha, x, y_0)$.

(iv) If $(x_1 - x_2)'\beta + \gamma y_0 \leq 0$, then $p_3(1,0,0|\alpha, x, y_0) \leq p_3(0,0,1|\alpha, x, y_0)$.

That is, in Honoré and Kyriazidou (2000) notation:

$$P(A|\alpha, x, y_0, x_2 = x_3) \gtreqless P(B|\alpha, x, y_0, x_2 = x_3) \text{ iff } (x_2 - x_1)'\beta + \gamma(y_3 - y_0) \gtreqless 0$$

Honoré and Kyriazidou (2000) obtain point identification from these inequalities under two additional assumptions: that conditional on α, x, y_0 distribution of u is absolutely continuous (CS2), and that $x_{21} \equiv x_2 - x_1$ contains a continuously distributed component (CS3). Theorem 4.2 shows that we can also get rid of the first "i" in "iid" assumed in HK by weakening it to conditional exchangeability.

Next, we combine 3-exchangeability with conditional stationarity of u_1, u_2 and u_3 to obtain the identified set that is sharp under general exchangeability assumption.

Theorem 4.2. *Let $\Theta_{I,cex}^{\{1,2,3\}}$ be the set of $\theta \in \Theta$ that satisfy restrictions in Proposition 4.1. Also, let $\Theta_{I,cex}^{\{1,3\}}(1)$ be the set of parameters that satisfy the following restrictions: if for some x, y_0,*

(1) $P(y_3 = 1|x, y_0) \geq P(y_1 = 1|x, y_0) \Rightarrow (x_3 - x_1)'\beta + \max\{0, \gamma\} - \gamma y_0 \geq 0$;

(2) $P(y_1 = 1|x, y_0) \geq P(y_3 = 1|x, y_0) \Rightarrow (x_3 - x_1)'\beta + \min\{0, \gamma\} - \gamma y_0 \leq 0$;

(3) $P(y_2 = 0, y_3 = 1|x, y_0) \geq P(y_1 = 1|x, y_0) \Rightarrow (x_3 - x_1)'\beta - \gamma y_0 \geq 0$;

(4) $P(y_2 = 0, y_3 = 0|x, y_0) \geq P(y_1 = 0|x, y_0) \Rightarrow (x_3 - x_1)'\beta - \gamma y_0 \leq 0$;

(5) $P(y_2 = 1, y_3 = 1|x, y_0) \geq P(y_1 = 1|x, y_0) \Rightarrow (x_3 - x_1)'\beta + \gamma - \gamma y_0 \geq 0$;

(6) $P(y_2 = 1, y_3 = 0|x, y_0) \geq P(y_1 = 0|x, y_0) \Rightarrow (x_3 - x_1)'\beta + \gamma - \gamma y_0 \leq 0$.

Finally, let $\Theta_{I,cex}^{\{2,3\}}(1)$ satisfy the restrictions for $\Theta_{I,stat}^{\{2,3\}}$ in Theorem 4.1, only with the conditional on x probabilities replaced by the conditional on $z = (x, y_0)$ probabilities. Then $\Theta_{I,cex}^{T=3} = \Theta_{I,cex}^{\{1,2,3\}} \cap \Theta_{I,cex}^{\{1,2\}}(1) \cap \Theta_{I,cex}^{\{2,3\}}(1) \cap \Theta_{I,cex}^{\{1,3\}}(1)$ is the sharp identified set for θ under the assumption that u_1, u_2, u_3 are identically distributed conditional on x and α.

4.2 Point Identification with $T = 3$

In this section we provide interpretable sufficient conditions for point identification of the parameters β, γ under the (STAT) and exchangeability assumptions in the case for $T = 3$. As in our proofs of point identification before we will condition on subsets of the regressor space, so we assume these subsets have positive measure. For the case where Assumption (STAT) holds and $T = 3$ we make the following assumptions

P3ID1 Define the following subsets of $\mathcal{X}$:

1. $\mathcal{X}3_1 = \{x \in \mathcal{X} : P(y_3 = 1|x) \geq P(y_1 = 1|x)\}$
2. $\mathcal{X}3_2 = \{x \in \mathcal{X} : P(y_1 = 1|x) \geq P(y_3 = 1|x)\}$
3. $\mathcal{X}3_3 = \{x \in \mathcal{X} : P(y_0 = 0, y_1 = 0|x) \geq P(y_3 = 0|x)$ or $P(y_2 = 1, y_3 = 1|x) \geq P(y_1 = 1|x)\}$

4. $\mathcal{X}3_4 = \{x \in \mathcal{X} : P(y_0 = 1, y_1 = 1|x) \geq P(y_3 = 1|x)$ or $P(y_2 = 0, y_3 = 0|x) \geq P(y_1 = 0|x)\}$
5. $\mathcal{X}3_5 = \{x \in \mathcal{X} : P(y_0 = 1, y_1 = 0|x) \geq P(y_3 = 0|x)$ or $P(y_2 = 0, y_3 = 1|x) \geq P(y_1 = 0|x)\}$
6. $\mathcal{X}3_6 = \{x \in \mathcal{X} : P(y_0 = 0, y_1 = 1|x) \geq P(y_3 = 1|x)$ or $P(y_2 = 1, y_3 = 0|x) \geq P(y_1 = 0|x)\}$
7. $\mathcal{X}3_7 = \{x \in \mathcal{X} : P(y_2 = 1, y_3 = 1|x) + P(y_0 = 1, y_1 = 0|x) \geq 1\}$
8. $\mathcal{X}3_8 = \{x \in \mathcal{X} : P(y_2 = 1, y_3 = 0|x) + P(y_0 = 1, y_1 = 1|x) \geq 1\}$
9. $\mathcal{X}3_9 = \{x \in \mathcal{X} : P(y_2 = 1, y_3 = 1|x) + P(y_0 = 0, y_1 = 0|x) \geq 1\}$
10. $\mathcal{X}3_{10} = \{x \in \mathcal{X} : P(y_2 = 1, y_3 = 0|x) + P(y_0 = 0, y_1 = 1|x) \geq 1\}$
11. $\mathcal{X}3_{11} = \{x \in \mathcal{X} : P(y_2 = 0, y_3 = 1|x) + P(y_0 = 1, y_1 = 0|x) \geq 1\}$
12. $\mathcal{X}3_{12} = \{x \in \mathcal{X} : P(y_2 = 0, y_3 = 0|x) + P(y_0 = 1, y_1 = 1|x) \geq 1\}$

Then we assume each of $\mathcal{X}3_j,\ j = 1 : 12$, has positive measure, as does $\mathcal{X}3_7 \cap \mathcal{X}3_8$ and $\mathcal{X}3_{11} \cap \mathcal{X}3_{12}$

P3ID2 For $x \in \mathcal{X}3_j$ $j = 1 : 12$, conditional on α, x_{11}, x_{12}, $\tilde{x}_2 - \tilde{x}_1$ has full column rank.

P3ID3 For $x \in \mathcal{X}3_j$ $j = 1 : 12$, conditional on $\tilde{x}_1, \tilde{x}_2, \alpha$, $x_{21} - x_{11}$ has support on the real line.

P3ID4 Conditional on $\{x : x \in \mathcal{X}3_{11} \cap \mathcal{X}3_{12} \cap \mathcal{X}_9 \cap \mathcal{X}_{10}\}$, $(x_2 - x_1)'\beta - (x_3 - x_1)'\beta$ has positive density in a neighborhood of 0.

Theorem 4.3. *Suppose Assumptions $A1 - A4$, $PID1 - PID3$ and $P3ID1 - P3ID4$ hold. Then both $\tilde{\beta}, \tilde{\gamma}$ are point identified.*

We compare the results from this Theorem to those attained with $T = 2$ and those attained in Honoré and Kyriazidou (2000) with $T = 3$.

Remark 4.2. *We see from Theorem 4.3 that both β and γ can be point identified under the Assumption 3.1, when $T = 3$. This is in contrast to the result from Theorem 3.3, where only β (but not γ) could be point identified under Assumption 3.1, when $T = 2$. Thus our results demonstrate the informational content of observing data over additional time periods.*

Remark 4.3. *Assumption P3ID4 in Theorem 4.3 also imposes an overlapping support condition of the indexes $(x_2 - x_1)'\beta$ and $(x_3 - x_1)'\beta$. This does not require the regressors themselves have overlapping support over time, and thus permits time trends as covariates.*

Remark 4.4. *Honoré and Kyriazidou (2000) also attain point identification results for both β and γ when $T = 3$. However, their result is based on stronger conditions than ours. They assume that ϵ_{it} (the disturbance term in their notation) is conditionally independent and identically distributed which is much more restrictive than our stationarity condition. They also assume that ϵ_{it} is independent of x which we do not. Furthermore, they assume x_{it} satisfies a serial overlap condition, effectively assuming the support of the regressors is constant over time. This too is a very restrictive condition, ruling out for example $x_{it} = t$, or time trends as regressors. Theorem 4.3 imposes no such conditions. In addition, when the support conditions in the Theorem do not hold, Theorem 4.1 provides the sharp set without any assumptions on the support of the regressors.*

We conclude this section by exploring point identification for the case when $T = 3$ and we impose our exchangeability condition as in Proposition 4.1. Recall that under the exchangeability condition and $T = 2$ we were able to attain point identification for both β and γ when γ was nonnegative. As we will see here, having an additional time period will enable us to attain point identification even when $\gamma < 0$. This will even be true without the index overlap condition we assumed when $T = 3$ and stationarity. Our point identification result is based on the following conditions:

P4ID1 Define the following subsets of $\mathcal{X}$:

1. $\mathcal{X}4_1 = \{x \in \mathcal{X} : P(y_1 = 0, y_2 = 1, y_3 = 0|y_0, x) \geq P(y_1 = 0, y_2 = 0, y_3 = 1|y_0, x)\}$
2. $\mathcal{X}4_2 = \{x \in \mathcal{X} : P(y_1 = 0, y_2 = 1, y_3 = 0|y_0, x) \leq P(y_1 = 0, y_2 = 0, y_3 = 1|y_0, x)\}$
3. $\mathcal{X}4_3 = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 0, y_3 = 0|y_0, x) \geq P(y_1 = 0, y_2 = 0, y_3 = 1|y_0, x)\}$
4. $\mathcal{X}4_4 = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 0, y_3 = 0|y_0, x) \leq P(y_1 = 0, y_2 = 0, y_3 = 1|y_0, x)\}$
5. $\mathcal{X}4_5 = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 0, y_3 = 0|y_0, x) \geq P(y_1 = 0, y_2 = 1, y_3 = 0|y_0, x)\}$
6. $\mathcal{X}4_6 = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 0, y_3 = 0|y_0, x) \leq P(y_1 = 0, y_2 = 1, y_3 = 0|y_0, x)\}$
7. $\mathcal{X}4_7 = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 0, y_3 = 1|y_0, x) \geq P(y_1 = 0, y_2 = 1, y_3 = 1|y_0, x)\}$
8. $\mathcal{X}4_8 = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 0, y_3 = 1|y_0, x) \leq P(y_1 = 0, y_2 = 1, y_3 = 1|y_0, x)\}$

9. $\mathcal{X}4_9 = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 1, y_3 = 0|y_0, x) \geq P(y_1 = 0, y_2 = 1, y_3 = 1|y_0, x)\}$
10. $\mathcal{X}4_{10} = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 1, y_3 = 0|y_0, x) \leq P(y_1 = 0, y_2 = 1, y_3 = 1|y_0, x)\}$
11. $\mathcal{X}4_{11} = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 1, y_3 = 0|y_0, x) \geq P(y_1 = 1, y_2 = 0, y_3 = 1|y_0, x)\}$
12. $\mathcal{X}4_{12} = \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 1, y_3 = 0|y_0, x) \leq P(y_1 = 1, y_2 = 0, y_3 = 1|y_0, x)\}$

Assume each of $\mathcal{X}4_j,\ j = 1 : 12$, has positive measure. Furthermore, assume the set $\mathcal{X}^* \equiv \mathcal{X}2_4 \cap \mathcal{X}2_5 \cap \mathcal{X}4_1^c \cap \{x : (x_2 - x_3)'\beta > 0, (x_1 - x_3)'\beta > 0\}$ has positive measure.

P4ID2 For $x \in \mathcal{X}^*$ conditional on α, x_{11}, x_{12}, $\tilde{x}_2 - \tilde{x}_1$ has full column rank.

P4ID3 For $x \in \mathcal{X}^*$, conditional on $\tilde{x}_1, \tilde{x}_2, \alpha$, $x_{21} - x_{11}$ has support on the real line.

Theorem 4.4. *Suppose Assumptions $A1-A4$, $P2ID1-P2ID3$ and $P4ID1-P4ID3$ hold. Then both $\tilde{\beta}, \tilde{\gamma}$ are point identified.*

5 Inference

Though the main contribution of the paper is the characterization of the identified sets in these dynamic discrete choice models, we suggest an approach to inference that is computationally attractive under the assumption that the regressor vector x has finite support.

All the identified sets in the paper use conditional choice probabilities of the form (as an example)

$$
\begin{aligned}
P(y_1 = 0, y_2 = 1|y_0, x) &\equiv p_2(0, 1|y_0, x) \\
P(y_1 = 1, y_2 = 0|y_0, x) &\equiv p_2(1, 0|y_0, x) \\
P(y_1 = 0, y_2 = 0|y_0, x) &\equiv p_2(0, 0|y_0, x)
\end{aligned}
$$

The idea the inference section is to first construct a confidence region for the choice probabilities above. Then, heuristically, a confidence region for the identified set can be constructed by using draws from the (standard) confidence region for the choice probabilities. The mechanics of this exercise exploits linear programs to check whether a particular parameter vector θ belongs to the identified set. We describe this procedure in more details next.

5.1 A Confidence Region for the Choice Probabilities

One way to construct a confidence region for $\vec{p}(y_0, x) = (p_2(0,0|y_0,x), p_2(0,1|y_0,x), p_2(1,0|y_0,x))'$ is as follows. Let $(y_0^1, x^1), \ldots, (y_0^J, x^J)$ denote the support of (y_0, x). Then, as sample size increases, we have

$$\sqrt{n}W(\vec{p}(\cdot)) \equiv \sqrt{n} \begin{pmatrix} (\frac{1}{n}\sum_i \hat{w}_i^{1,0}(y_0^1, x^1) - p_2(1,0|y_0^1, x^1))1\{0 < p_2(1,0|y_0^1, x^1) < 1\} \\ (\frac{1}{n}\sum_i \hat{w}_i^{0,1}(y_0^1, x^1) - p_2(0,1|y_0^1, x^1))1\{0 < p_2(0,1|y_0^1, x^1) < 1\} \\ (\frac{1}{n}\sum_i \hat{w}_i^{0,0}(y_0^1, x^1) - p_2(0,0|y_0^1, x^1))1\{0 < p_2(0,0|y_0^1, x^1) < 1\} \\ \cdots \\ (\frac{1}{n}\sum_i \hat{w}_i^{1,0}(y_0^J, x^J) - p_2(1,0|y_0^J, x^J))1\{0 < p_2(1,0|y_0^J, x^J) < 1\} \\ (\frac{1}{n}\sum_i \hat{w}_i^{0,1}(y_0^J, x^J) - p_2(0,1|y_0^J, x^J))1\{0 < p_2(0,1|y_0^J, x^J) < 1\} \\ (\frac{1}{n}\sum_i \hat{w}_i^{0,0}(y_0^J, x^J) - p_2(0,0|y_0^J, x^J))1\{0 < p_2(0,0|y_0^J, x^J) < 1\} \end{pmatrix} \Rightarrow N(0, \Sigma(\vec{p}(\cdot)))$$

where $\Sigma(\vec{p}(\cdot))$ is the variance-covariance matrix and

$$\hat{w}_i^{ds}(y_0, x) = \frac{1\{y_{1i} = d, y_{2i} = s, y_{0i} = y_0, x_i = x\}}{\hat{p}_z(y_0, x)} \text{ for } d, s \in \{0, 1\}$$

and

$$\hat{p}_z(y_0, x) = \frac{1}{n}\sum_i 1\{y_{0i} = y_0, x_i = x\}$$

Note that some rows and columns of $\Sigma(\vec{p}(\cdot))$ may be zero, so in general this matrix can be singular. Let $\Sigma^*(\vec{p}(\cdot))$ be a sub-matrix of $\Sigma(\vec{p}(\cdot))$ that corresponds to all non-zero rows and columns. Then $\Sigma^*(\vec{p}(\cdot))$ has full rank. Let $W^*(\vec{p}(\cdot))$ be a sub-vector of $W(\vec{p}(\cdot))$ that corresponds to those non-zero columns (rows). Then

$$\sqrt{n}W^*(\vec{p}(\cdot)) \Rightarrow N\left(0, \Sigma^*(\vec{p}(\cdot))\right)$$

and

$$T_n^{as}(\vec{p}(\cdot) \equiv nW^*(\vec{p}(\cdot))' \left(\Sigma^*(\vec{p}(\cdot))\right)^{-1} W^*(\vec{p}(\cdot)) \Rightarrow \chi^2_{q(\vec{p}(\cdot))}$$

where $q(\vec{p}(\cdot)) = \dim(W^*(\vec{p}(\cdot)))$.

Then, an asymptotic $100(1-\alpha)\%$ confidence set for $\vec{p}(y_0, x)$:

$$CS^p_{1-\alpha} = \Big\{\vec{p}(y_0, x) \geq 0 : \text{for all } (y_0, x), p_2(0,0|y_0, x) + p_2(0,1|y_0, x) + p_2(1,0|y_0, x) \leq 1 \\ \text{and } T^{as}_n(\vec{p}(\cdot)) \leq c^*_{1-\alpha}(\vec{p}(\cdot))\Big\} \tag{5.1}$$

where $c^*_{1-\alpha}(\vec{p}(\cdot))$ is the $(1-\alpha)$ quantile of χ^2 distribution with $q(\vec{p}(\cdot)) = \dim(W^*(\vec{p}(\cdot)))$ degrees of freedom (the number of probabilities in $\vec{p}(\cdot)$ that are strictly between 0 and 1).

One way to obtain a draw from this confidence region is to use the weighted bootstrap via a posterior distribution for these choice probabilities.

5.2 Confidence region for the identified set

We illustrate here how we map the CI for the choice probabilities to a CI for the identified set. We do it in the context of two simple examples that we call the "Stationary Example" and the "Exchangeable Example." These two examples showcase the issues that come up in a clean way. Let the parameter of interest $\theta = (\gamma, \beta)$ and consider the following examples:

"Stationarity Example"

$$\begin{aligned} p_1(1|y_0, x) \geq P(y_2 = 1|y_0, x) &\Rightarrow \Delta x'\beta + \min\{0, \gamma\} - \gamma y_0 \leq 0 \\ p_1(1|y_0, x) \leq P(y_2 = 1|y_0, x) &\Rightarrow \Delta x'\beta + \max\{0, \gamma\} - \gamma y_0 \geq 0 \\ p_2(0,1|y_0, x) \geq p_1(1|y_0, x) &\Rightarrow \Delta x'\beta - \gamma y_0 \geq 0 \\ p_2(1,0|y_0, x) \geq p_1(0|y_0, x) &\Rightarrow \Delta x'\beta + \gamma(1 - y_0) \geq 0 \end{aligned} \tag{5.2}$$

"Exchangeable Example":

$$\begin{aligned} -\Delta x'\beta + \gamma y_0 \geq 0 \text{ and } -\Delta x'\beta + \gamma y_0 - \gamma \geq 0 &\Rightarrow p_2(1,0|x, y_0) \geq p_2(0,1|x, y_0) \\ -\Delta x'\beta + \gamma y_0 \leq 0 \text{ and } -\Delta x'\beta + \gamma y_0 - \gamma \leq 0 &\Rightarrow p_2(1,0|x, y_0) \leq p_2(0,1|x, y_0) \end{aligned} \tag{5.3}$$

where at least one strict inequality on the left-hand side implies a strict inequality on the right-hand side (in both examples).

There is an important qualitative difference between the two "toy" models above: in the Stationary model, there is one sided inequalities and all the parameters that satisfy these implications for all (y_0, x). On the other hand, in the Exchangeable model, two conditions have to be satisfied for the implication to hold (notice the "and" in the implication) and that makes the implementation different. We explain this in details next.

Generally, a CI for the identified sets in both models above can simply be defined as follows and are based on the chi-squared approximation above.

The confidence set for $\theta = (\gamma, \beta)$ based on asymptotic chi-squared approximation:

$$CS^{\theta}_{1-\alpha} = \{\theta \in \Theta : \text{conditions (5.2) and (5.3) hold for some } \vec{p}(\cdot) \in CS^{p}_{1-\alpha}\}$$

where $CS^{p}_{1-\alpha}$ is defined in (5.1) above. It us computationally tedious to check whether the inequalities above are satisfied for a given vector of choice probabilities. However, it is possible to exploit the linearity in the model as follows.

5.2.1 Linear Program for solving Model (5.2):

Conditions (5.2) are straightforward to verify for a given $\vec{p}(\cdot) \in CS^{p}_{1-\alpha}$. The following is an algorithm to build a CS based on a linear program.

(i) Pick an element $\vec{p}_{(k)}(\cdot)$ from $CS^{p}_{1-\alpha}$. Alternatively, use the Bayesian bootstrap to get a draw $\vec{p}_{(k)}(\cdot)$ from the confidence ellipse for the choice probabilities. This can be done instantaneously.

(ii) Get $\Theta^{+}_{(k)}$, the set of parameters θ that solve

$$\max_{(\gamma,\beta)\in\Theta} c + 0 \cdot \gamma + \beta' \cdot 0$$

$$\text{subject to}$$

$$
\begin{cases}
\gamma \geq 0 \\
1\{p_{1,(k)}(1|y_0^1, x^1) \geq P_{(k)}(y_2 = 1|y_0^1, x^1)\}(-\Delta x^{1'}\beta + \gamma y_0^1) \geq 0 \\
1\{p_{1,(k)}(1|y_0^1, x^1) \leq P_{(k)}(y_2 = 1|y_0^1, x^1)\}(\Delta x^{1'}\beta + \gamma(1 - y_0^1)) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^1, x^1) \geq p_{1,(k)}(1|y_0^1, x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^1, x^1) \leq p_{1,(k)}(1|y_0^1, x^1)\}(\Delta x^{1'}\beta + \gamma(1 - y_0^1)) \geq 0 \\
1\{p_{1,(k)}(1|y_0^2, x^2) \geq P_{(k)}(y_2 = 1|y_0^2, x^2)\}(-\Delta x^{2'}\beta + \gamma y_0^2) \geq 0 \\
1\{p_{1,(k)}(1|y_0^2, x^2) \leq P_{(k)}(y_2 = 1|y_0^2, x^2)\}(\Delta x^{2'}\beta + \gamma(1 - y_0^2)) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^2, x^2) \geq p_{1,(k)}(1|y_0^2, x^2)\}(\Delta x^{2'}\beta - \gamma y_0^2) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^2, x^2) \leq p_{1,(k)}(1|y_0^2, x^2)\}(\Delta x^{2'}\beta + \gamma(1 - y_0^2)) \geq 0 \\
\dots \\
1\{p_{1,(k)}(1|y_0^J, x^J) \geq P_{(k)}(y_2 = 1|y_0^J, x^J)\}(-\Delta x^{J'}\beta + \gamma y_0^J) \geq 0 \\
1\{p_{1,(k)}(1|y_0^J, x^J) \leq P_{(k)}(y_2 = 1|y_0^J, x^J)\}(\Delta x^{J'}\beta + \gamma(1 - y_0^J)) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^J, x^J) \geq p_{1,(k)}(1|y_0^J, x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^J, x^J) \leq p_{1,(k)}(1|y_0^J, x^J)\}(\Delta x^{J'}\beta + \gamma(1 - y_0^J)) \geq 0
\end{cases}
\tag{5.4}
$$

(iii) Similarly, get $\Theta^-_{(k)}$, the set of parameters θ that solve

$$\max_{(\gamma,\beta)\in\Theta} c + 0 \cdot \gamma + \beta' \cdot 0$$

subject to

$$
\begin{cases}
\gamma \leq 0 \\
1\{p_{1,(k)}(1|y_0^1, x^1) \geq P_{(k)}(y_2 = 1|y_0^1, x^1)\}(-\Delta x^{1'}\beta + \gamma(y_0^1 - 1)) \geq 0 \\
1\{p_{1,(k)}(1|y_0^1, x^1) \leq P_{(k)}(y_2 = 1|y_0^1, x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^1, x^1) \geq p_{1,(k)}(1|y_0^1, x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^1, x^1) \leq p_{1,(k)}(1|y_0^1, x^1)\}(\Delta x^{1'}\beta + \gamma(1 - y_0^1)) \geq 0 \\
1\{p_{1,(k)}(1|y_0^2, x^2) \geq P_{(k)}(y_2 = 1|y_0^2, x^2)\}(-\Delta x^{2'}\beta + \gamma(y_0^2 - 1)) \geq 0 \\
1\{p_{1,(k)}(1|y_0^2, x^2) \leq P_{(k)}(y_2 = 1|y_0^2, x^2)\}(\Delta x^{2'}\beta - \gamma y_0^2) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^2, x^2) \geq p_{1,(k)}(1|y_0^2, x^2)\}(\Delta x^{2'}\beta - \gamma y_0^2) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^2, x^2) \leq p_{1,(k)}(1|y_0^2, x^2)\}(\Delta x^{2'}\beta + \gamma(1 - y_0^2)) \geq 0 \\
\dots \\
1\{p_{1,(k)}(1|y_0^J, x^J) \geq P_{(k)}(y_2 = 1|y_0^J, x^J)\}(-\Delta x^{J'}\beta + \gamma(y_0^J - 1)) \geq 0 \\
1\{p_{1,(k)}(1|y_0^J, x^J) \leq P_{(k)}(y_2 = 1|y_0^J, x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^J, x^J) \geq p_{1,(k)}(1|y_0^J, x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^J, x^J) \leq p_{1,(k)}(1|y_0^J, x^J)\}(\Delta x^{J'}\beta + \gamma(1 - y_0^J)) \geq 0
\end{cases}
\tag{5.5}
$$

(iv) Repeat the above M times

(v) A $(1-\alpha)$ CS for θ would be the $\left(\cap_{k\leq M}\Theta^+_{(k)}\right) \cup \left(\cap_{k\leq M}\Theta^-_{(k)}\right)$

The computationally tedious part in the above linear program is part (ii) which builds the set of all θ's for which the linear program is feasible. One approach for this is to get a grid for θ and check whether each point on this grid is feasible. Checking feasibility is simple even when J is large.

5.2.2 Linear Program for solving Model (5.3)

Now, building a CS for θ in the exchangeable model of (5.3) is more complicated since checking that both $-\Delta x'\beta + \gamma y_0 \geq 0$ *and* $-\Delta x'\beta + \gamma y_0 - \gamma \geq 0$ renders the program nonlinear. However, here, notice that the inequalities (5.3) are equivalent to (notice the "and" became an "or")

$$
\begin{aligned}
p_2(1,0|y_0,x) \leq p_2(0,1|y_0,x) &\Rightarrow \Delta x'\beta - \gamma y_0 \geq 0 \text{ or } \Delta x'\beta + \gamma(1-y_0) \geq 0 \\
p_2(1,0|y_0,x) \geq p_2(0,1|y_0,x) &\Rightarrow \Delta x'\beta - \gamma y_0 \leq 0 \text{ or } \Delta x'\beta + \gamma(1-y_0) \leq 0
\end{aligned}
$$

that is, any θ that belongs to the *compliment* of the set below belongs to the identified set:

$$
\begin{aligned}
p_2(1,0|y_0,x) \leq p_2(0,1|y_0,x) &\Rightarrow \Delta x'\beta - \gamma y_0 < 0 \text{ and } \Delta x'\beta + \gamma(1-y_0) < 0 \\
p_2(1,0|y_0,x) \geq p_2(0,1|y_0,x) &\Rightarrow \Delta x'\beta - \gamma y_0 > 0 \text{ and } \Delta x'\beta + \gamma(1-y_0) > 0
\end{aligned}
\tag{5.6}
$$

This can be posed as the constraints in a *linear* program and so checking whether θ is feasible (here belongs to the *complement* of the identified set) is equivalent to checking the feasibility of a linear program. For a given $p_{(k)}$ draw of the vector of choice probabilities, checking whether a parameter is feasible can be done for example using the following program:

$$\max_{(\gamma,\beta)\in\Theta} c + 0 \cdot \gamma + \beta' \cdot 0$$

subject to

$$
\begin{cases}
1\{p_{2,(k)}(1,0|y_0^1,x^1) \leq p_{2,(k)}(0,1|y_0^1,x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \leq 0 \\
1\{p_{2,(k)}(1,0|y_0^1,x^1) \leq p_{2,(k)}(0,1|y_0^1,x^1)\}(\Delta x^{1'}\beta + \gamma(1-y_0^1)) \leq 0 \\
1\{p_{2,(k)}(1,0|y_0^1,x^1) \geq p_{2,(k)}(0,1|y_0^1,x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \geq 0 \\
1\{p_{2,(k)}(1,0|y_0^1,x^1) \geq p_{2,(k)}(0,1|y_0^1,x^1)\}(\Delta x^{1'}\beta + \gamma(1-y_0^1)) \geq 0 \\
\ldots \\
1\{p_{2,(k)}(1,0|y_0^J,x^J) \leq p_{2,(k)}(0,1|y_0^J,x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \leq 0 \\
1\{p_{2,(k)}(1,0|y_0^J,x^J) \leq p_{2,(k)}(0,1|y_0^J,x^J)\}(\Delta x^{J'}\beta + \gamma(1-y_0^J)) \leq 0 \\
1\{p_{2,(k)}(1,0|y_0^J,x^J) \geq p_{2,(k)}(0,1|y_0^J,x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \geq 0 \\
1\{p_{2,(k)}(1,0|y_0^J,x^J) \geq p_{2,(k)}(0,1|y_0^J,x^J)\}(\Delta x^{J'}\beta + \gamma(1-y_0^J)) \geq 0
\end{cases}
\tag{5.7}
$$

Hence, we can then use the same algorithm as above by repeatedly drawing a vector of choice probabilities from the confidence ellipse and then collecting all the parameters that solve the above program for at least one of these draws. The closure of the complement of this set will give us a well defined confidence set for the identified set.

6 Finite Sample Properties

6.1 Simulation Results

In this section we conduct an extensive simulation study to explore how well our theoretical conclusions in previous sections hold up in simulated data. Our theoretical results were mainly tied to establishing the empirical content of varying assumptions in dynamic binary choice models.

In establishing our theoretical results we reached the following conclusions regarding identifying the structural parameters in the model:

- Regression coefficients on strictly exogenous variables were generally easier to identify than the coefficient on the lagged binary dependent variable, which was our measure of the persistence in the model.
- Restricting dependence structure on the idiosyncratic components of the model facilitated identification of the structural parameters.
- Increasing the richness of the support of the exogenous variables facilitates the identification.
- Increasing the length of the time series added informational content, as structural parameters could be identified under weaker conditions.
- The value of the parameters themselves could effect their identifiability. For example, it was shown that a negative value of the persistence parameter made its identification more difficult.

We will illustrate these results by simulating data from the following model:

$$y_{it} = I\{u_{it} \leq v_{it} + x_{it}\beta + \gamma y_{i,t-1} + \alpha_i\} \quad i = 1, 2, ...10000; \quad t = 0, 1, ...T \tag{6.1}$$

y_{it} is the observed binary dependent variable and $y_{i,t-1}$ is its lagged value. v_{it}, x_{it} are each observed scalar exogenous variables, the first whose coefficient is normalized to 1, and the second, whose coefficient β we aim to identify, along with persistence parameter γ. α_i, a scalar, denotes the unobserved individual specific effect and u_{it} denotes the unobserved scalar idiosyncratic term. The simulation exercise explores identification of β, γ under varying models, corresponding different values of T, different assumptions on u_{it}, varying support conditions on (v_{it}, x_{it}), and different values of γ. As will be explained below, this will correspond to 64 different designs and for each we demonstrate graphically the nature of the identified region of the structural parameters β, γ.

We demonstrate identification graphically with projections of three dimensional plots of our objective function. Specifically we look at values of the objective function of different values of β and γ along a grid of a two dimensional plane. Instead of constructing three dimensional plots, we show values of the parameters which attain the global maximum of the objective function. The objective functions used corresponded to the moment inequalities used in the main theorems. In models where point identification is attainable, a single value will be in the plot, whereas in partially identified models, a subset of the grid will be plotted.

6.1.1 Stationary Model, T=2

In this model we simulated data where v_{it}, x_{it} were each discretely distributed, with the number of support points for v_{it}, increasing from 2 to 7, and then continuously (standard normal) distributed. The number of support points for x_{it} was always two, though there where two distinct designs- one with identical support in each time period, and the other with strictly nonoverlapping support- $x_{it} = t \quad t = 1, 2$. The idiosyncratic terms u_{it} were bivariate normal, mean 0 variance 1, correlation 0.5, and the fixed effect α_i was standard normal. We assumed that all variables were mutually independent. The parameters where set to 1 for β and either 0.5 or -0.5 for γ.

Our plots for this model agree with our theoretical results. We note that when x_{it}, v_{it} are discrete, neither parameter is point identified. For example, in Figure 2, we have x is binary while v starts out as binary and then we add points of support ending with 14. This Figure is repeated for when true γ is negative. As we can see the identified set is not trivial. Its size shrinks in Figure 4 when v is normally distributed with increasing variance. Notice here that in all the plots, β appears well identified.

In Figure 6, we change x to a time trend ($x = t$) and in the top lhs plot, we have the identified set in the case when v is binary. Here, we cannot pin down the sign of γ. But, as we increase the points of support for v, the identified set shrinks and eventually it appears that the sign of γ is identified. The same story holds for when γ is negative. The next Figures allow for time trend in the case when v is normal.

Throughout, when v_{it} is continuously distributed, β is point identified, whereas γ is not. But the graph clearly demonstrates that its sign is.

6.1.2 Exchangeabiiity Model, T=2

Here we construct plots for the objective function based on the exchangeability assumption. Starting with Figure 10 we see that the model contains information with both x and v being binary, and that the identified set in this case is smaller than the one in Figure 2 under stationarity. What is more interesting here is Figure 12 where we get essentially point identification of *both* γ and β when v is standard normal. The identified set here shrinks to a point (up to numerical error) as the variance of v increases. The same story does not hold for the case when true γ is negative as confirmed in Figure 13. When γ is positive, even when x is a time trend, we get tight identified sets for β *and* γ when v is normal as can be seen in Figure 15 which is remarkable and points to the strength of the exchangeability assumption.

Now we see that when v_{it} is continuously distributed both β and *gamma* are point identified when $\gamma = 0.5$. However, only β is point identified when $\gamma = -0.5$, though we can see from the graph that it is negative. This agrees with all our theoretical conclusions from the exchangeability section of the paper.

6.1.3 Stationary Model, T=3

Here we simulated data with an extra time period, maintaining the stationarity assumption, so that u_{it} was trivariate normal with pairwise correlations of 0.25. The graphs now demonstrate that both β and γ will be point identified when even when γ is negative and $x_t = t$. This matches up with our theoretical conclusion that point identification can be achieved with *all* of nonoverlapping support, serial correlation and state dependence. In Figure 16 we provide the identified sets for a few designs. In the top, the two designs correspond to the case when $x_t = t$ and v is discrete (top left) and when v is standard normal (top right). The bottom of the figure plots the case when v is normal (variance 1 on the left and 2.5 on the right).

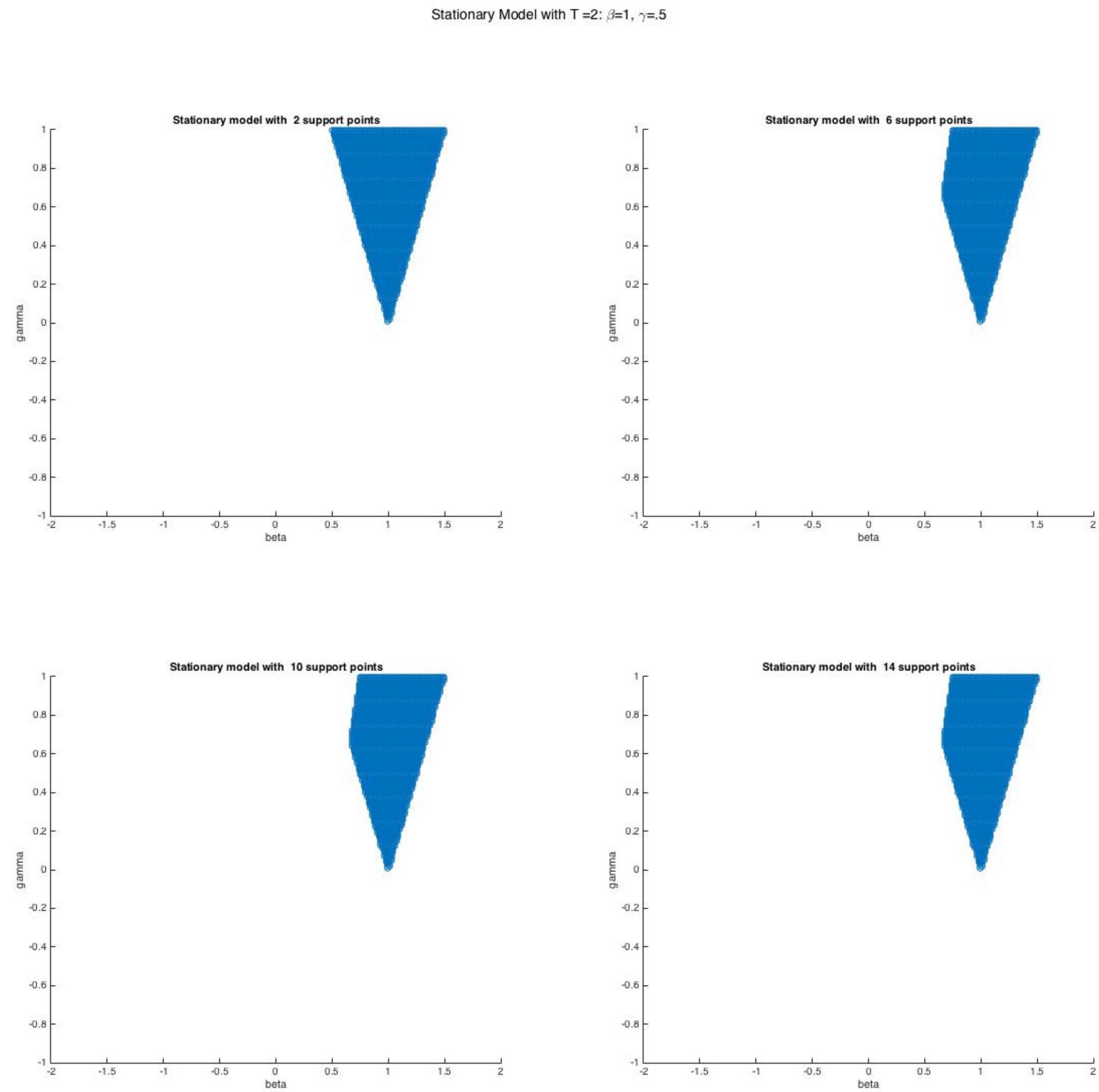

Figure 2: Stationary with $T = 2$ and Discrete Support with $\gamma = .5$

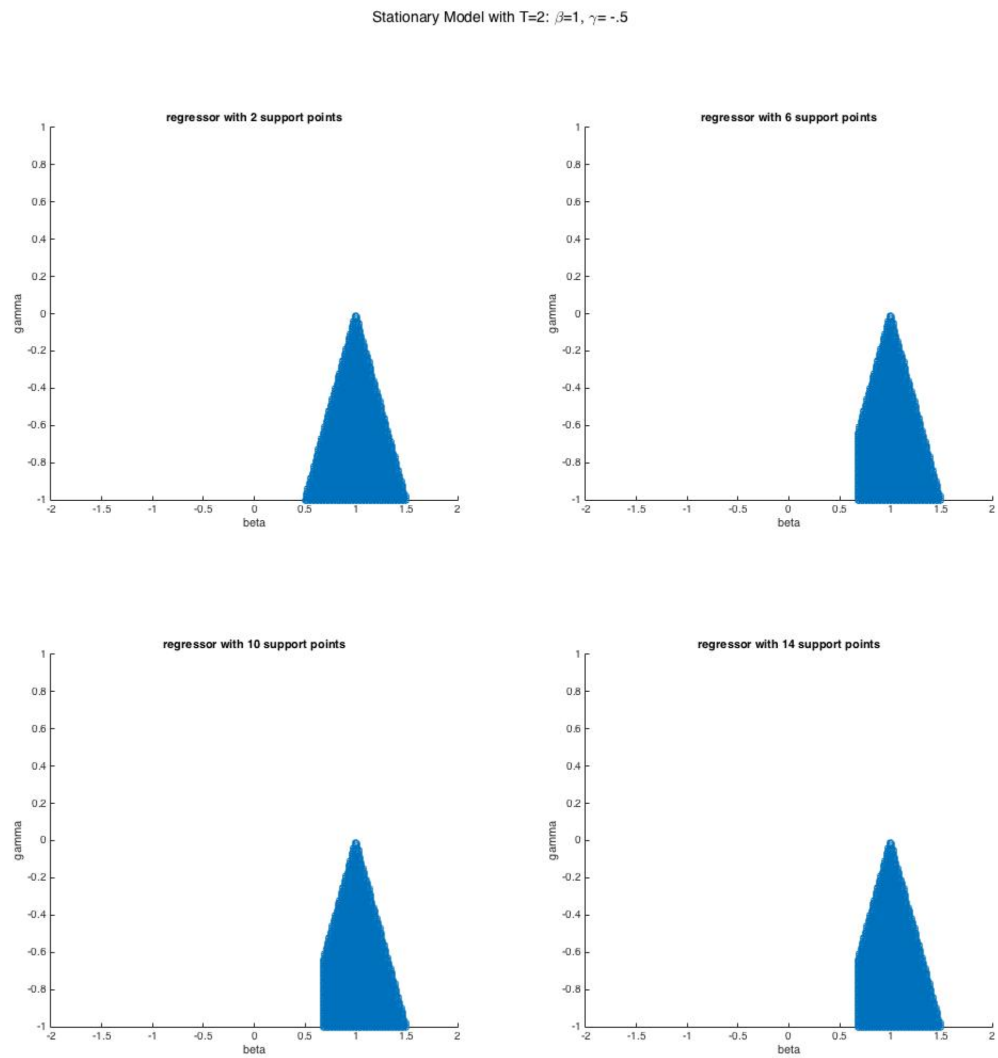

Figure 3: Stationary with $T = 2$ and Discrete Support with $\gamma = .5$

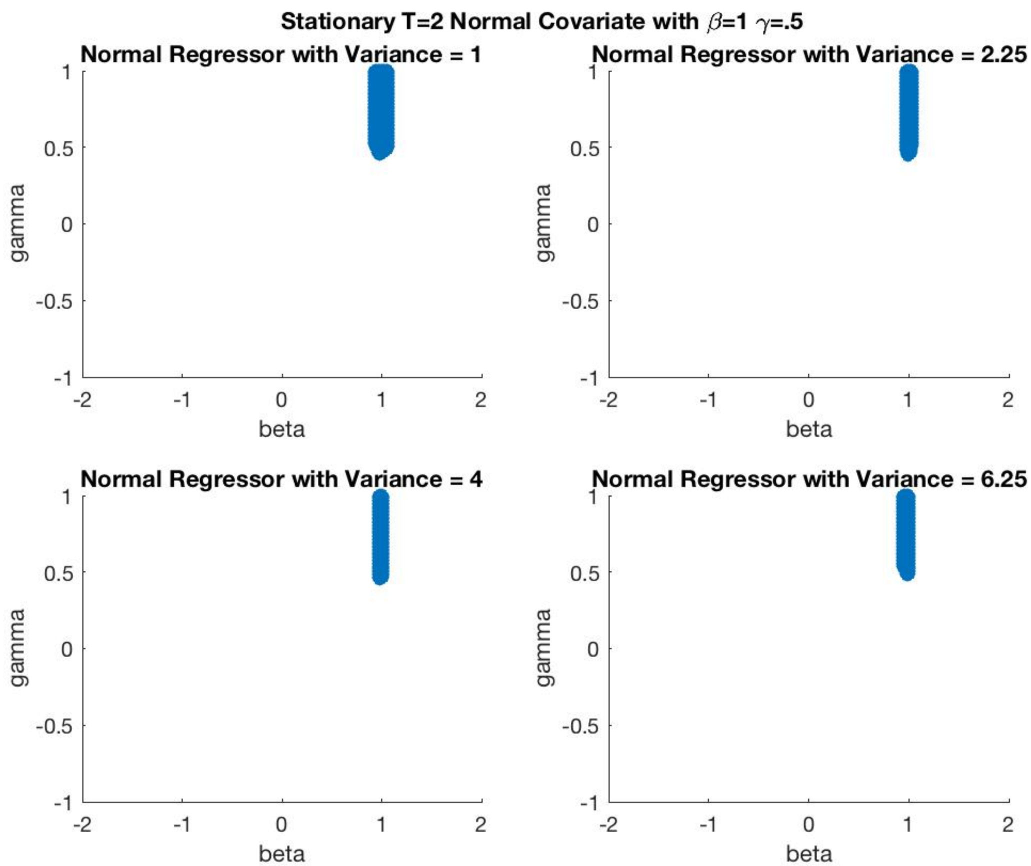

Figure 4: Stationary with $T = 2$ and Normal v with $\gamma = .5$

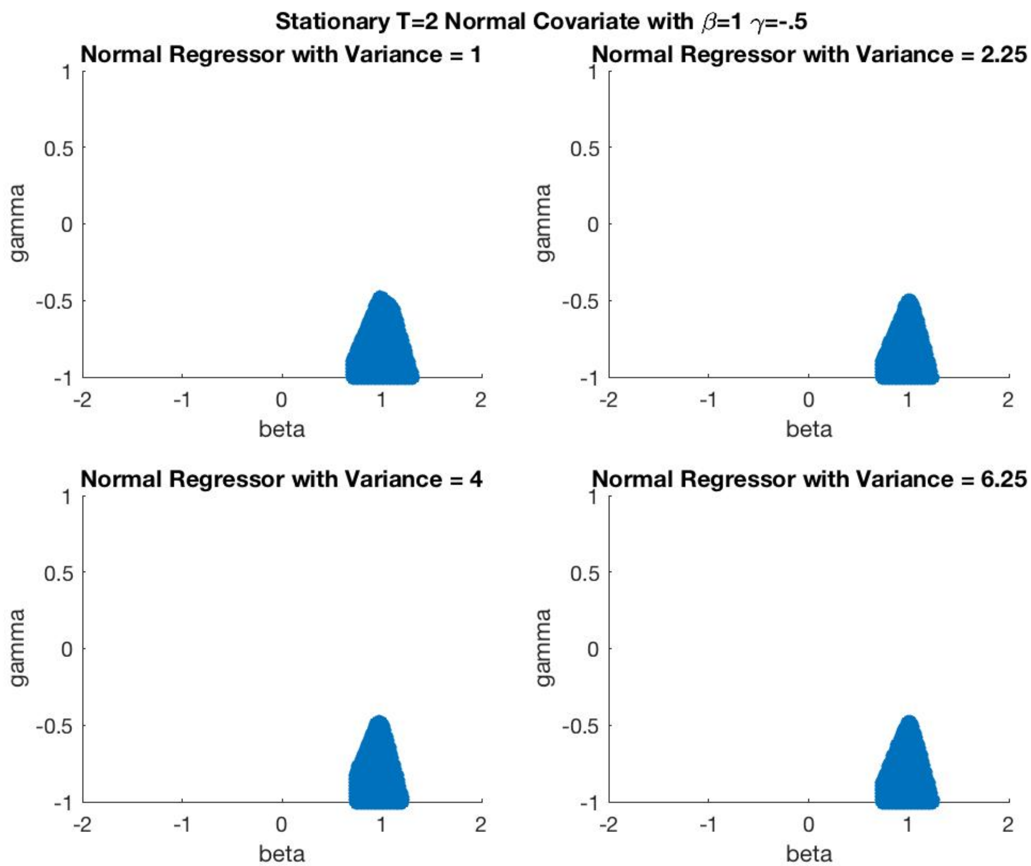

Figure 5: Stationary with $T = 2$ and Normal v with $\gamma = -.5$

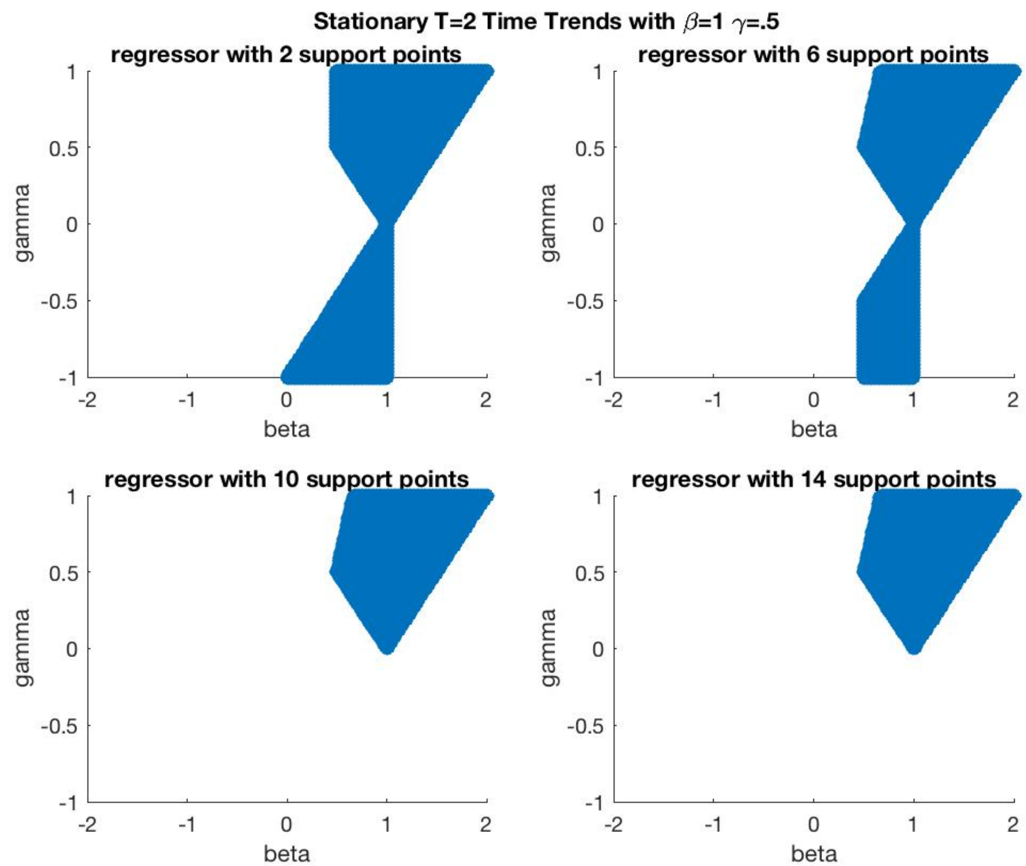

Figure 6: Stationary with $T = 2$ and Time Trend and Discrete Support for v with $\gamma = .5$

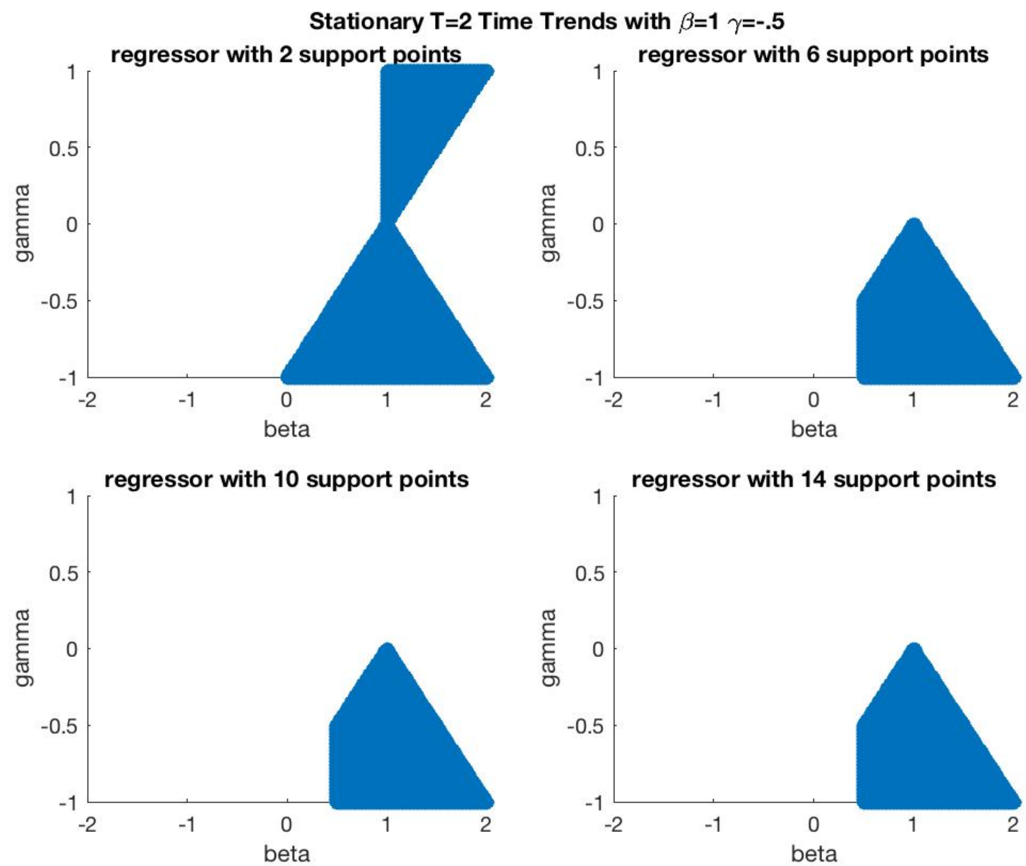

Figure 7: Stationary with $T = 2$ and Time Trend and Discrete Support for v with $\gamma = -.5$

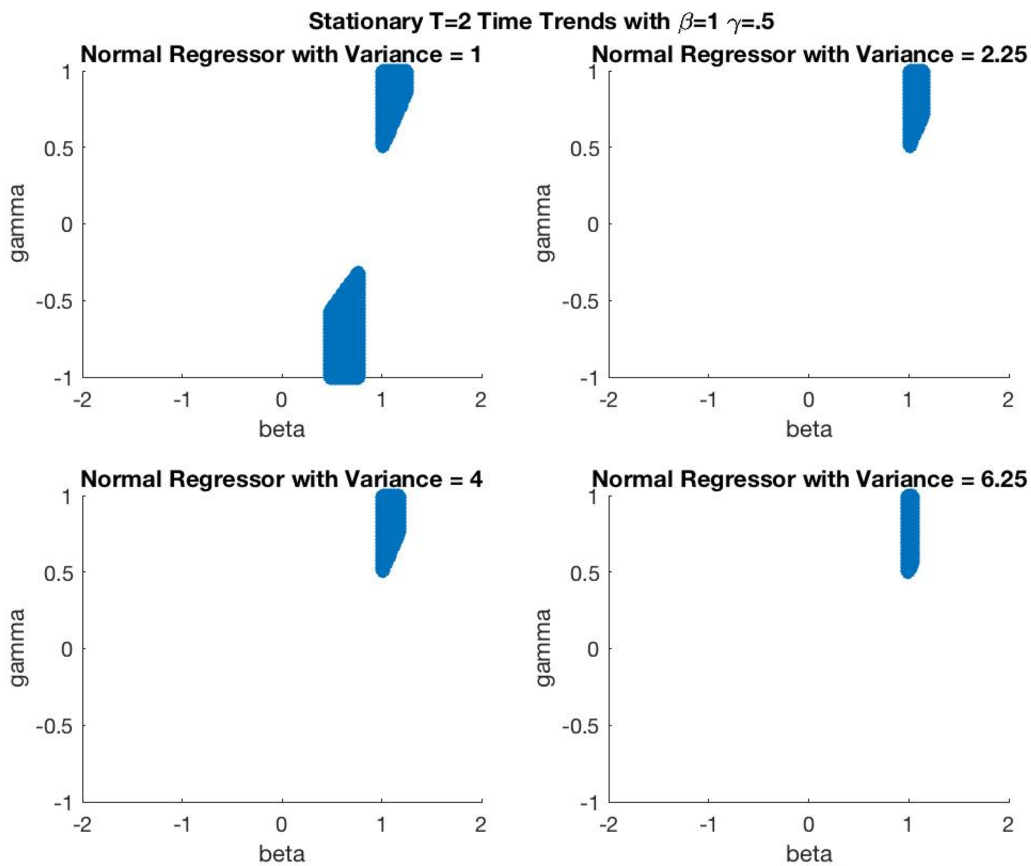

Figure 8: Stationary with $T = 2$ and Time Trend and Normal v with $\gamma = .5$

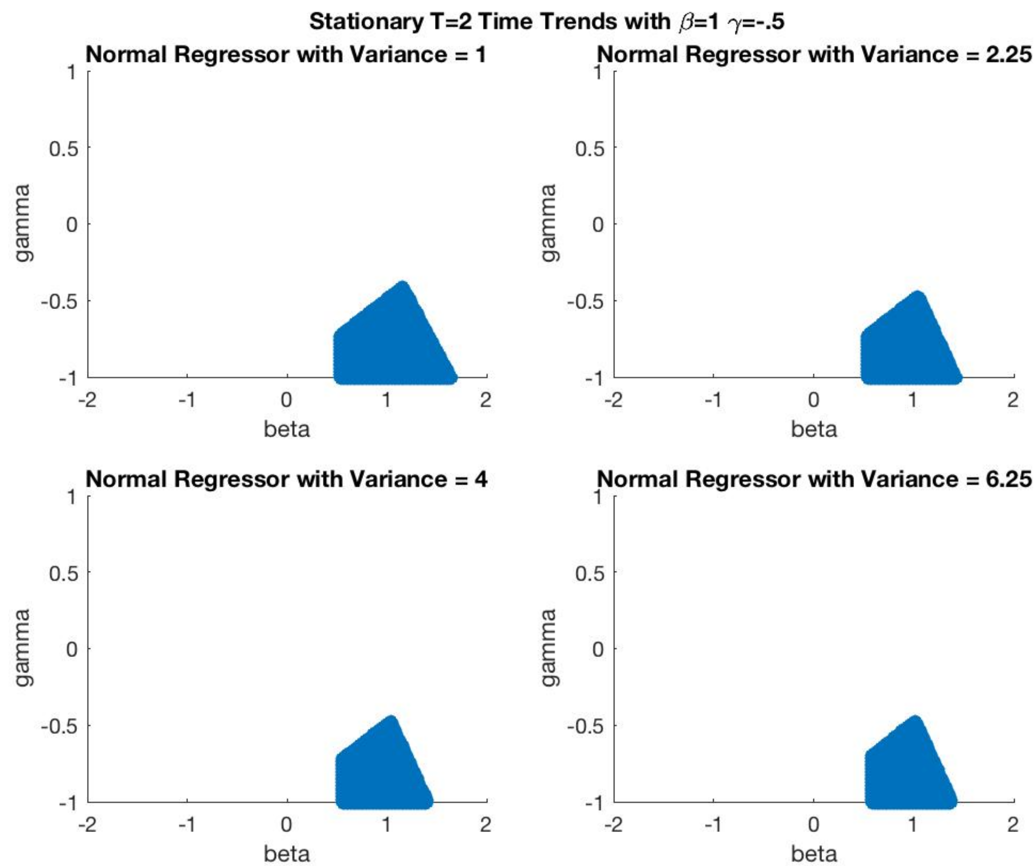

Figure 9: Stationary with $T = 2$ and Time Trend and Normal v with $\gamma = -.5$

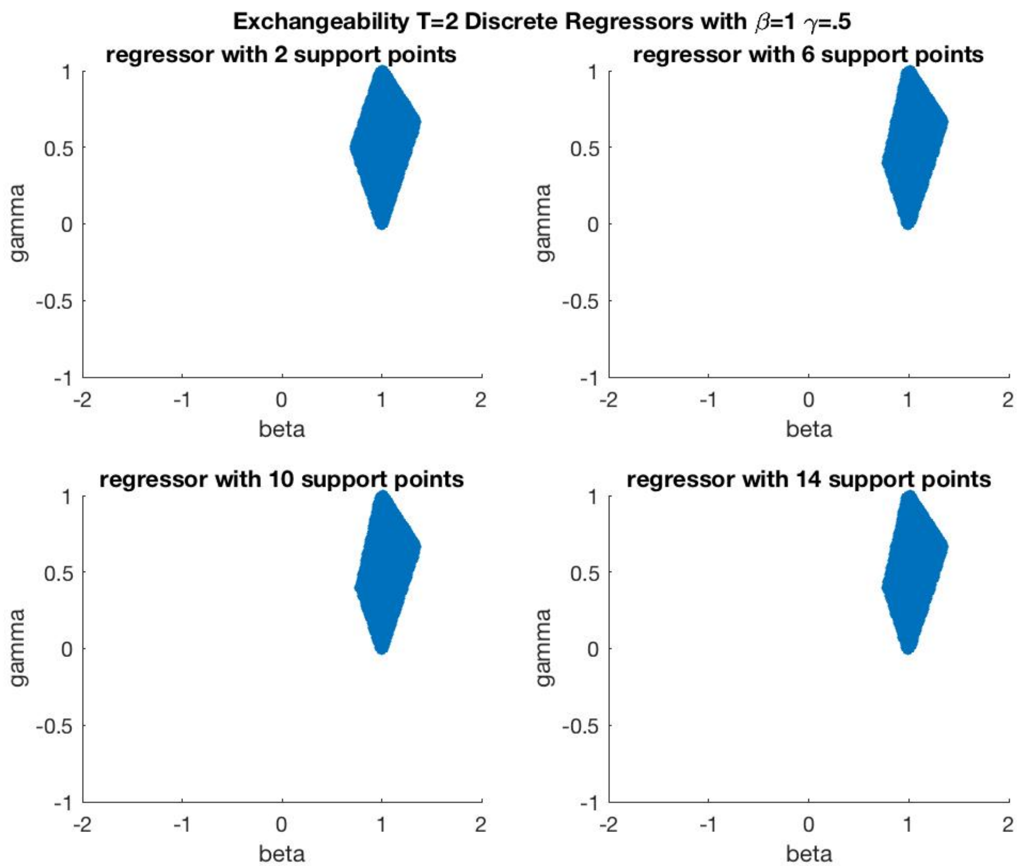

Figure 10: Exchangeability with $T = 2$ Discrete Support for v with $\gamma = .5$

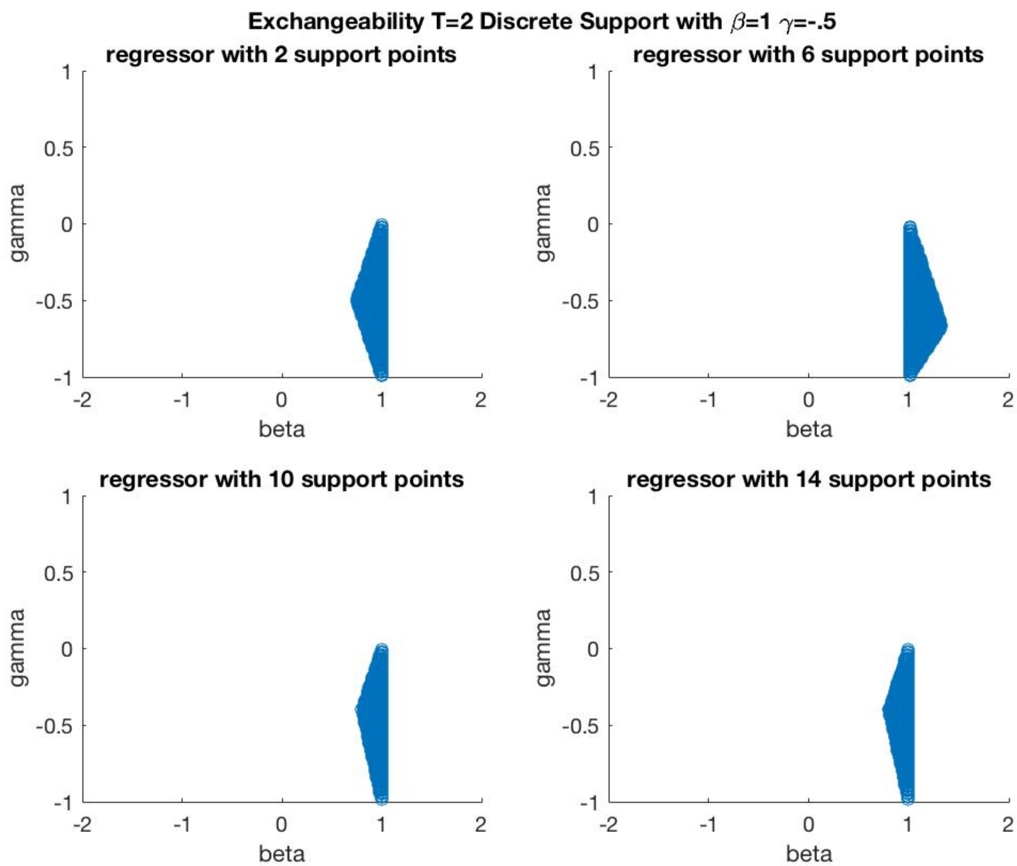

Figure 11: Stationary with $T = 2$ Discrete Support for v with $\gamma = -.5$

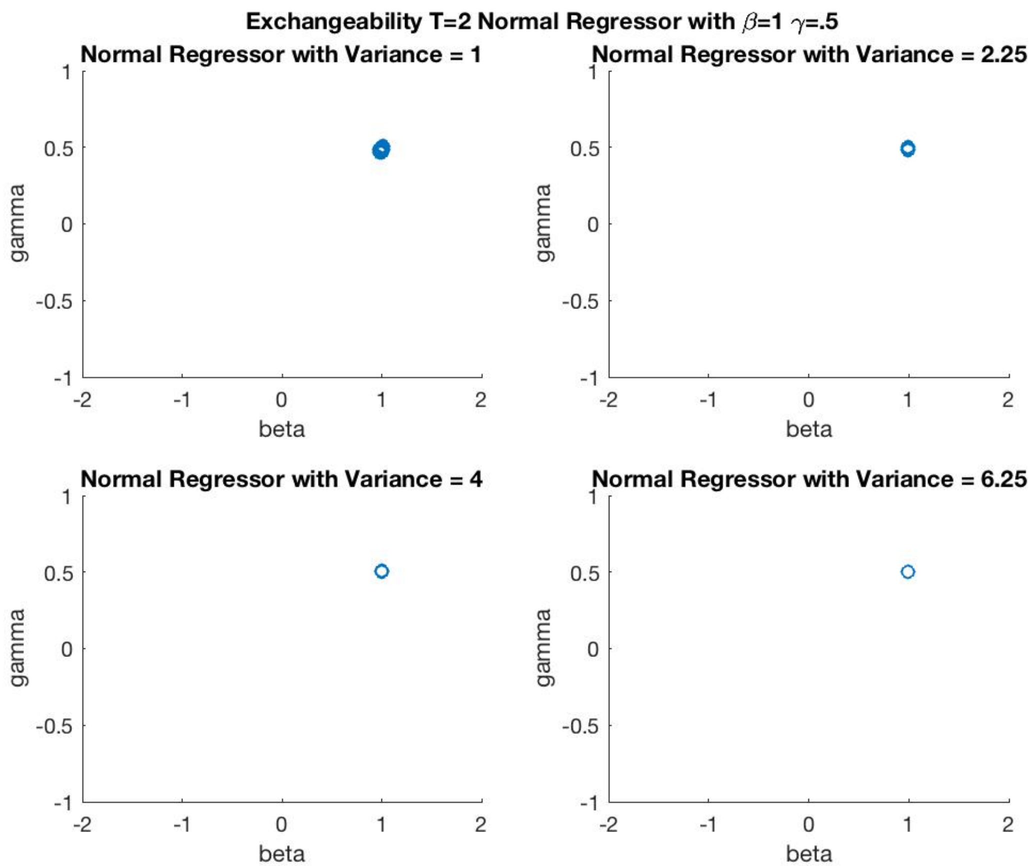

Figure 12: Exchangeability with $T = 2$ Discrete v with $\gamma = .5$

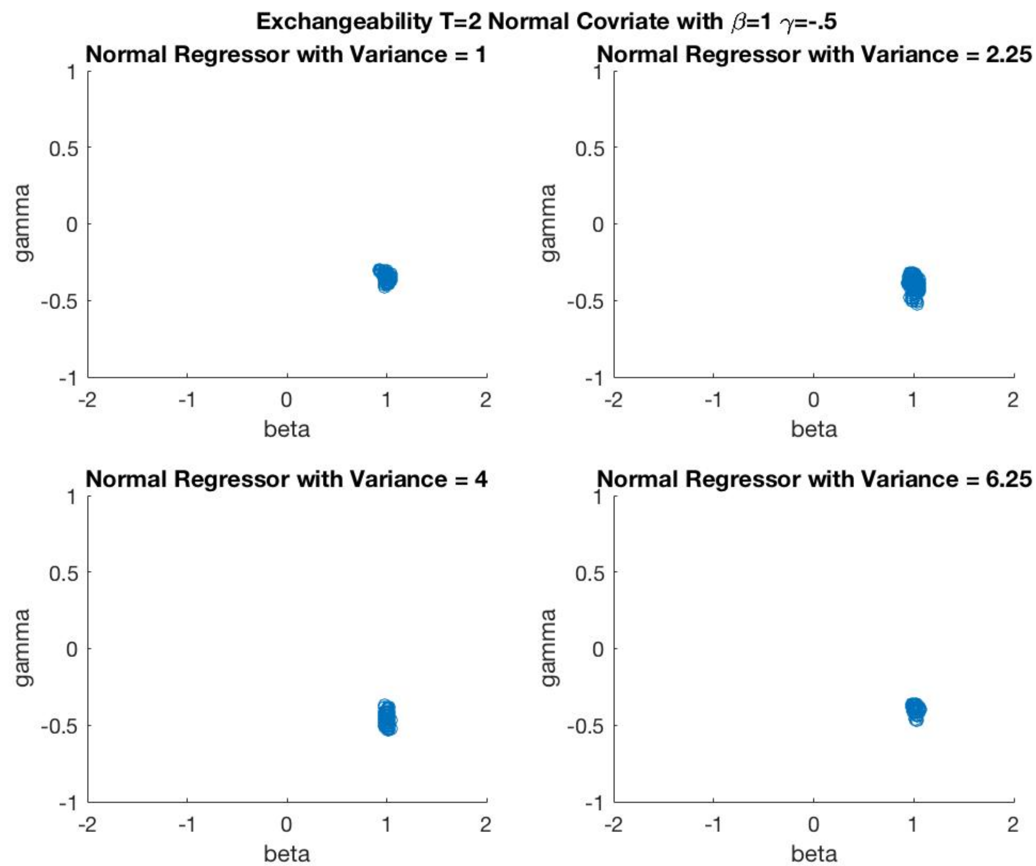

Figure 13: Stationary with $T = 2$ Normal for v with $\gamma = -.5$

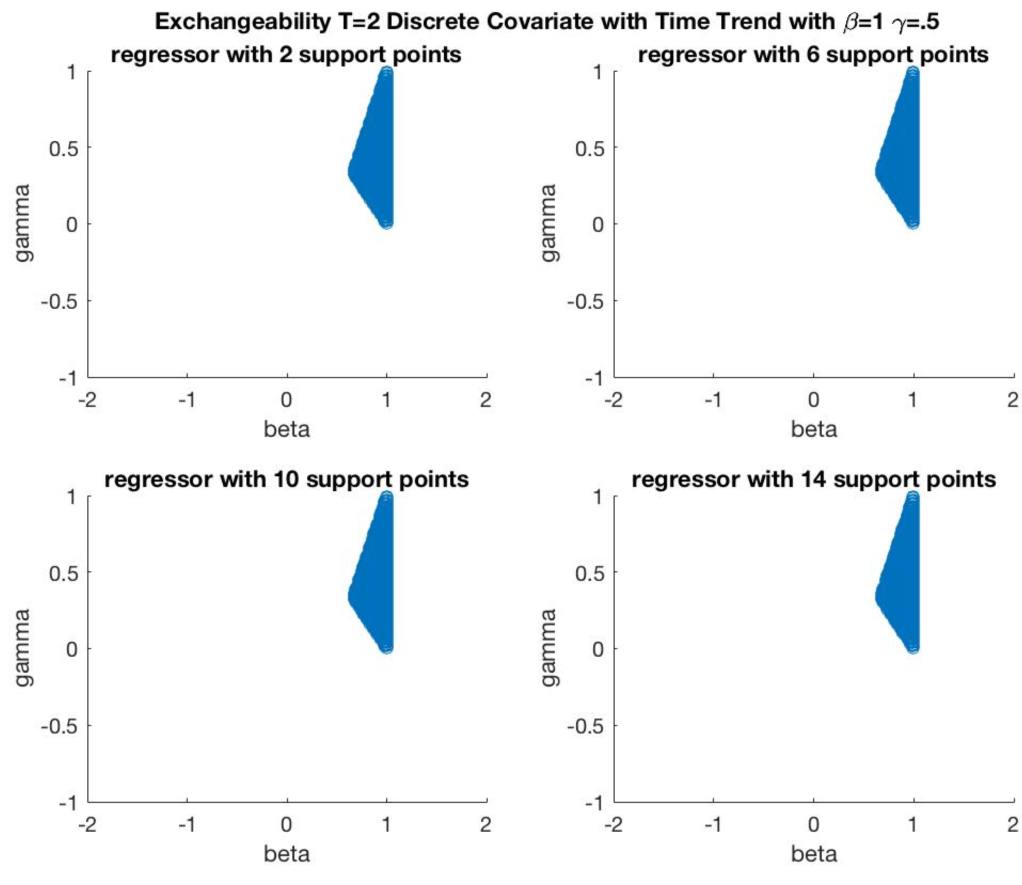

Figure 14: Exchangeability with $T = 2$ $x = t$ Discrete v with $\gamma = .5$

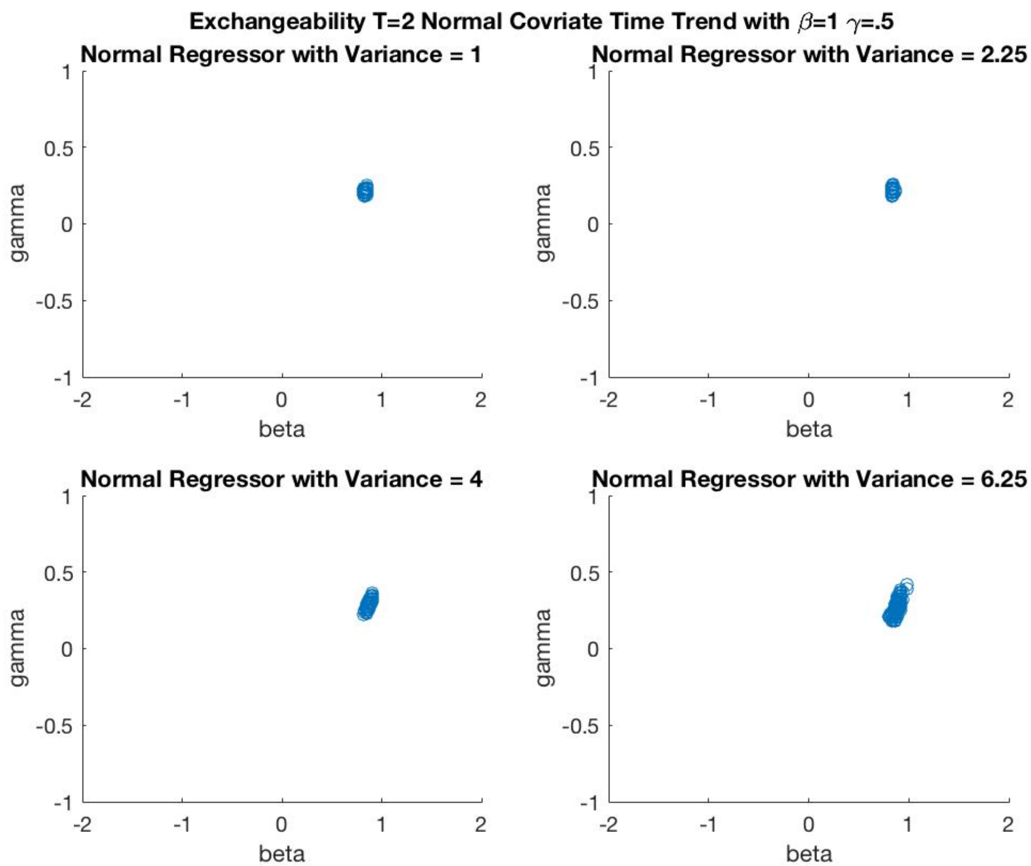

Figure 15: Stationary with $T = 2$, $x = t$ Normal for v with $\gamma = .5$

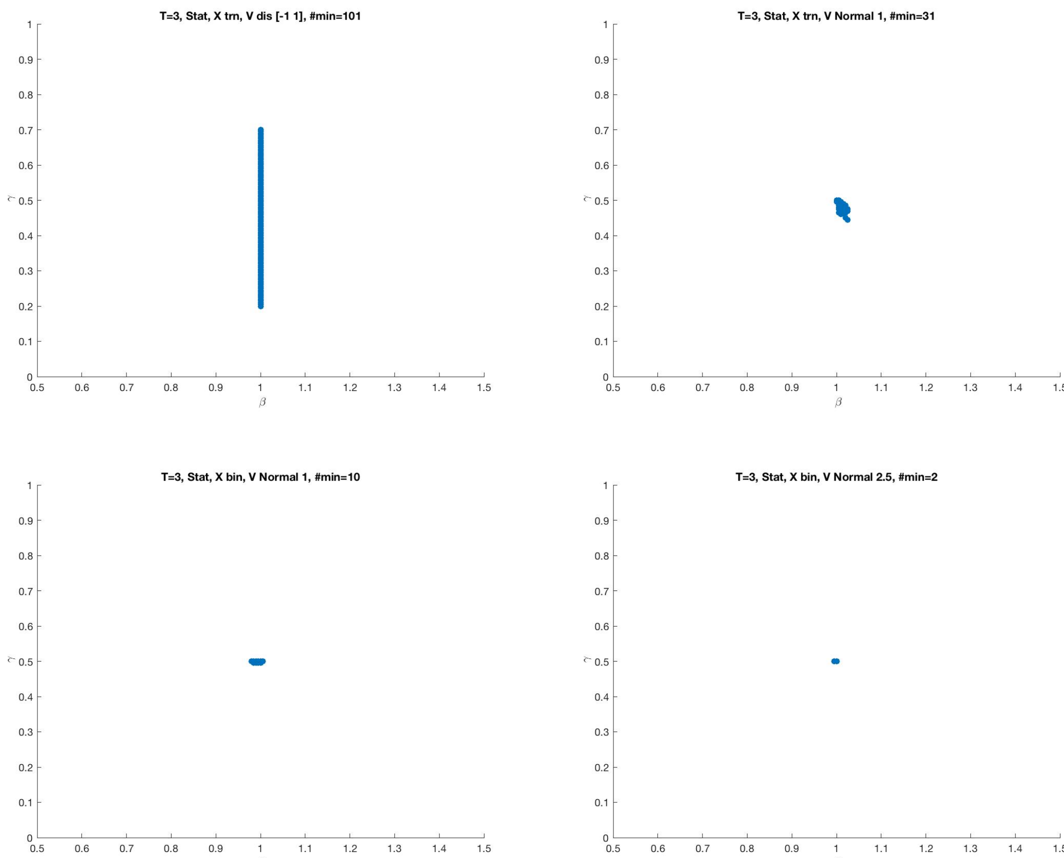

Figure 16: Stationarity with T=3: Various Designs

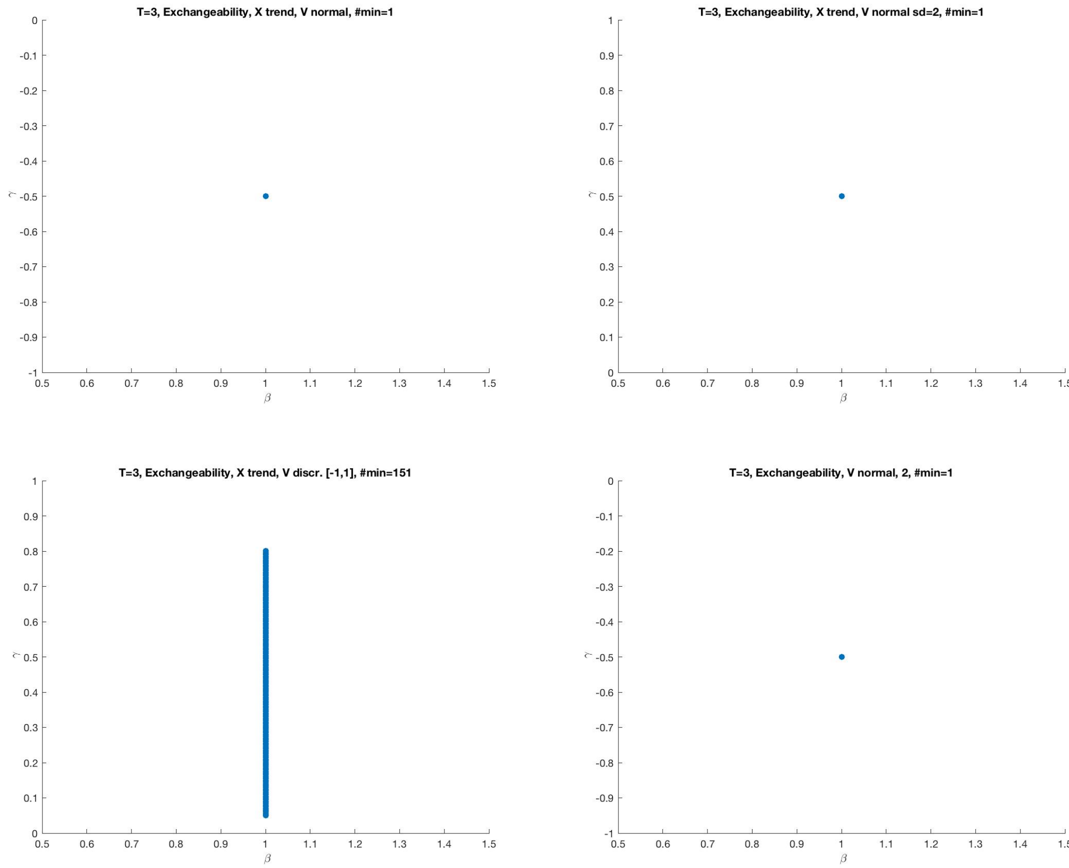

Figure 17: Exchangeability with T=3: Various Designs

6.1.4 Exchangeability Model, T=3

Here we graphed the objective function based on the exchangeability assumption and $T = 3$. It demonstrates we can achieve point identification of γ even when $\gamma = -0.5$. This was also the case when there was strictly nonoverlapping support conditions on x_{it}. In particular, Figure 17 shows that the identified set is essentially a point when v is normal, but γ is not point identified when both v and x are discrete.

7 Conclusion

This paper analyzes the identification of slope parameters in panel binary response models with lagged dependent variables under minimal assumptions. In particular, we consider stationarity and exchangeability and characterize the identified set under these two restrictions without making any assumptions on the fixed effect. We show that the characterization yields the sharp set. In addition, we provide sufficient conditions for point identification. The analysis is interesting and highlights the interplay between the strength of the assumptions, the number of time periods and the support of the exogenous regressors. Overall, we generalize many existing results for this model in various directions.

References

Aguirregabiria, V., J. Gu, and Y. Luo (2018): "Sufficient Statistics for Unobserved Heterogeneity in Structural Dynamic Logit Models," working paper, University of Toronto.

Altonji, J., and R. Matzkin (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053–1102.

Chamberlain, G. (1984): "Panel Data," in *Handbook of Econometrics, Vol. 2*, ed. by Z. Griliches, and M. Intriligator. North Holland.

——— (1985): "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in *Logitudinal Analysis of Labor Market Data*, ed. by J. Heckman, and B. Singer. Cambridge University Press.

——— (1993): "Feedback in Panel Data Models," mimeo., Harvard University.

Chamberlain, G. (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.

Dubé, J.-P., G. J. Hitsch, and P. E. Rossi (2010): "State dependence and alternative explanations for consumer inertia," *The RAND Journal of Economics*, 41(3), 417–445.

Handel, B. R. (2013): "Adverse selection and inertia in health insurance markets: When nudging hurts," *American Economic Review*, 103(7), 2643–82.

Heckman, J. (1981): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data*, ed. by C. Manski, and D. McFadden. MIT Press.

Honoré, B., and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68, 839–874.

Honoré, B., and A. Lewbel (2002): "Semiparametric Binary Choice Panel Data Models without Strictly Exogeneous Regressors," *Econometrica*, 70(5), 2053–2063.

Honoré, B., and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–629.

Honoré, B. E. (1993): "Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables," *Journal of Econometrics*, 59(1-2), 35–61.

Horowitz, J. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60(3).

Hu, L. (2002): "Estimation of a Censored Dynamic Panel Data Model," *Econometrica*, 70(6).

Illanes, G. (2016): "Switching Costs in Pension Plan Choice," *Unpublished manuscript.*

Ketcham, J. D., C. Lucarelli, and C. A. Powers (2015): "Paying attention or paying too much in Medicare Part D," *American Economic Review*, 105(1), 204–33.

Khan, S., M. Ponomareva, and E. Tamer (2011): "Sharpness in Randomly Censored Linear Models," *Economic Letters*, 113, 23–25.

——— (2016): "Identification of Panel Data Models with Endogenous Censoring," *Journal of Econometrics*, 94, 57–75.

Kyriazidou, E. (2001): "Estimation of dynamic panel data sample selection models," *The Review of Economic Studies*, 68(3), 543–572.

Manski, C. F. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55(2), 357–362.

Ouyang, F., S. Khan, and E. Tamer (2019): "Identification and Estimation of Dynamic Panel Data Multinomial Response Models," working paper.

Pakes, A., and J. Porter (2014): "Moment Inequalties for Multinomial Choice with Fixed Effects," Harvard University Working Paper.

Polyakova, M. (2016): "Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D," *American Economic Journal: Applied Economics*, 8(3), 165–95.

Raval, D., and T. Rosenbaum (2018): "Why Do Previous Choices Matter for Hospital Demand? Decomposing Switching Costs from Unobserved Preferences," *Review of Economics and Statistics*, forthcoming.

Shi, X., M. Shum, and W. Song (2018): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," *Econometrica*, 86(2), 737–761.

Torgovitsky, A. (2016): "Nonparametric inference on state dependence with applications to employment dynamics," working paper, University of Chicago.

A Appendix

Proof of Theorem 3.3: We establish point our conclusions sequentially. We first show β is point identified without having established point identification for γ. Next we explore identification for γ, assuming that β is point identified. For this, we will determine if the sign of γ is identified. Then, if the sign of γ is known, we show its magnitude generally *cannot* be identified. To show the first result by assumption $PID1$ we can find a subset of regressor values that lie in $\mathcal{X}_7 \cap \mathcal{X}_8$. By 3.1, $x_1'\beta \geq x_2'\beta$ and $x_1'\beta \leq x_2'\beta$. We can thus identify β by $(x_2 - x_1)'\beta \equiv \Delta x \beta = 0$ and Assumption $PID2, PID3$.

With β identified we can turn attention to the point identification of γ. We first establish if the sign of γ is identified. To to this, note by Assumption PID1, there exists a subset of $\mathcal{X}$ with positive density such that $P(y_2 = 1|x) = P(y_1 = 1|x)$ where $x = (x_1, x_2)$. For this subset $\gamma = 0$ if and only if $\Delta x \beta = 0$. Thus with the identification of β we identify whether or not $\gamma = 0$.

To infer if $\gamma \leq 0$ or not, condition on $\mathcal{X}_9$. If $\gamma > 0$, values of x in this set must satisfy $\Delta x'\beta \geq 0$. Thus if there exists an $x \in \mathcal{X}_9$ s.t. $\Delta x'\beta \leq 0$, one can conclude that $\gamma \leq 0$. By the same argument can conclude that $\gamma \geq 0$ by conditioning on the set $\mathcal{X}_5$. In this set , if $\gamma < 0$, values of x must satisfy $\Delta x'\beta \leq 0$. Thus if there exists an $x \in \mathcal{X}_5$ s.t. $\Delta x'\beta \geq 0$, one can conclude that $\gamma \geq 0$. So, overall, an additional condition which would be sufficient for identification of the sign of γ would be positive measure of the set $\mathcal{X}_{sgn} = \{x \in (\mathcal{X}_5 \cap \mathcal{X}_{1/2}) \cup (\mathcal{X}_9 \cap \mathcal{X}_{1/2'})\}$, where $\mathcal{X}_{1/2}, \mathcal{X}_{1/2'}$ are the half spaces $\{x \in \mathcal{X} : \Delta x'\beta \geq 0\}, \{x \in \mathcal{X} : \Delta x'\beta \leq 0\}$ respectively.

Finally, to attain identification of γ, with knowledge of its sign, we will the show result based on the assumption that $\gamma \geq 0$. This is without loss of generality and the same arguments we use here can be used if $\gamma \leq 0$, just by conditioning on different subsets of $\mathcal{X}$. For the case when $\gamma \geq 0$, we condition on values of x that lie in both $\mathcal{X}_6 \cap \mathcal{X}_9$. For these values of x, it must be the case that $\gamma = \Delta x'\beta$. With β already identified, we could then identify γ. However, we show here that it cannot be the case that x simultaneously lies in $\mathcal{X}_6$ and $\mathcal{X}_9$. Notationally, let P_0, P_1, P_2 denote $P(y_0 = 1|x), P(y_1 = 1|x), P(y_2 = 1|x)$,respectively. Assume joint events are independent for ease of illustration. So implication (6) in Theorem 3.1 can be expressed as $P_0 P_1 > P_2$, and implication (9) in Theorem 3.1 can be expressed as $P_0(1 - P_1) + (1 - P_1)(P_2) > 1$. If they are both true we can add the inequalities resulting in $P_0 + P_2(1 - P_1) > P_2 + 1$. But this cannot generally be true since $P_2(1 - P_1) < P_2$ and $P_0 < 1$, though both sides of both inequalities will be 0 and 1 respectively be at limiting values, $P_0 = 1, P_2 = 0, P_1 = 1$. This can result in an equality (and potential identification) at limiting values.

We focused on $\mathcal{X}_6, \mathcal{X}_9$ because they gave us opposite sign inequalities for γ which would be necessary to attain equalities involving γ. There are several other pairs of implications to work with yielding opposite sign inequalities. Similar arguments can be used to show that the pairs of implications cannot (generally) simultaneously be satisfied, establishing nonidentification.

So, in summary, under our conditions we can point identify β and the sign of γ, but generally not γ itself.

Proof of Theorem 3.4

Our approach will be similar to that used in the proof of Theorem 3.3. Specifically, we will show identification in 3 stages. We first show the regression coefficient(s) β are identified. This will then be used to fist identify the sign of γ, which can then be used to identify γ itself.

Define the following events:

1. $\mathcal{A} = (x_2 - x_1)'\beta - \gamma y_0 \leq 0$ and $(x_2 - x_1)'\beta - \gamma y_0 + \gamma \leq 0$
2. $\mathcal{B} = (x_2 - x_1)'\beta - \gamma y_0 \geq 0$ and $(x_2 - x_1)'\beta - \gamma y_0 + \gamma \geq 0$

From which we can define their respective complements:

1. $\mathcal{A}^c = (x_2 - x_1)'\beta - \gamma y_0 > 0$ or $(x_2 - x_1)'\beta - \gamma y_0 + \gamma > 0$
2. $\mathcal{B}^c = (x_2 - x_1)'\beta - \gamma y_0 < 0$ or $(x_2 - x_1)'\beta - \gamma y_0 + \gamma < 0$

To show identification of β we use Theorem 3.2, and work with the sets:

$$\begin{aligned}
\mathcal{X}2_3 &= \{x \in \mathcal{X} : P(y_1 = 1|y_0, x) \geq P(y_2 = 1|y_0, x)\} \\
\mathcal{X}2_4 &= \{x \in \mathcal{X} : P(y_1 = 1|y_0, x) \leq P(y_2 = 1|y_0, x)\} \\
\mathcal{X}2_5 &= \{x \in \mathcal{X} : P(y_1 = 0, y_2 = 1|y_0, x) \leq P(y_1 = 1|y_0, x)\} \\
\mathcal{X}2_6 &= \{x \in \mathcal{X} : P(y_1 = 1, y_2 = 0|y_0, x) \leq P(y_1 = 0|y_0, x)\}
\end{aligned}$$

which by Assumption $P2ID1$have positive measure. Note we can identify β independently of identifying γ. If we are in $\mathcal{X}2_5$ and $y_0 = 0$ and we are in $\mathcal{X}2_6$ and $y_0 = 1$ it must be the case that $\Delta x\beta = 0$ from which we can identify β.

Having identified β, we next turn attention to identifying the sign of γ. As was the case in our proof of Theorem 3.3, with β having been shown to be identified we can now treat its value as known, and as before we will focus on the region where $\Delta x\beta \approx 0$. This region has positive measure by our Assumption $P2ID3$ that $\Delta x'\beta$ has positive density on the real line after conditioning on being in the regions $\mathcal{X}2_j \;\; j = 1-6$. Thus, if we are in the complement of $\mathcal{X}2_1$, this implies $\mathcal{A}^c$, which implies after conditioning on $\Delta z'\beta = 0, y_0 = 0$, that $\gamma \geq 0$. This will also be true if we are in the complement of $\mathcal{X}2_2$, which implies $\mathcal{B}^c$, which we have if $\Delta x\beta = 0$ and $y_0 = 1$. Conversely, we conclude $\gamma \leq 0$ if we are in the complement of $\mathcal{X}2_2$ and $y_0 = 0, \Delta x\beta = 0$, or we are in the complement of $\mathcal{X}2_1$ and $y_0 = 1, \Delta x'\beta = 0$.

Next we focus on identification of γ. Recall the sign of γ is identified. If it is nonnegative, and we are in $\mathcal{X}2_3 \cap \mathcal{X}2_5$ and $y_0 = 1$, we have $\Delta x\beta - \gamma = 0$ from which we can identify γ given β has been identified.

This addresses point identification of γ for the case when $\gamma \geq 0$. We next turn attention to the case where $\gamma < 0$. Interestingly, while we **can** still identify β, we **cannot** point identify γ for this case.

To see why, we again look at the inequalities implied by each of the regions $\mathcal{X}3_3, \mathcal{X}3_4, \mathcal{X}3_5, \mathcal{X}3_6$. Note if $y_0 = 0$ this gives us two distinct inequalities $\Delta x\beta + \gamma \leq 0, \Delta x\beta \geq 0$. Also note if $y_0 = 1$, we have two distinct inequalities, $\Delta x\beta \leq 0, \Delta x\beta - \gamma \geq 0$. So we immediately see that we can identify β if $y_0 = 0$ and we are in $\mathcal{X}2_4 \cup \mathcal{X}2_5$ and $y_0 = 1$ and we are in $\mathcal{X}2_3 \cup \mathcal{X}2_6$.

However we cannot identify γ here because when $y_0 = 0$, $\mathcal{X}2_3, \mathcal{X}2_6$ imply $\Delta x\beta + \gamma \leq 0$, whereas when $y_0 = 1$, $\mathcal{X}2_4, \mathcal{X}2_5$ imply that $\Delta x\beta - \gamma \geq 0$. The fact that the former involves $-\gamma$ and the latter involves $+\gamma$ means we cannot combine them to attain the moment **equality** needed for point identification.

Proof of Theorem 4.3

Our approach will be virtually identical to that used in the proof of Theorem 3.3. Specifically, we will show identification in 3 stages- identification of β independent of γ, identification of sign of γ treating β as known, then identification of γ. The crucial difference is we will show that γ can be identified without relying on limiting probabilities as was the case in the proof of Theorem 3.3. This is what demonstrates the informational content of having an additional time period of data.

To proceed we can treat β and the sign of γ as known, simply by referring to our of proof of Theorem 3.3. So focusing on γ, we first look at the 12 implications stated in Theorem 4.1. These will involve inequalities involving γ. For example implications (11) and (12) gives us upper bounds and lower bounds for γ respectively. So if both implications are true we would have a moment equality to identify γ. Unfortunately as was the case encountered in the proof of Theorem 3.3, both implications cannot be true simultaneously unless we have limiting probability values of 0 and 1.

However, now we can use implications from Theorem 3.1. For example, implication (9) from Theorem 3.1 gives an upper bound of $(x_2 - x_1)'\beta$ for γ. Implication (12) from Theorem 4.1 gave us a lower bound of $(x_3 - x_1)'\beta$ for γ. This will give us a moment equality for γ if both implications are true and $(x_2 - x_1)'\beta = (x_3 - x_1)'\beta$, the latter holding by Assumption $P3ID4$. We note the latter can be true even if *some* of the regressors x have nonoverlapping support since we only need the indexes to match and not the regressors themselves. The former (both implications holding) can be true *without* relying on limiting probabilities.

Proof of Theorem 4.4

We note from Theorem 3.4 that β is point identified regardless of the sign of γ, and γ is identified when γ is nonnegative. We will thus only show that γ is point identified when it is negative. To so, we condition on the complement of $\mathcal{X}4_1$, denoted by $\mathcal{X}4_1^c$. This will imply the complement of the left handside of the first implication in Table 1. This complement is the intersection of two events, each of which is the union of three events. Focusing on the second of the two intersected events (which is the union of 3 events, $(x_2 - x_3)'\beta < 0, (x_2 - x_1)'\beta - \gamma y_0 < 0, (x_1 - x_3)'\beta + \gamma(y_0 - 1) < 0$. Set $y_0 = 1$ and $(x_2 - x_3)'\beta > 0$, $(x_1 - x_3)'\beta > 0$. So for the union of the three events to be true, it must be the case that $(x_2 - x_1)'\beta \leq \gamma$. But recall if we are in $\mathcal{X}2_4, \mathcal{X}2_5$, we must have $(x_2 - x_1)'\beta \geq \gamma$. Combined, this gives us the moment equality condition to identify γ.

If:	Then:
$(x_2-x_3)'\beta \geq 0$, $(x_2-x_3)'\beta \geq \gamma$ or $(x_2-x_3)'\beta \geq 0$, $(x_2-x_1)'\beta-\gamma y_0 \geq 0$, $(x_1-x_3)'\beta+\gamma(y_0-1) \geq 0$	$P(0,1,0\|x,y_0) \geq P(0,0,1\|x,y_0)$
$(x_2-x_3)'\beta \leq 0$, $(x_2-x_3)'\beta \leq \gamma$ or $(x_2-x_3)'\beta \leq 0$, $(x_2-x_1)'\beta-\gamma y_0 \leq 0$, $(x_1-x_3)'\beta+\gamma(y_0-1) \leq 0$	$P(0,1,0\|x,y_0) \leq P(0,0,1\|x,y_0)$
$(x_1-x_3)'\beta+\gamma y_0 \geq 0$, $(x_1-x_2)'\beta+\gamma(y_0-1) \geq 0$, $(x_2-x_3)'\beta \geq 0$ or $(x_1-x_3)'\beta+\gamma y_0 \geq 0$, $-\gamma \geq 0$	$P(1,0,0\|x,y_0) \geq P(0,0,1\|x,y_0)$
$(x_1-x_3)'\beta+\gamma y_0 \leq 0$, $(x_1-x_2)'\beta+\gamma(y_0-1) \leq 0$, $(x_2-x_3)'\beta \leq 0$ or $(x_1-x_3)'\beta+\gamma y_0 \leq 0$, $-\gamma \leq 0$	$P(1,0,0\|x,y_0) \leq P(0,0,1\|x,y_0)$
$(x_1-x_2)'\beta+\gamma(y_0-1) \geq 0$, $\gamma \geq 0$ or $(x_3-x_2)'\beta \geq 0$, $(x_1-x_3)'\beta+\gamma y_0 \geq 0$, $(x_1-x_2)'\beta+\gamma y_0 \geq 0$	$P(1,0,0\|x,y_0) \geq P(0,1,0\|x,y_0)$
$(x_1-x_2)'\beta+\gamma(y_0-1) \leq 0$, $\gamma \leq 0$ or $(x_3-x_2)'\beta \leq 0$, $(x_1-x_3)'\beta+\gamma y_0 \leq 0$, $(x_1-x_2)'\beta+\gamma y_0 \geq 0$	$P(1,0,0\|x,y_0) \leq P(0,1,0\|x,y_0)$
$(x_1-x_2)'\beta+\gamma y_0 \geq 0$, $-\gamma \geq 0$ or $(x_1-x_3)'\beta+\gamma(y_0-1) \geq 0$, $(x_3-x_2)'\beta \geq 0$	$P(1,0,1\|x,y_0) \geq P(0,1,1\|x,y_0)$
$(x_1-x_2)'\beta+\gamma y_0 \leq 0$, $-\gamma \leq 0$ or $(x_1-x_3)'\beta+\gamma(y_0-1) \leq 0$, $(x_3-x_2)'\beta \leq 0$	$P(1,0,1\|x,y_0) \leq P(0,1,1\|x,y_0)$
$(x_1-x_3)'\beta+\gamma(y_0-1) \geq 0$, $(x_1-x_2)'\beta+\gamma y_0 \geq 0$, $(x_2-x_3)'\beta \geq 0$ or $(x_1-x_3)'\beta+\gamma(y_0-1) \geq 0$, $\gamma \geq 0$	$P(1,1,0\|x,y_0) \geq P(0,1,1\|x,y_0)$
$(x_1-x_3)'\beta+\gamma(y_0-1) \leq 0$, $(x_1-x_2)'\beta+\gamma y_0 \leq 0$, $(x_2-x_3)'\beta \leq 0$ or $(x_1-x_3)'\beta+\gamma(y_0-1) \leq 0$, $\gamma \leq 0$	$P(1,1,0\|x,y_0) \leq P(0,1,1\|x,y_0)$
$(x_2-x_3)'\beta \geq 0$, $(x_2-x_3)'\beta+\gamma \geq 0$ or $(x_2-x_3)'\beta \geq 0$, $(x_1-x_3)'\beta+\gamma y_0 \geq 0$, $(x_2-x_1)'\beta+\gamma(1-y_0) \geq 0$	$P(1,1,0\|x,y_0) \geq P(1,0,1\|x,y_0)$
$(x_2-x_3)'\beta \leq 0$, $(x_2-x_3)'\beta+\gamma \leq 0$ or $(x_2-x_3)'\beta \leq 0$, $(x_1-x_3)'\beta+\gamma y_0 \leq 0$, $(x_2-x_1)'\beta+\gamma(1-y_0) \leq 0$	$P(1,1,0\|x,y_0) \leq P(1,0,1\|x,y_0)$

Table 1: Restrictions for $T = 3$ time periods under conditional exchangeability assumption.