Identification of Dynamic Panel Binary Response Models*

S. Khan† M. Ponomareva‡ E. Tamer §

March 2019

Abstract

We analyze identification in dynamic econometric models of binary choice with fixed effects under general conditions. This class of models is often used in the literature to distinguish between state dependence (invariably referred to in the recent literature as switching costs, inertia or stickiness) and heterogeneity. We first characterize the sharp set for parameters in a dynamic panel of binary choice under *conditional stationarity*. The identified set can be characterized by a union of convex polyhedrons. 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. We also provide sufficient conditions for point identification. For inference in cases with discrete regressors, we provide an approach to constructing confidence sets for the identified sets using a linear program that is simple to implement. The paper then provides simulation based evidence on the size and shape of the identified sets in varying designs to illustrate the informational content of different assumptions. We also illustrate the inference approach using a data set on women's labor supply decisions.

Keywords: Binary Choice, Dynamic Panel Data, Partial Identification.

JEL: C22, C23, C25.

*We are grateful for helpful comments from S. Chen, A. Chesher, H.Ichimura, T. Magnac, as well as conference and seminar participants at CUHK, Fudan, HKUST, Iowa State, LSE, Nankai, SMU, SUFE, Toulouse, UCL, UCLA, University of Connecticut, and Wuhan. X. Lin, P. Sarkis and Z. Wang 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 recent renewed interest in empirical economics in estimating models of discrete choice over time. This is partly motivated by empirical regularities: certain individuals are more likely to stay with a choice if they have experienced that choice in the past and this choice "stickiness" has been attributed variably in the literature to *inertia* or *switching costs*. For example, Handel (2013) estimates a model of health insurance choice in a large firm over time 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[1]. The 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 which is focused on the question of identification in these models.

The dynamic discrete choice model has appeared 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, i.e., a similar individual who did not experience the state in the past is less likely to experience it now. The current literature sometimes refers to state dependence as switching costs, inertia or stickiness and can be thought of as a *causal effect* of past occupancy of the state[2]. The second explanation advanced by Heckman is *heterogeneity*, whereby individuals are different in unobservable 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 prescriptions. Hence, the econometrics literature since then has focused on models under which we are able to empirically identify state dependence while allowing for serial correlation. These models differ in the kinds of assumptions used.

[1]See also Ketcham, Lucarelli, and Powers (2015) for work on quantifying 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. Finally, see Erdem, Imai, and Keane (2003) for a study on the importance of dynamics in discrete decision problems.

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

Concretely, the binary dynamic panel data 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

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

where γ is meant to measure the effect of state dependence (also switching costs or inertia). This parameter γ is treated as a fixed (but unknown) constant while the unobservables here take the standard form $u_t - \alpha$ where α is an individual specific and time independent and is meant to capture the systematic correlation of the unobservables over time[3]. 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 $x = (x_1', \ldots, x_T')'$. For important work on inference on (β, γ) here, see Heckman (1981) and Chamberlain (1984). Honoré and Kyriazidou (2000) provide some sufficient conditions for point identification in this model [4].

This paper's main focus is the question of identification of $\theta = (\gamma, \beta')'$ under minimal assumptions. The starting point is a class of models defined by weak assumptions on the distribution of $u|\alpha, x$, and the main contribution is the characterization of the sharp identification region[5] for θ. 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 these assumptions. These identified sets do not require any restrictions on the distribution of α, or restrictions on the support of the regressor vector x over time hence allowing 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 the support of x that provide new point identification results. In addition we provide sufficient conditions for identification of

[3]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.

[4]For other work on different dynamic models, see the thorough literature survey in Arellano and Honoré (2001), as well as the papers Honoré and Tamer (2006), Honoré and Lewbel (2002), Altonji and Matzkin (2005) , Chen, Khan, and Tang (2015)

[5]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 sign of γ in a model with and without covariates with $T = 2, 3$. Complementing our identification results, 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 allowing for partial identification. 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 clever 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. Their approach can be used in many (nonlinear) panel models with fixed effects to approximate the size of the identified set. Ouyang, Khan, and Tamer (2017) extend results in Honoré and Kyriazidou (2000) to cover point identified *multinomial models* with dynamics. Honoré and Kyriazidou (2019) calculate identified regions for parameters in panel data autoregression models (although they do not provide an explicit characterization of these identified regions). 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). In addition, Torgovitsky (2016) constructs identified sets for average causal effect of lagged outcomes in binary response models under minimal assumptions. Finally, the above linear autoregressive model has a direct link to the structural dynamic discrete decision problems in economics. See for example Merlo and Wolpin (2008) and references therein.

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 explores the finite sample properties of our procedures with an empirical application on female employment status as well as reporting 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) and Ouyang, Khan, and Tamer (2017).

2 Dynamic Panel Binary Choice Model

Recall our model of the form:

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

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):

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

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

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

(CID) Conditional i.i.d.:

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

(IND) Full independence, semiparametric:

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

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, (CIID) implies (CEX), and finally, (IND) implies (CIID). As we 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. Note that Assumption (CIID) can be further strengthened to $u_1, \dots, u_T$ being mutually independent and identically distributed *and also* independent from x and α (IND assumption). This latter version is used in Honoré and Kyriazidou (2000) along with other support condition to obtain point identification. We also require throughout the Assumption below.

Assumption 2.1. *Suppose that the following conditions hold for model 2.1:*

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

A2. *We observe* $i = 1, \dots, n$ *i.i.d. draws from (2.1):* $\{y_i = (y_{i0}, \dots, y_{iT}), x_i = (x_{i1}, \dots, x_{iT}), i = 1, \dots, n\}$.

These are common regularity conditions. The first allows us to use strict monotonicity of the distribution functions of various objects. A2 describes the sampling process. Finally, we assume that n is large relative to T, so any notions of asymptotics are derived under the assumption that $n \to \infty$, while T is fixed. 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 $t = 2$.

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)*

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

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 outcome y_0.

The result below (the proof of which is in the Appendix) 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 Assumption 2.1 hold. 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) $P(y_1=0,y_2=1|x)+P(y_0=1,y_1=0|x) \geq P(y_1=1|x)+P(y_2=0|x) \Rightarrow (x_2-x_1)'\beta - \min\{0,\gamma\} \geq 0$;

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

(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 3.1 with* $T=2$.

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 estimated from the data. 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 from the theorem is that the set based on the above inequalities, is the smallest set attainable based on the stated assumptions in the model (and the data).

3.2 Exchangeability with $T = 2$

In this section, we replace the conditional (on α and x) stationarity assumption with conditional (on α, x, y_0) *exchangeability*[6] of idiosyncratic error terms and investigate its identifying power.

Definition 1. *A sequence $u_1, u_2, \ldots, u_T$ is exchangeable conditional on α, x, y_0 if the following conditions hold:*

(i) $u_t \sim u_1|\alpha, x, y_0$ for all $t = 2, \ldots, T$.

(ii) For any $k \leq T$ and any permutation $(s_1, \ldots, s_k)$ of $(1, \ldots, k)$,

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

A simple example of an exchangeable sequence is a sequence of i.i.d. random variables: if $u_1, \ldots, u_T$ are i.i.d. conditional on α, x, y_0, then that sequence is also exchangeable. In general, conditional independence is a stronger assumption than exchangeability, but as we show in this paper, for the identification purposes in a dynamic panel binary choice model 2.1, the two assumptions are equivalent. Note also that the above exchangeability is stated conditional on the initial condition y_0. We do not think this is a strong assumption. In particular, the lemma below provides sufficient conditions on the structural components of model 2.1 that guarantee this conditional (on y_0) exchangeability.

Lemma 3.1. *Assume that model 2.1 has a beginning M periods back (no dynamic component), so that*

$$y_{-M} = 1\{u_{-M} \leq x'_{-M}\beta + \alpha\}$$

where $M \geq 0$. Also, assume that $(u_{-M}, \ldots, u_0, u_1, \ldots, u_T)$ is exchangeable conditional on $\alpha, x_{-M}, \ldots, x_0, x_1, \ldots, x_T$. Then $(u_1, \ldots, u_T)$ is exchangeable conditional on α, x, y_0 where $x = (x_1, \ldots, x_T)$.

The proof of this result is in the Appendix. The assumption that there is the beginning (or start) is crucial here, since it allows us to condition on a finite (although unobserved) history

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

$u_{-M}, \ldots, u_0$, $x_{-M}, \ldots, x_0$, and α; and at the same time y_0 (the first observable outcome) is a *deterministic* function of that history.

Given the above, we analyze the identifying power of the conditional exchangeability of idiosyncratic error terms u_t's, summarized in the following assumption:

Assumption 3.2. *(CEX): $u_1, \ldots, u_T$ are exchangeable conditional on α, x, y_0.*

An even stronger alternative to the stationarity Assumption 3.1 is the conditional independence assumption:

Assumption 3.3. *(CID): $u_1, \ldots, u_T$ are independent and identically distributed conditional on α, x, y_0.*

For $T = 2$, the assumption of exchangeability implies two things: that u_1 and u_2 have the same marginal distribution (that part is similar to the stationarity assumption 3.1), and also that *jointly*, (u_1, u_2) has the same distribution as its permutation (u_2, u_1). The conditional (on α, x, and y_0) stationarity of u_t's implies that u_t's are also stationary conditional on α and x only, so we expect (and show) that exchangeability provides stronger identifying results than the stationarity assumption discussed in the previous section. At the same time, we also show that when $T = 2$, the pair-exchangeability (part (ii) of Definition 1) does not add any identifying power on top of stationarity condition (part (i)). However, with $T = 3$, pair- and triplet- exchangeability has significant identifying power on top of stationarity (see Section 4).

The identifying power of Assumption 3.2 comes from two parts of Definition 1. Specifically, restrictions on parameters β and γ implied by part (ii) are given in the proposition below.

Proposition 3.1. *Assume that Assumption 2.1 holds, and suppose that $(u_1, u_2) \sim (u_2, u_1)$ conditional on (α, x, y_0) (so that part (ii) of Definition 1 holds for $T = 2$). Then parameters β and γ must satisfy the following conditions for every (x, y_0) in the support:*

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

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

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

Proof: Conditional exchangeability together with part A1 of Assumption 2.1 assumption imply 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 $P(y_1 = 1, y_2 = 0|\alpha,x,y_0) \leq P(y_1 = 0, y_2 = 1|\alpha,x,y_0)$

where strict inequalities in parameters imply strict inequalities in probabilities (since u_1, u_2 are absolutely continuous conditional on α, x, y_0. Integrating α out gives us conditions (1) and (2). ■

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 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[7] on the support of x.

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

Theorem 3.2. *Suppose that Assumption 2.1 holds. Let $\Theta_{I,cex}^{\{1,2\}}(2)$ be the set of parameters that satisfy conditions (1) and (2) of Proposition 3.1. Also let $\Theta_{I,cex}^{\{1,2\}}(1)$ satisfy the following restriction: if for some $z = (x, y_0)$*

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

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

(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 $\tilde{\theta} \in \Theta_{I,cex}^{\{1,2\}}(1)$. Then $\Theta_{I,cex}^{\{1,2\}} = \Theta_{I,cex}^{\{1,2\}}(2) \cap \Theta_{I,cex}^{\{1,2\}}(1)$ is the sharp identified set for θ under either conditional exchangeability assumption 3.2 or conditional independence assumption 3.3 when $T = 2$.

In this result, $\Theta_{I,cex}^{\{1,2\}}(2)$ 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 (which is an implication of conditional exchangeability). Conditional exchangeability assumption also implies that u_1 and u_2 are identically distributed, too, so $\Theta_{I,cex}^{\{1,2\}}(1)$ is the sharp identified set for θ that respects the condition where u_1 and u_2 are identically distributed conditional on y_0, x, and α. We show that with only two time periods, exchangeability assumption does not add anything on top of stationarity assumptions, as summarized in the following corollary.

Corollary 3.1. *Under Assumption 2.1, $\Theta_{I,cex}^{\{1,2\}}(1) \subset \Theta_{I,cex}^{\{1,2\}}(2)$.*

Proof: Since $P(y_1 = 1|z) = P(y_1 = 1, y_2 = 0|z) + P(y_1 = 1, y_2 = 1|z)$ and $P(y_2 = 1|z) = P(y_1 = 0, y_2 = 1|z) + P(y_1 = 1, y_2 = 1|z)$, we can write conditions (1) and (2) for set $\Theta_{I,cex}^{\{1,2\}}(1)$ as

[7] Note that in the Manski model, β can be point identified even with discrete regressors as long as choice probabilities are equal to $\frac{1}{2}$ on a sufficiently rich set. See Condition 3 above.

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

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

Since u_1, u_2 are strictly continuously distributed conditional on α, z, strict inequalities in parameters imply strict inequalities in probabilities in Proposition 3.1, so that we have:

(1) $(x_2 - x_1)'\beta - \gamma y_0 + \max\{0, \gamma\} < 0 \Rightarrow P(y_1 = 1, y_2 = 0|x, y_0) > P(y_1 = 0, y_2 = 1|x, y_0)$

(2) $(x_2 - x_1)'\beta - \gamma y_0 + \min\{0, \gamma\} > 0 \Rightarrow P(y_1 = 1, y_2 = 0|x, y_0) < P(y_1 = 0, y_2 = 1|x, y_0)$

which is equivalent to conditions (1) and (2) for $\Theta_{I,cex}^{\{1,2\}}(1)$. ■

3.2.1 Proof of Theorem 3.2

The proof of Theorem 3.2 above consists of three Lemmas. First, Lemma 3.2 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 3.3 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 (see for example 3) and 4) in Theorem 3.2). Lemma 3.4 provides a similar result under conditional iid assumption.

Some notation that will be useful here: $u = (u_1, \ldots, u_T)$ denotes the vector of error terms. Also, $F_{u,\alpha|z}$ denotes the distribution of unobservables conditional on $z = (y_0, x)$, and $p(y_0, y_1, y_2, x|\theta, F_{u,\alpha|z})$ denotes the distribution of observables in the model characterized by parameter θ and $F_{u,\alpha|z}$.

Lemma 3.2. *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 for any $d_1, \ldots, d_T \in \{0, 1\}$:*

$$P(\tilde{y}_1 = d_1, \ldots, \tilde{y}_T = d_T|y_0, x) = P(y_1 = d_1, \ldots, y_T = d_T|y_0, x)$$

Proof: Although Lemma 3.2 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 $(u_1 - \alpha, \ldots, u_T - \alpha)$ to a conditional iid model $(\tilde{u}_1 - \tilde{\alpha}, \ldots, \tilde{u}_T - \tilde{\alpha})$. So below we prove a general result for an arbitrary sequence $(a_1, \ldots, a_T)$. First, note that if $u_1, \ldots, u_T$ are exchangeable conditional on α and z, then $u_1 - \alpha, \ldots, u_T - \alpha$ 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 $\tilde{\alpha}$, $u_1 - \alpha, u_2 - \alpha, \ldots, u_T - \alpha$ are iid:

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

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

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

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

(i) both models are observationally equivalent:

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

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

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

The next lemma is a key result in the paper. For any parameter in $\Theta_{I,cex}^{\{1,2\}}$ it constructs a model

$$\tilde{y}_t = I\{\tilde{u}_t \le x_t'\tilde{\beta} + \tilde{\gamma}\tilde{y}_{t-1} + \tilde{\alpha},\ t = 1,2\}$$

where the distribution $\tilde{F}_{\tilde{u},\tilde{\alpha}|z}$ obeys the exchangeability assumption, and where the distribution of $(\tilde{y}_0, \tilde{y}_1, \tilde{y}_2, x)$ is the same as in the true model, i.e.

$$p(\tilde{y}_0, \tilde{y}_1, \tilde{y}_2, x|\tilde{\theta}, \tilde{F}_{\tilde{u},\tilde{\alpha}|z}) = p(y_0, y_1, y_2, x|\theta, F_{u,\alpha|z})$$

Lemma 3.3. *Let $\mathcal{F}_{cex}$ be the set of all distributions $F_{u,\alpha|z}$ such that $F_{u|\alpha,z}$ satisfies exchangeability assumption for all α and $z = (x, y_0)$. Then for any $\tilde{\theta} \in \Theta_{I,cex}^{\{1,2\}}$, $\theta \overset{o.e.}{\sim}_{\mathcal{F}} \tilde{\theta}$ under $\mathcal{F}_{cex}$.*

Proof: First, let $\tilde{y}_0 = y_0$ so both models begin identically. Let's pick an arbitrary parameter $\tilde{\theta} \in \Theta$ and consider the following model

$$\tilde{y}_t = I\{\tilde{u}_t \le x_t\tilde{\beta} + \tilde{\gamma}\tilde{y}_{t-1} + \tilde{\alpha}\}$$

where $\tilde{\alpha} = 0$ (zero fixed effects).

The discussion below is conditional on $z = (y_0, x)$. For each z, we define

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

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

An arbitrary choice for $\tilde{\theta} = (\tilde{\beta}, \tilde{\gamma})$ 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 $\tilde{q}_1(z), \tilde{q}_{20}(z), \tilde{q}_{21}(z)$.

A choice of $\tilde{\theta}$ corresponds to an ordering of $\tilde{q}_1(z), \tilde{q}_{20}(z), \tilde{q}_{21}(z)$. Consider a bivariate copula $\tilde{C}(\cdot,\cdot|z)$ such that

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

That is, that copula matches conditional probabilities of observable outcomes: 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 equation matches $P(y_1 = 0, y_2 = 1|z)$.

Copula $\tilde{C}$ is a non-negative, non-decreasing function where

$$\tilde{C}(\tilde{q}_1(z), 0|z) = 0 \text{ and } \tilde{C}(\tilde{q}_1(z), 1|z) = \tilde{q}_1(z)$$

Also, for any q_2, $\tilde{C}(\tilde{q}_1(z), q_2|z)$ 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 visually illustrate potential solutions to (3.2) and some restrictions that these solutions must satisfy, we turn to Figure 1. Here, 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 horizontal line $P(y_1 = 1, y_2 = 1|z)$, and the blue line $q_2 - P(y_1 = 0, y_2 = 1|z)$. If $\tilde{C}(P_1^1, \cdot)$ 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 $\tilde{q}_{21}(z) < \tilde{q}_{20}(z)$.

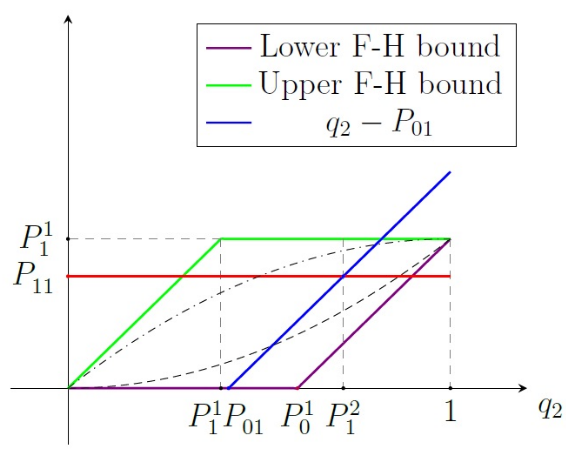

Figure 1: $P_j^t = P(y_t = j|z)$ and $P_{jk} = P(y_1 = j, y_2 = k|z)$. The dashed curves represent potential copula function $\tilde{C}(P_1^1, q_2)$ as a function of q_2.

Since copulas are non-negative, $\tilde{q}_{20}(z) \geq P_{01}$. So if $P(y_1 = 0, y_2 = 1|z) \geq P(y_1 = 1|z)$, then $\tilde{q}_{20}(z) > \tilde{q}_1(z)$ and so we need to place the following restrictions on $\tilde{\theta}$ to be able to find a solution to (3.2):

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

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 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 $P(y_1 = 0|z) < P(y_1 = 1|z) - P(y_1 = 1, y_2 = 1|z) \equiv P(y_1 = 1, y_2 = 0|z)$, 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'll be able to find a solution to (3.2) for a chosen $\tilde{\theta}$ only if

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

Similarly, if $P(y_1 = 1|z) \geq P(y_2 = 1|z)$, we'll be able to find a solution to (3.2) for a chosen $\tilde{\theta}$ only if

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

Note that if $\tilde{\theta}$ satisfies these four restrictions above, then there exists a solution to (3.2): for example, we can choose a bivariate Fréchet copula that is a convex combination of upper and lower Fréchet-Hoeffding bounds (which are proper copulas in a bivariate case) and the bivariate independence copula:

$$\tilde{C}(q_1, q_2) = \delta_1 \min\{q_1, q_2\} + \delta_2 \max\{q_1 + q_2 - 1, 0\} + \delta_3 q_1 q_2$$

$$\text{where } \delta_1 + \delta_2 + \delta_3 = 1 \text{ and } d_j \geq 0$$

A bivariate Fréchet copula is symmetric. If the joint distribution of $\tilde{u}_1$ and $\tilde{u_2}$ is defined by a symmetric copula, i.e. if

$$\tilde{F}_{\tilde{u}_1,\tilde{u}_2|z} = \tilde{C}(\tilde{F}, \tilde{F})$$

then $\tilde{u}_1$ and $\tilde{u}_2$ are exchangeable.

Given a solution $q_1(z)$, $q_{20}(z)$ and $q_{21}(z)$ to (3.2) described above, we can always find a differentiable cumulative distribution function $\tilde{F}$ such that (3.1) holds, so that $\tilde{u}_t$ is absolutely continuous w.r.t. Lebesgue measure μ.

Finally, $\tilde{\theta}$ satisfies the four restrictions described above if and only if $\tilde{\theta} \in \Theta^{\{1,2\}}_{I,cex}$. That is, for any $\tilde{\theta} \in \Theta^{\{1,2\}}_{I,cex}$ we are able to find the distribution $\tilde{F}_{\tilde{u},\tilde{\alpha}|z} \in \mathcal{F}_{cex}$ such that

$$p(\tilde{y}_0, \tilde{y}_1, \tilde{y}_2, x|\tilde{\theta}, \tilde{F}_{\tilde{u},\tilde{\alpha}|z}) = p(y_0, y_1, y_2, x|\theta, F_{u,\alpha|z})$$

which means $\tilde{\theta}$ is observationally equivalent to the true θ under $\mathcal{F}_{cex}$. ∎

Next, we show that the same set $\Theta^{\{1,2\}}_{I,cex}$ gives us the set of parameters that are observationally equivalent to the true parameter under the conditional iid assumption.

Lemma 3.4. *Let $\mathcal{F}_{cex}$ be the set of all distributions $F_{u,\alpha|z}$ such that $F_{u|\alpha,z}$ satisfies exchangeability assumption for all α and $z = (x, y_0)$. Then for any $\tilde{\theta} \in \Theta^{\{1,2\}}_{I,cex}$, $\theta \overset{o.e.}{\sim}_{\mathcal{F}} \tilde{\theta}$ under $\mathcal{F}_{ciid}$.*

Proof: Now suppose that $F \in \mathcal{F}_{ciid}$. As before, everything below is conditional on z. Suppose that $\tilde{\theta} \in \Theta^{\{1,2\}}_{I,cex}$ and $\tilde{C}$ is a Fréchet copula that solves equation (3.2) with marginal distribution $\tilde{F}$ for that $\tilde{\theta}$. Yang, Qi, and Wang (2009) show for any $T > 2$ there exists a T-variate copula $\tilde{C}_T$ such that

(i) all of its two-dimensional margins are given by the bivariate Fréchet copula $\tilde{C}$;

(ii) there exists a sequence of T uniform $[0, 1]$ random variables $Q_1, \ldots, Q_T$ and a uniform $[0, 1]$ random variable $\tilde{Q}$ such that $Q_1, \ldots, Q_T$ are iid conditional on $\tilde{Q}$ and the joint distribution of $Q_1, \ldots, Q_T$ is given by $\tilde{C}$.

Now let's define $\tilde{u}_t = \tilde{F}^{-1}(Q_t)$. Then the sequence $\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_T$ is exchangeable for any T, which in turn implies that the sequence $\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_T$ is infinitely exchangeable (see

Konstantopoulos and Yuan (2018)). Finally, Lemma 3.2 guarantees that there exists $\tilde{F}_{\tilde{u},\tilde{\alpha}|z}$ such that $\tilde{u}_1, \dots, \tilde{u}_t$ are iid conditional on $\tilde{\alpha}$ and z, and by construction,

$$p(\tilde{y}_0, \tilde{y}_1, \tilde{y}_2, x|\tilde{\theta}, \tilde{F}_{\tilde{u},\tilde{\alpha}|z}) = p(y_0, y_1, y_2, x|\theta, F_{u,\alpha|z})$$

■

Finally, to complete the proof of sharpness of $\Theta_{I,cex}^{\{1,2\}}$, we need to show that θ is identified relative to any $\tilde{\theta} \notin \Theta_{I,cex}^{\{1,2\}}$ under either conditional exchangeability or conditional iid assumptions. Assume that for a given $\tilde{\theta} \notin \Theta_{I,cex}^{\{1,2\}}$ e.g. condition (1) of Theorem 3.2 does not hold. That is, there exists some $z = (y_0, x)$ such that

$$P(y_1 = 1|z) \geq P(y_2|z) \text{ and } (x_2 - x_1)'\tilde{\beta} + \min\{0, \tilde{\gamma}\} - \tilde{\gamma} y_0 > 0$$

However, if there exists $\tilde{F}_{\tilde{u},\tilde{\alpha}|z} \in \mathcal{F}_{cex}$ such that

$$p(\tilde{y}_0, \tilde{y}_1, \tilde{y}_2, x|\tilde{\theta}, \tilde{F}_{\tilde{u},\tilde{\alpha}|z}) = p(y_0, y_1, y_2, x|\theta, F_{u,\alpha|z})$$

then it must be the case that (see the proof of Lemma 3.3)

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

And since $\mathcal{F}_{cex}$ and $\mathcal{F}_{ciid}$ are observationally equivalent here, this completes the proof of 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) and that $\Delta u = u_1 - u_2$ is absolutely continuous. Then for pairs of (x, y_0) and $(\tilde{x}, \tilde{y}_0)$, the parameters β and γ must satisfy the following restrictions:*

(1) If $P(y_1 = 1, y_2 = 0|x, y_0) \geq 1 - P(y_1 = 0, y_2 = 1|\tilde{x}, \tilde{y}_0)$, then $((x_1 - x_2) - (\tilde{x}_1 - \tilde{x}_2))'\beta + \gamma(y_0 - \tilde{y}_0 - 1) \geq 0$.

(2) If $P(y_1 = 1, y_2 = 0|x, y_0) > 1 - P(y_1 = 0, y_2 = 1|\tilde{x}, \tilde{y}_0)$, then $((x_1 - x_2) - (\tilde{x}_1 - \tilde{x}_2))'\beta + \gamma(y_0 - \tilde{y}_0 - 1) > 0$.

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

Given that $y_0, \tilde{y}_0$ are binary, by looking at (1) and (2) in the Proposition 3.2 above, we see 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:

(1) If $P(y_1 = 1, y_2 = 0|x, y_0) \geq 0.5$, then $(x_1 - x_2)'\beta + \gamma(y_0 - 1) \geq 0$.

(2) If $P(y_1 = 0, y_2 = 1|x, y_0) \geq 0.5$, then $(x_1 - x_2)'\beta + \gamma y_0 \leq 0$.

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).

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 set in Theorem 3.2.

3.4 Point Identification when $T = 2$

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 but its sign can be. This is in contrast to the $T = 2$ with exchangeability where both *beta* 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 Stationarity

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. Let $\mathcal{X} \subseteq \mathbb{R}^{2k}$ denote the support of $x = (x_1, x_2)$. Following Theorem 3.1, we define ten subsets (one for each inequality) of $\mathcal{X}$ that can help us to identify the sign of γ:

$$
\begin{aligned}
\Delta\mathcal{X}_1 &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_1 = 1|x) > P(y_2 = 1|x)\} \\
\Delta\mathcal{X}_2 &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_2 = 1|x) > P(y_1 = 1|x)\} \\
\Delta\mathcal{X}_3 &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that} \\
&\qquad P(y_1 = 0, y_2 = 1|x) + P(y_0 = 1, y_1 = 0|x) > P(y_1 = 1|x) + P(y_2 = 0|x)\} \\
\Delta\mathcal{X}_4 &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that} \\
&\qquad P(y_0 = 1, y_1 = 0|x) + P(y_0 = 0, y_1 = 1|x) > P(y_1 = 0|x) + P(y_2 = 1|x)\} \\
\Delta\mathcal{X}_5 &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_0 = 1, y_1 = 1|x) > P(y_2 = 0|x)\} \\
\Delta\mathcal{X}_6 &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_0 = 1, y_1 = 1|x) > P(y_2 = 1|x)\} \\
\Delta\mathcal{X}_7 &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_0 = 0, y_1 = 0|x) + P(y_1 = 0, y_2 = 1|x) > 1\} \\
\Delta\mathcal{X}_8 &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_0 = 1, y_1 = 1|x) + P(y_1 = 1, y_2 = 0|x) > 1\} \\
\Delta\mathcal{X}_9 &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_0 = 1, y_1 = 0|x) + P(y_1 = 0, y_2 = 1|x) > 1\} \\
\Delta\mathcal{X}_{10} &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_0 = 0, y_1 = 1|x) + P(y_1 = 1, y_2 = 0|x) > 1\}
\end{aligned}
$$

Similarly, we define the two sets that can be used to point identify β, and their union:

$$
\begin{aligned}
\mathcal{X}_7 &= \{x \in \mathcal{X} \text{ such that } P(y_0 = 0, y_1 = 0|x) + P(y_1 = 0, y_2 = 1|x) \geq 1\} \\
\mathcal{X}_8 &= \{x \in \mathcal{X} \text{ such that } P(y_0 = 1, y_1 = 1|x) + P(y_1 = 1, y_2 = 0|x) \geq 1\} \\
\mathcal{X}_{7,8} &= \mathcal{X}_7 \cup \mathcal{X}_8
\end{aligned}
$$

Theorem 3.1 implies that $\Delta\mathcal{X}_7 \subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta > 0\}$ and $\Delta\mathcal{X}_8 \subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta < 0\}$. If these two set are large enough (as formalized in the assumption below), we will be able to identify β from score conditions.

Assumption 3.4. *Suppose that for the sets defined above, the following holds:*

PID-STAT1. $\Delta\mathcal{X}_{7,8} = \{\Delta x = x_2 - x_1 : x = (x_1, x_2) \in \mathcal{X}_{7,8}\}$ *is not contained in any proper linear subspace of* $\mathbb{R}^k$.

PID-STAT2. *There exists at least one* $j \in \{1, \ldots, k\}$ *such that* $\beta_j \neq 0$ *and for any* $\Delta x \in \Delta\mathcal{X}_{7,8}$ *the support of* $\Delta x_j = x_{2j} - x_{1j}$ *is the whole real line (*$x_{2j} - x_{1j}$ *has everywhere positive Lebesgue measure conditional on* $\Delta x_{-j} = x_{2,-j} - x_{1,-j}$*).*

Conditions PID-STAT1 and PID-STAT2 require that there is at least one covariate with large support. This assumption is common in the literature and is e.g. used in Manski (1985) for the cross-sectional semiparametric binary choice model or in Manski (1987) for the static panel data binary choice model.

Under these assumptions we can attain point identification for β, though not γ (although the sign of γ potentially can be identified), as stated in the following theorem that gives sufficient conditions for point identification of β and the sign of γ (the proof of this result is delegated to the Appendix).

Theorem 3.3. *Suppose that Assumptions 2.1, 3.1, and 3.4 hold. Then* β *is point identified (up to scale). Further,*

(1) If $(\Delta\mathcal{X}_1 \cup \Delta\mathcal{X}_5) \cap \Delta\mathcal{X}_{10} \neq \varnothing$ *or* $(\Delta\mathcal{X}_2 \cup \Delta\mathcal{X}_6) \cap (\Delta\mathcal{X}_9) \neq \varnothing$ *or* $\Delta\mathcal{X}_3 \cap \Delta\mathcal{X}_8 \neq \varnothing$ *or* $\Delta\mathcal{X}_4 \cap \Delta\mathcal{X}_7 \neq \varnothing$*, then* $\gamma < 0$.

(2) If $\Delta\mathcal{X}_5 \cap \Delta\mathcal{X}_8 \neq \varnothing$ *or* $\Delta\mathcal{X}_6 \cap \Delta\mathcal{X}_7 \neq \varnothing$*, then* $\gamma > 0$.

(3) If sets in both (1) and (2) have a non-empty intersection, then γ *is zero (so it is point identified).*

(4) Finally, when β *is point identified, we can bound* γ *as follows:*

$$\begin{aligned} |\gamma| &\geq \max\{-m_1, M_2\} \\ \gamma &\leq \min\{m_9, -M_{10}\} \end{aligned} \tag{3.3}$$

where for $j = 1, 2, 9, 10$*:*

$$m_j = \inf_{\Delta x \in \Delta \mathcal{X}_j} \Delta x'\beta, \; M_j = \sup_{\Delta x \in \Delta \mathcal{X}_j} \Delta x'\beta$$

.

Note that the identification of the sign of γ in this result does not rely on β being point identified. However, when the sign of γ is identified, we can weaken Assumption 3.4. In particular, if γ is positive, then we can replace $\mathcal{X}_7$ and $\mathcal{X}_8$ in Assumption 3.4 with $\mathcal{X}_3 \cup \mathcal{X}_7$ and $\mathcal{X}_4 \cup \mathcal{X}_8$, respectively, where

$$\mathcal{X}_3 = \{x \in \mathcal{X} \text{ such that } P(y_1 = 0, y_2 = 1|x) + P(y_0 = 1, y_1 = 0|x) \geq P(y_1 = 1|x) + P(y_2 = 0|x)\}$$
$$\mathcal{X}_4 = \{x \in \mathcal{X} \text{ such that } P(y_0 = 1, y_1 = 0|x) + P(y_0 = 0, y_1 = 1|x) \geq P(y_1 = 0|x) + P(y_2 = 1|x)\}$$

If γ is negative, then we can replace $\mathcal{X}_7$ and $\mathcal{X}_8$ with $\mathcal{X}_5 \cup \mathcal{X}_7$ and $\mathcal{X}_6 \cup \mathcal{X}_8$, respectively, where

$$\mathcal{X}_5 = \{x \in \mathcal{X} \text{ such that } P(y_0 = 1, y_1 = 1|x) \geq P(y_2 = 0|x)\}$$
$$\mathcal{X}_6 = \{x \in \mathcal{X} \text{ such that } P(y_0 = 1, y_1 = 1|x) \geq P(y_2 = 1|x)\}$$

Note also that if 0 belongs to the support of $(x_2 - x_1)$ and if there exists $\tilde{x} = (\tilde{x}_1, \tilde{x}_2 = \tilde{x}_1)'$ such that $P(y_0 = 1, y_1 = 0|x = \tilde{x}) + P(y_1 = 0, y_2 = 1|x = \tilde{x}) \geq 1$, then $\gamma > 0$. Similarly, if there exists $\tilde{x} = (\tilde{x}_1, \tilde{x}_2 = \tilde{x}_1)'$ such that $P(y_0 = 0, y_1 = 1|x = \tilde{x}) + P(y_1 = 1, y_2 = 0|x = \tilde{x}) \geq 1$, then $\gamma < 0$.

3.4.2 Point Identification under Exchangeabililty

As was the case under the stationarity 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. Again, we define the following sets:

$$
\begin{aligned}
\Delta\mathcal{X}_1(y_0) &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_1 = 1|x, y_0) > P(y_2 = 1|x, y_0)\} \\
\Delta\mathcal{X}_2(y_0) &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_2 = 1|x, y_0) > P(y_1 = 1|x, y_0)\} \\
\Delta\mathcal{X}_3(y_0) &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_1 = 0, y_2 = 1|x, y_0) > P(y_1 = 1|x, y_0)\} \\
\Delta\mathcal{X}_4(y_0) &= \{\Delta x \in \mathbb{R}^k : \exists x = (x_1, x_1 + \Delta x) \in \mathcal{X} \text{ such that } P(y_1 = 1, y_2 = 0|x, y_0) > P(y_1 = 0|x, y_0)\} \\
\mathcal{X}_3(y_0) &= \{x \in \mathcal{X} \text{ such that } P(y_1 = 0, y_2 = 1|x, y_0) \geq P(y_1 = 1|x, y_0)\} \\
\mathcal{X}_4(y_0) &= \{x \in \mathcal{X} \text{ such that } P(y_1 = 1, y_2 = 0|x, y_0) \geq P(y_1 = 0|x, y_0)\}
\end{aligned}
$$

Now Theorem 3.2 guarantees that $\Delta\mathcal{X}_1(y_0) \subset \{\Delta x \in \mathbb{R}^k : \Delta x'\beta + \min\{0, \gamma\} - \gamma y_0 < 0\}$, $\Delta\mathcal{X}_2(y_0) \subset \{\Delta x \in \mathbb{R}^k : \Delta x'\beta + \max\{0, \gamma\} - \gamma y_0 > 0\}$, $\Delta\mathcal{X}_3(y_0) \subset \{\Delta x \in \mathbb{R}^k : \Delta x'\beta - \gamma y_0 > 0\}$, and $\Delta\mathcal{X}_4(y_0) \subset \{\Delta x \in \mathbb{R}^k : \Delta x'\beta - \gamma y_0 < 0\}$. The last two sets will allow us to point identify β is $\mathcal{X}_3(0)$ and $\mathcal{X}_4(1)$ have sufficiently large support, as summarized in the assumption below.

Assumption 3.5. *Suppose that for the sets defined above, the following holds:*

PID-CEX1. $\Delta\mathcal{X}_{3,4} = \{\Delta x = x_2 - x_1 : x = (x_1, x_2) \in \mathcal{X}_3(0) \cup \mathcal{X}_4(1)\}$ *is not contained in any proper linear subspace of* $\mathbb{R}^k$ *(where* $x_t = (x_{t1}, \dots, x_{tk})'$*).*

PID-CEX.2 *There exists at least one* $j \in \{1, \dots, k\}$ *such that* $\beta_j \neq 0$ *and for any* $\Delta x \in \Delta\mathcal{X}_{3,4}$ *the support of* $\Delta x_j = x_{2j} - x_{1j}$ *is the whole real line (*$x_{2j} - x_{1j}$ *has everywhere positive Lebesgue measure conditional on* $\Delta x_{-j} = x_{2,-j} - x_{1,-j}$*).*

Note that conditions of Assumptions 3.5 are weaker than those of Assumption 3.4. Specifically, $\mathcal{X}_7 \subset \mathcal{X}_3$ and $\mathcal{X}_8 \subset \mathcal{X}_4$, so that Assumptions 3.5 can hold even when Assumption 3.4 does not.

We state sufficient conditions for point identification of β and identification of γ in the next theorem (the proof is in the Appendix).

Theorem 3.4. *Let Assumptions 2.1, 3.2 and 3.5 hold. Then* β *is point identified (up to scale). Further,*

(1) If $\Delta\mathcal{X}_1(y_0) \cup \Delta\mathcal{X}_3(y_0) \neq \varnothing$ *or* $\Delta\mathcal{X}_2(y_0) \cup \Delta\mathcal{X}_4(y_0) \neq \varnothing$ *or* $\Delta\mathcal{X}_1(0) \cup \Delta\mathcal{X}_2(1) \neq \varnothing$ *or* $\Delta\mathcal{X}_3(0) \cup \Delta\mathcal{X}_4(1) \neq \varnothing$*, then* $\gamma < 0$*.*

(2) If $\Delta\mathcal{X}_2(0) \cup \Delta\mathcal{X}_4(1) \neq \varnothing$ or $\Delta\mathcal{X}_3(0) \cap \Delta\mathcal{X}_1(1) \neq \varnothing$, then $\gamma > 0$.

(3) If sets in both (1) and (2) have a non-empty intersection, then γ is zero (so it's point identified).

(4) Finally, if β is point identified, we can bound γ from above:

$$\begin{aligned} |\gamma| &\geq \max\{Q_1, -q_2\} \\ \gamma &\leq \min\{-Q_4, q_3\} \end{aligned} \tag{3.4}$$

where

$$\begin{aligned} Q_1 &= \sup_{\Delta x \in \Delta\mathcal{X}_1(0) \cup \Delta\mathcal{X}_1(1)} \Delta x'\beta,\ Q_4 = \sup_{\Delta x \in \Delta\mathcal{X}_4(0)} \Delta x'\beta \\ q_2 &= \inf_{\Delta x \in \Delta\mathcal{X}_2(0) \cup \Delta\mathcal{X}_2(1)} \Delta x'\beta,\ q_3 = \inf_{\Delta x \in \Delta\mathcal{X}_3(1)} \Delta x'\beta \end{aligned}$$

Remark 3.2. *Note that in the above, it is possible that γ is point identified when it is nonnegative. Given enough variation in the supports of the sets in (3.4), the upper and lower bounds can collapse to a point. See the Monte Carlo section for such a design where we demonstrate that γ is point identified. However, we can only get an upper bound for γ when γ is negative. So then, generally, γ is not point identified unless 1) it is positive, and 2) we have sufficient variation such that the above upper and lower bounds on γ collapse. Note on the other hand that β is always point identified under the sufficient conditions in Theorem 3.4.*

4 Extensions

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 those models in various directions and show how the identified regions can change. We will extend the model in the time dimension. Specifically we will explore the identifying power of time, by considering models where $T = 3$. As we will show, there is identifying power in observing agents across more time periods. Also, we show how our identification results hold up to the case without covariates.

4.1 Stationarity and Exchangeability with $T = 3$

In this section we explore the informational content of observing data over additional time periods, specifically we assume that $T = 3$. 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 under rather mild support restrictions in the $T = 3$ case but not the $T = 2$ setting. Here we will derive the identified region for β, γ for $T = 3$ under (unconditional on y_0) stationarity and (conditional on y_0) exchangeability assumptions we defined previously. We start with the identified set for $\theta \equiv (\beta, \gamma)$ under the (STAT) assumption which is given in the result below.

Theorem 4.1. *Suppose that Assumptions 2.1 holds. 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) $P(y_2 = 0, y_3 = 1|x) + P(y_0 = 1, y_1 = 0|x) \geq 1 \Rightarrow (x_3 - x_1)'\beta - \gamma \geq 0$;

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

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

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 Assumption 3.1 (stationarity).*

With three time periods, we are able to shrink the identified set under $T = 2$ and stationarity assumption (set $\Theta_{I,stat}^{\{1,2\}}$ defined in Theorem 3.1). Specifically, sets $\Theta_{I,stat}^{\{1,3\}}$ and $\Theta_{I,stat}^{\{2,3\}}$ provide additional restrictions on the parameters of interests β and γ associated with the third time period: $\Theta_{I,stat}^{\{1,3\}}$ gives us the set of parameters that are observationally equivalent to the true parameter $\theta = (\beta', \gamma)'$ under the assumption that u_1 and u_3 are identically distributed, while $\Theta_{I,stat}^{\{2,3\}}$ does the same under the assumption that u_2 and u_3 are identically distributed.

Next, we provide the identified set in the conditional exchangeability case with $T = 3$ where we also condition on the initial value y_0. For $T = 2$ we demonstrated that adding exchangeability assumption on top of assuming that u_1 and u_2 are stationary (identically distributed) conditional on α, x and y_0 did not provide any extra identifying power. However, unlike the two-period case, we will see that with $T = 3$ exchangeability assumption does indeed help to further shrink the identified set (beyond what stationarity assumption alone does). Specifically, when $T = 3$, conditional exchangeability assumption now involves 24 moment inequalities of the "if, then" restrictions that are summarized in Table 3 in the Appendix.

Some of restrictions in Table 3 are the same restrictions we have in Proposition 3.1 for $T = 2$ case (namely, pairs of restrictions 5.a and 6.a, and 7.a and 8.a). However, the remaining 20 inequality conditions are new and never appeared before. The result below is the analog of Proposition 3.1 for $T = 2$;

Proposition 4.1. *Suppose that Assumption 2.1 holds and* $u = (u_1, u_2, u_3)$ *satisfy Assumption 3.2 (exchangeability). Then parameters* β *and* γ *must satisfy the restrictions in Table 3 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 3 reduce to the following set:*

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

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

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

(4) 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 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). However, they mention that similar result can be obtained under weaker conditions of conditional iid assumption. Theorem 4.2 shows that we can also get rid of the first "i" in "iid" by weakening it to conditional exchangeability.

Now 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 either Assumption 3.2 (exchangeability) or Assumption 3.3 (conditional independence).*

Here the intersection of sets $\Theta_{I,stat}^{\{1,2\}}$, $\Theta_{I,stat}^{\{2,3\}}$ and $\Theta_{I,stat}^{\{1,3\}}$ gives us the set of parameters that are observationally equivalent to the true parameter under the assumption that u_1, u_2 and u_3 are identically distributed conditional on x, y_0, and α. Set $\Theta_{I,cex}^{\{1,2,3\}}$ gives us the set of parameters that are observationally equivalent to the true parameter under conditional (on x, y_0, and α) exchangeability of u_1, u_2, u_3. Note that unlike the case with $T = 2$, some of the exchangeability restrictions are not implied but any of the stationarity restrictions. For example, exchangeability-based restriction 1.a in Table 3 is a stronger version the following stationarity-based restriction for $\Theta_{I,cex}^{\{2,3\}}(1)$:

$$P(y_1 = 0, y_2 = 0|x, y_0) \geq P(y_3 = 0|x, y_0) \Rightarrow (x_3 - x_2)'\beta + \max\{0, \gamma\} \geq 0$$

and so on.

4.2 Point Identification with $T = 3$

In this section we provide sufficient conditions for point identification of the parameters β, γ under the stationarity and exchangeability assumptions in the case for $T = 3$. Point identification under stationarity will rely on the result in Theorem 4.1, while point identification under exchangeability will be based on Theorem 4.2.

We start with point identification under stationarity. Theorem 4.1 provides 6 conditions that involve β only, so we can use these conditions to point identify β in a similar way we did for $T = 2$. In particular, we define the following sets

$$\begin{aligned}
\mathcal{X}_7^{\{1,2\}} &= \{x \in \mathcal{X} \text{ such that } P(y_0 = 0, y_1 = 0|x) + P(y_1 = 0, y_2 = 1|x) \geq 1\} \\
\mathcal{X}_8^{\{1,2\}} &= \{x \in \mathcal{X} \text{ such that } P(y_0 = 1, y_1 = 1|x) + P(y_1 = 1, y_2 = 0|x) \geq 1\} \\
\mathcal{X}_7^{\{1,3\}} &= \{x \in \mathcal{X} \text{ such that } P(y_0 = 1, y_1 = 0|x) + P(y_2 = 1, y_3 = 1|x) \geq 1\} \\
\mathcal{X}_8^{\{1,3\}} &= \{x \in \mathcal{X} \text{ such that } P(y_0 = 1, y_1 = 1|x) + P(y_2 = 1, y_3 = 0|x) \geq 1\} \\
\mathcal{X}_7^{\{2,3\}} &= \{x \in \mathcal{X} \text{ such that } P(y_1 = 0, y_2 = 0|x) + P(y_2 = 0, y_3 = 1|x) \geq 1\} \\
\mathcal{X}_8^{\{2,3\}} &= \{x \in \mathcal{X} \text{ such that } P(y_1 = 1, y_2 = 1|x) + P(y_2 = 1, y_3 = 0|x) \geq 1\} \\
\Delta\mathcal{X}_{7,8}^{\{1,2\}} &= \{\Delta x = x_2 - x_1 : x \in \mathcal{X}_7^{\{1,2\}} \cup \mathcal{X}_8^{\{1,2\}}\} \\
\Delta\mathcal{X}_{7,8}^{\{1,3\}} &= \{\Delta x = x_3 - x_1 : x \in \mathcal{X}_7^{\{1,3\}} \cup \mathcal{X}_8^{\{1,3\}}\} \\
\Delta\mathcal{X}_{7,8}^{\{2,3\}} &= \{\Delta x = x_3 - x_2 : x \in \mathcal{X}_7^{\{2,3\}} \cup \mathcal{X}_8^{\{2,3\}}\}
\end{aligned}$$

and make the following assumption:

Assumption 4.1. *Suppose that for the sets defined above, the following holds:*

PID3-STAT1. $\Delta\mathcal{X}_{7,8} = \Delta\mathcal{X}_{7,8}^{\{1,2\}} \cup \Delta\mathcal{X}_{7,8}^{\{1,3\}} \cup \Delta\mathcal{X}_{7,8}^{\{2,3\}}$ *is not contained in any proper linear subspace of* $\mathbb{R}^k$ *(where* $x_t = (x_{t1}, \ldots, x_{tk})'$*).*

PID3-STAT2. *There exists at least one* $j \in \{1, \ldots, k\}$ *such that* $\beta_j \neq 0$ *and for any* $\Delta x \in \Delta\mathcal{X}_{7,8}$ *the support of* Δx_j *is the whole real line (*Δx_j *has everywhere positive Lebesgue measure conditional on* $\Delta x_{-j} = (\Delta x_1, \ldots, \Delta x_{j-1}, \Delta x_{j+1}, \ldots, \Delta x_k)'$*).*

Note that with $T = 3$ this assumption is more likely to hold than a similar assumption for $T = 2$ (Assumption 3.4). Similar to $T = 2$ case, β is point identified if Assumption 4.1 holds.

We won't present here identification results for the sign of γ (again, these are very similar to Theorem 3.3). Instead, we focus on discussing what can be learned about sign and magnitude of γ with that one extra period of observation. In particular, in contrast to the result in Theorem 3.3, we can bound γ directly both from above and from below. In particular, we now have the following restrictions on γ when β is point identified:

$$\begin{aligned}
|\gamma| &\geq \max\{-m_1^{\{1,2\}}, -m_1^{\{1,3\}}, -m_1^{\{2,3\}}, M_2^{\{1,2\}}, M_2^{\{1,3\}}, M_2^{\{2,3\}}\} \\
\gamma &\leq \min\{m_9^{\{1,2\}}, m_9^{\{1,3\}}, m_9^{\{2,3\}}, -M_{10}^{\{1,2\}}, -M_1^{\{1,3\}}, -M_{10}^{\{2,3\}}\} \qquad (4.1) \\
\gamma &\geq \max\{M_{11}^{\{1,3\}}, -m_{12}^{\{2,3\}}\}
\end{aligned}$$

where $m_j^{\{t,s\}}$ and $M_j^{\{t,s\}}$ are defined similar to Theorem 3.3.

In comparison to $T = 2$ case in Theorem 3.3 where we only had an upper bound on γ, with $T = 3$ we now also can bound γ from below (if a certain set is not empty), while the upper bound becomes more tight as well.

4.3 Identification in a Model without Covariates

Here, we consider identification of the sign of γ in the following model:

$$y_t = I\{u_t \leq \gamma y_{t-1} + \alpha\} \quad t = 1, 2, ...T$$

Although the scale of γ cannot be identified in that model, its sign sometimes can be identified. Below we characterize the conditions under which this is possible to do.

We start with $T = 2$ and stationarity Assumption 3.1. The last two inequalities in Theorem 3.1 allow us (sometimes, if a particular relationship between conditional probabilities of certain events holds) to tell that γ is negative. In particular, without covariates x conditions (9) and (10) of Theorem 3.1 become:

$$\begin{aligned}(9) &: P(y_0 = 1, y_1 = 0) + P(y_1 = 0, y_2 = 1) \geq 1 \Rightarrow \gamma \leq 0 \\ (10) &: P(y_0 = 0, y_1 = 1) + P(y_1 = 1, y_2 = 0) \geq 1 \Rightarrow \gamma \leq 0\end{aligned} \tag{4.2}$$

Similarly, Theorem 3.2 allows to potentially identify the sign of γ under conditional exchangeability Assumption 3.2:

$$\begin{aligned}(3) &: P(y_1 = 0, y_2 = 1|y_0 = 1) \geq P(y_1 = 1|y_0 = 1) \Rightarrow \gamma \leq 0 \\ (4) &: P(y_1 = 1, y_2 = 0|y_0 = 0) \geq P(y_1 = 0|y_0 = 0) \Rightarrow \gamma \leq 0\end{aligned} \tag{4.3}$$

Note that conditions (3) and (4) in 4.3 are a weaker set of restrictions on probabilities of observing certain events than conditions (9) and (10) in 4.2. For example, condition (3) can be re-written as

$$P(y_0 = 1, y_1 = 0) + P(y_1 = 0, y_2 = 1) > P(y_0 = 0, y_1 = 0, y_2 = 1) + P(y_0 = 1)$$

The right-hand side of this inequality is less than one, so if the sign of γ is identified under stationarity restriction (i.e. condition (9) in 4.2 holds), then the sign of γ is also be identified under the exchangeability restriction (condition (3) in 4.3 holds), but the reverse is not true.

When $T = 3$, it is sometimes possible to identify the sign of γ even when γ is positive (unlike in $T = 2$ case). We start with stationarity Assumption 3.1: under that assumption, the sharp identified set for β and γ is given by Theorem 4.1. In the absence of covariates, this result (in addition to the restrictions on γ described above for $T = 2$) places the following restrictions based on set $\Theta^{\{1,3\}}_{I,stat}$:

$$\begin{aligned}(9) &: P(y_0 = 1, y_1 = 0) + P(y_2 = 0, y_3 = 1) \geq 1 \Rightarrow \gamma \leq 0 \\ (10) &: P(y_0 = 0, y_1 = 1) + P(y_2 = 1, y_3 = 0) \geq 1 \Rightarrow \gamma \leq 0 \\ (11) &: P(y_0 = 1, y_1 = 1) + P(y_2 = 0, y_3 = 0) \geq 1 \Rightarrow \gamma \geq 0 \\ (12) &: P(y_0 = 0, y_1 = 0) + P(y_2 = 1, y_3 = 1) \geq 1 \Rightarrow \gamma \geq 0\end{aligned} \tag{4.4}$$

and based on set $\Theta^{\{2,3\}}_{I,stat}$:

$$\begin{aligned}(9) &: P(y_1 = 1, y_2 = 0) + P(y_2 = 0, y_3 = 1) \geq 1 \Rightarrow \gamma \leq 0 \\ (10) &: P(y_1 = 0, y_2 = 1) + P(y_2 = 1, y_3 = 0) \geq 1 \Rightarrow \gamma \leq 0\end{aligned} \tag{4.5}$$

Under the stationarity assumption, there's still a possibility that the sign of γ is not identified even if $T = 3$. However, this is no longer a case under the conditional exchangeability assumption: with three time periods, we always can identify the sign of γ by comparing probabilities of different sequences of the three consecutive outcomes. For example:

$$P(y_1 = 0, y_2 = 0, y_3 = 1|y_0) \lesseqgtr P(y_1 = 0, y_2 = 1, y_3 = 0|y_0) \iff \gamma \lesseqgtr 0$$

Tables 4 and 5 in the Appendix provide a full set of identifying restrictions under Assumption 3.2 that is based on the result in Table 3.

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)

$$P(y_1 = 0, y_2 = 1|y_0, x) \equiv p_2(0, 1|y_0, x)$$
$$P(y_1 = 1, y_2 = 0|y_0, x) \equiv p_2(1, 0|y_0, x)$$
$$P(y_1 = 0, y_2 = 0|y_0, x) \equiv p_2(0, 0|y_0, x)$$

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 $\vec{p}(y_0, x)$:

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

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 "2-Exchangeability 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":

$$
\begin{aligned}
&P(y_1 = 1|y_0, x) \geq P(y_2 = 1|y_0, x) \Rightarrow \Delta x'\beta + \min\{0, \gamma\} - \gamma y_0 \leq 0 \\
&P(y_1 = 1|y_0, x) \leq P(y_2 = 1|y_0, x) \Rightarrow \Delta x'\beta + \max\{0, \gamma\} - \gamma y_0 \geq 0 \\
&P(y_1 = 0, y_2 = 1|y_0, x) \geq P(y_1 = 1|y_0, x) \Rightarrow \Delta x'\beta - \gamma y_0 \geq 0 \\
&P(y_1 = 1, y_2 = 0|y_0, x) \geq P(y_1 = 0|y_0, x) \Rightarrow \Delta x'\beta + \gamma(1 - y_0) \geq 0
\end{aligned}
\tag{5.2}
$$

"2-Exchangeability Example":

$$
\begin{aligned}
&\left.\begin{aligned} -\Delta x'\beta + \gamma y_0 \geq 0 \\ -\Delta x'\beta + \gamma y_0 - \gamma \geq 0 \end{aligned}\right\} \Rightarrow P(y_1 = 1, y_2 = 0|x, y_0) \geq P(y_1 = 0, y_2 = 1|x, y_0) \\
&\left.\begin{aligned} -\Delta x'\beta + \gamma y_0 \leq 0 \\ -\Delta x'\beta + \gamma y_0 - \gamma \leq 0 \end{aligned}\right\} \Rightarrow P(y_1 = 1, y_2 = 0|x, y_0) \leq P(y_1 = 0, y_2 = 1|x, y_0)
\end{aligned}
\tag{5.3}
$$

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 is 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 "posterior". This can be done instantaneously by exploiting the Dirichlet prior approach in multinomials[8].

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

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

subject to

$$\mathbf{M}_p(\mathbf{x},\beta,\gamma) = \begin{bmatrix} \gamma \\ 1\{P_{(k)}(y_1=1|y_0^1,x^1) \geq P_{(k)}(y_2=1|y_0^1,x^1)\}(-\Delta x^{1\prime}\beta + \gamma y_0^1) \\ 1\{P_{(k)}(y_1=1|y_0^1,x^1) \leq P_{(k)}(y_2=1|y_0^1,x^1)\}(\Delta x^{1\prime}\beta + \gamma(1-y_0^1)) \\ 1\{P_{(k)}(y_1=0,y_2=1|y_0^1,x^1) \geq P_{(k)}(y_1=1|y_0^1,x^1)\}(\Delta x^{1\prime}\beta - \gamma y_0^1) \\ 1\{P_{(k)}(y_1=0,y_2=1|y_0^1,x^1) \leq P_{(k)}(y_1=1|y_0^1,x^1)\}(\Delta x^{1\prime}\beta + \gamma(1-y_0^1)) \\ 1\{P_{(k)}(y_1=1|y_0^2,x^2) \geq P_{(k)}(y_2=1|y_0^2,x^2)\}(-\Delta x^{2\prime}\beta + \gamma y_0^2) \\ 1\{P_{(k)}(y_1=1|y_0^2,x^2) \leq P_{(k)}(y_2=1|y_0^2,x^2)\}(\Delta x^{2\prime}\beta + \gamma(1-y_0^2)) \\ 1\{P_{(k)}(y_1=0,y_2=1|y_0^2,x^2) \geq P_{(k)}(y_1=1|y_0^2,x^2)\}(\Delta x^{2\prime}\beta - \gamma y_0^2) \\ 1\{P_{(k)}(y_1=0,y_2=1|y_0^2,x^2) \leq P_{(k)}(y_1=1|y_0^2,x^2)\}(\Delta x^{2\prime}\beta + \gamma(1-y_0^2)) \\ \dots \\ 1\{P_{(k)}(y_1=1|y_0^J,x^J) \geq P_{(k)}(y_2=1|y_0^J,x^J)\}(-\Delta x^{J\prime}\beta + \gamma y_0^J) \\ 1\{P_{(k)}(y_1=1|y_0^J,x^J) \leq P_{(k)}(y_2=1|y_0^J,x^J)\}(\Delta x^{J\prime}\beta + \gamma(1-y_0^J)) \\ 1\{P_{(k)}(y_1=0,y_2=1|y_0^J,x^J) \geq P_{(k)}(y_1=1|y_0^J,x^J)\}(\Delta x^{J\prime}\beta - \gamma y_0^J) \\ 1\{P_{(k)}(y_1=0,y_2=1|y_0^J,x^J) \leq P_{(k)}(y_1=1|y_0^J,x^J)\}(\Delta x^{J\prime}\beta + \gamma(1-y_0^J)) \end{bmatrix} \geq \mathbf{0}$$

(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

[8] See Chamberlain and Imbens (2003) for an example of the Bayesian Bootstrap in the context of moment condition models.

$$
\mathbf{M}_n(\mathbf{x},\beta,\gamma) = \begin{bmatrix}
-\gamma \\
1\{P_{(k)}(y_1=1|y_0^1,x^1) \geq P_{(k)}(y_2=1|y_0^1,x^1)\}(-\Delta x^{1'}\beta + \gamma(y_0^1-1)) \\
1\{P_{(k)}(y_1=1|y_0^1,x^1) \leq P_{(k)}(y_2=1|y_0^1,x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \\
1\{P_{(k)}(y_1=0,y_2=1|y_0^1,x^1) \geq P_{(k)}(y_1=1|y_0^1,x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \\
1\{P_{(k)}(y_1=0,y_2=1|y_0^1,x^1) \leq P_{(k)}(y_1=1|y_0^1,x^1)\}(\Delta x^{1'}\beta + \gamma(1-y_0^1)) \\
1\{P_{(k)}(y_1=1|y_0^2,x^2) \geq P_{(k)}(y_2=1|y_0^2,x^2)\}(-\Delta x^{2'}\beta + \gamma(y_0^2-1)) \\
1\{P_{(k)}(y_1=1|y_0^2,x^2) \leq P_{(k)}(y_2=1|y_0^2,x^2)\}(\Delta x^{2'}\beta - \gamma y_0^2) \\
1\{P_{(k)}(y_1=0,y_2=1|y_0^2,x^2) \geq P_{(k)}(y_1=1|y_0^2,x^2)\}(\Delta x^{2'}\beta - \gamma y_0^2) \\
1\{P_{(k)}(y_1=0,y_2=1|y_0^2,x^2) \leq P_{(k)}(y_1=1|y_0^2,x^2)\}(\Delta x^{2'}\beta + \gamma(1-y_0^2)) \\
\dots \\
1\{P_{(k)}(y_1=1|y_0^J,x^J) \geq P_{(k)}(y_2=1|y_0^J,x^J)\}(-\Delta x^{J'}\beta + \gamma(y_0^J-1)) \\
1\{P_{(k)}(y_1=1|y_0^J,x^J) \leq P_{(k)}(y_2=1|y_0^J,x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \\
1\{P_{(k)}(y_1=0,y_2=1|y_0^J,x^J) \geq P_{(k)}(y_1=1|y_0^J,x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \\
1\{P_{(k)}(y_1=0,y_2=1|y_0^J,x^J) \leq P_{(k)}(y_1=1|y_0^J,x^J)\}(\Delta x^{J'}\beta + \gamma(1-y_0^J))
\end{bmatrix} \geq \mathbf{0}
$$

(iv) Repeat steps (i)-(iii) above for $k = 1, \dots, M$.

(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.

But, one computationally trivial approach and the approach we recommend is to target *linear functionals* of (β,γ), such as any scalar subvector like β_1, or γ or generally $l'\begin{bmatrix}\beta \\ \gamma\end{bmatrix}$. For example, to get the identified set (interval in this case) for γ, one can use the following linear program for γ positive:

$$
\max / \min_{(\gamma,\beta)\in\Theta} \gamma
$$

$$
\text{subject to} \quad \mathbf{M}_p(\mathbf{x},\beta,\gamma) \geq 0
$$

and for γ negative:

$$
\max / \min_{(\gamma,\beta)\in\Theta} \gamma
$$

$$
\text{subject to} \quad \mathbf{M}_n(\mathbf{x},\beta,\gamma) \geq 0
$$

Then, take the union of the two intervals. This approach is simple and easy to implement especially if the vector β is of high dimensions.

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")

$$P(y_1 = 1, y_2 = 0|y_0, x) \leq P(y_1 = 0, y_2 = 1|y_0, x) \Rightarrow \Delta x'\beta - \gamma y_0 \geq 0 \text{ or } \Delta x'\beta + \gamma(1 - y_0) \geq 0$$
$$P(y_1 = 1, y_2 = 0|y_0, x) \geq P(y_1 = 0, y_2 = 1|y_0, x) \Rightarrow \Delta x'\beta - \gamma y_0 \leq 0 \text{ or } \Delta x'\beta + \gamma(1 - y_0) \leq 0$$

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

$$P(y_1 = 1, y_2 = 0|y_0, x) \leq P(y_1 = 0, y_2 = 1|y_0, x) \Rightarrow \Delta x'\beta - \gamma y_0 < 0 \text{ and } \Delta x'\beta + \gamma(1 - y_0) < 0$$
$$P(y_1 = 1, y_2 = 0|y_0, x) \geq P(y_1 = 0, y_2 = 1|y_0, x) \Rightarrow \Delta x'\beta - \gamma y_0 > 0 \text{ and } \Delta x'\beta + \gamma(1 - y_0) > 0$$

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

$$
\begin{cases}
1\{P_{(k)}(y_1=1,y_2=0|y_0^1,x^1) \leq P_{(k)}(y_1=0,y_2=1|y_0^1,x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \leq 0 \\
1\{P_{(k)}(y_1=1,y_2=0|y_0^1,x^1) \leq P_{(k)}(y_1=0,y_2=1|y_0^1,x^1)\}(\Delta x^{1'}\beta + \gamma(1-y_0^1)) \leq 0 \\
1\{P_{(k)}(y_1=1,y_2=0|y_0^1,x^1) \geq P_{(k)}(y_1=0,y_2=1|y_0^1,x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \geq 0 \\
1\{P_{(k)}(y_1=1,y_2=0|y_0^1,x^1) \geq P_{(k)}(y_1=0,y_2=1|y_0^1,x^1)\}(\Delta x^{1'}\beta + \gamma(1-y_0^1)) \geq 0 \\
\dots \\
1\{P_{(k)}(y_1=1,y_2=0|y_0^J,x^J) \leq P_{(k)}(y_1=0,y_2=1|y_0^J,x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \leq 0 \\
1\{P_{(k)}(y_1=1,y_2=0|y_0^J,x^J) \leq P_{(k)}(y_1=0,y_2=1|y_0^J,x^J)\}(\Delta x^{J'}\beta + \gamma(1-y_0^J)) \leq 0 \\
1\{P_{(k)}(y_1=1,y_2=0|y_0^J,x^J) \geq P_{(k)}(y_1=0,y_2=1|y_0^J,x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \geq 0 \\
1\{P_{(k)}(y_1=1,y_2=0|y_0^J,x^J) \geq P_{(k)}(y_1=0,y_2=1|y_0^J,x^J)\}(\Delta x^{J'}\beta + \gamma(1-y_0^J)) \geq 0
\end{cases}
$$

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.

Here too if one is interested in a *scalar* parameter (such as γ, a subvector of β or a linear combination of (γ, β)), one can directly minimize and maximize that linear functional subject to the constraints above. This is simple to do and obviates the need to computationally obtain the set of feasible parameters for every draw from the CI for the choice probabilities.

6 Finite Sample Properties

We first illustrate the finite sample properties using a limited empirical illustration. We then provide a set of simulated identified sets to showcase the identifying power of the various assumptions and how this changes with the support of the regressors.

6.1 Empirical Illustration

The $Panel\,Data\,for\,Women's\,Labor\,Force\,Participation$ contains data on N=5663 married women over $T = 5$ periods, where the periods are spaced four months apart. The response variable lfp_{it} is a binary indicator for labor force participation. The key explanatory variables we use are $kids_{it}$ (number of children under 18) and $lhinc_{it}$=log($hinc_{it}$), where husband's income, $hinc_{it}$, is in dollar per month and is positive for all i and t. There are also time-constant variables $educ$, $black$,age and $agesq$, these variables are dropped out when using fixed effects estimators.

In the following analysis, a binary version of $lhinc_{it}$ and $kids_{it}$, $newlhinc_{it}$ and $newkids_{it}$ respectively, enter regressions as explanatory variables. These are defined as:

$$newkids_{it} = \begin{cases} 0, lhinc_{it} = 0\, or\, 1, \\ 1, lhinc_{it} > 1. \end{cases} \tag{6.1}$$

$$newlhinc_{it} = \begin{cases} 0, lhinc_{it} \leq Median\,(lhinc_t), \\ 1, lhinc_{it} > Median\,(lhinc_t). \end{cases} \tag{6.2}$$

We also use 3 time periods. A table of brief descriptive statistics is provided below.

Table 1: Summary Statistics

	obs	mean	sd	min	max
lfp	16989	.6812643	.466	0	1
newkids	16989	.4139737	.4925584	0	1
newlhinc	16989	.4997351	.5000146	0	1

We compare our estimates to three dynamic models. First, we use two random effects Probit: the *reprobit* where the random effect is mean zero and independent of the regressors, and another *reprobitci* where the random effect is a function of the vector of covariates at all time periods. The third model is the dynamic Logit FE model of Honoré and Kyriazidou (2000). Table 2 reports the estimates along with confidence regions. The FE logit model uses the following conditional likelihood

$$\sum_{i=1}^{n} 1[x_{i2} = x_{i3}]1[y_{i1} \neq y_{i2}] \times log\left(\frac{exp((x_{i1} - x_{i2})'b + g(y_{i0} - y_{i3}))^{y_{i1}}}{1 + exp((x_{i1} - x_{i2})'b + g(y_{i0} - y_{i3}))}\right) \tag{6.3}$$

We found that the above objective function was easy to optimize and is robust to starting values. On the other hand, the last column provides our estimate of the identified set for scalar parameters along with a confidence region that covers the identified set with 95% probability.

Implementation Via Linear Program: To implement the procedure described in the inference section above and obtain a confidence region, we require that one obtains draws from the confidence region of the choice probabilities in (5.1) above. One computationally automatic way to get such draws is to use the Bayesian bootstrap which is equivalent to drawing from the posterior distribution of a multinomial with the usual Dirichlet priors[9]. For each draw from this posterior, we solve the linear program for min / max of the scalar $a = l'(\gamma, \beta)$. This allows us to get CIs for every scalar component of the parameter vector for example, including[10] γ. Using 1000 draws from the posterior of the choice probabilities, we obtain 1000 copies of the identified set. Then, we report in the table below the smallest set that contains 95% of the intervals. This procedure is simple to compute even with thousands of inequalities.

Table 2: Dynamic Models: RE, Logit FE (HK), Stationary FE/T=2 using Linear Programs

	(1)	(2)	(3)	(4)
	reprobit	reprobitci	HK Logit FE	Stationary FE - Tolerance =0
newkids	-0.139	-0.443	-.063	-.5
	(0.037)	(0.214)	(.374)	
newlhinc	-0.161	-0.150	-.273	$[.5, 3]$
	(0.036)	(0.095)	(.72)	$[-.5, 3.2]$
lag_lfp	2.475	1.163	2.331	$[.5, 3.1]$
	(0.034)	(0.134)	(.53)	$[-.2, 3.2]$

Notice here that consistently across all the models, the γ coefficient appears positive. In the Stationary model, we fix the parameter on *newkids* to $-.5$ for normalization[11] and it is clear that the other parameter is not significant.

6.2 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

[9] In cases when the regressors have many support points, one can use a "reduced form" estimator for the choice probabilities such as a multinomial logit and use that model to get draws from its posterior predictive distribution.

[10] It may be that with real data, the set of inequalities that define the stationary set does not have a nonempty interior. In this case, we first add a tolerance parameter t to each inequality (so now the inequalities are less than a positive t rather than less then 0), and in the first pass through the linear program we minimize t subject to the constraints that define the problem (optimizing over (γ, β)) to obtain a feasible tolerance t^*. Then, we fix the tolerance at t^* when computing the confidence set. In our data setup, we obtained a tolerance $t^* = 0$.

[11] We normalized at the point estimates from the random effects probit model.

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:

$$y_{it} = I\{u_{it} \leq v_{it} + x_{it}\beta + \gamma y_{i,t-1} + \alpha_i\} \quad i = 1, 2, ...10000; \quad t = 0, 1, ...T \tag{6.4}$$

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.

6.2.1 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.

6.2.2 Exchangeability 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 γ 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.

6.2.3 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).

6.2.4 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 even in models that have time trends as regressors, which is ruled out in Honoré and Kyriazidou (2000). 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 interesting 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.

Arellano, M., and B. Honoré (2001): "Panel Data Models: Some Recent Developments," *Handbook of econometrics. Volume 5*, pp. 3229–96.

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.

Chamberlain, G., and G. W. Imbens (2003): "Nonparametric applications of Bayesian inference," *Journal of Business & Economic Statistics*, 21(1), 12–18.

Chen, S., S. Khan, and X. Tang (2015): "Informational Content in Static and Dynamic Discrete Response Panel Data Models," U Penn Working Paper.

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.

Erdem, T., S. Imai, and M. P. Keane (2003): "Brand and quantity choice dynamics under price uncertainty," *Quantitative Marketing and economics*, 1(1), 5–64.

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.

——— (2019): "Identification In Binary Response Panel Data Models: Is Point-Identification More Common Than We Thought?," working paper.

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.

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.

Konstantopoulos, T., and L. Yuan (2018): "On the Extendibility of Finitely Exchangeable Probability Measures," *Transactions of the American Mathematical Society*, 10(371), 7067–7092.

Manski, C. F. (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27(3), 313–33.

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

Merlo, A., and K. I. Wolpin (2008): "The transition from school to jail: Youth crime and high school completion among black males," .

Olshen, R. (1973): "A Note on Exchangable Sequences," Technical Report No 79, Department of Statistics, Stanford University.

Ouyang, F., S. Khan, and E. Tamer (2017): "Identification and Estimation of Dynamic Panel Data Multinomial Response Models," working paper, Duke University.

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.

Yang, J., Y. Qi, and R. Wang (2009): "A Class of Multivariate Copulas with Bivariate Fréchet Marginal Copulas," *Insurance: Mathematics and Economics*, pp. 139–147.

A Figures

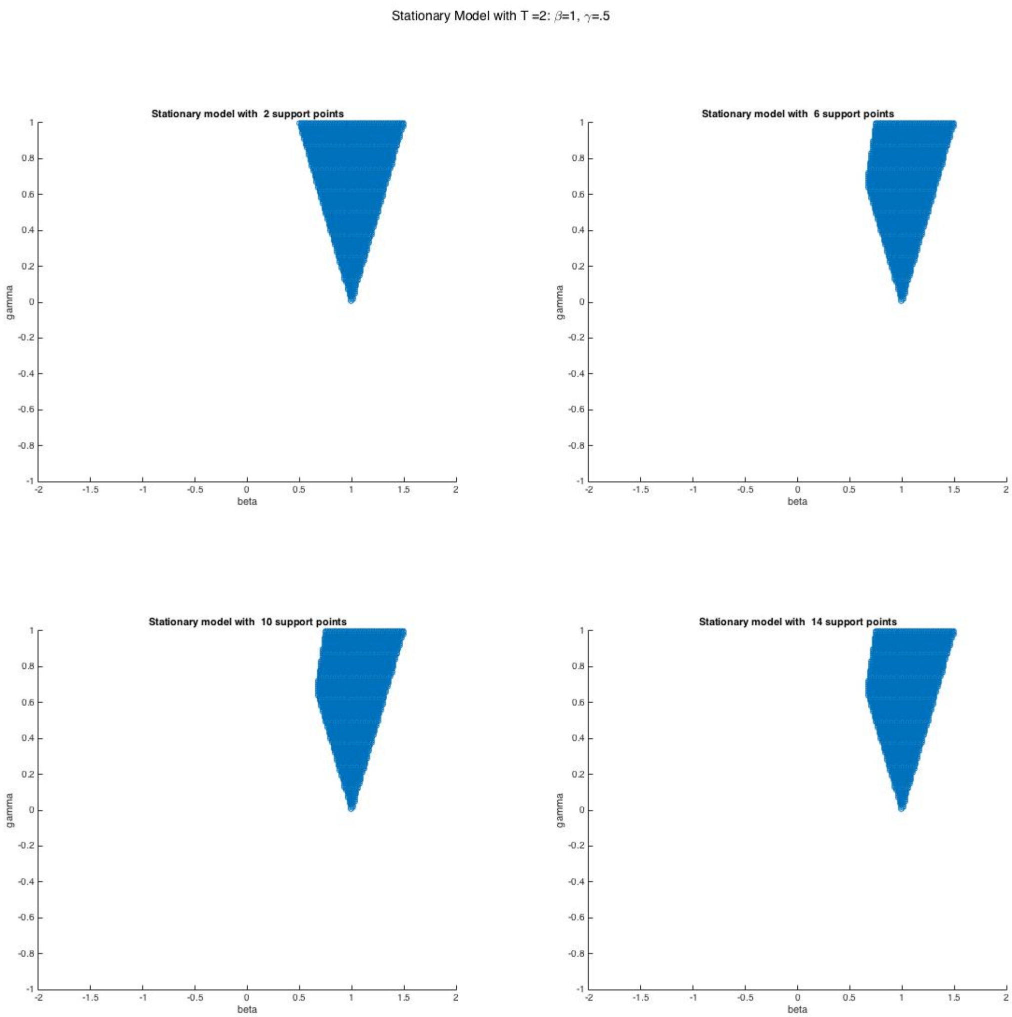

Figure 2: Stationary with $T = 2$ and Discrete Support with $\gamma = .5$

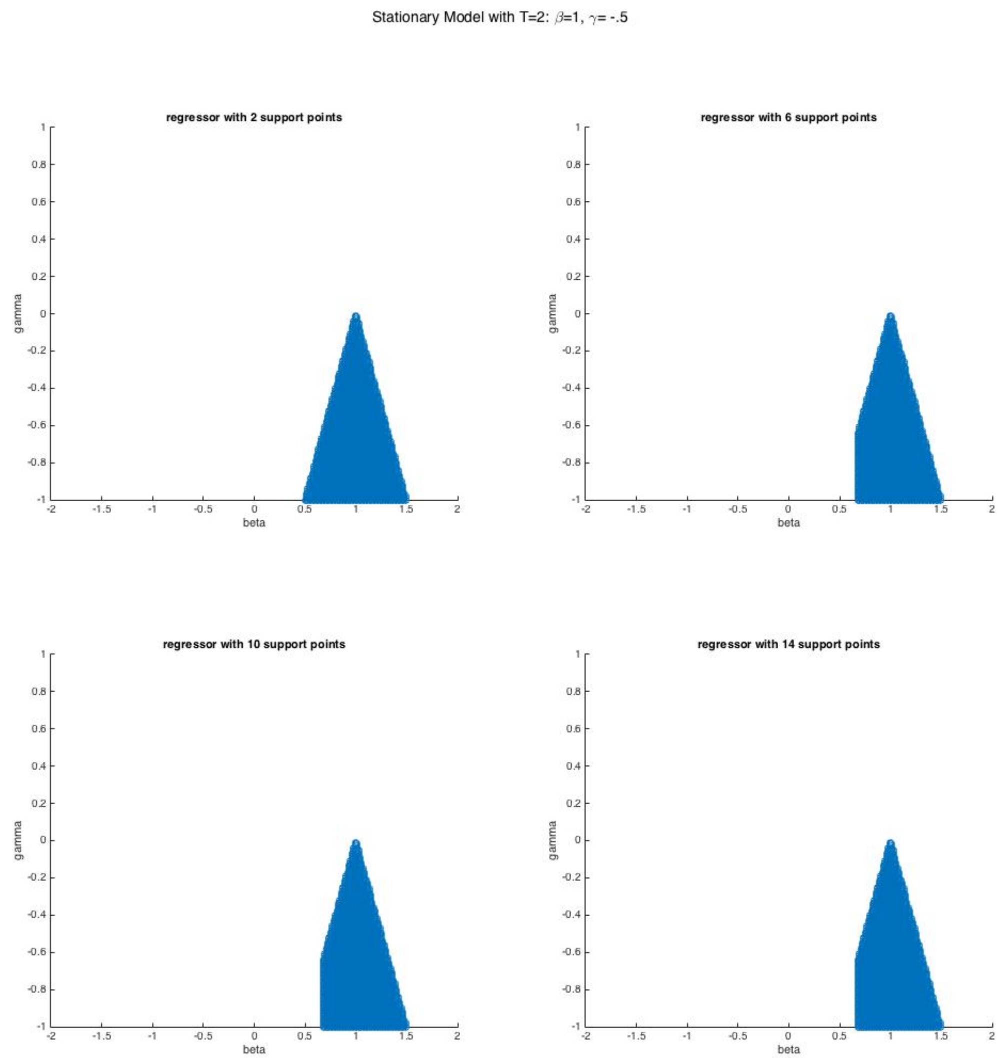

Figure 3: Stationary with $T = 2$ and Discrete Support with $\gamma = .5$

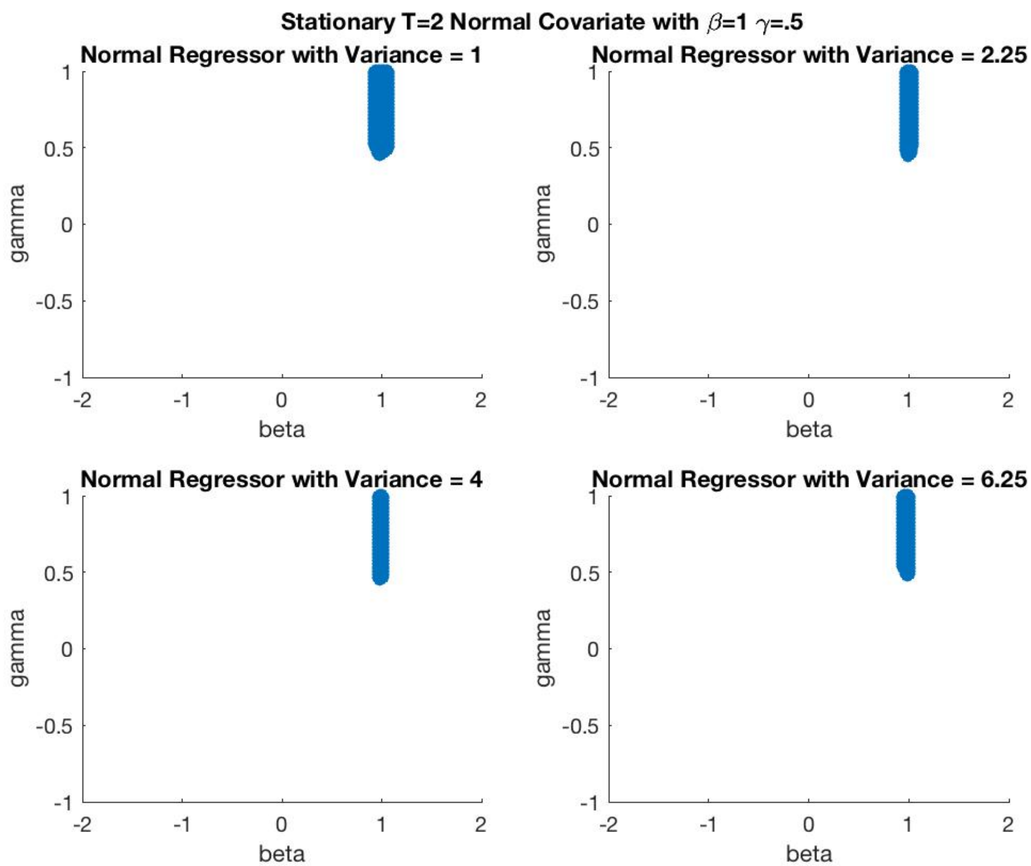

Figure 4: Stationary with $T = 2$ and Normal v with $\gamma = .5$

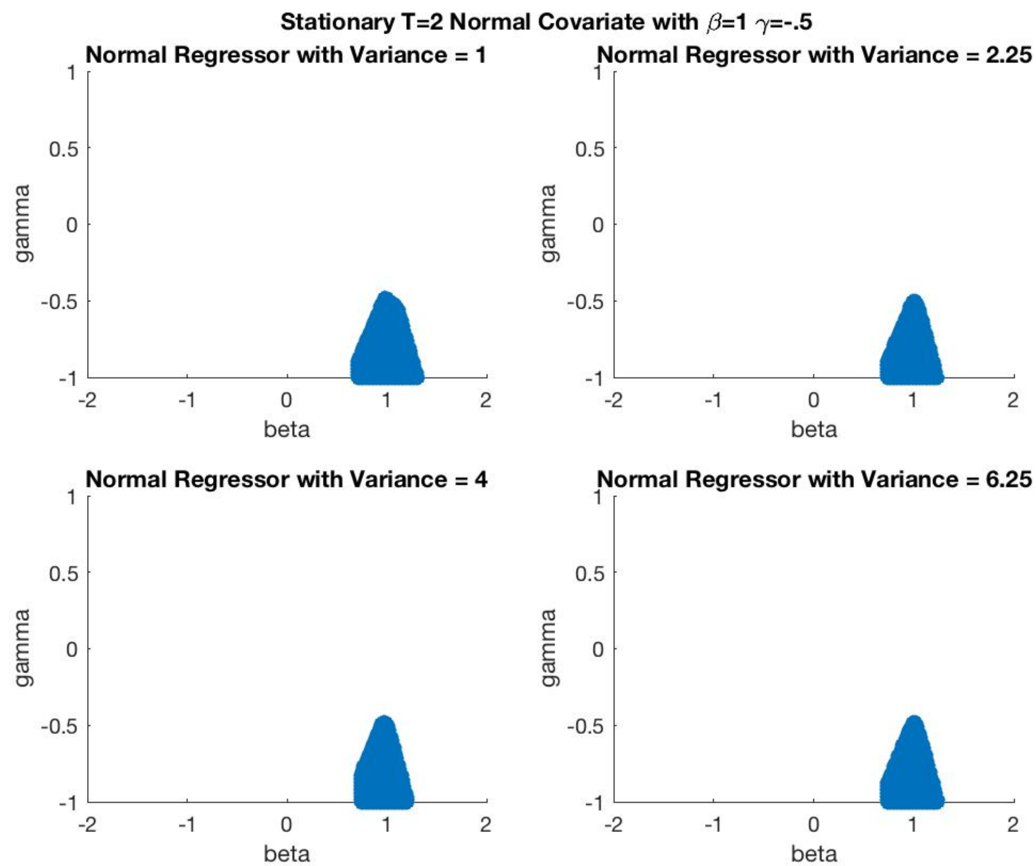

Figure 5: Stationary with $T = 2$ and Normal v with $\gamma = -.5$

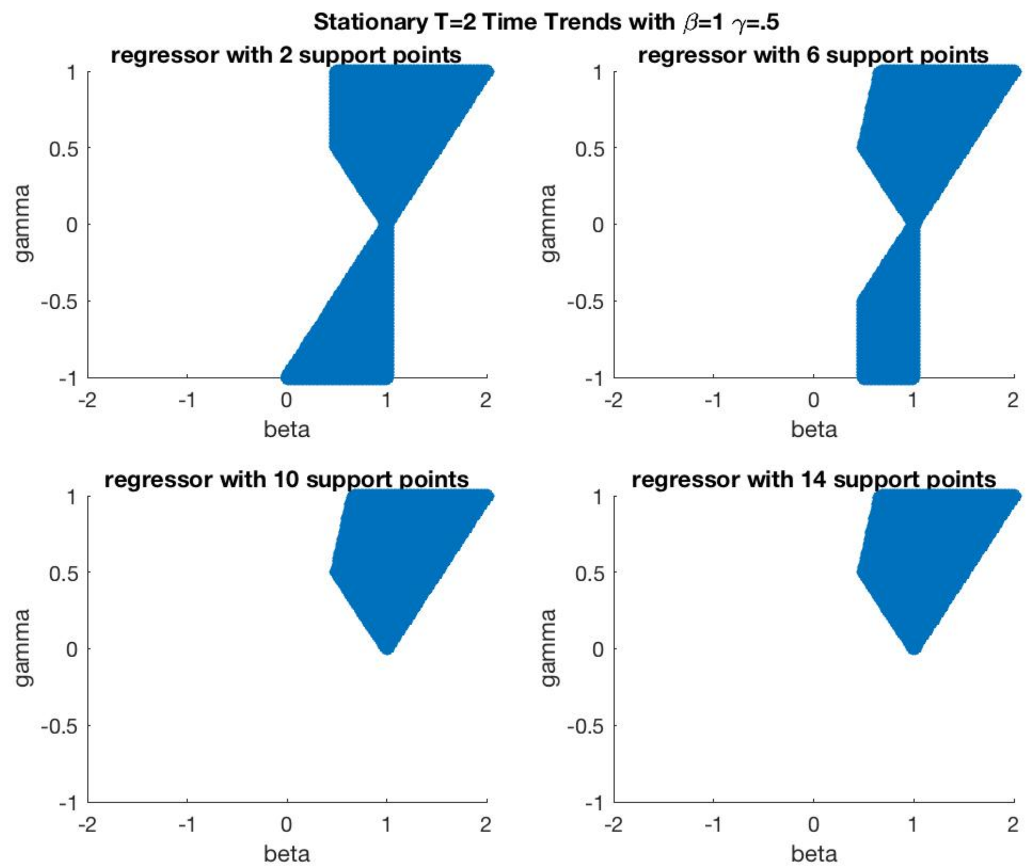

Figure 6: Stationary with $T = 2$ and Time Trend and Discrete Support for v with $\gamma = .5$

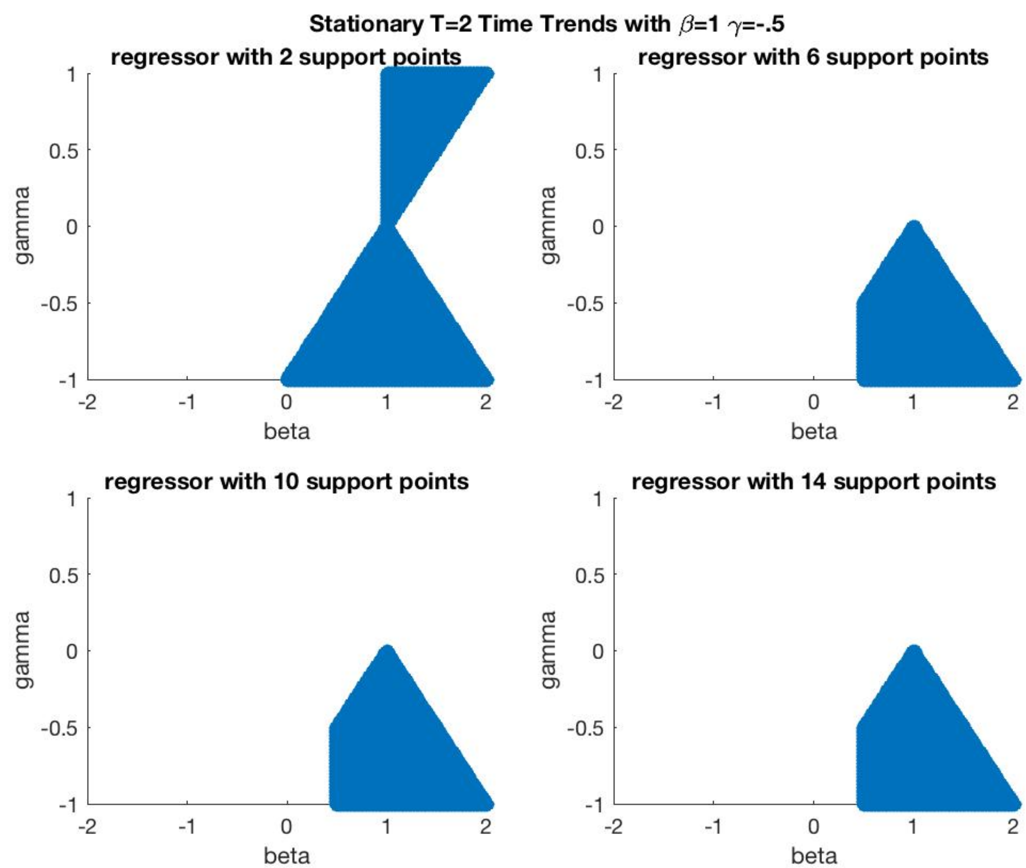

Figure 7: Stationary with $T = 2$ and Time Trend and Discrete Support for v with $\gamma = -.5$

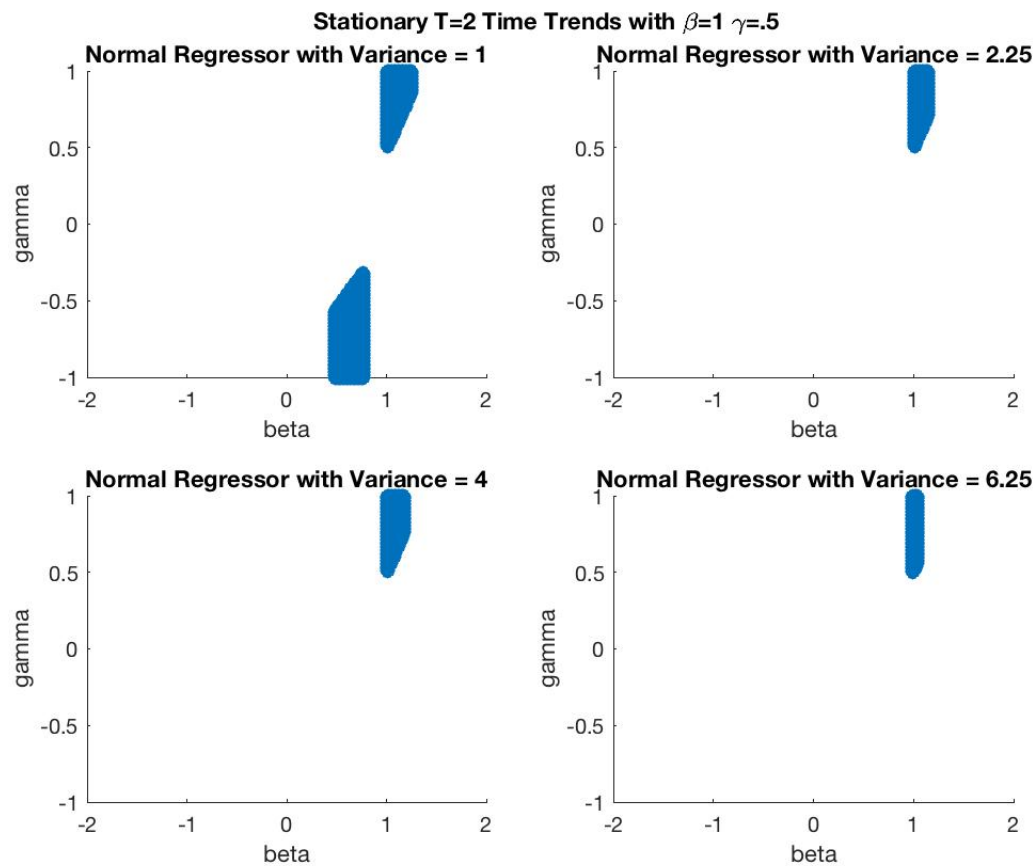

Figure 8: Stationary with $T = 2$ and Time Trend and Normal v with $\gamma = .5$

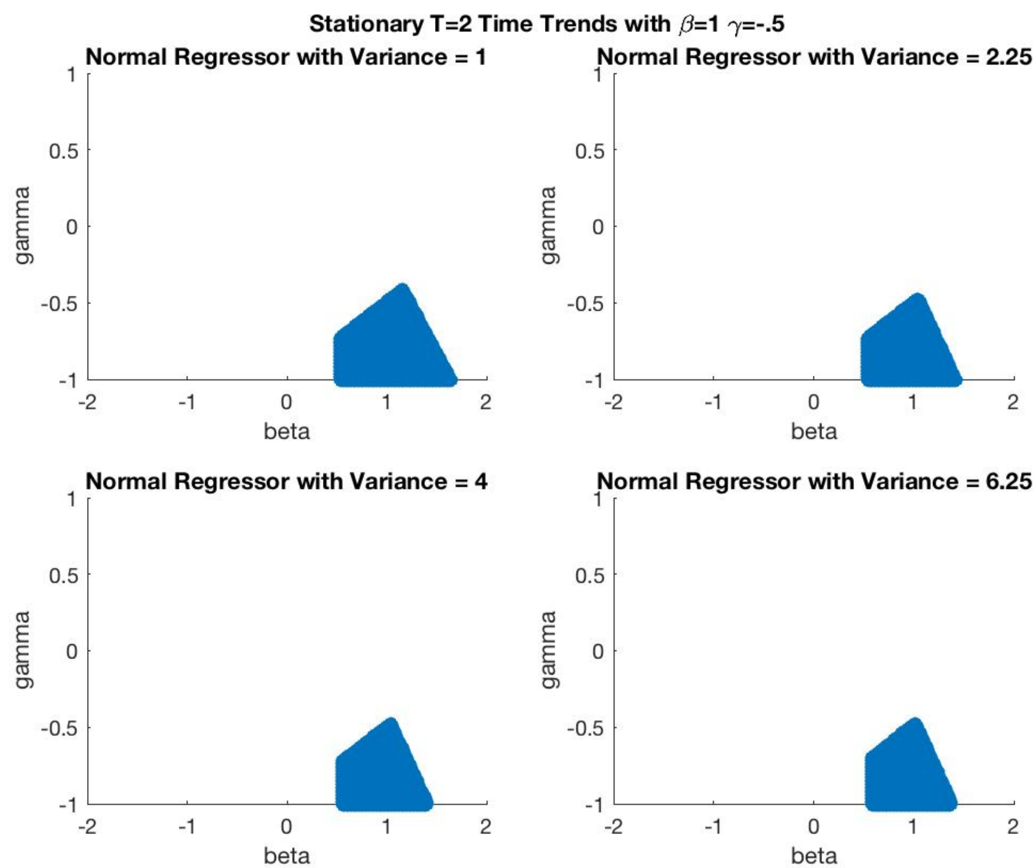

Figure 9: Stationary with $T = 2$ and Time Trend and Normal v with $\gamma = -.5$

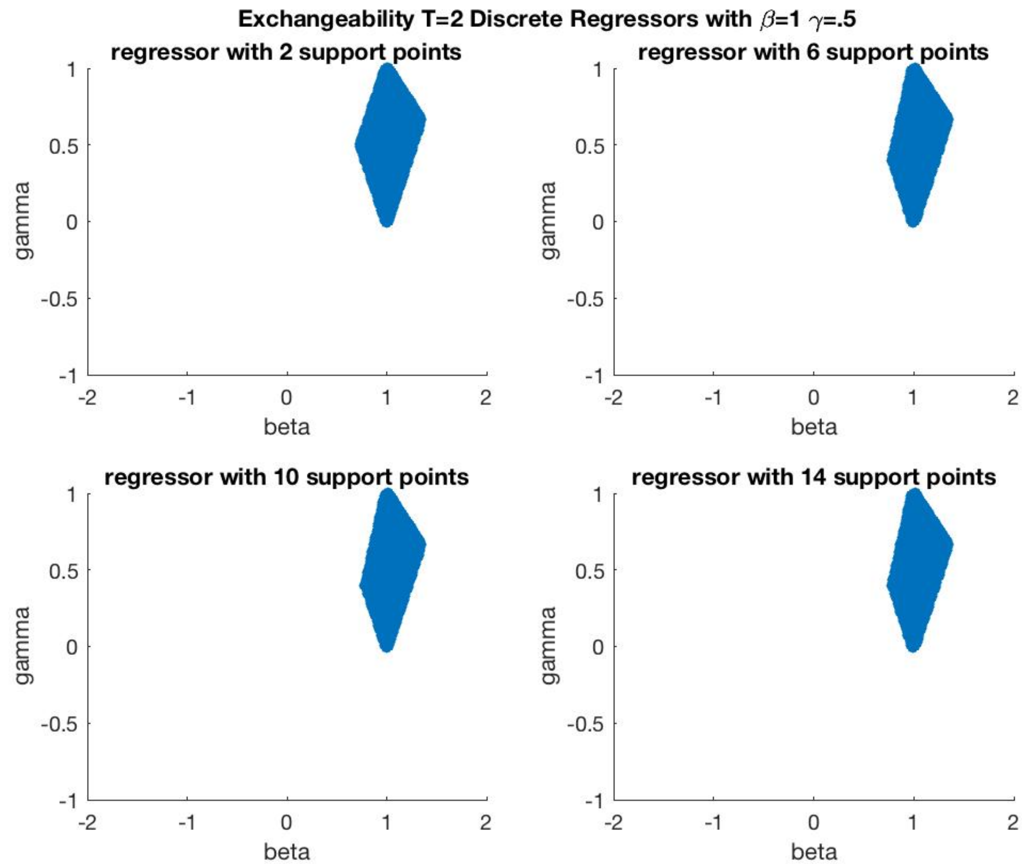

Figure 10: Exchangeability with $T = 2$ Discrete Support for v with $\gamma = .5$

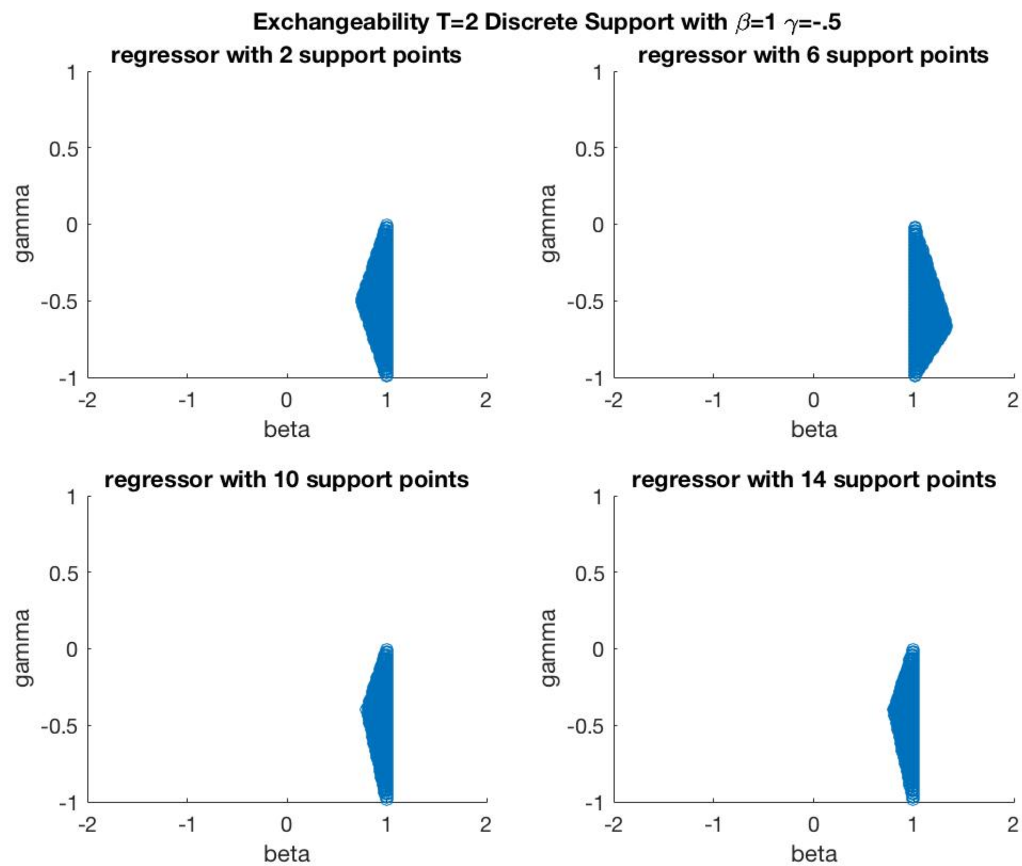

Figure 11: Stationary with $T = 2$ Discrete Support for v with $\gamma = -.5$

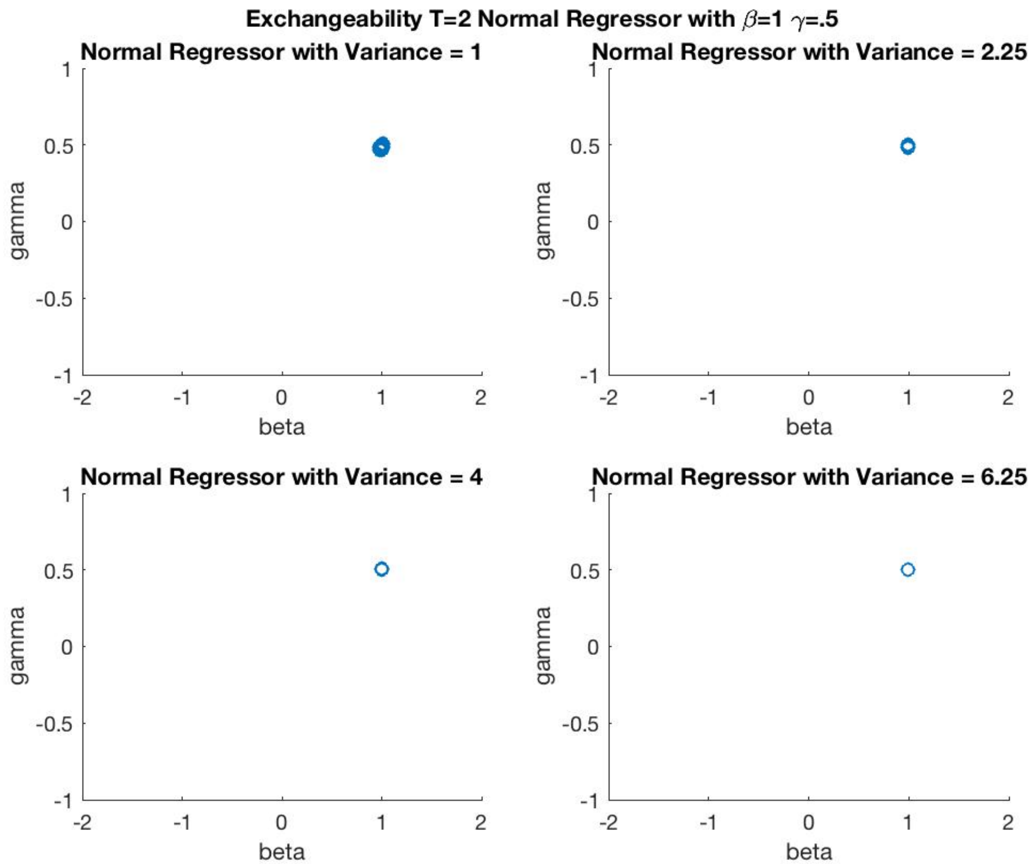

Figure 12: Exchangeability with $T = 2$ Discrete v with $\gamma = .5$

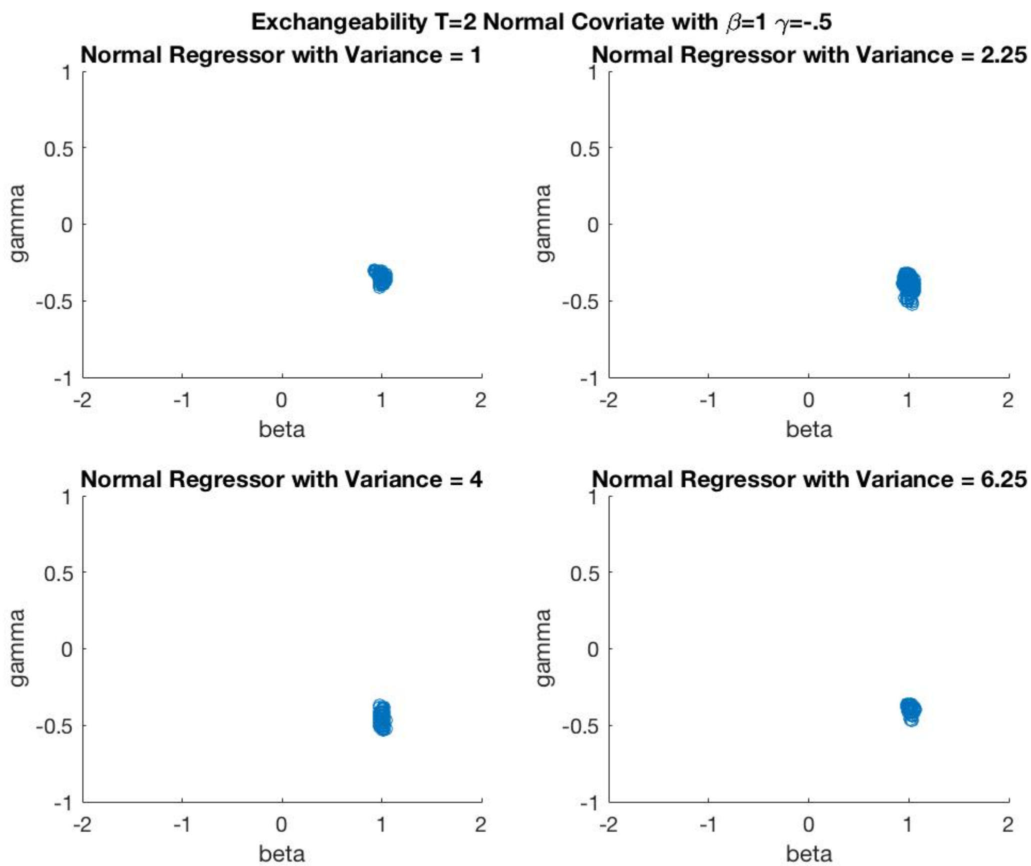

Figure 13: Stationary with $T = 2$ Normal for v with $\gamma = -.5$

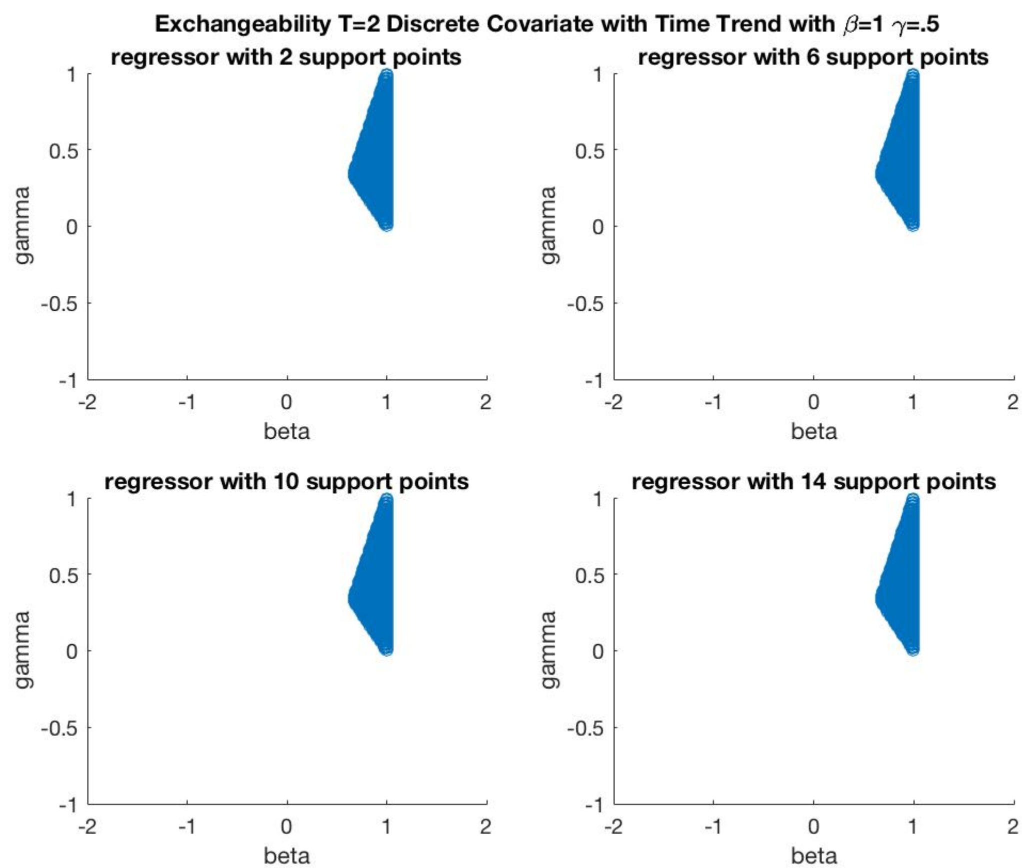

Figure 14: Exchangeability with $T = 2$ $x = t$ Discrete v with $\gamma = .5$

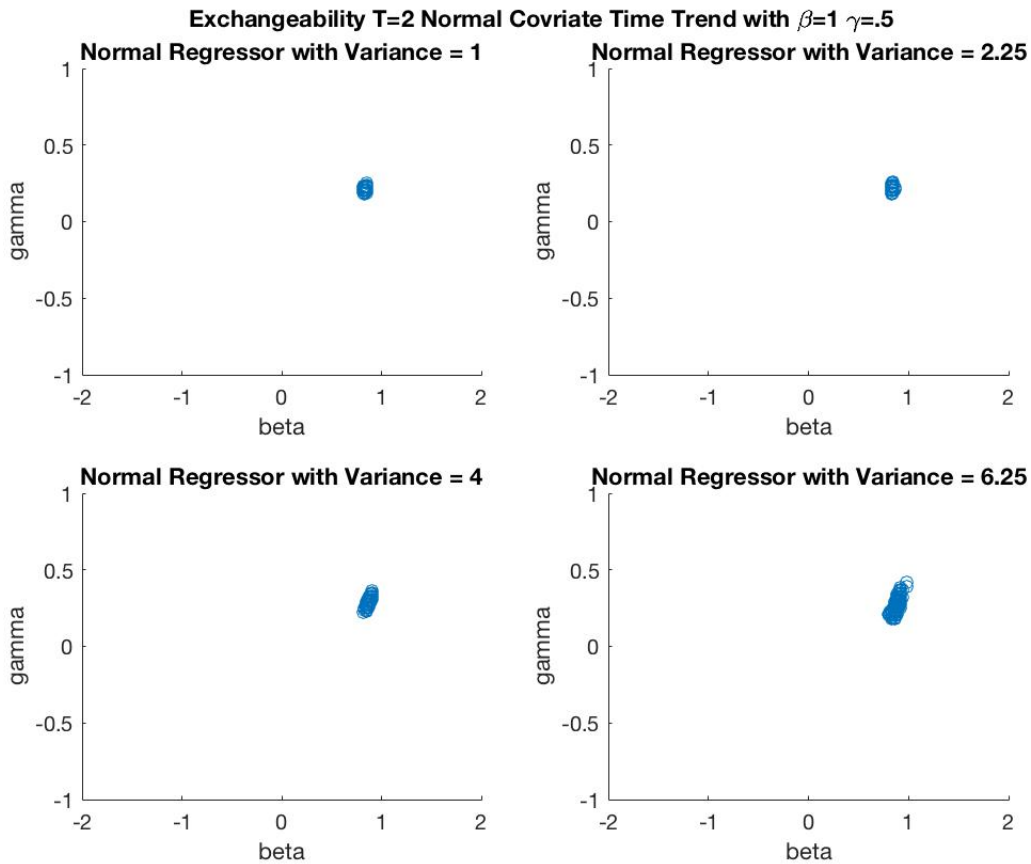

Figure 15: Stationary with $T = 2$, $x = t$ Normal for v with $\gamma = .5$

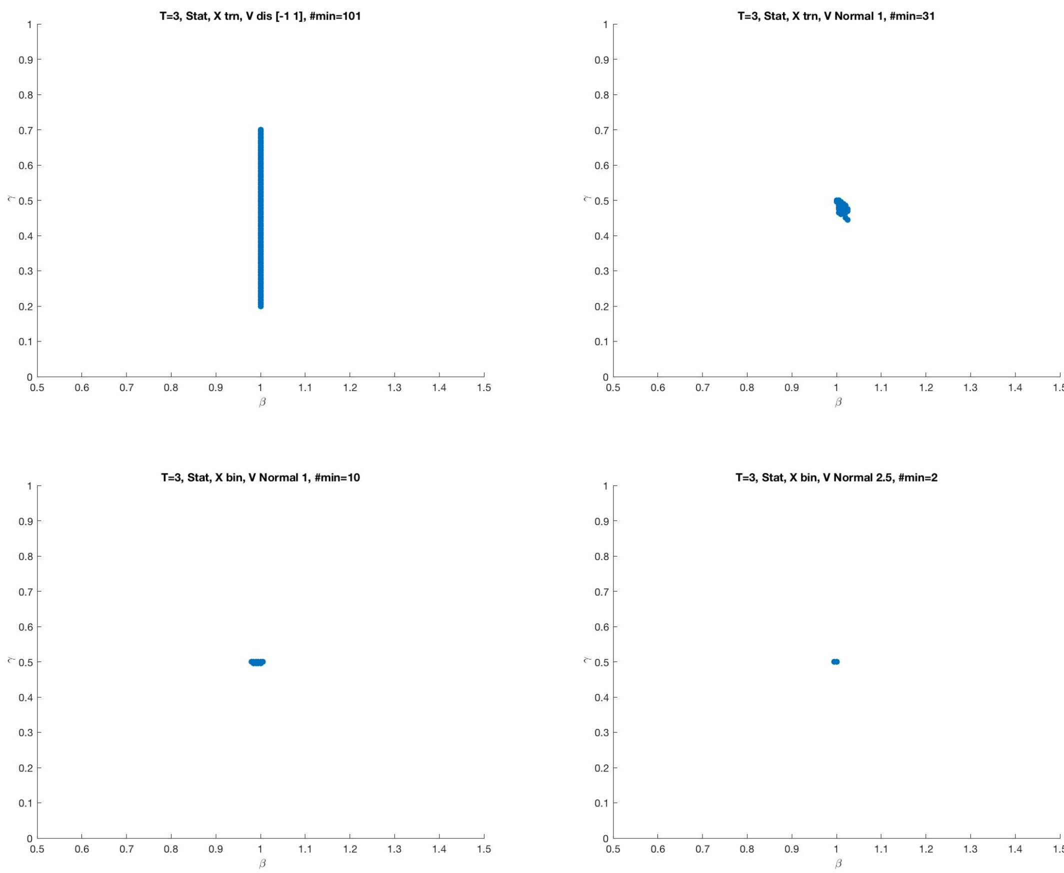

Figure 16: Stationarity with T=3: Various Designs

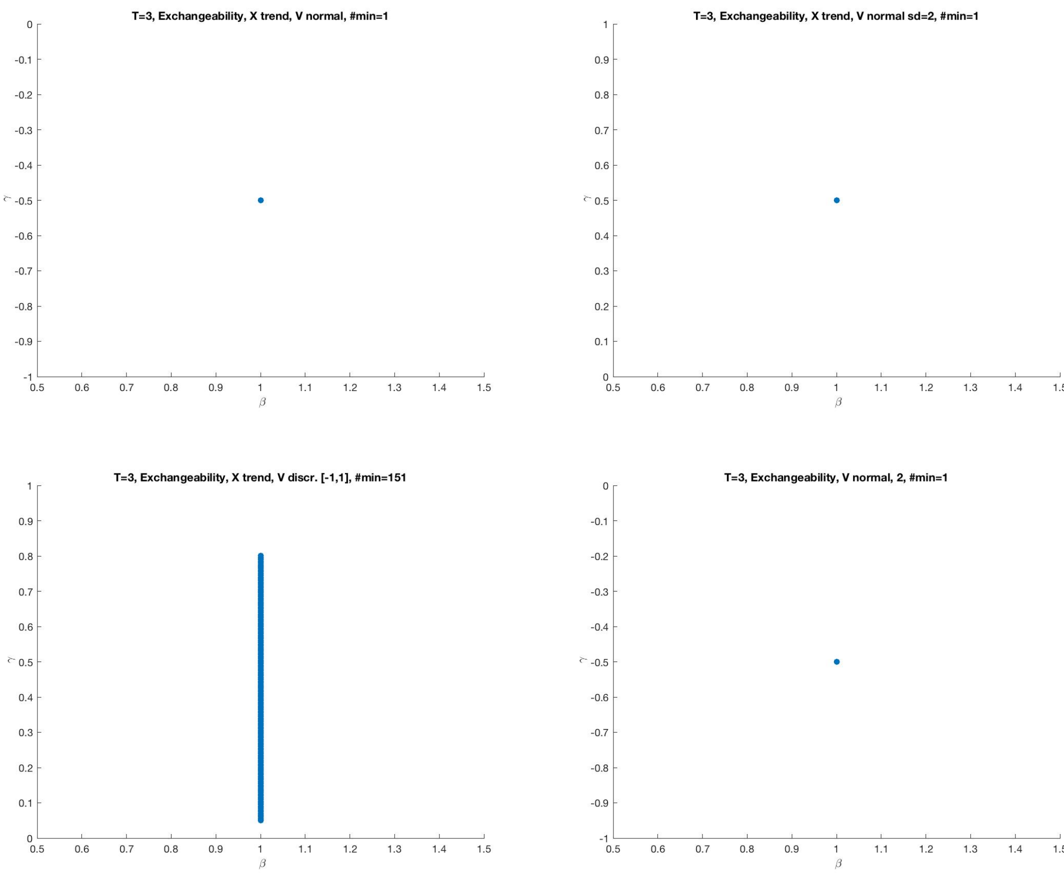

Figure 17: Exchangeability with T=3: Various Designs

B Proof

B.1 Proof of Lemma 3.1

Since $(u_{-M}, \ldots, u_0, u_1, \ldots, u_T)$ is exchangeable conditional on $\alpha, x_{-M}, \ldots, x_0, x_1, \ldots, x_T$, Theorem 3 in Olshen (1973) implies that there exists a random variable ξ such that error terms $u_{-M}, \ldots, u_0, u_1, \ldots, u_T$ are i.i.d. conditional on $\xi, \alpha, x_{-M}, \ldots, x_0, x_1, \ldots, x_T$. In turn, this implies that $u_1, \ldots, u_T$ are i.i.d. conditional on $\xi, \alpha, u_{-M}, \ldots, u_0, x_{-M}, \ldots, x_0, x_1, \ldots, x_T$. In particular, for any $(a_1, \ldots, a_T) \in \mathbb{R}^T$, we have

$$\begin{aligned} &P(u_1 \leq a_1, \ldots, u_T \leq a_T | \xi, \alpha, u_{-M}, \ldots, u_0, x_{-M}, \ldots, x_0, x_1, \ldots, x_T) = \\ &\qquad P(u_{s_1} \leq a_1, \ldots, u_{s_T} \leq a_T | \xi, \alpha, u_{-M}, \ldots, u_0, x_{-M}, \ldots, x_0, x_1, \ldots, x_T) \end{aligned}$$

where $\{s_1, \ldots, s_T\}$ is an arbitrary permutation of $\{1, \ldots, T\}$. Integrating ξ out we get

$$\begin{aligned} &P(u_1 \leq a_1, \ldots, u_T \leq a_T | \alpha, u_{-M}, \ldots, u_0, x_{-M}, \ldots, x_0, x_1, \ldots, x_T) = \\ &\qquad = P(u_{s_1} \leq a_1, \ldots, u_{s_T} \leq a_T | \alpha, u_{-M}, \ldots, u_0, x_{-M}, \ldots, x_0, x_1, \ldots, x_T) \end{aligned} \tag{B.1}$$

Since there's a beginning M periods back, we are able to write y_0 as a function of $\alpha, x_{-M}, \ldots, x_0, u_{-M}, \ldots, u_0$ only. Applying the LIE we get:

$$\begin{aligned} &P(u_{s_1} \leq a_1, \ldots, u_{s_T} \leq a_T | \alpha, y_0, x_1, \ldots, x_T) = \\ &\quad \overset{LIE}{=} E\left(P(u_{s_1} \leq a_1, \ldots, u_{s_T} \leq a_T | \alpha, u_{-M}, \ldots, u_0, x_{-M}, \ldots, x_0, x_1, \ldots, x_T) | \alpha, y_0, x_1, \ldots, x_T\right) \\ &\quad = E(P(u_1 \leq a_1, \ldots, u_T \leq a_T | \alpha, u_{-M}, \ldots, u_0, x_{-M}, \ldots, x_0, x_1, \ldots, x_T) | \alpha, y_0, x_1, \ldots, x_T) \\ &\quad \overset{LIE}{=} P(u_1 \leq a_1, \ldots, u_T \leq a_T | \alpha, y_0, x_1, \ldots, x_T) \end{aligned}$$

where the second equality follows from B.1. That is, distribution of $(u_1, \ldots, u_T)$ and the distribution of $(u_{s_1}, \ldots, u_{s_T})$ are identical conditional on α, x, y_0. ■

B.2 Proof of Theorem 3.1

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 $\Theta_{I,cex}^{\{1,2\}}$ in Lemma 3.3 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.

The rest of the proof relies on the identification result for conditional exchangeability (Theorem 3.2), but without the requirement that the trivariate copula $\tilde{C}(q_0, q_2, q_2)$ must be symmetric. ■

Proof of Proposition 3.2: 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,

$$P(y_1 = 1, y_2 = 0|\alpha, x, y_0) \leq P(u_1 - u_2 \leq (x_1 - x_2)'\beta + \gamma(y_0 - 1)|\alpha, x, y_0)$$

Integrating α out, we get

$$P(y_1 = 1, y_2 = 0|x, y_0) \leq P(u_1 - u_2 \leq (x_1 - x_2)'\beta + \gamma(y_0 - 1)|x, y_0) = F_{12}((x_1 - x_2)'\beta + \gamma(y_0 - 1))$$

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:

$$P(y_1 = 1, y_2 = 0|x, y_0) \leq F_{12}((x_1 - x_2)'\beta + \gamma(y_0 - 1))$$

Similarly, we can show that

$$P(y_1 = 0, y_2 = 1|\alpha, x, y_0) \leq P(u_1 - u_2 > (x_1 - x_2)'\beta + \gamma y_0|\alpha, x, y_0)$$

so that we have the following:

$$F_{12}((x_1 - x_2)'\beta + \gamma y_0) \leq 1 - P(y_1 = 0, y_2 = 1|x, y_0)$$

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

$$P(y_1 = 1, y_2 = 0|x, y_0) \leq F_{12}((x_1 - x_2)'\beta + \gamma(y_0 - 1))$$
$$F_{12}((x_1 - x_2)'\beta + \gamma y_0) \leq 1P(y_1 = 0, y_2 = 1|x, y_0)$$

Conditions (1) and (2) immediately follow from the fact that $F_{12}(\cdot)$ is a strictly increasing function. ■

B.3 Proof of Theorem 3.3

We establish our conclusions sequentially. We first show $\tilde{\beta}$ is point identified without having established point identification for $\tilde{\gamma}$. Next we explore identification for $\tilde{\gamma}$, assuming that $\tilde{\beta}$ is point identified. For this, we will first show the sign of $\tilde{\gamma}$ is identified. Then, assuming the sign of $\tilde{\gamma}$ is known, we show its magnitude generally *cannot* be identified. To show the first result, suppose that $\tilde{\beta} \neq \lambda\beta$ for any $\lambda > 0$. Note that in this case, conditions $PID - Stat1$ and $PID - Stat2$ in Assumption 3.4 imply that $P(\text{sign}(\Delta x\tilde{\beta}) \neq \text{sign}(\Delta x\beta)|x \in \mathcal{X}_7 \cap \mathcal{X}_8) > 0$ (see Lemma 2 in Manski (1985)). That is, there exist a subset of $\mathcal{X}_7 \cap \mathcal{X}_8$ (that has a positive probability measure) where $\text{sign}(\Delta x\tilde{\beta}) \neq \text{sign}(\Delta x\beta)$. For example, let $x^* \in \mathcal{X}_7 \cap \mathcal{X}_8$ be such that $\Delta x^*\tilde{\beta} > 0$ and $\Delta x^*\beta < 0$. Since x^* belongs to the union of $\mathcal{X}_7$ and $\mathcal{X}_8$ and $\Delta x^*\beta < 0$, Theorem 3.1 implies that it must be that $P(y_0 = 1, y_1 = 1|x^*) + P(y_1 = 1, y_2 = 0|x^*) > 1$ holds, which in turn rules out any $\tilde{\beta}$ such that $\Delta x^*\tilde{\beta} > 0$. Similar argument applies if $\Delta x^*\tilde{\beta} < 0$ abut $\Delta x^*\beta > 0$. Note that the above reasoning does not work when $\tilde{\beta} = \lambda\beta$ for some $\lambda > 0$, so β is point identified (only up to scale) on $\mathcal{X}_7 \cap \mathcal{X}_8$ under Assumption 3.4.

With β identified we can turn attention to the point identification of γ. We first establish when the sign of γ can be identified. First note that if $\gamma \geq 0$, then Theorem 3.1 implies that

$$\begin{aligned}
\Delta\mathcal{X}_1 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta + \gamma > 0\} \\
\Delta\mathcal{X}_2 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta - \gamma < 0\} \\
\Delta\mathcal{X}_3 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta > 0\} \\
\Delta\mathcal{X}_4 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta < 0\} \\
\Delta\mathcal{X}_5 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta + \gamma > 0\} \\
\Delta\mathcal{X}_6 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta - \gamma < 0\} \\
\Delta\mathcal{X}_7 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta > 0\} \\
\Delta\mathcal{X}_8 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta < 0\} \\
\Delta\mathcal{X}_9 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta - \gamma > 0\} \\
\Delta\mathcal{X}_{10} &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta + \gamma < 0\}
\end{aligned}$$

So if $(\Delta\mathcal{X}_1 \cup \Delta\mathcal{X}_5) \cap \Delta\mathcal{X}_{10} \neq \varnothing$ or $(\Delta\mathcal{X}_2 \cup \Delta\mathcal{X}_6) \cap (\Delta\mathcal{X}_9) \neq \varnothing$ or $\Delta\mathcal{X}_3 \cap \Delta\mathcal{X}_8 \neq \varnothing$ or $\Delta\mathcal{X}_4 \cap \Delta\mathcal{X}_7 \neq \varnothing$, then γ cannot be non-negative.

Similarly, if $\gamma \leq 0$, then we have (from Theorem 3.1)

$$\begin{aligned}
\Delta\mathcal{X}_1 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta - \gamma > 0\} \\
\Delta\mathcal{X}_2 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta + \gamma < 0\} \\
\Delta\mathcal{X}_3 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta - \gamma > 0\} \\
\Delta\mathcal{X}_4 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta + \gamma < 0\} \\
\Delta\mathcal{X}_5 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta > 0\} \\
\Delta\mathcal{X}_6 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta < 0\} \\
\Delta\mathcal{X}_7 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta > 0\} \\
\Delta\mathcal{X}_8 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta < 0\} \\
\Delta\mathcal{X}_9 &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta - \gamma > 0\} \\
\Delta\mathcal{X}_{10} &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x\beta + \gamma < 0\}
\end{aligned}$$

So if $\Delta\mathcal{X}_5 \cap \Delta\mathcal{X}_8 \neq \varnothing$ or $\Delta\mathcal{X}_6 \cap \Delta\mathcal{X}_7 \neq \varnothing$, then γ cannot be non-positive. Finally, if γ both cannot be positive or negative, it has to be zero (so it's point identified).

Finally, result in part (4) follows directly from Theorem 3.1. ■

B.4 Proof of Theorem 3.4

The proof of point identification of β follows closely the proof of a similar result for stationarity (Theorem 3.3). The identification result for γ follows from noticing that when $\gamma \geq 0$:

$$\begin{aligned}
\Delta\mathcal{X}_1(y_0) &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x'\beta - \gamma y_0 < 0\} \\
\Delta\mathcal{X}_2(y_0) &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x'\beta + \gamma - \gamma y_0 > 0\} \\
\Delta\mathcal{X}_3(y_0) &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x'\beta - \gamma y_0 > 0\} \\
\Delta\mathcal{X}_4(y_0) &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x'\beta + \gamma - \gamma y_0 < 0\}
\end{aligned}$$

Similarly, if $\gamma \leq 0$:

$$\begin{aligned}
\Delta\mathcal{X}_1(y_0) &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x'\beta + \gamma - \gamma y_0 < 0\} \\
\Delta\mathcal{X}_2(y_0) &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x'\beta - \gamma y_0 > 0\} \\
\Delta\mathcal{X}_3(y_0) &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x'\beta - \gamma y_0 > 0\} \\
\Delta\mathcal{X}_4(y_0) &\subseteq \{\Delta x \in \mathbb{R}^k : \Delta x'\beta + \gamma - \gamma y_0 < 0\}
\end{aligned}$$

The rest follows from arguments similar to the ones used to prove Theorem 3.3. ■

B.5 Proof of Theorem 4.1

This 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):

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

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

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

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

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

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

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

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

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

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

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

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

■

B.6 Proof of Theorem 4.2

We only give a sketch of the proof since it closely follows the proof of Theorem 4.2. First, for any $\tilde{\theta} \in \Theta_{I,cex}^{\{1,2\}}(1)$, $\tilde{\theta} \in \Theta_{I,cex}^{\{1,3\}}(1)$ and $\tilde{\theta} \in \Theta_{I,cex}^{\{2,3\}}(1)$ we can construct corresponding bivariate Fréchet copulas (with corresponding marginal distributions $\tilde{F}$) such that, respectively:

(1) $P(y_1, y_2|z)$ are matched;

(2) $P(y_1, y_3|z)$ are matched;

(3) $P(y_2, y_3|z)$ are matched.

If $\tilde{\theta} \in \Theta_{I,cex}^{\{1,2\}}(1) \cap \Theta_{I,cex}^{\{1,3\}}(1) \cap \Theta_{I,cex}^{\{2,3\}}(1)$, the we'll be able to find the same marginal distribution $\tilde{F}$ that works for all three copulas. And finally, if $\tilde{\theta} \in \Theta_{I,cex}^{\{1,2,3\}}$, then we'll be able to use the same Fréchet copula in (1)-(3) and find a trivariate copula with two-dimensional marginals given by that common Fréchet copula that allows us to match the distribution of observables.

■

C Tables

	If:	Then:
1.a	$(x_3 - x_2)'\beta + \max\{0, \gamma\} \leq 0$	$P(0,0,1\|z) \leq P(0,1,0\|z)$
1.b	$(x_3 - x_2)'\beta \leq 0$, $(x_3 - x_1)'\beta - \gamma y_0 + \gamma \leq 0$, $(x_1 - x_2)'\beta + \gamma y_0 \leq 0$	
2.a	$(x_3 - x_2)'\beta + \min\{0, \gamma\} \geq 0$	$P(0,0,1\|z) \geq P(0,1,0\|z)$
2.b	$(x_3 - x_2)'\beta \geq 0$, $(x_3 - x_1)'\beta - \gamma y_0 + \gamma \geq 0$, $(x_1 - x_2)'\beta + \gamma y_0 \geq 0$	
3.a	$(x_3 - x_1)'\beta - \gamma y_0 \leq 0$, $(x_2 - x_1)'\beta - \gamma y_0 + \gamma \leq 0$, $(x_3 - x_2)'\beta \leq 0$	$P(0,0,1\|z) \leq P(1,0,0\|z)$
3.b	$(x_3 - x_1)'\beta - \gamma y_0 \leq 0$, $\gamma \leq 0$	
4.a	$(x_3 - x_1)'\beta - \gamma y_0 \geq 0$, $(x_2 - x_1)'\beta - \gamma y_0 + \gamma \geq 0$, $(x_3 - x_2)'\beta \geq 0$	$P(0,0,1\|z) \geq P(1,0,0\|z)$
4.b	$(x_3 - x_1)'\beta - \gamma y_0 \geq 0$, $\gamma \geq 0$	
5.a	$(x_2 - x_1)'\beta - \gamma y_0 + \gamma \leq 0$, $(x_2 - x_1)'\beta - \gamma y_0 \leq 0$, $-\gamma \leq 0$	$P(0,1,0\|z) \leq P(1,0,0\|z)$
5.b	$(x_3 - x_1)'\beta - \gamma y_0 \leq 0$, $(x_2 - x_1)'\beta - \gamma y_0 \leq 0$, $(x_2 - x_3)'\beta \leq 0$	
6.a	$(x_2 - x_1)'\beta - \gamma y_0 + \gamma \geq 0$, $(x_2 - x_1)'\beta - \gamma y_0 \geq 0$, $-\gamma \geq 0$	$P(0,1,0\|z) \geq P(1,0,0\|z)$
6.b	$(x_3 - x_1)'\beta - \gamma y_0 \geq 0$, $(x_2 - x_1)'\beta - \gamma y_0 \geq 0$, $(x_2 - x_3)'\beta \geq 0$	
7.a	$(x_2 - x_1)'\beta - \gamma y_0 \leq 0$, $(x_2 - x_1)'\beta - \gamma y_0 + \gamma \leq 0$, $\gamma \leq 0$	$P(0,1,1\|z) \leq P(1,0,1\|z)$
7.b	$(x_3 - x_1)'\beta - \gamma y_0 + \gamma \leq 0$, $(x_2 - x_1)'\beta - \gamma y_0 + \gamma \leq 0$, $(x_2 - x_3)'\beta \leq 0$	
8.a	$(x_2 - x_1)'\beta - \gamma y_0 \geq 0$, $(x_2 - x_1)'\beta - \gamma y_0 + \gamma \geq 0$, $\gamma \geq 0$	$P(0,1,1\|z) \geq P(1,0,1\|z)$
8.b	$(x_3 - x_1)'\beta - \gamma y_0 + \gamma \geq 0$, $(x_2 - x_1)'\beta - \gamma y_0 + \gamma \geq 0$, $(x_2 - x_3)'\beta \geq 0$	
9.a	$(x_3 - x_1)'\beta - \gamma y_0 + \gamma \leq 0$, $(x_2 - x_1)'\beta - \gamma y_0 \leq 0$, $(x_3 - x_2)'\beta \leq 0$	$P(0,1,1\|z) \leq P(1,1,0\|z)$
9.b	$(x_3 - x_1)'\beta - \gamma y_0 + \gamma \leq 0$, $-\gamma \leq 0$	
10.a	$(x_3 - x_1)'\beta - \gamma y_0 + \gamma \geq 0$, $(x_2 - x_1)'\beta - \gamma y_0 \geq 0$, $(x_3 - x_2)'\beta \geq 0$	$P(0,1,1\|z) \geq P(1,1,0\|z)$
10.b	$(x_3 - x_1)'\beta - \gamma y_0 + \gamma \geq 0$, $-\gamma \geq 0$	
11.a	$(x_3 - x_2)'\beta - \min\{0, \gamma\} \leq 0$	$P(1,0,1\|z) \leq P(1,1,0\|z)$
11.b	$(x_3 - x_2)'\beta \leq 0$, $(x_3 - x_1)'\beta - \gamma y_0 \leq 0$, $(x_1 - x_2)'\beta + \gamma y_0 - \gamma \leq 0$	
12.a	$(x_3 - x_2)'\beta - \max\{0, \gamma\} \geq 0$	$P(1,0,1\|z) \geq P(1,1,0\|z)$
12.b	$(x_3 - x_2)'\beta \geq 0$, $(x_3 - x_1)'\beta - \gamma y_0 \geq 0$, $(x_1 - x_2)'\beta + \gamma y_0 - \gamma \geq 0$	

Table 3: Restrictions for $T = 3$ time periods under conditional exchangeability assumption. Strict inequalities in parameters imply strict inequalities in probabilities.

	If:	Then:
1	$\gamma \leq 0$	$P(0,0,1\|y_0=0) \leq P(0,1,0\|y_0=0)$
2	$\gamma \geq 0$	$P(0,0,1\|y_0=0) \geq P(0,1,0\|y_0=0)$
3	$\gamma \leq 0$	$P(0,0,1\|y_0=0) \leq P(1,0,0\|y_0=0)$
4	$\gamma \geq 0$	$P(0,0,1\|y_0=0) \geq P(1,0,0\|y_0=0)$
7	$\gamma \leq 0$	$P(0,1,1\|y_0=0) \leq P(1,0,1\|y_0=0)$
8	$\gamma \geq 0$	$P(0,1,1\|y_0=0) \geq P(1,0,1\|y_0=0)$
9.a	$\gamma \leq 0$	$P(0,1,1\|y_0=0) \leq P(1,1,0\|y_0=0)$
10.a	$\gamma \geq 0$	$P(0,1,1\|y_0=0) \geq P(1,1,0\|y_0=0)$
11.b	$\gamma \geq 0$	$P(1,0,1\|y_0=0) \leq P(1,1,0\|y_0=0)$
12.b	$\gamma \leq 0$	$P(1,0,1\|y_0=0) \geq P(1,1,0\|y_0=0)$

Table 4: Identifying restrictions for $T = 3$ and conditional exchangeability assumption. Strict inequalities for γ imply strict inequalities in probabilities.

	If:	Then:
1	$\gamma \leq 0$	$P(0,0,1\|y_0=1) \leq P(0,1,0\|y_0=1)$
2	$\gamma \geq 0$	$P(0,0,1\|y_0=1) \geq P(0,1,0\|y_0=1)$
3.a	$\gamma \geq 0$	$P(0,0,1\|y_0=1) \leq P(1,0,0\|y_0=1)$
4.a	$\gamma \leq 0$	$P(0,0,1\|y_0=1) \geq P(1,0,0\|y_0=1)$
5	$\gamma \geq 0$	$P(0,1,0\|y_0=1) \leq P(1,0,0\|y_0=1)$
6	$\gamma \leq 0$	$P(0,1,0\|y_0=1) \geq P(1,0,0\|y_0=1)$
9	$\gamma \geq 0$	$P(0,1,1\|y_0=1) \leq P(1,1,0\|y_0=1)$
10	$\gamma \leq 0$	$P(0,1,1\|y_0=1) \geq P(1,1,0\|y_0=1)$
11.b	$\gamma \geq 0$	$P(1,0,1\|y_0=1) \leq P(1,1,0\|y_0=1)$
12.b	$\gamma \leq 0$	$P(1,0,1\|y_0=1) \geq P(1,1,0\|y_0=1)$

Table 5: Identifying restrictions for $T = 3$ and conditional exchangeability assumption. Strict inequalities for γ imply strict inequalities in probabilities.