Identification of Dynamic Binary Response Models*

S. Khan† M. Ponomareva‡ E. Tamer §

September 3, 2021

Abstract

We consider identification of parameters in dynamic binary response models with panel data under minimal assumptions. This model is prominent in empirical economics as it has been used to infer state dependence in the presence of unobserved heterogeneity. The main results in the paper are characterizations of the identified set under weak assumptions. Relative to the results in the seminal work of Honoré and Kyriazidou (2000) (HK) we make several advances: 1) our identified set do not require any restrictions on the support of the observables; for example, unlike HK, we allow for time trends, time dummies, and/or only discrete covariates; 2) our main results are derived under the assumption that the idiosyncratic error terms are stationary over time conditional on only the fixed effect and the covariates (and without conditioning on initial conditions); this is in contrast to most dynamic nonlinear panel data models including HK that use stronger restrictions on the initial conditions and require independence over time; 3) we show that it is possible to get point identification in some cases even with $T = 2$ (two time periods). For inference, and in cases with discrete regressors, we provide a linear programming approach to constructing confidence sets for the parameters of interest that is simple to implement. We also simulate the size and shape of identified sets in varying designs to illustrate the informational content of different assumptions. As illustration, we estimate women's labor force participation using the data set from Chay and Hyslop (2014).

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

JEL: C22, C23, C25.

*We are grateful for helpful comments from S. Chen, A. Chesher, B. Honoré, H.Ichimura, T. Magnac, as well as conference and seminar participants at several institutions. 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

We consider the binary dynamic panel data model that relates the outcome in period t for individual i, y_{it}, to its lagged value $y_{i,t-1}$ in the following way

$$y_{it} = I\{u_{it} \leq x'_{it}\beta + \gamma y_{i,t-1} + \alpha_i\} \quad t = 1, 2, ...T \tag{1}$$

We are interested in learning about $\theta = (\gamma, \beta')'$ using and iid sample on n entities for T time periods. This parameter θ is treated as a fixed (but unknown) constant vector while the unobservables here take the standard form $u_{it} - \alpha_i$ where α_i is an individual specific and time independent "fixed effect," and is meant to capture the systematic correlation of the unobservables over time[1], while u_{it} is an idiosyncratic error term that is both time and entity specific. The parameter γ is of special interest as it measures the effect of state dependence (also switching costs or inertia). A fixed effect approach treats α_i as possibly arbitrarily correlated with the regressor vector $x_i = (x'_{i1}, \ldots, x'_{iT})'$. The challenge is to identify θ under general assumptions on the conditional distribution of $u_i = (u_{i1}, \ldots, u_{iT})'$. Another serious complication in this model is the treatment of y_{i0}, or the initial condition, and in particular its relationship to the unobservables over time.

In econometric theory, the literature that has studied this model is vast (see Section 1.1 below), but the benchmark theoretical results are in Honoré and Kyriazidou (2000) (HK). In its most general form, HK considers a particular version of the above model that maintains the following: 1) u_i is independent of (x_i, α_i), 2) u_i is iid over time, i.e., u_{it} is independent and identically distributed with $u_{it'}$ for all $t, t' \leq T, \quad t \neq t'$ and in particular 3) u_i is independent of y_{i0}. In addition, HK's identification results, focused on obtaining point identification, require that one is able to match regressors for the same individual over time, i.e., point identification (and for that matter consistency of their estimator) is only guaranteed if there is *overlapping support* in regressors, and hence there is a set of observations where $x_{it} = x_{it'}$ for $t \neq t'$. So, this point identification strategy rules out models where the vector x_{it} contains time trends or time dummies. This paper substantively relaxes all these assumptions. First, we characterize the identified set of the model when a weak stationarity assumption is maintained on u_i. We require that u_{i1} has the same distribution as u_{it} for all $1 < t \leq T$ conditional on $(\alpha_i, x'_i)'$. This weak stationarity restriction is in

[1]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.

line with the stationarity restriction used in Manski (1987) for the analysis of the *static* treatment of this model and so we allow for serial correlation, and heteroskedasticity. More importantly, this stationarity restriction does not condition on the initial condition y_{i0}. Also and importantly, we do not make any restrictions on the support of x_i and so allow for time trends, time dummies, etc. Finally, our identified sets, which are unions of convex polytopes, are characterized by sets of linear inequalities that hold for any T and are simple to compute. These identified sets contain information about θ even in the case when $T = 2$.

The paper also contains a set of additional results. We first strengthen the stationarity restriction, but in such a way that the model is still less restrictive than HK. Specifically we require that stationarity now holds conditional on y_0, whereby, similarly to x_i, the initial condition is now strictly exogenous. We then derive our new identified set under this model. Conditioning on the initial condition adds identifying power, and so by comparing this set to the one obtained without conditioning on y_0 provides information on the strength of this assumption. This allows one to determine the sensitivity of our inferences to the initial condition problem.

Given these identified sets, we are able to study sufficient conditions under which the models lead to *point identification* of θ. For our main model under stationarity, and when $T = 2$, we derive a set of interesting results. The parameter β and the sign of γ can always be point identified. In addition, γ can be point identified when it is non negative. When γ is negative, we obtain a sharp upper bound on it. When $T = 3$, it is possible to point identify θ, regardless of the sign of γ.

Given the linear structure of our identified sets, we propose a simple and insightful linear program that solves for min/max of linear functions of θ (such as γ) and show how that leads to a simple inference strategy. The paper contains insightful numerical experiments in models with discrete regressors, time trends, and/or time dummies for $T = 2, 3$ that show that the model under stationarity (conditional and unconditional) contains information about θ (the identified set is not trivial). Finally, we illustrate our approach using a data set on women's labor supply.

1.1 Literature

In empirical economics, there is a recent renewed interest 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 (like a particular health insurance plan) if they have experienced that choice in the past. 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 where today's choice depends on last period's and where he documents *inertia* in choices overtime. Dubé, Hitsch, and Rossi (2010) empirically find that this "inertia" in packaged goods markets is likely caused by brand loyalty. Pakes, Porter, Shepard, and Wang (2019) also study the choice of health insurance over time using panel data[2]. Moreover the recent availability of 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 an empirical motivation for this paper.

More broadly, the dynamic discrete choice model has appeared prominently in both the applied and theoretical econometrics literature. In fundamental work, Heckman (1981) who considered a version of model (1) 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[3]. 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 prescription. Hence, the econometrics literature since then has focused on models under which we are able to empirically identify state dependence while allowing

[2]See also Polyakova (2016) that studies the question of quantifying the effect of switching costs in Medicare Part D markets and its relation to adversely selected plans. Also, Ketcham, Lucarelli, and Powers (2015) quantifies 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.

[3]See the recent work in Torgovitsky (2019) that provides characterization of this causal effect under minimal assumptions, but unlike the work here does not identify parameters such as θ under general conditions

for serial correlation. These models differ in the kinds of assumptions used. For important work on inference on θ here, and in addition to HK, see Heckman (1981) and Chamberlain (1984)[4]. 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[5]. 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. Khan, Ouyang, and Tamer (2019) extend results in HK to cover point identified *multinomial models* with dynamics. Honoré and Kyriazidou (2019) calculate identified regions for parameters such as θ above in panel data autoregressive models (although they do not provide an explicit characterization of these identified regions). Aristodemou (2019) contains informative outer sets under the stronger independence assumptions. See also Gao and Li (2019) for recent results on identification of parameter in panel data nonlinear models. 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 (2019) constructs identified sets for average causal effect of lagged outcomes in binary response models under minimal assumptions.
Finally, this paper takes a direct approach to proving that the sets we propose are sharp. This means that for every parameter in the bounds we propose, we can construct a dgp that obeys the model restrictions and generate the (observed) data. For other approaches that can be used to construct identified sets (and conduct inference), see Beresteanu, Molchanov, and Molinari (2011) and Chesher and Rosen (2017).

The rest of our paper is organized as follows. Section 2 introduces the dynamic binary

[4]For other work on different dynamic models, see the thorough literature surveys in Arellano and Honoré (2001) and Arellano and Bonhomme (2011), as well as the papers Honoré and Tamer (2006), Honoré and Lewbel (2002), Altonji and Matzkin (2005) , Chen, Khan, and Tang (2019). 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) and Honoré and Hu (2020). For recent new developments on the dynamic Logit model with fixed effects, see Kitazawa (2021), Honoré and Weidner (2020). These two recent papers rely heavily on the logistic distribution assumption for their moment conditions, and hence are just as sensitive to distributional misspecification as the early work in this literature, e.g. Andersen (1970).

[5]Some of their results were extended and expanded in Pakes, Porter, Shepard, and Wang (2019).

choice model and addresses the identification of its structural parameters under the assumption of stationarity as was introduced in Manski (1987) for the static binary response panel data model. Section 3 provides sufficient conditions for point identification of these structural parameters in $T = 2$ and $T = 3$ cases. Section 4 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 5 considers a range of simulation studies that provide insight into the shape and size of identified region across the different data generating processes, while Section 6 uses the approach proposed in the paper to investigate persistence in female labor supply. Section 7 concludes with a summary of results and discussions on areas for future research.

2 Dynamic Panel Binary Choice Model

Assume a random sample $i = 1, \ldots, n$ over T time periods where we observe for each i, a binary outcome $y_{it} \in \{0, 1\}$, a vector of covariates x_{it}. In addition, the outcome y_{i0} is also observed at the initial time period $t = 0$ (but no covariates are observed at that time). We also consider the following binary choice model for the outcome y_{it}:

$$y_{it} = 1\{u_{it} \leq x_{it}'\beta + \gamma y_{i,t-1} + \alpha_i\} \tag{2}$$

where both u_{it} and α_i are unobserved and represent unobserved heterogeneity. u_{it} denotes an idiosyncratic individual and time dependent term while α_i is individual/entity-specific and captures unobserved constant over time heterogeneity. Let $y_i = (y_{i0}, \ldots, y_{iT})'$ and $x_i = (x_{i1}', \ldots, x_{iT}')'$. Throughout, we assume that observations $\{(y_i, x_i), i = 1, \ldots, n\}$ are independent and identically distributed.

The unknown scalar parameter γ measures persistence (or the degree of state dependence) in this model. The objective of this paper is to explore the identifiability of parameters β and γ under relatively weak conditions. This mapping between the restrictions on the conditional joint distribution of $(u_{i1}, \ldots, u_{iT}, \alpha_i)$ (the model) and information about γ (and β) is of preliminary interest. We do not want to restrict the way individual-specific effects α_i can vary with covariates, neither do we want to impose *any* restrictions on the support of x_i (we do not rule out time trends, time dummies, or discrete regressors for example). The

main restriction we maintain is that the idiosyncratic error terms $u_{i1}, u_{i2}, \ldots, u_{iT}$ all have the same marginal distribution conditional on observed covariates and fixed effects. This stationarity restriction is the main assumption we make in the paper.

Assumption 1. ***(Stationarity)*** $u_{it}|x_i, \alpha_i \overset{d}{=} u_{i1}|x_i, \alpha_i$ *for all* $t = 2, 3, \ldots, T$.

Remarks on Assumption 1

This assumption requires that conditional on x_i and α_i, the distribution of u_{it} remains the same over time. We note that this assumption does not require conditioning on y_0 (nor does it require *independence from* y_{i0}) since conditioning on y_{i0} may rule out certain feedback from past outcomes. The conditioning on the whole vector of x_i is common in panel data models and is related to a version of *strict exogeneity* of the regressors. This stationarity (or identical distribution of idiosyncratic error terms) is the key identifying assumption in Manski (1987) for the *static* semiparametric binary response model with fixed effects. It is worth noting that stationarity allows for dependence between u_{it}'s over time, and naturally, it is satisfied when idiosyncratic error terms are independent and identically distributed as assumed in HK. Note also that this assumption does not rule out possible serial correlation in some components of x_{it}, nor does it rule out time-varying explanatory variables including time trend or time dummies. This model also does not impose any restriction on the distribution of α_i conditional on x_i. Finally, it is worth repeating that benchmark theoretical results in this literature (such as HK's) assume that the u_{it}'s are independent of (x_i, α_i, y_{i0}) and also are iid over time.

To study the identification power of the assumption above, we also require the conditions below to hold.

Assumption 2. *Suppose that the following conditions hold for model 2:*

A1. $u_i = (u_{i1}, \ldots, u_{iT})$ *is absolutely continuous conditional on* x_i, α_i.

A2. *We observe* n *i.i.d. draws from (2):* $\{(y_i, x_i), i = 1, \ldots, n\}$.

A3. *Parameter space* $\Theta = \{\theta = (\gamma, \beta')' \in \mathbb{R}^{k+1} : \|\theta\| = 1\}$

Remarks on Assumption 2

The first part of the above requires us to use strict monotonicity of the distribution functions of various objects. The second part **A2** describes the sampling process. The third part, **A3**, is a normalization and is common in semiparametric binary choice models literature where the parameter can only be identified up to location and scale. 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 stationarity assumption 1 that is imposed on the error terms is pretty weak and so we cannot expect to be able to pinpoint the true parameters β and γ. However, there is still some information content in this assumption that allows us to restrict the set of all possible parameter values to some subset of values that are compatible with the distribution of observables. What we mean by that is that any value in this set could have generated the distribution of observables while maintaining assumptions 1 and 2. The result in Theorem 1 below is the main result of this paper: it constructs the identified set for the parameter of interest $\theta = (\gamma, \beta')'$ and shows that this identified set is *sharp* under the stationarity assumption 1. That is, for any parameter value in that set there exists a distribution of unobservables that follows the stationarity assumption and dynamic binary choice model (2) such that it produces exactly the same distribution of (y_i, x_i) as the true parameter.

Let $\mathcal{X} = supp(x_i)$ be the support of x_i. We define Θ_I as the set of all $\theta = (\gamma, \beta')' \in \Theta$ such that for all $x \in \mathcal{X}$ and $t, s = 1, \ldots, T$ with $t \neq s$, the following hold (all probability statements below are conditional on $x_i = x$ where $x_i = (x_{i1}', \ldots, x_{iT}')'$ and $x = (x_1', \ldots, x_T')'$):

(i) If $P(y_{it} = 1) \geq P(y_{is} = 1)$, then $(x_t - x_s)'\beta + |\gamma| \geq 0$.

(ii) If $P(y_{it} = 1) \geq 1 - P(y_{i,s-1} = 1, y_{is} = 0)$, then $(x_t - x_s)'\beta - \min\{0, \gamma\} \geq 0$.

(iii) If $P(y_{it} = 1) \geq 1 - P(y_{i,s-1} = 0, y_{is} = 0)$, then $(x_t - x_s)'\beta + \max\{0, \gamma\} \geq 0$.

(iv) If $P(y_{i,t-1} = 1, y_{it} = 1) \geq P(y_{is} = 1)$, then $(x_t - x_s)'\beta + \max\{0, \gamma\} \geq 0$.

(v) If $P(y_{i,t-1} = 1, y_{it} = 1) \geq 1 - P(y_{i,s-1} = 1, y_{is} = 0)$, then $(x_t - x_s)'\beta \geq 0$.

(vi) If $P(y_{i,t-1} = 1, y_{it} = 1) \geq 1 - P(y_{i,s-1} = 0, y_{is} = 0)$, then $(x_t - x_s)'\beta + \gamma \geq 0$.

(vii) If $P(y_{i,t-1} = 0, y_{it} = 1) \geq P(y_{is} = 1 | x_i = x)$, then $(x_t - x_s)'\beta - \min\{0, \gamma\} \geq 0$.

(viii) If $P(y_{i,t-1}=0, y_{it}=1) \geq 1 - P(y_{i,s-1}=1, y_{is}=0)$, then $(x_t - x_s)'\beta - \gamma \geq 0$.

(ix) If $P(y_{i,t-1}=0, y_{it}=1) \geq 1 - P(y_{i,s-1}=0, y_{is}=0)$, then $(x_t - x_s)'\beta \geq 0$.

Remark on Θ_I

The above set is defined in terms of a set of inequalities conditional on regressors x, and the conditional probabilities in these inequalities are observed in the data. These inequalities are obtained by first bounding the distribution of the composite error term $u_{it} - \alpha_i$ with probabilities of observed outcomes (so, y_{it}'s), and second, ensuring that the "intersection" of these bounds is not empty in the sense that there exists a distribution function such that it passes between these bounds in all time periods. The inequalities obtained this way are linear in θ which is helpful since the identified set can then be characterized as a solution to a set of linear half spaces, a union of convex polyhedrons. Note that some of the inequalities compare marginal probabilities to joint (bivariate) which is interesting and is related to the model containing one lagged dependent variable. Note also that the comparison is always between the value of the outcome at t vs at s where $t \neq s$. So, when $T = 2$, Θ_I is defined by 18 inequalities (9 inequalities for $t = 1$ and $s = 2$ and 9 inequalities for $t = 2$ and $s = 1$). For an arbitrary T, Θ_I is defined by $9 \cdot 2 \cdot \binom{T}{2}$ inequalities, so the number of these inequalities increases at rate $T!$. The set Θ_I defined above is a *sharp identified set* for θ. This result, the main finding in this paper, is stated in the next Theorem.

Theorem 1. *Θ_I defined above is the sharp identified set for θ under stationarity assumption (1) and regularity condition A1 in assumption 2.*

Theorem 1 (proof in the Appendix) provides a constructive way to characterize the identified set under the assumption of stationarity. In other words, we can use the above inequalities to construct set estimation and inference procedures for the parameters β, γ. It turns out that comparisons of joint to marginals are also relevant in the proof of the result. Note also how the comparisons work as T is increased. For instance, only bivariate comparisons show up and not joints of say 3 or more outcomes in cases when $T = 2$. This is so because inequalities that include 3 outcomes are implied by bivariate inequalities for any $T > 2$. Also, note that the result in Theorem 1 holds even when the "index" in (2) is nonlinear in θ. The only complication that results in this case would be that characterizing the identified set becomes harder with the added nonlinearities.

The proof of the Theorem 1 does not require the fixed effect α_i's to be additively separated from u_{it}'s: a model

$$y_{it} = 1\{v_{it} \leq x'_{it}\beta + \gamma y_{i,t-1}\}$$

where $v_{it} \sim v_{i1}$ conditional on x_i for $t = 2, \ldots, T$ produces exactly the same identified set for θ.[6] The model with additively separated fixed effects implies that if u_{it}'s are serially uncorrelated, the correlation between v_{it} and v_{is} is positive and the same for all t and s. However, the stationarity assumption also allows for negative correlation between v_{it} and v_{is}.

Remarks on the Sharp Identified Set Θ_I

The above set is defined in terms of a set of inequalities conditional on regressors x, and the conditional probabilities in these inequalities are observed in the data. The inequalities are also linear in θ which is helpful since the identified set is a solution to a set of linear half spaces, a union of convex polyhedrons. Note that some of the inequalities compare marginal probabilities to joint (bivariate) which is interesting and is related to the model containing one lagged dependent variable. Note also that the comparison is always between the value of the outcome at t vs at s where $t \neq s$. So, when $T = 2$, then $t = 1, s = 2$ or $t = 2, s = 1$. Also note that the above inequalities characterize the identified set for any T (so no need to re-derive the identified set for different T's). The number of these inequalities increases with T: the larger T is the more inequalities we get.

2.1 Stationarity Conditional on Initial Condition

In the identification analysis above, no assumptions are made about the distribution of the initial conditions, y_{i0}. However, a common assumption that is (almost) always assumed in the nonlinear panel data literature is that y_{i0}, similar to x_i, is strictly exogenous; see e.g. Honoré and Kyriazidou (2000) for a dynamic binary response model and where they assume serial independence in u_{it} as well as independence between u_{it} and x_i, Hu (2002) for a censored dynamic panel data model, or Ahn and Schmidt (1997) for a linear dynamic

[6] We thank the anonymous reviewer for pointing that out.

(linear) panel data model[7]. In this section, we do not consider full exogeneity of y_0, but rather, we maintain stationarity as in the previous section but make it conditional on y_0 (in addition to covariates and fixed effects). This effectively makes y_0 strictly exogenous in the same sense as x being strictly exogenous. One important reason to consider stationarity conditional on the initial condition is to compare the size of the identified set in Theorem 1 above -when we do not condition on y_0- to the one that results when we condition on y_0 (leaving all other assumptions the same). This will clarify the importance - in terms of identification power- of conditioning on the initial condition. A much tighter identified set would mean that conditioning on y_0 is a key restriction and hence is worthy of discussion.

Specifically, we modify Assumption 1 as follows:

Assumption 3. ***(Stationarity Conditional on Initial Condition)***

$u_{it}|y_{i0}, x_i, \alpha_i \overset{d}{=} u_{i1}|y_{i0}, x_i, \alpha_i$ *for all* $t = 2, 3, \ldots, T$.

Again, this assumption is weaker than assuming that $u_{it}|y_{i0}, x_i, \alpha_i$ are independent and identically distributed and independent of (x_i, α_i) and y_{i0} (a common assumption employed in almost all the parametric and semi-parametric dynamic binary response models such as HK).

Following our approach in the previous Section, we define two sets: Θ_1 and Θ_{2+}. Specifically, Θ_1 is defined as the set of all $(\beta, \gamma) \in \Theta$ such that for all $x \in \mathcal{X}$ and $t = 2, \ldots, T$, the following hold (all probability statements below are conditional on $x_i = x$ where $x_i = (x_{i1}', \ldots, x_{iT}')'$ in addition to y_{i0}):

(i) If $P(y_{i1} = 1|y_{i0} = d_0) \geq P(y_{it} = 1|y_{i0} = d_0)$, then $(x_t - x_1)'\beta + \min\{0, \gamma\} - \gamma d_0 \leq 0$.

(ii) If $P(y_{i1} = 1|y_{i0} = d_0) \leq P(y_{it} = 1|y_{i0} = d_0)$, then $(x_t - x_1)'\beta + \max\{0, \gamma\} - \gamma d_0 \geq 0$.

(iii) If $P(y_{i1} = 1|y_{i0} = d_0) \geq 1 - P(y_{i,t-1} = 1, y_{it} = 0|y_{i0} = d_0)$, then $(x_t - x_1)'\beta + \gamma(1 - d_0) \leq 0$.

(iv) If $P(y_{i1} = 1|y_{i0} = d_0) \geq 1 - P(y_{i,t-1} = 0, y_{it} = 0|y_{i0} = d_0)$, then $(x_t - x_1)'\beta - \gamma d_0 \leq 0$.

(v) If $P(y_{i1} = 1|y_{i0} = d_0) \leq P(y_{i,t-1} = 1, y_{it} = 1|y_{i0} = d_0)$, then $(x_t - x_1)'\beta + \gamma(1 - d_0) \geq 0$.

[7] Chay and Hyslop (2014) show that in a context of dynamic binary response model, estimates of the degree of state dependence, γ, are sensitive to the assumption that initial conditions are exogenous.

(vi) If $P(y_{i1} = 1|y_{i0} = d_0) \leq P(y_{i,t-1} = 0, y_{it} = 1|y_{i0} = d_0)$, then $(x_t - x_1)'\beta - \gamma d_0 \geq 0$.

The second set, Θ_{2+}, is defined in the same way as Θ_I, with two corrections:

- Probabilities are conditional on $x_i = x, y_{i0} = d_0$ (instead of just $x_i = x$);
- Time periods $t, s \geq 2$.

Remark on Θ_1 and Θ_{2+}

Note that here we condition on the initial value y_{i0} which allows us to obtain a (conditional on y_{i0}) version of the inequalities that defined Θ_I. In addition, we also have extra inequalities that form Θ_1. The intersection between Θ_1 and Θ_{2+} gives us a *sharp identified set* for θ under Assumption 3, as stated by Theorem 2 below. Conditioning on y_{i0} seems to have substantive identifying power in that it introduces *additional* restrictions that do not appear in the Θ_{2+} (the conditional on y_{i0} version of Θ_I from Theorem 1).

Theorem 2. *Let $\Theta_{I,0} = \Theta_1 \cap \Theta_{2+}$. Then, $\Theta_{I,0}$ is the (sharp) identified set for θ under stationarity assumption 3 with exogenous initial conditions and regularity condition A1 in assumption 2.*

The main usefulness of the identified in the Theorem 2 is its size relative to the identified set in Theorem 1. A comparison between these two sets indicates sensitivity to the exogeneity assumption on the initial condition. Indeed, if the identified set under Theorem 2 is empty while the set under Theorem 1 is not would mean that the conditional on initial condition distribution is false. Note also that Theorems 1 and 2 provide sharp sets without any restrictions on the covariate distribution, such as whether the covariates vary over time, whether there is overlap in their support, or whether these covariates admit discrete or continuous distribution, etc.

Finally, it remains interesting to ask under what conditions do the identified sets in Theorems 1 and 2 shrink to a point. Typically, these (sufficient) point identification conditions are made on the support of the regressors. We examine this question of point identification in the next Section. We also analyze whether anything can be learned about γ (its sign in this case) in a model without any covariates.

3 Point Identification

Here we explore the question of whether and under what conditions do the sharp sets characterized in Theorem 1 shrink to a single 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 $T = 2$, β can be point identified as well as the sign of γ. In addition, γ can be point identified when it is non-negative.

3.1 Point Identification with $T = 2$

Even when only two periods of data are available (in addition to the initial value y_{i0}), there are implications of Theorem 1 that can be explored to study whether and under what conditions would the identified set shrinks to a point (point identification). Most notably, we establish here that under certain support conditions on x the parameter vector β can be *point* identified (up to scale), and so can the sign of γ. Also, when γ is non negative, it can be point identified.

When $T = 2$ and considering both $t = 1$, $s = 2$ and $t = 2$, $s = 1$ we end up with the following restrictions (again all probability statements are conditional on $x_i = x$):

(1) If $P(y_{i2} = 1) \geq P(y_{i1} = 1)$, then $(x_2 - x_1)'\beta + |\gamma| \geq 0$;

(2) If $P(y_{i1} = 1) \geq P(y_{i2} = 1)$, then $(x_2 - x_1)'\beta - |\gamma| \leq 0$;

(3) If $P(y_{i1} = 0, y_{i2} = 1) \geq P(y_{i1} = 1)$ or $P(y_{i0} = 1, y_{i1} = 0) \geq P(y_{i2} = 0)$, then $(x_2 - x_1)'\beta - \min\{0, \gamma\} \geq 0$;

(4) If $P(y_{i1} = 1, y_{i2} = 0) \geq P(y_{i1} = 0)$ or $P(y_{i0} = 0, y_{i1} = 1) \geq P(y_{i2} = 1)$, then $(x_2 - x_1)'\beta + \min\{0, \gamma\} \leq 0$;

(5) If $P(y_{i0} = 0, y_{i1} = 0) \geq P(y_{i2} = 0)$, then $(x_2 - x_1)'\beta + \max\{0, \gamma\} \geq 0$;

(6) If $P(y_{i0} = 1, y_{i1} = 1) \geq P(y_{i2} = 1)$, then $(x_2 - x_1)'\beta - \max\{0, \gamma\} \leq 0$;

(7) If $P(y_{i0} = 0, y_{i1} = 0) + P(y_{i1} = 0, y_{i2} = 1) \geq 1$, then $(x_2 - x_1)'\beta \geq 0$;

(8) If $P(y_{i0} = 1, y_{i1} = 1) + P(y_{i1} = 1, y_{i2} = 0) \geq 1$, then $(x_2 - x_1)'\beta \leq 0$;

(9) If $P(y_{i0}=1,y_{i1}=0)+P(y_{i1}=0,y_{i2}=1)\geq 1$, then $(x_2-x_1)'\beta-\gamma\geq 0$;

(10) If $P(y_{i0}=0,y_{i1}=1)+P(y_{i1}=1,y_{i2}=0)\geq 1$, then $(x_2-x_1)'\beta+\gamma\leq 0$

Based on these restrictions, we define subsets (one for each restriction) of $\mathcal{X}$ that can help us to identify the sign of γ (as before, $\mathcal{X}=supp(x_i=(x_{i1},\ldots,x_{iT}))$):

$$\begin{aligned}
\Delta\mathcal{X}_1 &= \{\Delta x\in\mathbb{R}^k : \exists x=(x_1,x_1+\Delta x)\in\mathcal{X} \text{ such that } P(y_{i1}=1|x_i=x)>P(y_{i2}=1|x_i=x)\}\\
\Delta\mathcal{X}_2 &= \{\Delta x\in\mathbb{R}^k : \exists x=(x_1,x_1+\Delta x)\in\mathcal{X} \text{ such that } P(y_{i2}=1|x_i=x)>P(y_{i1}=1|x_i=x)\}\\
\Delta\mathcal{X}_3 &= \{\Delta x\in\mathbb{R}^k : \exists x=(x_1,x_1+\Delta x)\in\mathcal{X} \text{ such that}\\
&\qquad P(y_{i1}=0,y_{i2}=1|x_i=1)\geq P(y_{i1}=1|x_i=x) \text{ or } P(y_{i0}=1,y_{i1}=0|x_i=x)\geq P(y_{i2}=0|x_i=x)\}\\
\Delta\mathcal{X}_4 &= \{\Delta x\in\mathbb{R}^k : \exists x=(x_1,x_1+\Delta x)\in\mathcal{X} \text{ such that}\\
&\qquad P(y_{i1}=1,y_{i2}=0|x_i=x)\geq P(y_{i1}=0|x_i=x) \text{ or } P(y_{i0}=0,y_{i1}=1|x_i=x)\geq P(y_{i2}=1|x_i=x)\}\\
\Delta\mathcal{X}_5 &= \{\Delta x\in\mathbb{R}^k : \exists x=(x_1,x_1+\Delta x)\in\mathcal{X} \text{ such that } P(y_{i0}=1,y_{i1}=1|x_i=x)>P(y_{i2}=0|x_i=x)\}\\
\Delta\mathcal{X}_6 &= \{\Delta x\in\mathbb{R}^k : \exists x=(x_1,x_1+\Delta x)\in\mathcal{X} \text{ such that } P(y_{i0}=1,y_{i1}=1|x_i=x)>P(y_{i2}=1|x_i=x)\}\\
\Delta\mathcal{X}_7 &= \{\Delta x\in\mathbb{R}^k : \exists x=(x_1,x_1+\Delta x)\in\mathcal{X} \text{ such that}\\
&\qquad P(y_{i0}=0,y_{i1}=0|x_i=x)+P(y_{i1}=0,y_{i2}=1|x_i=x)>1\}\\
\Delta\mathcal{X}_8 &= \{\Delta x\in\mathbb{R}^k : \exists x=(x_1,x_1+\Delta x)\in\mathcal{X} \text{ such that}\\
&\qquad P(y_{i0}=1,y_{i1}=1|x_i=x)+P(y_{i1}=1,y_{i2}=0|x_i=x)>1\}\\
\Delta\mathcal{X}_9 &= \{\Delta x\in\mathbb{R}^k : \exists x=(x_1,x_1+\Delta x)\in\mathcal{X} \text{ such that}\\
&\qquad P(y_{i0}=1,y_{i1}=0|x_i=x)+P(y_{i1}=0,y_{i2}=1|x_i=x)>1\}\\
\Delta\mathcal{X}_{10} &= \{\Delta x\in\mathbb{R}^k : \exists x=(x_1,x_1+\Delta x)\in\mathcal{X} \text{ such that}\\
&\qquad P(y_{i0}=0,y_{i1}=1|x_i=x)+P(y_{i1}=1,y_{i2}=0|x_i=x)>1\}
\end{aligned}$$

Similarly, define the two sets that can be used to point identify β:

$$\begin{aligned}
\mathcal{X}_7 &= \{x\in\mathcal{X} \text{ such that } P(y_{i1}=1,y_{i2}=1)+P(y_{i0}=1,y_{i1}=0)\geq 1\}\\
\mathcal{X}_8 &= \{x\in\mathcal{X} \text{ such that } P(y_{i0}=1,y_{i1}=1|x_i=x)+P(y_{i1}=1,y_{i2}=0|x_i=x)\geq 1\}\\
\mathcal{X}_{7,8} &= \mathcal{X}_7\cup\mathcal{X}_8
\end{aligned}$$

Theorem 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 sets are large enough (as formalized in the assumption below), we will be able to identify β.

Assumption 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* $\Delta x_{-j} = x_{2,-j} - x_{1,-j}$ *where* $x_{2,-j}$ *denotes all the other components of* x_2 *besides the* j^{th} *one).*

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. This assumption is both necessary and sufficient for point identification of β: first, if $x_2 - x_1$ has a discrete support, then the number of inequalities that define the identified set is finite and does not have to shrink to a single point. For a simple set up of a semiparametric binary choice model of Manski (1985), Komarova (2013) illustrates what the identified set looks like when the distribution of covariates is discrete. Second, the inequalities that define the identified set are based on comparing $(x_2 - x_1)'\beta$ to 0, so for any $\tilde{\beta} \neq \beta$ we want to be able to find a point in the support such that the signs of $(x_2 - x_1)'\beta$ and $(x_2 - x_1)'\tilde{\beta}$ are different, which may be difficult without PID-STAT1 part of Assumption 4.

Under these assumptions we can attain point identification for β and the sign of γ, as stated in the following Theorem 3 that gives sufficient conditions for point identification of β and the sign of γ. Also, the Theorem 3 contains upper and lower bounds that can be used to obtain point identification of γ when it is positive.

Theorem 3. *The following hold.*

1. *Let Assumptions 1, 2, and 4 hold. Then* β *is point identified (up to scale).*
2. *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:

$$\begin{aligned} |\gamma| &\geq \max\{-m_1, M_2\} \\ \gamma &\leq \min\{m_9, -M_{10}\} \end{aligned} \tag{3}$$

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

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

.

Remarks: Some Implications of Theorem 3

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 4. In particular, if γ is positive, then we can replace $\mathcal{X}_7$ and $\mathcal{X}_8$ in Assumption 4 with $\mathcal{X}_3 \cup \mathcal{X}_7$ and $\mathcal{X}_4 \cup \mathcal{X}_8$, respectively, where

$\mathcal{X}_3 = \{x \in \mathcal{X}$ such that
$P(y_{i1} = 0, y_{i2} = 1|x_i = x) + P(y_{i0} = 1, y_{i1} = 0|x_i = x) \geq P(y_{i1} = 1|x_i = x) + P(y_{i2} = 0|x_i = x)\}$
$\mathcal{X}_4 = \{x \in \mathcal{X}$ such that
$P(y_{i0} = 1, y_{i1} = 0|x_i = x) + P(y_{i0} = 0, y_{i1} = 1|x_i = x) \geq P(y_{i1} = 0|x_i = x) + P(y_{i2} = 1|x_i = x)\}$

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

$$\begin{aligned} \mathcal{X}_5 &= \{x \in \mathcal{X} \text{ such that } P(y_{i0} = 1, y_{i1} = 1|x_i = x) \geq P(y_{i2} = 0|x_i = x)\} \\ \mathcal{X}_6 &= \{x \in \mathcal{X} \text{ such that } P(y_{i0} = 1, y_{i1} = 1|x_i = x) \geq P(y_{i2} = 1|x_i = x)\} \end{aligned}$$

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 $P(y_{i0} = 1, y_{i1} = 0|x_i = x = \tilde{x}) + P(y_{i1} = 0, y_{i2} = 1|x_i = x = \tilde{x}) \geq 1$, then $\gamma > 0$. Similarly, if there exists $\tilde{x} = (\tilde{x}_1, \tilde{x}_2 = \tilde{x}_1)'$ such that $P(y_{i0} = 0, y_{i1} = 1|x_i = x = \tilde{x}) + P(y_{i1} = 1, y_{i2} = 0|x_i = x = \tilde{x}) \geq 1$, then $\gamma < 0$.

Finally, note that from the inequalities in (3) it is possible that γ can be point identified

when it is non negative given enough variation on the support of the regressors. Heuristically, when γ is non negative, if the lhs and the rhs of these inequalities are the same for some x, then γ is point identified. On the other hand, the situation is not the same when γ is negative. This is so because inequalities in (3) place only an upper bound on γ when it is negative. Indeed, looking at the identified set in Theorem 1, we see that the inequalities do not place a *lower bound* for γ when it is negative when $T = 2$.

3.1.1 Point Identification of θ with T=2 under Conditional Exogeneity

Under Assumption 3 the identified set for θ is a subset of the one derived under Assumption 1. So, the results stated in Theorem 3 on point identification apply also for the case when $T = 2$ under conditional exogeneity of y_0 (with a weaker set of restrictions on the support of x). Also, the additional inequalities that constitute Θ_1 do not place a lower bound on γ when γ is negative[8]. So, under either Assumption 1 and 3, γ can be point identified when $T = 2$ if it is non negative. The situation is different when $T > 2$. We illustrate with the case where $T = 3$ next.

3.2 Point Identification with $T = 3$

Next, we provide sufficient conditions for point identification of the parameters β, γ under *only* the stationarity assumption in the case for $T = 3$. With 3 time periods, in addition to the ten restrictions defined in the previous section, we get two more sets of restrictions. Specifically, for $t, s \in \{1, 3\}$ (again all probabilities are conditional on $x_i = x$):

(1) If $P(y_{i3} = 1) \geq P(y_{i1} = 1)$, then $(x_3 - x_1)'\beta + |\gamma| \geq 0$;

(2) If $P(y_{i1} = 1) \geq P(y_{i3} = 1)$, then $(x_3 - x_1)'\beta - |\gamma| \leq 0$;

(3) If $P(y_{i0} = 0, y_{i1} = 0) \geq P(y_{i3} = 0)$ or $P(y_{i2} = 1, y_{i3} = 1) \geq P(y_{i1} = 1)$, then $(x_3 - x_1)'\beta + \max\{0, \gamma\} \geq 0$;

(4) If $P(y_{i0} = 1, y_{i1} = 1) \geq P(y_{i3} = 1)$ or $P(y_{i2} = 0, y_{i3} = 0) \geq P(y_{i1} = 0)$, then $(x_3 - x_1)'\beta - \max\{0, \gamma\} \leq 0$;

[8]Note for example that inequality (ν) in Definition 2 which can be used to bound γ from below (when $d_0 = 0$) will never hold as when $t = 2$, the lhs $P(y_{i1} = 1)$ is never less than $P(y_{i1} = 1; y_{i2} = 1|d_0)$.

(5) If $P(y_{i0} = 1, y_{i1} = 0) \geq P(y_{i3} = 0)$ or $P(y_{i2} = 0, y_{i3} = 1) \geq P(y_{i1} = 1)$, then $(x_3 - x_1)'\beta - \min\{0, \gamma\} \geq 0$;

(6) If $P(y_{i0} = 0, y_{i1} = 1) \geq P(y_{i3} = 1)$ or $P(y_{i2} = 1, y_{i3} = 0) \geq P(y_{i1} = 0)$, then $(x_3 - x_1)'\beta + \min\{0, \gamma\} \leq 0$;

(7) If $P(y_{i2} = 1, y_{i3} = 1) + P(y_{i0} = 1, y_{i1} = 0) \geq 1$, then $(x_3 - x_1)'\beta \geq 0$;

(8) If $P(y_{i2} = 1, y_{i3} = 0) + P(y_{i0} = 1, y_{i1} = 1) \geq 1$, then $(x_3 - x_1)'\beta \leq 0$;

(9) If $P(y_{i2} = 0, y_{i3} = 1) + P(y_{i0} = 1, y_{i1} = 0) \geq 1$, then $(x_3 - x_1)'\beta - \gamma \geq 0$;

(10) If $P(y_{i2} = 1, y_{i3} = 0) + P(y_{i0} = 0, y_{i1} = 1) \geq 1$, then $(x_3 - x_1)'\beta + \gamma \leq 0$;

(11) If $P(y_{i2} = 0, y_{i3} = 0) + P(y_{i0} = 1, y_{i1} = 1) \geq 1$, then $(x_3 - x_1)'\beta - \gamma \leq 0$;

(12) If $P(y_{i2} = 1, y_{i3} = 1) + P(y_{i0} = 0, y_{i1} = 0) \geq 1$, then $(x_3 - x_1)'\beta + \gamma \geq 0$.

and for $t, s \in \{2, 3\}$:

(1) If $P(y_{i3} = 1) \geq P(y_{i2} = 1)$, then $(x_3 - x_2)'\beta + |\gamma| \geq 0$;

(2) If $P(y_{i2} = 1) \geq P(y_{i3} = 1)$, then $(x_3 - x_2)'\beta - |\gamma| \leq 0$;

(3) If $P(y_{i2} = 0, y_{i3} = 1) + P(y_{i0} = 1, y_{i2} = 0) \geq P(y_{i2} = 1|x_i = x) + P(y_{i3} = 0)$, then $(x_3 - x_2)'\beta - \min\{0, \gamma\} \geq 0$;

(4) If $P(y_{i2} = 1, y_{i3} = 0) + P(y_{i0} = 0, y_{i2} = 1) \geq P(y_{i2} = 0) + P(y_{i3} = 1)$, then $(x_3 - x_2)'\beta + \min\{0, \gamma\} \leq 0$;

(5) If $P(y_{i0} = 0, y_{i2} = 0) \geq P(y_{i3} = 0)$, then $(x_3 - x_2)'\beta + \max\{0, \gamma\} \geq 0$;

(6) If $P(y_{i0} = 1, y_{i2} = 1) \geq P(y_{i3} = 1)$, then $(x_3 - x_2)'\beta - \max\{0, \gamma\} \leq 0$;

(7) If $P(y_{i0} = 0, y_{i2} = 0) + P(y_{i2} = 0, y_{i3} = 1) \geq 1$, then $(x_3 - x_2)'\beta \geq 0$;

(8) If $P(y_{i0} = 1, y_{i2} = 1) + P(y_{i2} = 1, y_{i3} = 0) \geq 1$, then $(x_3 - x_2)'\beta \leq 0$;

(9) If $P(y_{i0} = 1, y_{i2} = 0) + P(y_{i2} = 0, y_{i3} = 1) \geq 1$, then $(x_3 - x_2)'\beta - \gamma \geq 0$;

(10) If $P(y_{i0} = 0, y_{i2} = 1) + P(y_{i2} = 1, y_{i3} = 0) \geq 1$, then $(x_3 - x_2)'\beta + \gamma \leq 0$

Theorem 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

$$\mathcal{X}_7^{\{1,2\}} = \{x \in \mathcal{X} \text{ such that } P(y_{i0} = 0, y_{i1} = 0 | x_i = x) + P(y_{i1} = 0, y_{i2} = 1 | x_i = x) \geq 1\}$$
$$\mathcal{X}_8^{\{1,2\}} = \{x \in \mathcal{X} \text{ such that } P(y_{i0} = 1, y_{i1} = 1 | x_i = x) + P(y_{i1} = 1, y_{i2} = 0 | x_i = x) \geq 1\}$$
$$\mathcal{X}_7^{\{1,3\}} = \{x \in \mathcal{X} \text{ such that } P(y_{i0} = 1, y_{i1} = 0 | x_i = x) + P(y_{i2} = 1, y_{i3} = 1 | x_i = x) \geq 1\}$$
$$\mathcal{X}_8^{\{1,3\}} = \{x \in \mathcal{X} \text{ such that } P(y_{i0} = 1, y_{i1} = 1 | x_i = x) + P(y_{i2} = 1, y_{i3} = 0 | x_i = x) \geq 1\}$$
$$\mathcal{X}_7^{\{2,3\}} = \{x \in \mathcal{X} \text{ such that } P(y_{i1} = 0, y_{i2} = 0 | x_i = x) + P(y_{i2} = 0, y_{i3} = 1 | x_i = x) \geq 1\}$$
$$\mathcal{X}_8^{\{2,3\}} = \{x \in \mathcal{X} \text{ such that } P(y_{i1} = 1, y_{i2} = 1 | x_i = x) + P(y_{i2} = 1, y_{i3} = 0 | x_i = x) \geq 1\}$$
$$\Delta\mathcal{X}_{7,8}^{\{1,2\}} = \{\Delta x = x_2 - x_1 : x \in \mathcal{X}_7^{\{1,2\}} \cup \mathcal{X}_8^{\{1,2\}}\}$$
$$\Delta\mathcal{X}_{7,8}^{\{1,3\}} = \{\Delta x = x_3 - x_1 : x \in \mathcal{X}_7^{\{1,3\}} \cup \mathcal{X}_8^{\{1,3\}}\}$$
$$\Delta\mathcal{X}_{7,8}^{\{2,3\}} = \{\Delta x = x_3 - x_2 : x \in \mathcal{X}_7^{\{2,3\}} \cup \mathcal{X}_8^{\{2,3\}}\}$$

and make the following assumption:

Assumption 5. *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 4) as the set on which for example the full rank condition must hold is not richer. Similar to the $T = 2$ case, β is point identified if Assumption 5 holds.

We will not present here identification results for the sign of γ (again, these are very similar to Theorem 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, we can bound γ both from above and from below. In particular, we

now have the following restrictions on γ when β is point identified:

$$\begin{aligned} |\gamma| &\geq \max\{-m_1^{\{1,2\}}, -m_1^{\{1,3\}}, -m_1^{\{2,3\}}, M_2^{\{1,2\}}, M_2^{\{1,3\}}, M_2^{\{2,3\}}\} \\ \gamma &\leq \min\{m_9^{\{1,2\}}, m_9^{\{1,3\}}, m_9^{\{2,3\}}, -M_{10}^{\{1,2\}}, -M_1^{\{1,3\}}, -M_{10}^{\{2,3\}}\} \\ \gamma &\geq \max\{M_{11}^{\{1,3\}}, -m_{12}^{\{2,3\}}\} \end{aligned} \tag{4}$$

where $m_j^{\{t,s\}}$ and $M_j^{\{t,s\}}$ are defined similar to Theorem 3.

In comparison to the $T = 2$ case in Theorem 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. Enough variation in the support of the covariates could lead to point identification of γ. Point identification of γ was attained in Honoré and Kyriazidou (2000) but under much stronger restrictions, such as serial independence in u_{it} and strict overlap in the support of x_{it}, which as we mentioned previously, rules out time trends.

3.3 Identification in a Model without Covariates

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

$$y_{it} = I\{u_{it} \leq \gamma y_{i,t-1} + \alpha_i\} \quad t = 1, 2, ...T$$

Although the scale of γ cannot be identified in this model (without covariates), 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 1. With $T = 2$, the identified set in Theorem 1 is given by these inequalities (without covariates where we eliminated ones where γ does not show up):

(1) If $P(y_{i2} = 1) \geq P(y_{i1} = 1)$, then $|\gamma| \geq 0$;

(2) If $P(y_{i1} = 1) \geq P(y_{i2} = 1)$, then $-|\gamma| \leq 0$;

(3) If $P(y_{i1} = 0, y_{i2} = 1) + P(y_{i0} = 1, y_{i1} = 0) \geq P(y_{i1} = 1) + P(y_{i2} = 0)$, then $-\min\{0, \gamma\} \geq 0$;

(4) If $P(y_{i1}=1, y_{i2}=0)+P(y_{i0}=0, y_{i1}=1) \geq P(y_{i1}=0)+P(y_{i2}=1)$, then $\min\{0,\gamma\} \leq 0$;

(5) If $P(y_{i0}=0, y_{i1}=0) \geq P(y_{i2}=0)$, then $\max\{0,\gamma\} \geq 0$;

(6) If $P(y_{i0}=1, y_{i1}=1) \geq P(y_{i2}=1)$, then $\max\{0,\gamma\} \leq 0$;

(7) If $P(y_{i0}=1, y_{i1}=0) + P(y_{i1}=0, y_{i2}=1) \geq 1$, then $-\gamma \geq 0$;

(8) If $P(y_{i0}=0, y_{i1}=1) + P(y_{i1}=1, y_{i2}=0) \geq 1$, then $\gamma \leq 0$

In the absence of covariates, any inequalities above that involve $|\gamma|$, $\max\{0,\gamma\}$ and $\min\{0,\gamma\}$ are trivial. But the last two inequalities here allow us (sometimes, if a particular relationship between conditional probabilities of certain events holds) to tell if γ is negative. Specifically, we summarize our results in the corralary below.

Corollary 1. *Under the conditions of Theorem 1 (with $T=2$), if*

$$\max\left(P(y_0=1, y_1=0) + P(y_1=0, y_2=1),\ P(y_0=0, y_1=1) + P(y_1=1, y_2=0)\right) \geq 1$$

then γ is negative.

When $T=3$, it is possible to identify the sign of γ when γ is positive (unlike in $T=2$ case) so when $T=3$ it is possible to identify the sign of γ. In the absence of covariates, an extra time period in Theorem 1 adds nontrivial restrictions on γ to what we have above with $T=2$ (again, any restrictions involving $|\gamma|$, $\max\{0,\gamma\}$ and $\min\{0,\gamma\}$ are trivial and therefore non-informative without conditioning covariates). We state our results for the identification of the sign of γ in the case with $T=3$ in the following Corollary.

Corollary 2. *Let the conditions of Theorem 1 hold. Let $T=3$. Then, the following holds.*

$$\begin{aligned}
&\textit{If } P(y_0=1, y_1=0) + P(y_2=0, y_3=1) \geq 1, \textit{ then } \gamma \leq 0 \\
&\textit{If } P(y_0=0, y_1=1) + P(y_2=1, y_3=0) \geq 1, \textit{ then } \gamma \leq 0 \\
&\textit{If } P(y_0=1, y_1=1) + P(y_2=0, y_3=0) \geq 1, \textit{ then } \gamma \geq 0 \\
&\textit{If } P(y_0=0, y_1=0) + P(y_2=1, y_3=1) \geq 1, \textit{ then } \gamma \geq 0 \\
&\textit{If } P(y_1=1, y_2=0) + P(y_2=0, y_3=1) \geq 1, \textit{ then } \gamma \leq 0 \\
&\textit{If } P(y_1=0, y_2=1) + P(y_2=1, y_3=0) \geq 1, \textit{ then } \gamma \leq 0
\end{aligned} \tag{5}$$

4 Inference

Though the main contribution of the paper is the characterization of the identified sets in dynamic discrete choice models under weak assumptions, we suggest an approach to conduct inference that is computationally attractive under the assumption that the regressor vector x has finite support. This inference approach leads to a confidence region for the identified set. The idea is to first construct a confidence region for the choice probabilities, which is a vector of multinomial probabilities. Then, heuristically, a confidence region for the identified set can be constructed by using draws from this confidence region for the choice probabilities. The mechanics of this exercise is computationally simple as it exploits the linear (in (β, γ)) nature of the inequalities and so linear programs can be used to check whether a particular parameter vector θ belongs to the identified set. We describe this procedure in more details next. In the discussion below, we focus on $T = 2$ for simplicity of exposure.

4.1 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 (here we illustrate this for the case where we condition on y_0 wlog). 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

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

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

$$\begin{aligned} CS^p_{1-\alpha} = \Big\{ &\vec{p}(y_0,x) \geq 0 : \text{for all } (y_0,x), p_2(0,0|y_0,x) + p_2(0,1|y_0,x) + p_2(1,0|y_0,x) \leq 1 \\ &\text{and } T_n^{as}(\vec{p}(\cdot)) \leq c^*_{1-\alpha}(\vec{p}(\cdot)) \Big\} \end{aligned} \tag{6}$$

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 simple way to obtain a draw from this confidence region is to use the weighted bootstrap via a posterior distribution for these choice probabilities[9]. This can be done by exploiting properties of the interplay between the gamma distribution and the dirichlet priors. See for example Kline and Tamer (2016) for details on how to implement this.

4.2 Confidence Region for θ

We illustrate here how we map the confidence region for the choice probabilities to a confidence region for the identified set. We will be using the $T = 2$ case with the strict exogeneity of initial conditions to illustrate our approach and use the following 4 inequalities to illustrate:

[9] Another approach that is easy to compute is to use a sup test to contruct a rectangular CI. These rectangles are such $P(p \in [a,b]) = P(p_1 \in [a_1, b_1], \ldots, p_J \in [a_J, b_J]) \to 1-\alpha$. A simple way to obtain these a, b's is to simulate a cutoff for the *sup* statistic $\|\mathcal{N}(0, U^{-1/2}\Sigma U^{-1/2})\|_\infty$ and use rectangles of the form $[a_i, b_i] = [\hat{p}_i \pm \hat{c}\sqrt{\hat{\Sigma}_{ii}/n}]$. We find that getting draws from the posterior via a Bayesian Bootstrap to be simpler to implement in this case.

$$
\begin{aligned}
p_1(1|y_0,x) \geq P(y_2=1|y_0,x) &\Rightarrow \Delta x'\beta + \min\{0,\gamma\} - \gamma y_0 \leq 0 \\
p_1(1|y_0,x) \leq P(y_2=1|y_0,x) &\Rightarrow \Delta x'\beta + \max\{0,\gamma\} - \gamma y_0 \geq 0 \\
p_2(0,1|y_0,x) \geq p_1(1|y_0,x) &\Rightarrow \Delta x'\beta - \gamma y_0 \geq 0 \\
p_2(1,0|y_0,x) \geq p_1(0|y_0,x) &\Rightarrow \Delta x'\beta + \gamma(1-y_0) \geq 0
\end{aligned} \tag{7}
$$

where $\Delta x = x_2 - x_1$ and where at least one strict inequality on the left-hand side implies a strict inequality on the right-hand side.

Generally, a confidence set for the partially identified θ can be constructed based on the chi-squared approximation above as follows:

$$
CS^{\theta}_{1-\alpha} = \{\theta \in \Theta : \text{conditions (7) hold for some } \vec{p}(\cdot) \in CS^{p}_{1-\alpha}\}
$$

where $CS^{p}_{1-\alpha}$ is defined in (6) 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 (7).

Conditions (7) are straightforward to verify for a given $\vec{p}(\cdot) \in CS^{p}_{1-\alpha}$. The following algorithm builds a confidence region for θ based on a linear program[10] (a similar approach is used in Honoré and Tamer (2006)).

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

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

$$
\max_{(\gamma,\beta)\in\Theta} c
$$

subject to

[10]Since the identified set is one where the set of contraints are feasible, we augment these constraints with a an objective of maximizing "'c"' which is just a scalar and is not useful. Essentially, for a given set of parameters, if the LP is *feasible* then the parameter belongs to the identified set and so the optimal value of c is not relevant.

$$
\begin{cases}
\gamma \geq 0 \\
1\{p_{1,(k)}(1|y_0^1, x^1) \geq P_{(k)}(y_2 = 1|y_0^1, x^1)\}(-\Delta x^{1'}\beta + \gamma y_0^1) \geq 0 \\
1\{p_{1,(k)}(1|y_0^1, x^1) \leq P_{(k)}(y_2 = 1|y_0^1, x^1)\}(\Delta x^{1'}\beta + \gamma(1 - y_0^1)) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^1, x^1) \geq p_{1,(k)}(1|y_0^1, x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^1, x^1) \leq p_{1,(k)}(1|y_0^1, x^1)\}(\Delta x^{1'}\beta + \gamma(1 - y_0^1)) \geq 0 \\
1\{p_{1,(k)}(1|y_0^2, x^2) \geq P_{(k)}(y_2 = 1|y_0^2, x^2)\}(-\Delta x^{2'}\beta + \gamma y_0^2) \geq 0 \\
1\{p_{1,(k)}(1|y_0^2, x^2) \leq P_{(k)}(y_2 = 1|y_0^2, x^2)\}(\Delta x^{2'}\beta + \gamma(1 - y_0^2)) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^2, x^2) \geq p_{1,(k)}(1|y_0^2, x^2)\}(\Delta x^{2'}\beta - \gamma y_0^2) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^2, x^2) \leq p_{1,(k)}(1|y_0^2, x^2)\}(\Delta x^{2'}\beta + \gamma(1 - y_0^2)) \geq 0 \\
\dots \\
1\{p_{1,(k)}(1|y_0^J, x^J) \geq P_{(k)}(y_2 = 1|y_0^J, x^J)\}(-\Delta x^{J'}\beta + \gamma y_0^J) \geq 0 \\
1\{p_{1,(k)}(1|y_0^J, x^J) \leq P_{(k)}(y_2 = 1|y_0^J, x^J)\}(\Delta x^{J'}\beta + \gamma(1 - y_0^J)) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^J, x^J) \geq p_{1,(k)}(1|y_0^J, x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \geq 0 \\
1\{p_{2,(k)}(0, 1|y_0^J, x^J) \leq p_{1,(k)}(1|y_0^J, x^J)\}(\Delta x^{J'}\beta + \gamma(1 - y_0^J)) \geq 0
\end{cases}
\tag{8}
$$

(3) Similarly, get $\Theta_{(k)}^{-}$, the set of parameters θ that solve

$$\max_{(\gamma,\beta)\in\Theta} c$$

subject to

$$
\begin{cases}
\gamma \leq 0 \\
1\{p_{1,(k)}(1|y_0^1, x^1) \geq P_{(k)}(y_2 = 1|y_0^1, x^1)\}(-\Delta x^{1'}\beta + \gamma(y_0^1 - 1)) \geq 0 \\
1\{p_{1,(k)}(1|y_0^1, x^1) \leq P_{(k)}(y_2 = 1|y_0^1, x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \geq 0 \\
1\{p_{2,(k)}(0,1|y_0^1, x^1) \geq p_{1,(k)}(1|y_0^1, x^1)\}(\Delta x^{1'}\beta - \gamma y_0^1) \geq 0 \\
1\{p_{2,(k)}(0,1|y_0^1, x^1) \leq p_{1,(k)}(1|y_0^1, x^1)\}(\Delta x^{1'}\beta + \gamma(1 - y_0^1)) \geq 0 \\
1\{p_{1,(k)}(1|y_0^2, x^2) \geq P_{(k)}(y_2 = 1|y_0^2, x^2)\}(-\Delta x^{2'}\beta + \gamma(y_0^2 - 1)) \geq 0 \\
1\{p_{1,(k)}(1|y_0^2, x^2) \leq P_{(k)}(y_2 = 1|y_0^2, x^2)\}(\Delta x^{2'}\beta - \gamma y_0^2) \geq 0 \\
1\{p_{2,(k)}(0,1|y_0^2, x^2) \geq p_{1,(k)}(1|y_0^2, x^2)\}(\Delta x^{2'}\beta - \gamma y_0^2) \geq 0 \\
1\{p_{2,(k)}(0,1|y_0^2, x^2) \leq p_{1,(k)}(1|y_0^2, x^2)\}(\Delta x^{2'}\beta + \gamma(1 - y_0^2)) \geq 0 \\
\dots \\
1\{p_{1,(k)}(1|y_0^J, x^J) \geq P_{(k)}(y_2 = 1|y_0^J, x^J)\}(-\Delta x^{J'}\beta + \gamma(y_0^J - 1)) \geq 0 \\
1\{p_{1,(k)}(1|y_0^J, x^J) \leq P_{(k)}(y_2 = 1|y_0^J, x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \geq 0 \\
1\{p_{2,(k)}(0,1|y_0^J, x^J) \geq p_{1,(k)}(1|y_0^J, x^J)\}(\Delta x^{J'}\beta - \gamma y_0^J) \geq 0 \\
1\{p_{2,(k)}(0,1|y_0^J, x^J) \leq p_{1,(k)}(1|y_0^J, x^J)\}(\Delta x^{J'}\beta + \gamma(1 - y_0^J)) \geq 0
\end{cases}
\tag{9}
$$

(4) Repeat the above M times

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

The computationally tedious part in the above linear program is part (2) which builds the set of all θ's for which the linear program is feasible. One approach for this is to get a grid for θ and check whether each point on this grid is feasible. Checking feasibility is simple even when J is large. 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.

A computationally simple approach for linear functionals of $\theta = (\beta', \gamma)'$:

Suppose that one is interested in the linear functional $a'\theta$ where $\theta = (\beta', \gamma)'$. So, for example, if a is a column of zeros except for the last entry, then $a'\theta = \gamma$. Then, we can replace the objective function in (2) and (3) above with

$$\min_{\theta} / \max \quad a'\theta$$

$$\text{subject to} \ldots$$

the same constraints as in (2) and (3) above. This simplifies the computations tremendously as no need for a grid search to construct the identified set corresponding to a given draw from the confidence region. The algorithm in this case would be: 1) take a draw from a CI for the choice probabilities, 2) solve linear program for min/max of $a'\theta$ which would be an interval, 3) repeat. One can then take as the CI the interval that contains 95% of all the intervals. We find that this method works very well in practice.

5 Simulation Results

In this section we perform a simulation study to explore the shapes and sizes for identified sets in simple designs. We view this as an important exercise since with weak assumptions it is possible that identified sets are the trivial ones (i.e. the model does not restrict the parameters in anyway). So, it is important to offer some evidence as to whether the stationary model in our paper contains any information. To that end, the sets of inequalities that define our identified sets are used in simple designs to construct the identified sets. Again, the designs are noteworthy as they give an idea of the size of the identified sets in short panels, such as $T = 2$ and $T = 3$, and in cases with time trends and time dummies.

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

- Regression coefficients on strictly exogenous variables (β) were generally easier to identify than the coefficient on the lagged binary dependent variable (γ), which was our measure of the persistence in the model.
- Allowing for the initial condition to be strictly exogenous adds information and hence in principle should reduce the size of the identified set.
- Increasing the richness of the support of the exogenous variables facilitates the identification.

- Increasing the length of the time series added informational content. This is evident from the fact that with larger T's we get more inequalities.
- The value of the parameters themselves could effect their identifiability. For example, a negative value of the persistence parameter made its identification more difficult.

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

$$y_{it} = I\{u_{it} \leq v_{it} + x_{it}\beta + \gamma y_{i,t-1} + \alpha_i\} \quad i = 1, 2, ...10000; \quad t = 0, 1, ...T \tag{10}$$

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, with $T = 2, 3$, varying support conditions on (v_{it}, x_{it}), and different values of 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 in a two dimensional plane. 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.

5.1 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 were two distinct designs- one with identical support in each time period, and the other with strictly nonoverlapping support- $x_{it} = t$, with $t = 1, 2$, i.e. a time trend. Recall that this typse of design (or any design with only discrete covariates) could not be handled by any existing methods. 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 1, 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 the trivial set. The same design is replicated in Figure 2 with $\gamma = -.5$.

Now, when v is normally distributed, the size of the identified set shrinks. This is illustrated for $\gamma = .5$ in Figure 3 and in Figure 4 with $\gamma = -.5$ with v is normally distributed with increasing variance. Notice here that in all the plots, β appears well identified relative to γ.

In Figure 5, 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. For instance, in Figure 7 we plot the case with a time trend and a normal covariate with increasing variance when $\gamma = .5$. We repeat the exercise

We also simulate the identified set for the $T = 2$ case for the model that conditions on y_0. In this case, the model is isomorphic to an "exchangeable model," a label we use in the Figures. In Figure 9, we start with the discrete regressors for $\gamma = .5$ and the next figure replicates it for $\gamma = -.5$. These should be compared to Figures 1 and 2 in the case without conditioning on the initial condition. In Figures 11 and 12 we plot the identified set when one of the regressors is normal with different variances. As we can see from these plots, the idenfied set for (β, γ) is very small.
In Figures 13 and 14 we plot the identified sets in the case with time trends when the other regressor is discrete and when it is normal respectively. Notice here that the identified set with a normal covariate is very tight.

5.2 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 15 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 Empirical Illustration

Here we illustrate our methods with the *Panel Data for Women's Labor Force Participation* used in Chay and Hyslop (2014), which contains data on N=5663 married women over $T = 5$ periods, where the periods are spaced four months apart[11]. 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:

$$newkids_{it} = \begin{cases} 0, kids_{it} = 0 \, or \, 1, \\ 1, kids_{it} > 1. \end{cases} \tag{11}$$

$$newlhinc_{it} = \begin{cases} 0, lhinc_{it} \leq Median\,(lhinc_t), \\ 1, lhinc_{it} > Median\,(lhinc_t). \end{cases} \tag{12}$$

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

[11]The data set can be downloaded from the website that accompanies Wooldridge (2010).

We compare our estimates of the model introduced in this paper 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

$$\sum_{i=1}^{n} 1[x_{i2} = x_{i3}]1[y_{i1} \neq y_{i2}] \times log\left(\frac{exp((x_{i1} - x_{i2})'b + g(y_{i0} - y_{i3}))^{y_{i1}}}{1 + exp((x_{i1} - x_{i2})'b + g(y_{i0} - y_{i3}))}\right) \tag{13}$$

We found that the above objective function was easy to optimize and is robust to starting values. The 95% CI for γ from this model (assumes that y_0 is strictly exogeneous in addition to assuming that the u_{it} is iid logistic) is approximately $[1.3, 3.4]$ which is large and positive meaning that being employed last period is highly predictive of being employed this period even controlling for unobserved time invariant fixed effects. This model provides support for state dependence of past employment.

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 (6) 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[12]. For each draw from this posterior, we solve the linear program for min / max of the scalar $a = l'(\gamma, \beta')'$. For example, the maginal CIs for γ would take $a = \gamma = (1, 0, 0) * (\gamma, \beta')'$ as the objective function to optimize using the linear program subject to the linear constraints. This allows us to get (marginal) CIs for every scalar component of the parameter vector[13]. 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

[12]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.

[13]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, one can add add a tolerance parameter t to each inequality (so now the inequalities are less than a positive t rather than less then 0 and this tolerance can be weighted by the standard error), and in the first pass through the linear program one can minimize t subject to the constraints that define the problem (optimizing over (γ, β)) to obtain a feasible tolerance t^*. Then, one can then fix the tolerance at t^* when computing the confidence set. In our data setup, our inequalities had a nonempty interior and so the linear program was feasible ($t^* = 0$).

intervals. This procedure is simple to compute even with many of inequalities.

	(1) reprobit	(2) reprobitci	(3) HK Logit FE	(4) Stationary FE - Tolerance =0
newkids	-0.139	-0.063	-.443	-.5
	(0.037)	(0.374)	(.214)	
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)	$[-.2, 3.2]$

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

Notice here that consistently across all the models, the γ coefficient appears positive. In the model only assuming stationarity, we fix the parameter on *newkids* to $-.5$ for normalization to match the point estimate from the HK FE model. Compared to HK's estimates, our CI for γ is wider on the lhs and covers zero while HK's does not. Note here that our model also provides a consistent estimator for the identified set in addition to a CI for this identidied set. Moreover, computing such CIs is computationally trivial since for every draw, it involves solving a linear program. Of course in more complicated models with lots of covariate values, the size of the linear program may increase. We leave exploring in more details this empirical model and its computational task for future work.

7 Conclusion

This paper analyzes the identification of slope parameters in panel binary response models with lagged dependent variables under minimal assumptions on the distribution of idiosyncratic error terms. In particular, we consider two versions of stationarity assumptions: with and without strictly exogenous initial conditions, and provide the identified set under these two restrictions without making any assumptions on the fixed effect. We show that the characterization yields the sharp set that can easily be characterized through a certain linear programming problem. Our identification approach is quite flexible in that it does not rely on conditioning on a sub-population for whom covariates do not change between time

periods (in contrast to Honoré and Kyriazidou (2000)), so it can cover dynamic binary response models with time trend and, in general, explanatory variables that grow over time.

In addition, we provide sufficient conditions for point identification in models with $T = 2$ and $T = 3$ time periods. Overall, our analysis highlights the interplay between the strength of the assumptions, the number of time periods and the support of the exogenous regressors.

The work here suggests many areas for future research. One such direction would be to establish identified sets for the same parameters under alternative sets of restrictions in the models. Two examples we already considered in companion work were based on the assumptions of *nonstationarity* and *independence*. In the former setting we allowed the distribution of the idiosyncratic error terms u_{it} to vary over time but imposed cross sectional homoskedasticity, resulting in a class of models non nested with the ones studied here. In the latter setting we considered a model with serially independent, identically distributed and homoskedastic error terms (with unknown distribution) and showed how much smaller the identified sets were for this class of models nested by the ones considered in this paper.

Another direction for future research would be to consider models with a more general time series structure, such as $AR(p)$ models where $p > 1$ is a known integer, and derive the identified set for the larger parameter $\theta_p \equiv (\beta, \gamma_1, ...\gamma_p)$ under our weak conditions of serial correlation and cross sectional heteroskedasticity. We leave this and conducting inference on this larger parameter set for future work.

References

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

Ahn, S. C., and P. Schmidt (1997): "Efficient Estimation of Dynamic Panel Data Models: Alternative Assumptions and Simplified Estimation," *Journal of Econometrics*, 76(1), 309 – 321.

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

Andersen, E. (1970): "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society Series B*, 32(2), 283–301.

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

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

Aristodemou, E. (2019): "Semiparametric Identification in Panel Data Discrete Response Models," *Journal of Econometrics*, forthcoming.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp identification regions in models with convex moment predictions," *Econometrica*, 79(6), 1785–1821.

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

Chay, K. Y., and D. R. Hyslop (2014): "Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence Using Alternative Approaches," in *Safety Nets and Benefit Dependence*, ed. by S. Carcillo, H. Immervoll, S. P. Jenkins, S. Knigs, and K. Tatsiram, vol. 39 of *Research in Labor Economics*, pp. 1–39. Emerald Publishing Ltd.

Chen, S., S. Khan, and X. Tang (2019): "Exclusion Restrictions in Dynamic Binary Choice Panel Data Models: Comment on Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors," *Econometrica*, 87, 1781–1785.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81(2), 535–580.

Chesher, A., and A. M. Rosen (2017): "Generalized instrumental variable models," *Econometrica*, 85(3), 959–989.

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.

Gao, W. Y., and M. Li (2019): "Robust Semiparametric Estimation in Panel Multinomial Choice Models," *Available at SSRN 3282293*.

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 L. Hu (2020): "Selection without Exclusion," forthcoming, *Econometrica*.

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.

Honoré, B., and M. Weidner (2020): "Moment Conditions for Dynamic Panel Logit Models with Fixed Effects," Working Paper.

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

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

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

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

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.

Kitazawa, Y. (2021): "Exploration of dynamic fixed effects logit models from a traditional angle," *Journal of Econometrics*, forthcoming.

Kline, B., and E. Tamer (2016): "Bayesian inference in a class of partially identified models," *Quantitative Economics*, 7(2), 329–366.

Komarova, T. (2013): "Binary choice models with discrete regressors: Identification and misspecification," *Journal of Econometrics*, 177(1), 14–33.

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.

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

Pakes, A., J. Porter, M. Shepard, and S. Wang (2019): "Fixed Effects, State Dependence, and Health Insurance Choice," 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. (2019): "Nonparametric inference on state dependence with applications to employment dynamics," *Econometrica*, 87, 1475–1505.

Wooldridge, J. (2010): *Econometric analysis of cross section and panel data*. MIT press.

A Figures

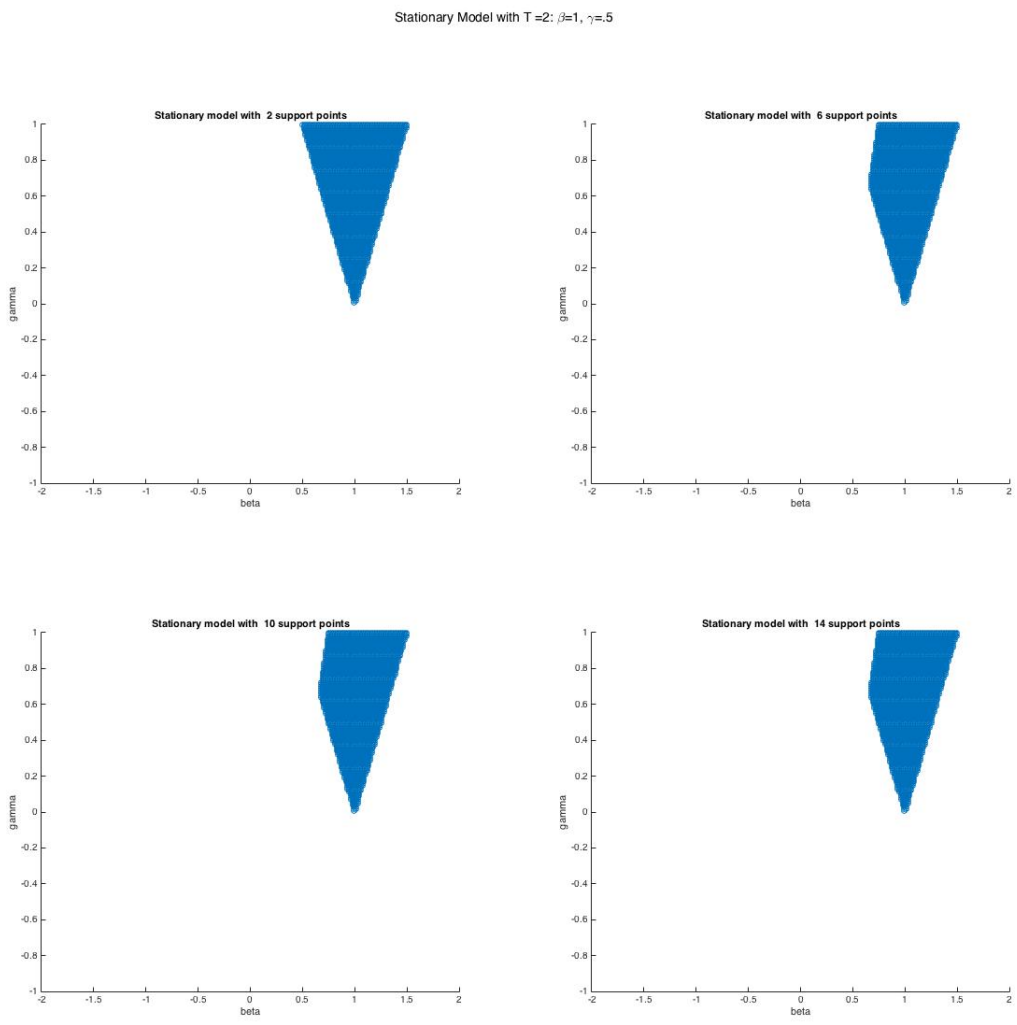

Figure 1: Stationary with $T = 2$ and Discrete Support with $\gamma = .5$

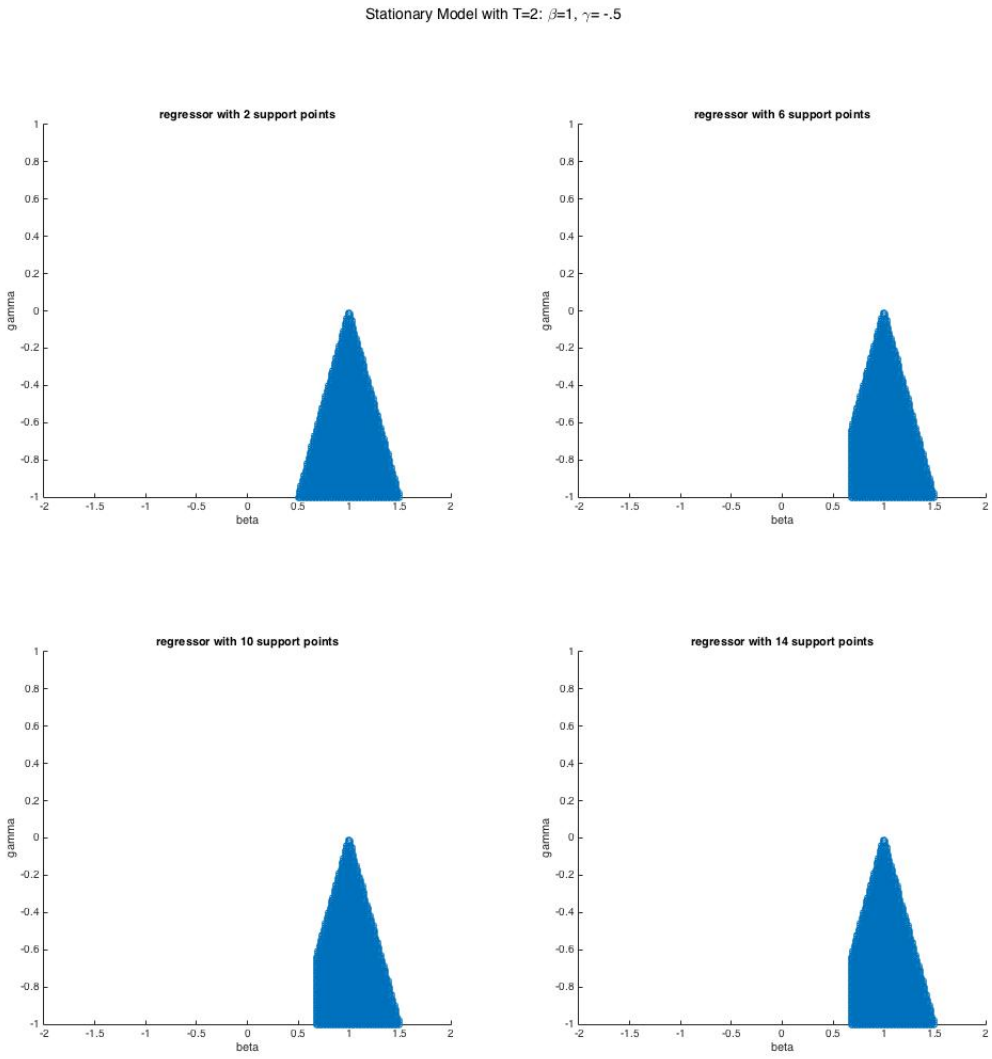

Figure 2: Stationary with $T = 2$ and Discrete Support with $\gamma = -.5$

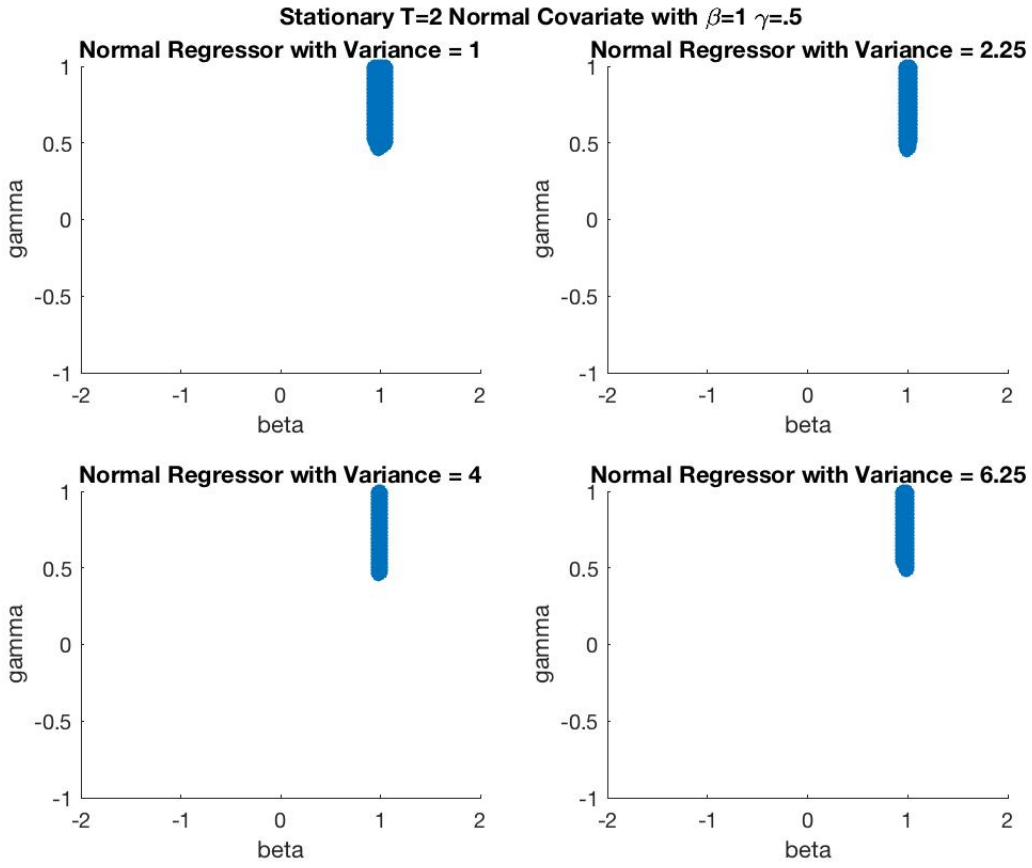

Figure 3: Stationary with $T = 2$ and Normal v with $\gamma = .5$

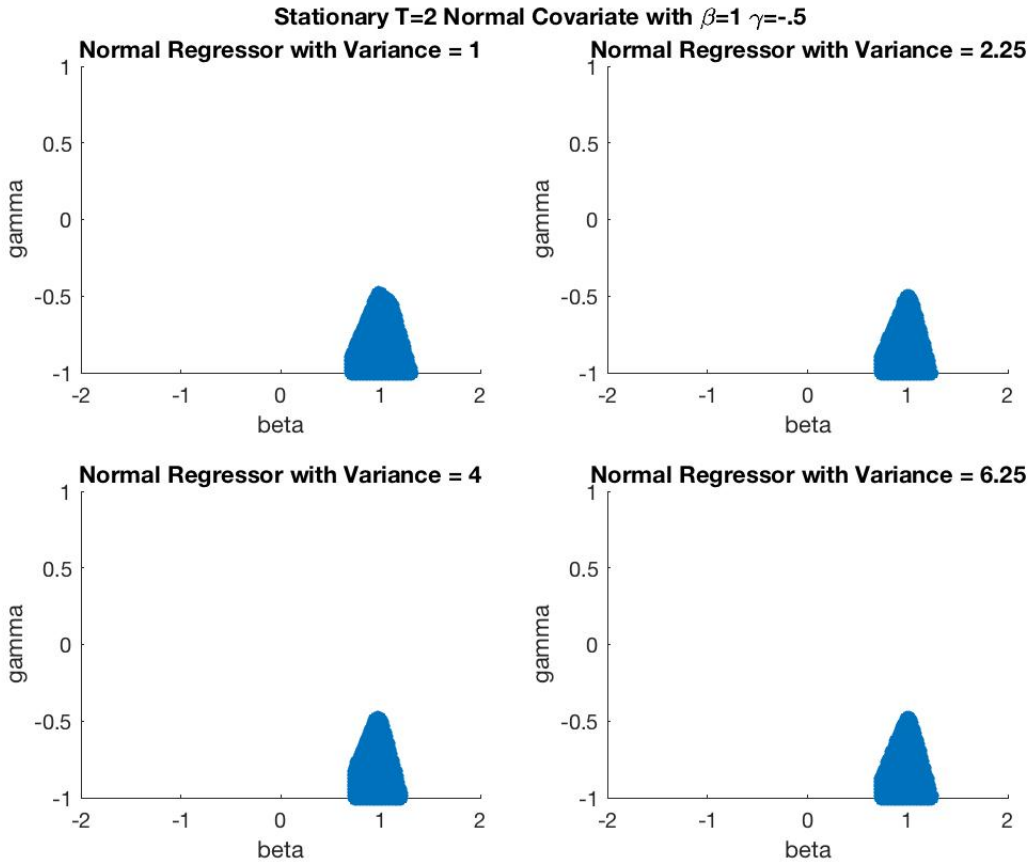

Figure 4: Stationary with $T = 2$ and Normal v with $\gamma = -.5$

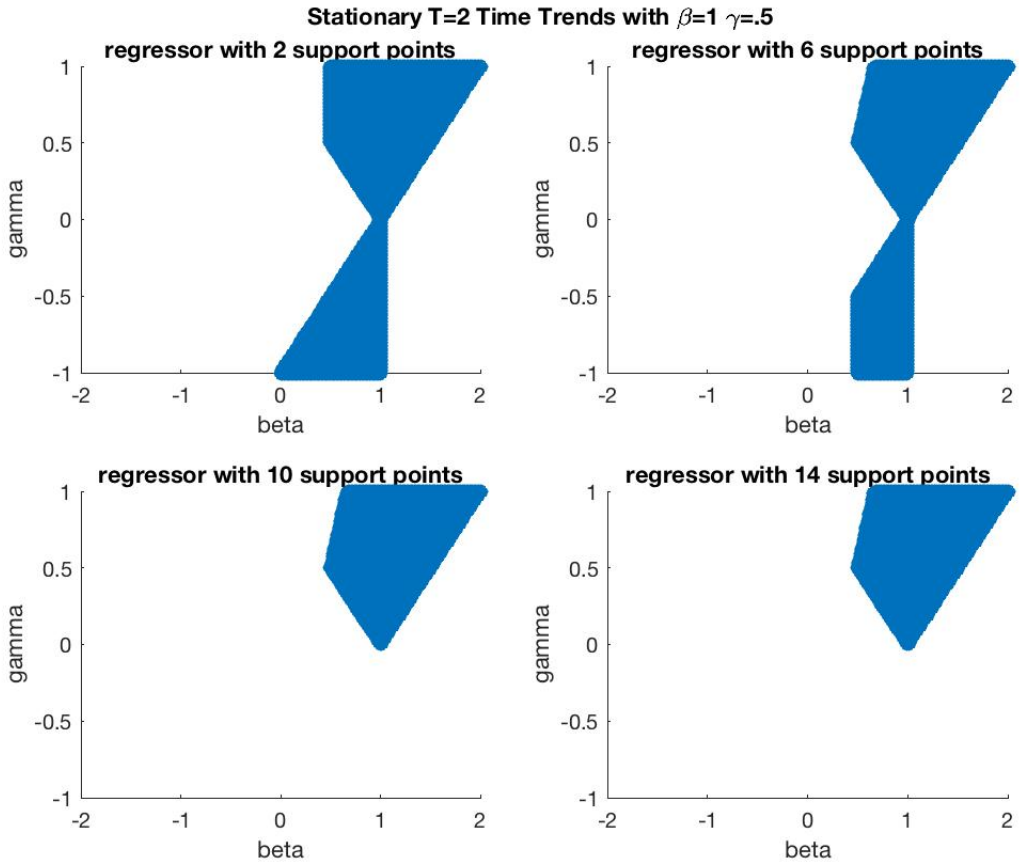

Figure 5: Stationary with $T = 2$ and Time Trend and Discrete Support for v with $\gamma = .5$

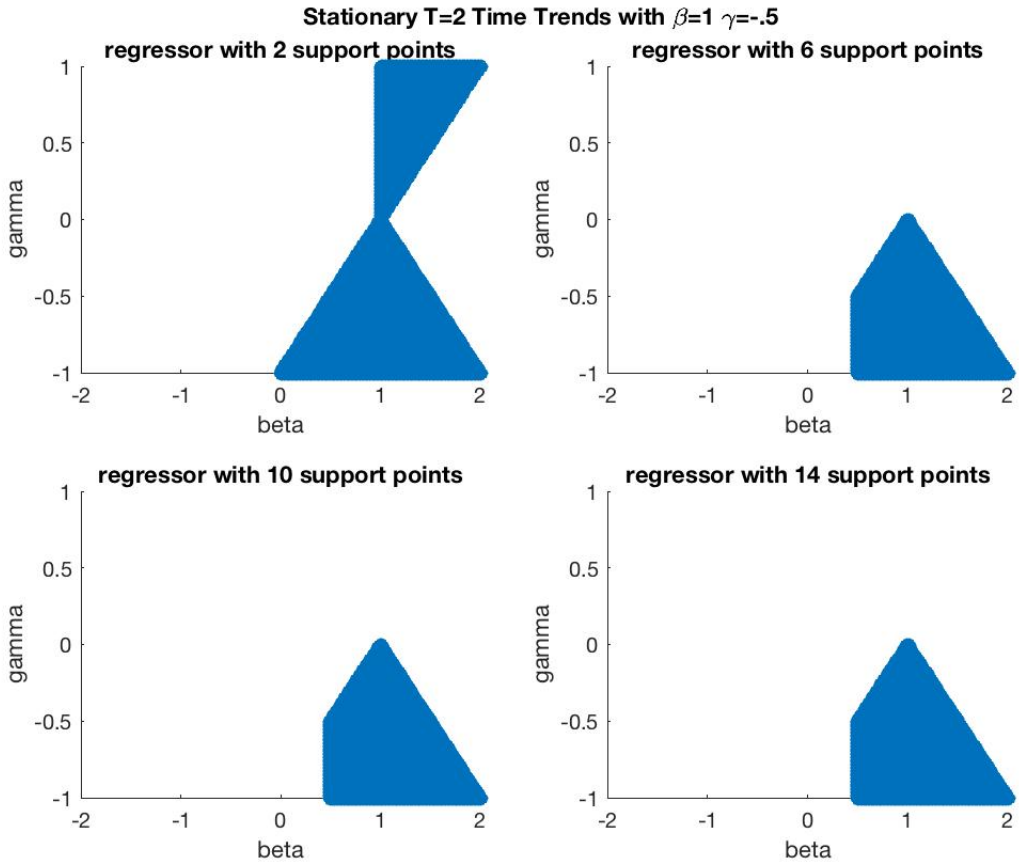

Figure 6: Stationary with $T = 2$ and Time Trend and Discrete Support for v with $\gamma = -.5$

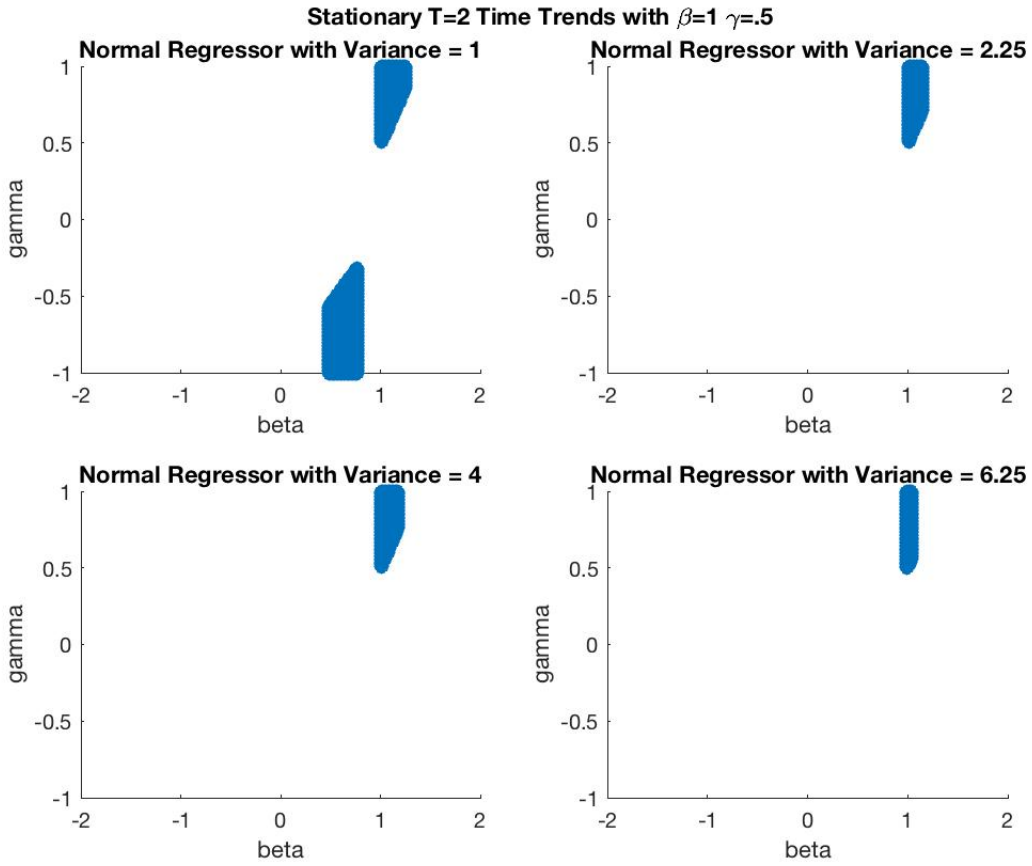

Figure 7: Stationary with $T = 2$ and Time Trend and Normal v with $\gamma = .5$

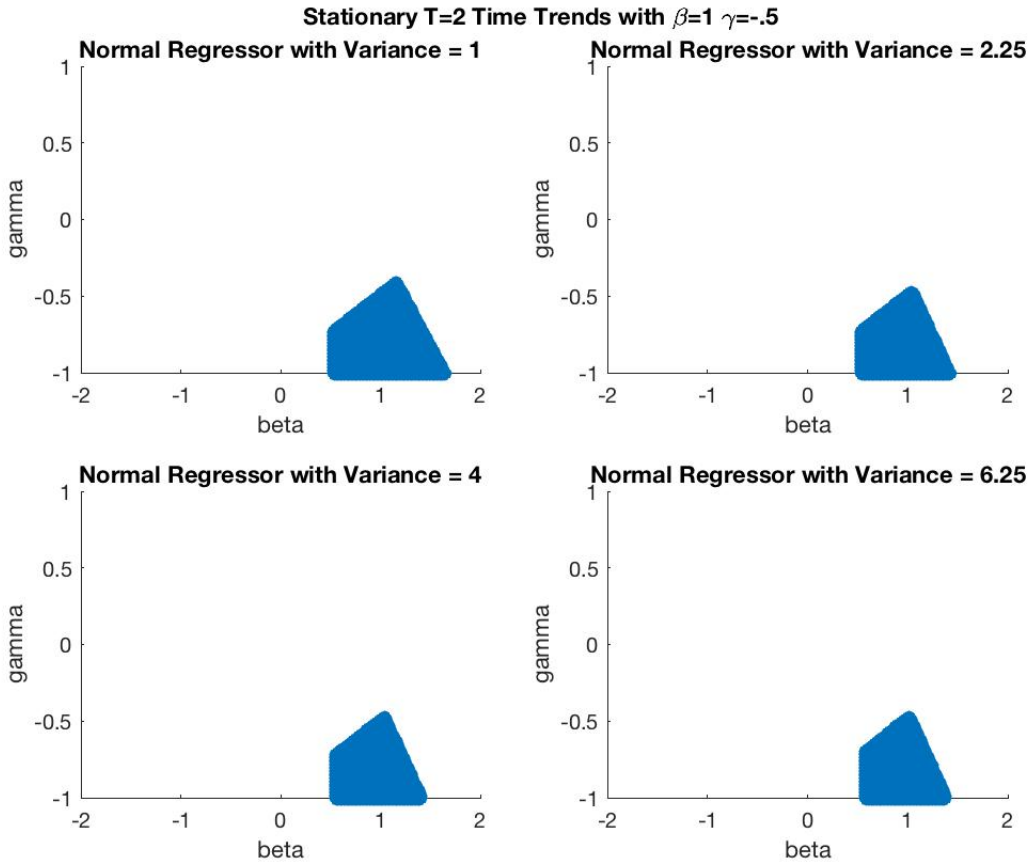

Figure 8: Stationary with $T = 2$ and Time Trend and Normal v with $\gamma = -.5$

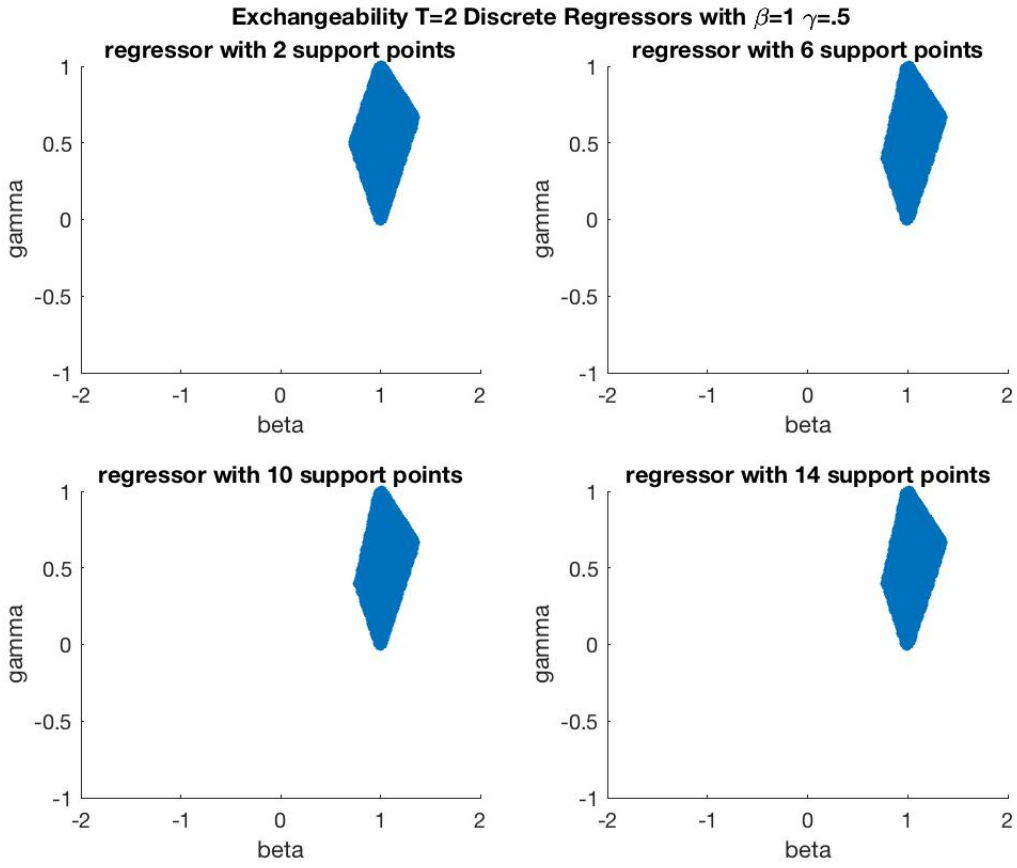

Figure 9: Exchangeability with $T = 2$ Discrete Support for v with $\gamma = .5$

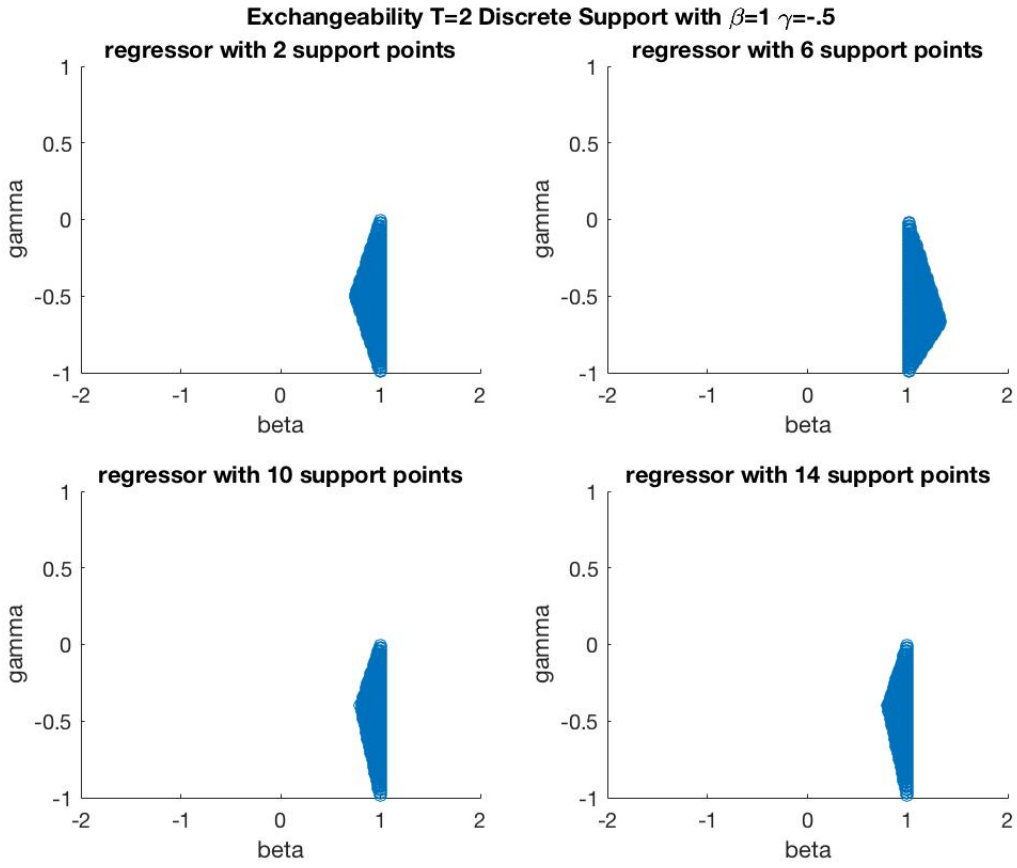

Figure 10: Stationary with $T = 2$ Discrete Support for v with $\gamma = -.5$

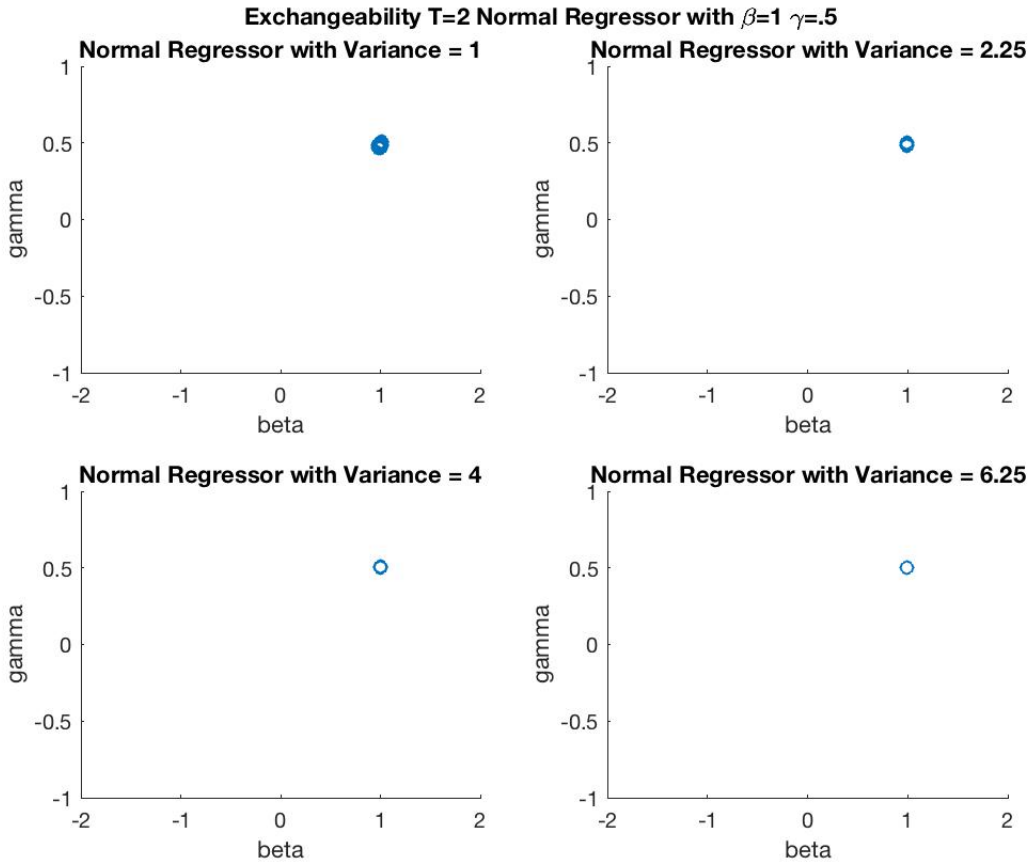

Figure 11: Exchangeability with $T = 2$ Normal v with $\gamma = .5$

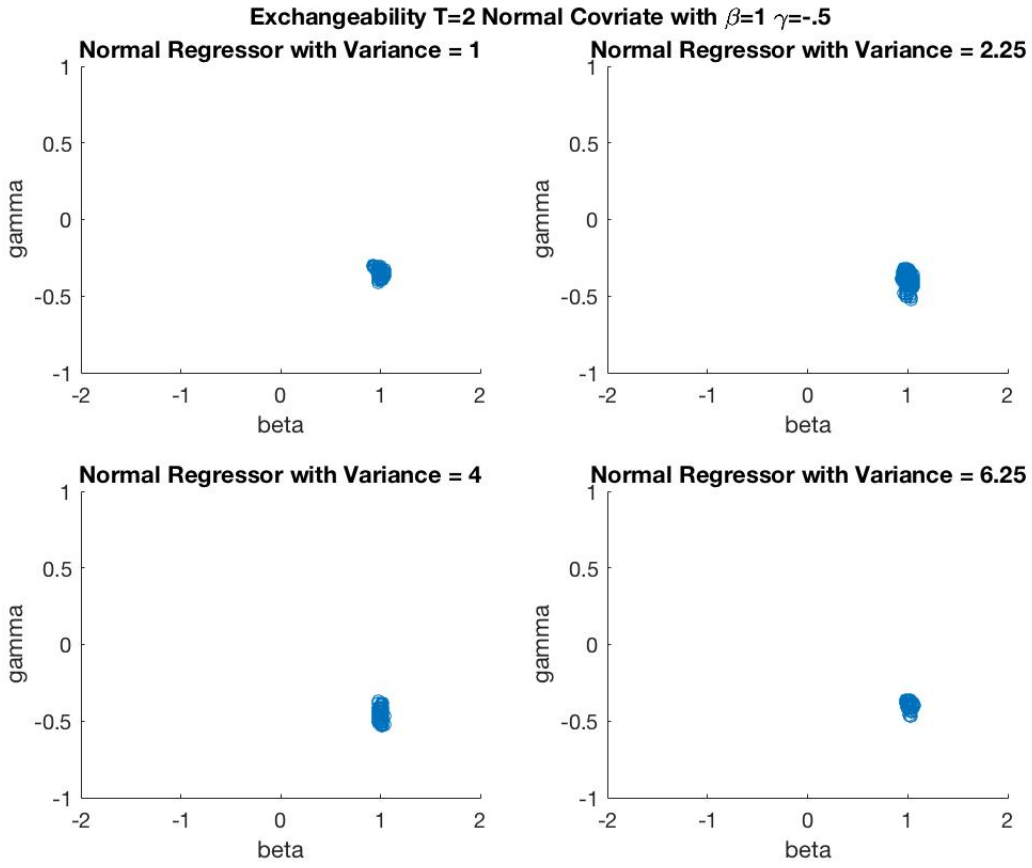

Figure 12: Exchangeability with $T = 2$ Normal v with $\gamma = -.5$

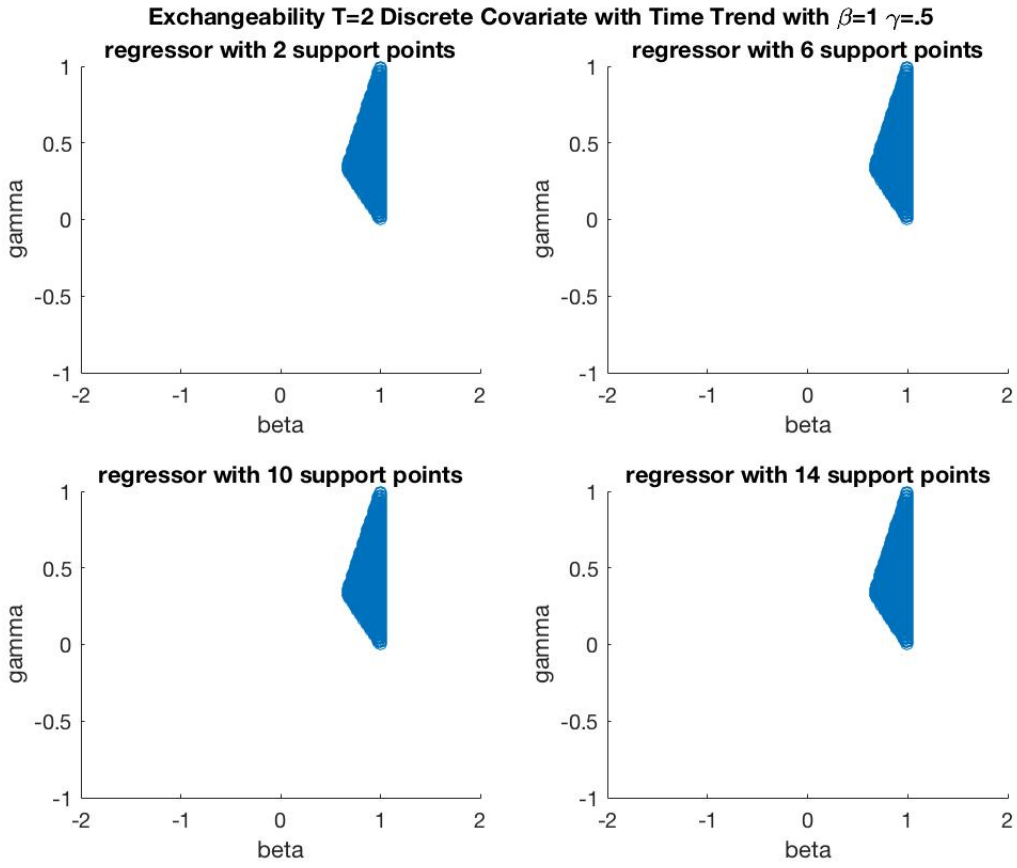

Figure 13: Exchangeability with $T = 2$ $x = t$ Discrete v with $\gamma = .5$

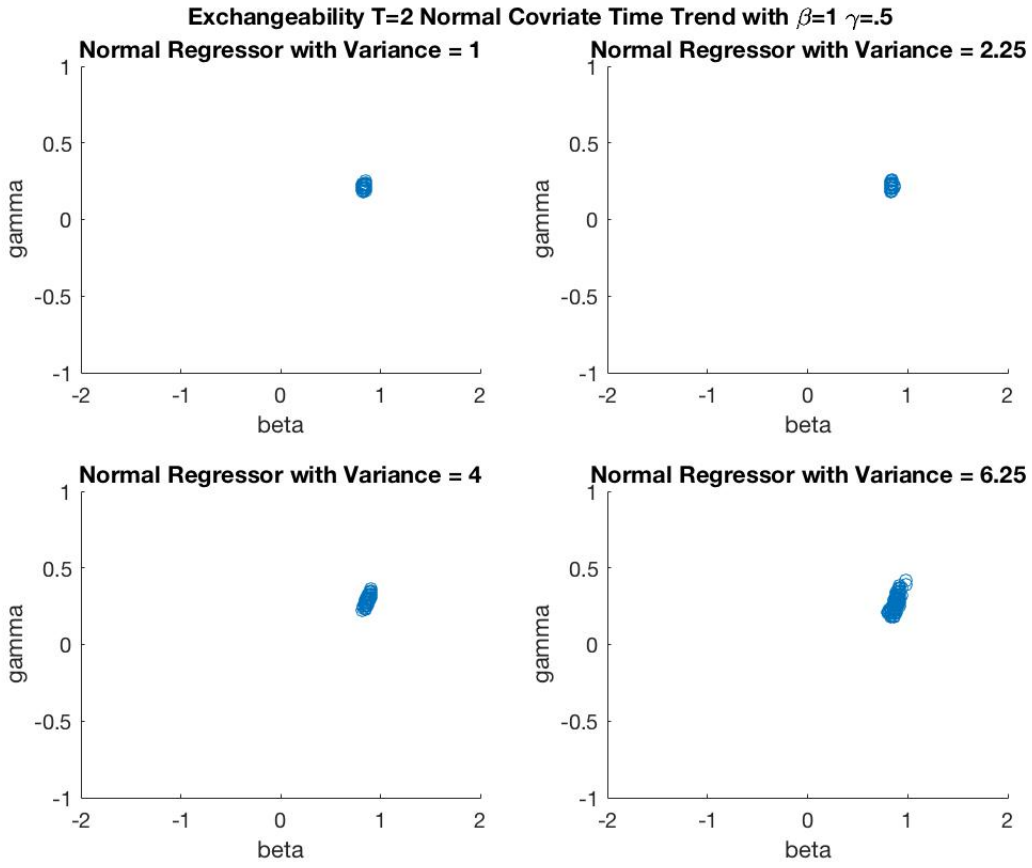

Figure 14: Exchangeability with $T = 2$, $x = t$ Normal v with $\gamma = .5$

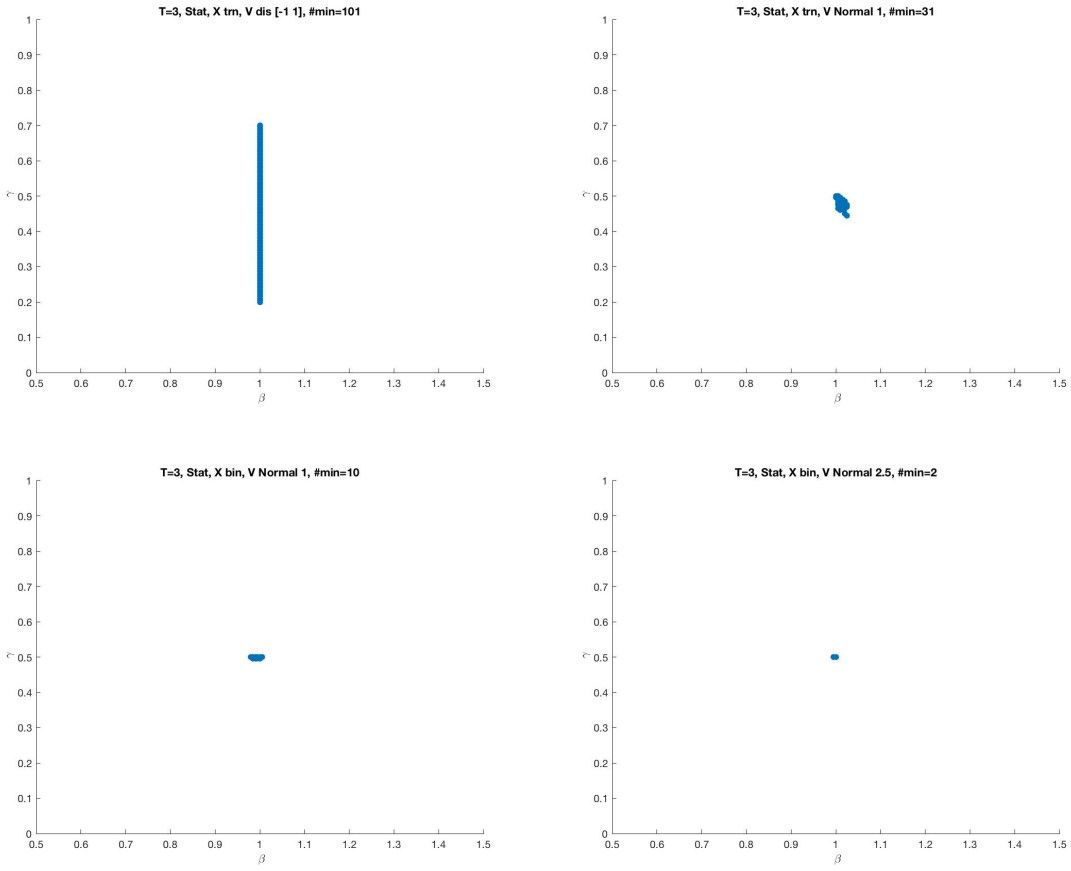

Figure 15: Stationarity with T=3: Various Designs

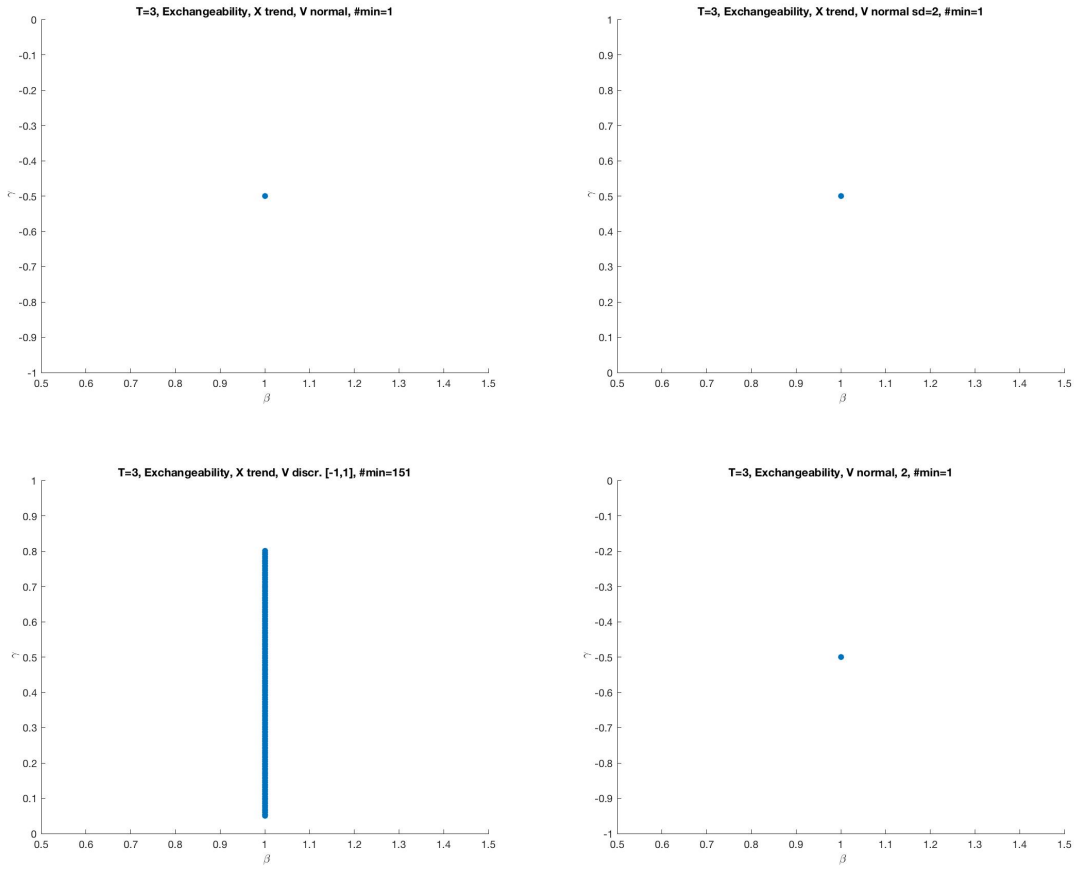

Figure 16: Exchangeability with T=3: Various Designs

B Proofs

B.1 Proof of Theorem 1 and Theorem 2

We start with the proof of our main result (Theorem 1) to demonstrate our approach. The proof has two parts: First, we show that the true parameter belongs to Θ_I (i.e. it satisfies the conditions of Theorem 1). Next, we show that for any parameter $\tilde{\theta} \in \Theta_I$ we can construct the distribution of unobservable $\tilde{\alpha}_i$'s and $\tilde{u}_{it}$'s that follows assumptions 1 and 2 so that in the dynamic binary choice model

$$\tilde{y}_{it} = 1\{\tilde{u}_{it} \leq x_{it}'\beta + \gamma \tilde{y}_{it-1} + \tilde{\alpha}_i\}$$

the distribution of $(\tilde{y}_i, x_i)$ is identical to the distribution of (y_i, x_i). In that case, we say that $\tilde{\theta}$ is *observationally equivalent* to θ.

True parameter belongs to Θ_I:

Let $v_{it} = u_{it} - \alpha_i$ and $v_{is} = u_{is} - \alpha_i$. Note that if u_{it} and u_{is} are identically distributed conditional on x_i and α_i then v_{it} and v_{is} must be identically distributed conditional on x_i. Let $F_v(\cdot|x)$ denote the (conditional on $x_i = x$) marginal distribution of v_{it} for $t = 1, 2, \ldots, T$. Then we have the following restrictions on $F_v(\cdot|x)$ for time period t:

$$\begin{aligned}
P(y_{it} = 1|x_i = x) \leq & F_v(x_t'\beta + \max\{0, \gamma\}|x) \\
P(y_{it-1} = 1, y_{it} = 1|x_i = x) \leq & F_v(x_t'\beta + \gamma|x) \\
P(y_{it-1} = 0, y_{it} = 1|x_i = x) \leq & F_v(x_t'\beta|x) \\
F_v(x_t'\beta + \min\{0, \gamma\}|x) \leq & P(y_{it} = 1|x_i = x) \\
F_v(x_t'\beta + \gamma|x) \leq & 1 - P(y_{it-1} = 1, y_{it} = 0|x_i = x) \\
F_v(x_t'\beta|x) \leq & 1 - P(y_{it-1} = 0, y_{it} = 0|x_i = x)
\end{aligned} \tag{14}$$

and for time period s:

$$\begin{aligned} P(y_{is}=1|x_i=x) \leq & F_v(x_s'\beta+\max\{0,\gamma\}) \\ P(y_{is-1}=1,y_{is}=1|x_i=x) \leq & F_v(x_s'\beta+\gamma|x) \\ P(y_{is-1}=0,y_{is}=1|x_i=x) \leq & F_v(x_s'\beta|x) \\ & F_v(x_s'\beta+\min\{0,\gamma\}|x) \leq P(y_{is}=1|x_i=x) \\ & F_v(x_s'\beta+\gamma|x) \leq 1-P(y_{is-1}=1,y_{is}=0|x_i=x) \\ & F_v(x_s'\beta|x) \leq 1-P(y_{is-1}=0,y_{is}=0|x_i=x) \end{aligned} \tag{15}$$

Part A1 of assumption 2 implies that $F_v(\cdot|x)$ is a strictly increasing function, and conditions (1)-(9) in Theorem 1 immediately follow from restrictions in (14) and (15).

Any parameter in Θ_I is observationally equivalent to the true parameter:

Now let $\tilde{\theta}=(\tilde{\gamma},\tilde{\beta}')'\in\Theta_I$, where Θ_I is characterized by the conditions (1)-(9) in Theorem 1. In what follows, we construct a sequence of random variables $\{\tilde{v}_{i1},\ldots,\tilde{v}_{iT}\}$ such that for all $(d_0,\ldots,d_T)\in\{0,1\}^{T+1}$ and all $x\in\mathcal{X}$,

$$P(y_{i0}=d_0,\ldots,y_{iT}=d_T|x_i=x)\equiv P(\tilde{y}_{i0}=d_0,\ldots,\tilde{y}_{iT}=d_T|x_i=x)$$

where

$$\tilde{y}_{it}=1\{\tilde{v}_{it}\leq x_{it}'\tilde{\beta}+\tilde{\gamma}\tilde{y}_{it-1}\}$$

and $\tilde{v}_{it}$ satisfies a stationarity property.

We start with $\tilde{y}_{i0}\equiv y_{i0}$. Suppose that we already found a sequence $\{\tilde{v}_{i1},\ldots,\tilde{v}_{it-1}\}$ that matches the distribution of the first t outcomes: for all $(d_0,\ldots,d_{t-1})\in\{0,1\}^t$ and all $x\in\mathcal{X}$,

$$P(y_{i0}=d_0,\ldots,y_{it}=d_{t-1}|x_i=x)\equiv P(\tilde{y}_{i0}=d_0,\ldots,\tilde{y}_{it-1}=d_{t-1}|x_i=x)$$

There exists a random variable $\tilde{v}_{it}$ such that

$$\begin{aligned} P(\tilde{v}_{it}\leq x_{it}'\tilde{\beta}|\tilde{y}_{it-1}=0,\tilde{y}_{it-2}=d_{it-2},\ldots,\tilde{y}_0=d_0,x_i=x)= \\ =P(y_{it}=1|y_{it-1}=0,y_{it-2}=d_{it-2},\ldots,y_{i0}=d_0,x_i=x) \\ P(\tilde{v}_{it}\leq x_{it}'\tilde{\beta}+\tilde{\gamma}|\tilde{y}_{it-1}=1,\tilde{y}_{it-2}=d_{it-2},\ldots,\tilde{y}_{i0}=d_0,x_i=x)= \\ =P(y_{it1}=1|y_{it-1}=1,y_{it-2}=d_{it-2},\ldots,y_{i0}=d_0,x_i=x) \end{aligned}$$

Then the sequence $\{\tilde{v}_{i1}, \ldots, \tilde{v}_{it}\}$ matches the distribution of the first $t+1$ outcomes:

$$P(y_{i0} = d_0, \ldots, y_{it} = d_t | x_i = x) \equiv P(\tilde{y}_{i0} = d_0, \ldots, \tilde{y}_{it} = d_t | x_i = x)$$

We can continue this procedure until we match the distribution of all $T+1$ outcomes.

Next step is to verify that the marginal distributions of $\tilde{v}_{it}$ and $\tilde{v}_{i1}$ can be made identical for all $t = 1, \ldots, T$, conditional on x_i. That is, conditional on x_i, $\tilde{v}_{it}$ is stationary.

For each t, the construction of the sequence $\tilde{y}_{i0}, \ldots, \tilde{y}_{iT}$ places restrictions on $P(\tilde{v}_{it} \leq v | x_i = x)$ only at these two points: $v = x_{it}'\tilde{\beta}$ and $v = x_{it}'\tilde{\beta} + \tilde{\gamma}$. Specifically,

$$\begin{aligned}
&P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} | x_i = x) = \\
&\quad = \sum_{(d_0,\ldots,d_{t-2}) \in \{0,1\}^{t-1}} P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} | \tilde{y}_{it-1} = 1, \ldots, \tilde{y}_{i0} = d_0, x_i = x) P(\tilde{y}_{it-1} = 1, \ldots, \tilde{y}_{i0} = d_0 | x_i = x) \\
&\qquad + \sum_{(d_0,\ldots,d_{t-2}) \in \{0,1\}^{t-1}} P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} | \tilde{y}_{it-1} = 0, \ldots, \tilde{y}_{i0} = d_0, x_i = x) P(\tilde{y}_{it-1} = 1, \ldots, \tilde{y}_{i0} = d_0 | x_i = x) \\
&\quad = \sum_{(d_0,\ldots,d_{t-2}) \in \{0,1\}^{t-1}} P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} | \tilde{y}_{it-1} = 1, \ldots, \tilde{y}_{i0} = d_0, x_i = x) P(\tilde{y}_{it-1} = 1, \ldots, \tilde{y}_{i0} = d_0 | x_i = x) \\
&\qquad + P(y_{it-1} = 0, y_{it} = 1 | x_i = x)
\end{aligned}$$

and similarly,

$$\begin{aligned}
&P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \tilde{\gamma} | x_i = x) = \\
&\quad = \sum_{(d_0,\ldots,d_{t-2}) \in \{0,1\}^{t-1}} P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \tilde{\gamma} | \tilde{y}_{it-1} = 0, \ldots, \tilde{y}_{i0} = d_0, x_i = x) P(\tilde{y}_{it-1} = 0, \ldots, \tilde{y}_{i0} = d_0 | x_i = x) \\
&\qquad + P(y_{it-1} = 1, y_{it} = 1 | x_i = x)
\end{aligned}$$

Note that probabilities $P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} | \tilde{y}_{it-1} = 1, \ldots, \tilde{y}_{i0} = d_0, x_i = x)$ and $P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \tilde{\gamma} | \tilde{y}_{it-1} = 0, \ldots, \tilde{y}_{i0} = d_0, x_i = x)$ are not restricted by the sequential process of constructing $\tilde{y}_{i0}, \ldots, \tilde{y}_{iT}$ described above, and so these probabilities can be anything between 0 and 1. Setting all these probabilities to 0 bounds $P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} | x_i = x)$ and $P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \tilde{\gamma} | x_i = x)$

from below by

$$P(y_{it-1}=0,y_{it}=1|x_i=x)\leq P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}|x_i=x)$$
$$P(y_{it-1}=1,y_{it}=1|x_i=x)\leq P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}+\tilde{\gamma}|x_i=x)$$

and setting all these probabilities to 1 gives us the upper bounds on $P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}|x_i=x)$ and $P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}+\tilde{\gamma}|x_i=x)$

$$\begin{aligned}P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}|x_i=x)\leq& P(y_{it-1}=0,y_{it}=1|x_i=x)+P(y_{it-1}=1|x_i=x)\\ &=1-P(y_{it-1}=0,y_{it}=0|x_i=x)\\ P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}+\tilde{\gamma}|x_i=x)\leq& P(y_{it-1}=1,y_{it}=1|x_i=x)+P(y_{it-1}=0|x_i=x)\\ &=1-P(y_{it-1}=1,y_{it}=0|x_i=x)\end{aligned}$$

Additionally, if $\tilde{\gamma}$ is positive, then

$$\begin{aligned}&\sum_{(d_0,\ldots,d_{t-2})\in\{0,1\}^{t-1}} P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}|\tilde{y}_{it-1}=1,\ldots,\tilde{y}_{i0}=d_0,x_i=x)P(\tilde{y}_{it-1}=1,\ldots,\tilde{y}_{i0}=d_0|x_i=x)\\ \leq&\sum_{(d_0,\ldots,d_{t-2})\in\{0,1\}^{t-1}} P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}+\tilde{\gamma}|\tilde{y}_{it-1}=1,\ldots,\tilde{y}_{i0}=d_0,x_i=x)P(\tilde{y}_{it-1}=1,\ldots,\tilde{y}_{i0}=d_0|x_i=x)\\ &=P(y_{it-1}=1,y_{it}=1|x_i=x)\end{aligned}$$

and so we bound $P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}|x_i=x)$ from above by:

$$P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}|x_i=x)\leq P(y_{it-1}=1,y_{it}=1|x_i=x)+P(y_{it-1}=0,y_{it}=1|x_i=x)=P(y_{it}=1|x_i=x)$$

Similarly,

$$\begin{aligned}&\sum_{(d_0,\ldots,d_{t-2})\in\{0,1\}^{t-1}} P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}+\tilde{\gamma}|\tilde{y}_{it-1}=0,\ldots,\tilde{y}_{i0}=d_0,x_i=x)P(\tilde{y}_{it-1}=0,\ldots,\tilde{y}_{i0}=d_0|x_i=x)\\ \geq&\sum_{(d_0,\ldots,d_{t-2})\in\{0,1\}^{t-1}} P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}|\tilde{y}_{it-1}=0,\ldots,\tilde{y}_{i0}=d_0,x_i=x)P(\tilde{y}_{it-1}=0,\ldots,\tilde{y}_{i0}=d_0|x_i=x)\\ &=P(y_{it-1}=0,y_{it}=1|x_i=x)\end{aligned}$$

and so we bound $P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}+\tilde{\gamma}|x_i=x)$ from below by:

$$P(\tilde{v}_{it}\leq x'_{it}\tilde{\beta}+\tilde{\gamma}|x_i=x)\geq P(y_{it-1}=0,y_{it}=1|x_i=x)+P(y_{it-1}=1,y_{it}=1|x_i=x)=P(y_{it}=1|x_i=x)$$

Using a similar approach for negative $\tilde{\gamma}$, we get the following bounds on $P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta}|x_i = x)$ and $P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \tilde{\gamma}|x_i = x)$:

$$P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta}|x_i = x) \geq P(y_{it} = 1|x_i = x)$$

and

$$P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \tilde{\gamma}|x_i = x) \leq P(y_{it} = 1|x_i = x)$$

That is, we have the following:

$$P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \min\{0, \tilde{\gamma}\}|x_i = x) \leq P(y_{it} = 1|x_i = x) \leq P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \max\{0, \tilde{\gamma}\}|x_i = x)$$

To summarize: our sequential construction of $\tilde{y}_{i0}, \ldots, \tilde{y}_{iT}$ only bounds each $P(\tilde{v}_{it} \leq v|x_i = x)$ for $t = 1, \ldots, T$ in the following ways:

$$\begin{aligned}
&P(y_{it-1} = 0, y_{it} = 1|x_i = x) \leq P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta}|x_i = x) \\
&P(y_{it-1} = 1, y_{it} = 1|x_i = x) \leq P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \tilde{\gamma}|x_i = x) \\
&P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta}|x_i = x) \leq 1 - P(y_{it-1} = 0, y_{it} = 0|x_i = x) \\
&P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \tilde{\gamma}|x_i = x) \leq 1 - P(y_{it-1} = 1, y_{it} = 0|x_i = x) \\
&P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \min\{0, \tilde{\gamma}\}|x_i = x) \leq P(y_{it} = 1|x_i = x) \\
&P(\tilde{v}_{it} \leq x_{it}'\tilde{\beta} + \max\{0, \tilde{\gamma}\}|x_i = x) \geq P(y_{it} = 1|x_i = x)
\end{aligned} \tag{16}$$

Other than having to satisfy the restrictions in (16), the distribution of $\tilde{v}_{it}$ conditional on x_i can move freely by varying conditional probabilities $P(\tilde{v}_{it} \leq v|\tilde{y}_{it-1} = d_{t-1}, \ldots, \tilde{y}_{i0} = d_0, x_i = x)$.

The restrictions in (16) are identical to the restrictions in (14). Since $(\tilde{\beta}, \tilde{\gamma}) \in \Theta_I$, there exists a (conditional on $x_i = x$) probability distribution $F(\cdot|x)$ such that for all $t = 1, \ldots, T$

the following holds:

$$\begin{aligned}
&P(y_{it-1}=0,y_{it}=1|x_i=x)\leq F(x_{it}'\tilde{\beta}|x)\\
&P(y_{it-1}=1,y_{it}=1|x_i=x)\leq F(x_{it}'\tilde{\beta}+\tilde{\gamma}|x)\\
&F(x_{it}'\tilde{\beta}|x)\leq 1-P(y_{it-1}=0,y_{it}=0|x_i=x)\\
&F(x_{it}'\tilde{\beta}+\tilde{\gamma}|x)\leq 1-P(y_{it-1}=1,y_{it}=0|x_i=x)\\
&F(x_{it}'\tilde{\beta}+\min\{0,\tilde{\gamma}\}|x)\leq P(y_{it}=1|x_i=x)\\
&F(x_{it}'\tilde{\beta}+\max\{0,\tilde{\gamma}\}|x)\geq P(y_{it}=1|x_i=x)
\end{aligned} \tag{17}$$

It remains to verify that $F(\cdot|x_i=x)$ can be the marginal distribution of $\{\tilde{v}_{i1},\ldots,\tilde{v}_{iT}\}$. We start with $\tilde{v}_{i1}$: the marginal distribution of $\tilde{v}_{i1}$ evaluated at $x_{i1}'\tilde{\beta}$ is given by

$$\begin{aligned}
P(\tilde{v}_{i1}\leq x_{i1}'\tilde{\beta}|x_i=x)=&P(y_{i0}=0,y_{i1}=1|x_i=x)\\
&+P(\tilde{v}_{i1}\leq x_{i1}'\tilde{\beta}|\tilde{y}_{i0}=1,x_i=x)P(\tilde{y}_{i0}=1|x_i=x)\\
=&P(y_{i0}=0,y_{i1}=1|x_i=x)\\
&+P(\tilde{v}_{i1}\leq x_{i1}'\tilde{\beta}|\tilde{y}_{i0}=1,x_i=x)P(y_{i0}=1|x_i=x)
\end{aligned}$$

where the last equality holds because by construction, $P(\tilde{y}_{i0}=1|x_i=x)=P(y_{i0}=1|x_i=x)$.

If $P(y_{i0}=1|x_i=x)=0$, then the first and third restrictions in (17) imply that $F(x_{i1}'\tilde{\beta}|x)=P(y_{i0}=0,y_{i1}=1,|x_i=x)$ and so we can set $P(\tilde{v}_{i1}\leq x_{i1}'\tilde{\beta}|\tilde{y}_{i0}=1,x_i=x)$ to be anything between 0 and 1.

If $P(y_{i0}=1|x_i=x)>0$, the we can set

$$P(\tilde{v}_{i1}\leq x_{i1}'\tilde{\beta}|\tilde{y}_{i0}=1,x_i=x)=\frac{F(x_{i1}'\tilde{\beta}|x)-P(y_{i0}=0,y_{i1}=1|x_i=x)}{P(y_{i0}=1|x_i=x)}$$

Note that the first and third restrictions in (17) ensures that the right-hand side of the above equation is between 0 and 1.

By setting these conditional probabilities as described we guarantee that

$$P(\tilde{v}_{i1}\leq x_{i1}'\tilde{\beta}|x_i=x)=F(x_{i1}'\tilde{\beta}|x)$$

Similarly, the marginal distribution of $\tilde{v}_{i1}$ evaluated at $x_{i1}'\tilde{\beta}+\tilde{\gamma}$ is given by

$$\begin{aligned}P(\tilde{v}_{i1}\leq x_{i1}'\tilde{\beta}+\tilde{\gamma}|x_i=x)=&P(y_{i0}=1,y_{i1}=1|x_i=x)\\&+P(\tilde{v}_{i1}\leq x_{i1}'\tilde{\beta}+\tilde{\gamma}|\tilde{y}_{i0}=0,x_i=x)P(y_{i0}=0|x_i=x)\end{aligned}$$

If $P(y_{i0}=0|x_i=x)=0$, then the second and fourth restrictions in (17) imply that $F(x_{i1}'\tilde{\beta}+\tilde{\gamma}|x)=P(y_{i0}=1,y_{i1}=1,|x_i=x)$ and so we are free to set $P(\tilde{v}_{i1}\leq x_{i1}'\tilde{\beta}+\tilde{\gamma}|\tilde{y}_{i0}=1,x_i=x)$ to be anything between 0 and 1.

If $P(\tilde{y}_{i0}=0|x_i=x)>0$, the we can set

$$P(\tilde{v}_{i1}\leq x_{i1}'\tilde{\beta}+\tilde{\gamma}|\tilde{y}_{i0}=0,x_i=x)=\frac{F(x_{i1}'\tilde{\beta}+\tilde{\gamma}|x)-P(y_{i0}=1,y_{i1}=1|x_i=x)}{P(y_{i0}=0|x_i=x)}$$

Here the second and fourth restrictions in (17) ensures that the right-hand side of the above equation is between 0 and 1.

By setting these conditional probabilities as described we guarantee that

$$P(\tilde{v}_{i1}\leq x_{i1}'\tilde{\beta}+\tilde{\gamma}|x_i=x)=F(x_{i1}'\tilde{\beta}+\tilde{\gamma}|x)$$

We can repeat this construction step for any $t>2$. Specifically, for an arbitrary $t>2$ we have

$$\begin{aligned}P(\tilde{v}_{it}\leq x_{it}'\tilde{\beta}|x_i=x)=&P(\tilde{y}_{it-1}=0,\tilde{y}_{it}=1|x_i=x)\\&+P(\tilde{v}_{it}\leq x_{it}'\tilde{\beta}|\tilde{y}_{it-1}=1,x_i=x)P(y_{it-1}=1|x_i=x)\end{aligned}$$

and

$$\begin{aligned}P(\tilde{v}_{it}\leq x_{it}'\tilde{\beta}+\tilde{\gamma}|x_i=x)=&P(\tilde{y}_{it-1}=1,\tilde{y}_{it}=1|x_i=x)\\&+P(\tilde{v}_{it}\leq x_{it}'\tilde{\beta}+\tilde{\gamma}|\tilde{y}_{it-1}=0,x_i=x)P(y_{it-1}=0|x_i=x)\end{aligned}$$

If $P(y_{it-1}=1|x_i=x)=0$, then we already have that $P(\tilde{v}_{it}\leq x_{it}'\tilde{\beta}|x_i=x)=F(x_{it}'\tilde{\beta}|x_i=x)$. Similarly, if $P(y_{it-1}=0|x_i=x)=0$, then we already have $P(\tilde{v}_{it}\leq x_{it}'\tilde{\beta}+\tilde{\gamma}|x_i=x)=F(x_{it}'\tilde{\beta}+\tilde{\gamma}|x)$.

If $P(y_{it-1} = 1|x_i = x) > 0$, then we can set

$$P(\tilde{v}_{it} \leq x'_{it}\tilde{\beta}|\tilde{y}_{it-1} = 1, x_i = x) = \frac{F(x'_{it}\tilde{\beta}|x) - P(y_{it} = 1, y_{it-1} = 0|x_i = x)}{P(y_{it-1} = 1|x_i = x)}$$

so that $P(\tilde{v}_{it} \leq x'_{it}\tilde{\beta}|x_i = x) = F(x'_{it}\tilde{\beta}|x)$.

And if $P(y_{it-1} = 0|x_i = x) > 0$, then we can set

$$P(\tilde{v}_{it} \leq x'_{it}\tilde{\beta} + \tilde{\gamma}|\tilde{y}_{it-1} = 0, x_i = x) = \frac{F(x'_{it}\tilde{\beta} + \tilde{\gamma}|x) - P(y_{it} = 1, y_{it-1} = 1|x_i = x)}{P(y_{it-1} = 0|x_i = x)}$$

so that $P(\tilde{v}_{it} \leq x'_{it}\tilde{\beta} + \tilde{\gamma}|x_i = x) = F(x'_{it}\tilde{\beta} + \tilde{\gamma}|x)$

For any $v \notin \{x'_{it}\tilde{\beta}, x'_{it}\tilde{\beta} + \tilde{\gamma}\}$ the construction of sequence $\{\tilde{y}_{i0}, \ldots, \tilde{y}_{iT}\}$ doesn't restrict the distribution of $\tilde{v}_{it}$ evaluated at v in any way, so we can set

$$P(\tilde{v}_{it} \leq v|x_i = x) = F(v|x)$$

To sum up: for a model $\tilde{y}_{it} = 1\{\tilde{v}_{it} \leq x'_{it}\tilde{\beta} + \tilde{\gamma}\tilde{y}_{it-1}\}$, we found a distribution of $\{\tilde{v}_{i1}, \ldots, \tilde{v}_{iT}\}$ such that the following holds:

- This model is observationally equivalent to the original model: $P(y_{i0} = d_0, \ldots, y_{iT} = d_T|x_i) = P(\tilde{y}_{i0} = d_0, \ldots, \tilde{y}_{iT} = d_T|x_i)$ for all $(d_0, \ldots, d_T) \in \{0, 1\}^{T+1}$
- Distribution of error terms $\tilde{v}_{it}$'s in this model is stationary: $P(\tilde{v}_{it} \leq v|x_i = x) = F(v|x)$ for all $t = 1, \ldots, T$.

Now let $\tilde{\alpha}_i = 0$ and let $\tilde{u}_{it} = \tilde{v}_{it}$. Then

$$\tilde{y}_{it} = 1\{\tilde{u}_{it} \leq x'_{it}\tilde{\beta} + \tilde{\gamma}\tilde{y}_{it-1} + \tilde{\alpha}_i\}$$

where $\tilde{u}_{it}|x_i, \tilde{\alpha}_i \overset{d}{=} \tilde{u}_{i1}|x_i, \tilde{\alpha}_i$ for all $t = 2, \ldots, T$ and where

$$P(y_{i0} = d_0, \ldots, y_{it} = d_T|x_i = x) \equiv P(\tilde{y}_{i0} = d_0, \ldots, \tilde{y}_{it} = d_T|x_i = x)$$

for all $(d_0, \ldots, d_T) \in \{0,1\}^{T+1}$ and all $x \in \mathcal{X}$. That is, any $\tilde{\theta} = (\tilde{\gamma}, \tilde{\beta}')' \in \Theta_I$ is observationally equivalent to the true parameter θ. This completes the proof of Theorem 1.

The proof of Theorem 2 closely follows the proof of Theorem 1 with the these two differences: conditioning on x_i, y_{i0} rather than just x_i, and matching $P(y_{i1} = 1|x_i, y_{i0})$ directly:

$$P(\tilde{v}_{i1} \leq x_1'\tilde{\beta} + \tilde{\gamma} d_0 | x_i = x, y_{i0} = d_0) = P(y_{i1} = 1 | x_i = x, y_{i0} = d_0)$$

■

B.2 Proof of Theorem 3

We establish point 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 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 1 implies that it must be that $P(y_{i0} = 1, y_{i1} = 1|x_i = x^*) + P(y_{i1} = 1, y_{i2} = 0|x_i = x^*) > 1$ 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 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 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 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 1. ■